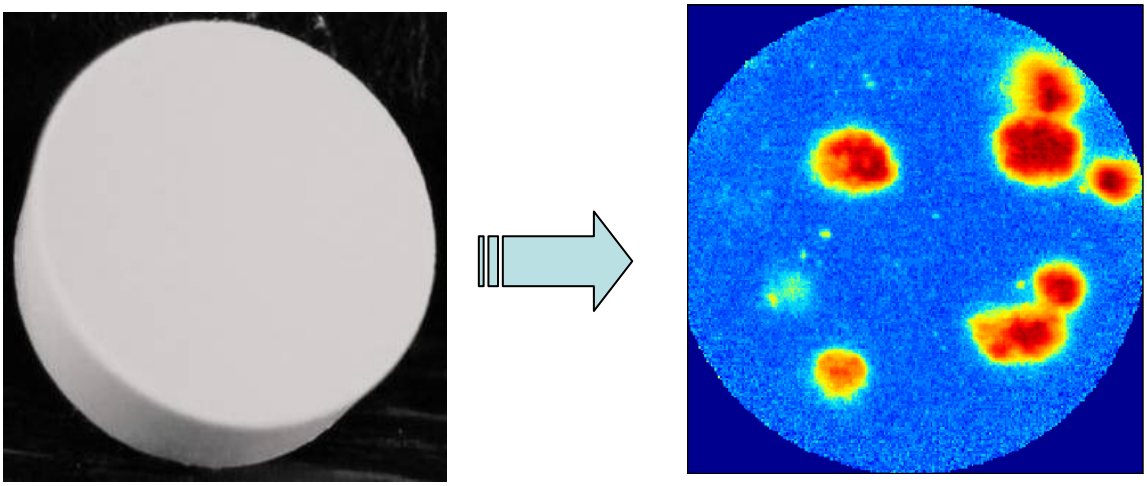
Near-infrared Chemical Imaging in Formulation Development of Solid Dosage Forms

PhD thesis

by

Carsten Ravn



Industrial PhD project

Department of Pharmaceutics and Analytical Chemistry

Faculty of Pharmaceutical Sciences

University of Copenhagen, Denmark

&

Novo Nordisk A/S

January 2009

NEAR-INFRARED CHEMICAL IMAGING IN FORMULATION DEVELOPMENT OF SOLID DOSAGE FORMS

Carsten Ravn[a]
Industrial PhD Thesis, January 2009
Department of Pharmaceutics and Analytical Chemistry
Faculty of Pharmaceutical Sciences
University of Copenhagen, Denmark
&
Novo Nordisk A/S, Måløv, Denmark

Abstract

The subject of this thesis is the use of near-infrared chemical imaging (NIR-CI) for the study of pharmaceutical tablets in formulation development. NIR-CI is an emerging technology in pharmaceutical analysis and solid dosage form development, as it provides insight into the structure and function of tablets, which cannot readily be obtained by any other technique. This information is highly valuable, as many quality attributes of tablets are closely related to the size and distribution of not only the active component (API) but also the excipients within the tablet matrix.

Near-infrared chemical imaging is an analytical technique based on near-infrared (NIR) spectroscopy, but in addition to identification and quantification of bulk components from NIR spectroscopy it provides information on the spatial distribution of *all* the spectroscopically active chemical components in a sample. A description of the basic principles and instrumentation for NIR-CI is shortly given in this thesis.

NIR-CI generates a huge amount of spectral data, and the use of multivariate and image analysis methods are critical to the translation and extraction of these data into accurate and relevant chemical information. Optimally, identification, quantification and location of the chemical components within the sample matrix will be obtained, and an image of the size, shape and distribution of each component within the sample created (chemical image).

The primary purpose was originally to explore the use of NIR-CI in different types of formulation development applications and evaluate the increased process and product understanding that could be gained from these experiments. However, during the initial project phase it became apparent that fundamental work on the NIR-CI technology was still needed, for instance to develop reliable data and image processing methods. Therefore, a major part of this thesis describes data and image processing methodologies for the analysis and extraction of useful and accurate information from NIR-CI data meas-

[a] *Author contact details:* cacaravn@yahoo.dk

urements. Each of the typical steps involved in NIR-CI analysis is explained, and this thesis can in fact be used as a guidance to perform such data analysis.

In order to develop reliable and accurate NIR-CI data processing methods, it is necessary to evaluate them with regard to their ability to generate accurate chemical images. In this context, three common data processing approaches were compared. A calibration data set was constructed and analysed. The accuracies of the three methods were compared according to their ability to generate accurate chemical images, which was evaluated by their concentration prediction ability.

Furthermore, results from the application of developed NIR-CI data processing and image processing tools to a pharmaceutical formulation development experiment are presented. Chemical images, histogram statistics, binary images, domain size statistics and a method for evaluating tablet homogeneity were all used to evaluate tablets from the formulation development study. The study demonstrates the unique and invaluable information that the NIR-CI technology is capable of providing for the formulation scientist. The presented approach will be applicable in many types of development studies for new tablet or other solid dosage formulations.

The NIR-CI technology is still maturing, and a discussion of some of the limitations and cautions is provided. Important aspects that require improvement for the NIR-CI technology to be widely adopted and deployed in pharmaceutical development are also discussed. Hence, the full potential on the use of the NIR-CI technology in the pharmaceutical industry has yet to be revealed but this thesis adds to its knowledge foundation. The unique information that can be obtained by near-infrared chemical imaging and the valuable role it can play in the development of new solid dosage formulations is demonstrated in this thesis.

PREFACE

The following thesis entitled "Near-infrared chemical imaging in formulation development of solid dosage forms" was submitted to the Faculty of Pharmaceutical Sciences, University of Copenhagen (Denmark), as part of the requirements to obtain the PhD degree. Being an Industrial PhD project supported by The Danish Agency for Science, Technology and Innovation, the thesis will also subsequently be send to the Agency together with a business report to obtain an Industrial PhD certificate.

This thesis presents the research performed during the industrial PhD project which started in September 2005. A Danish Industrial PhD project is a three-year research project with an industrial focus conducted jointly by a private company and a university (more information on: http://en.fi.dk/research). This project was a joint cooperation between Novo Nordisk A/S and The Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen (Denmark). Participating as third party institute The Department of Food Science, Faculty of Life Sciences, University of Copenhagen (Denmark) also made a major contribution to the project.

The project was at its beginning anchored in the company's Formulation Development department, whose mission was to develop solid dosage forms of drug molecules manufactured by synthesis in a world wide competitive manner regarding speed and quality. The PhD project was considered part of the department's development strategy to continuously develop and optimise existing technologies and core competences e.g. by implementing new innovative approaches. The objective of the project was to increase understanding of solid dosage formulation processes and products and to integrate this new near-infrared chemical imaging (NIR-CI) technology in future company development projects. Undoubtedly, the increasing promotion of the process analytical technology concept (read more in section 1.2 and 4.7) by the regulatory authorities also facilitated initiation of this project. Equally important, the project was well in line with the research profile of relevant groups at The Department of Pharmaceutics and Analytical Chemistry. The involvement of The Department of Food Science was due to their major expertise in multivariate data analysis, which was crucial for the completion of this project.

Halfway through the project it was due to management strategic decisions decided to discontinue the company's 'small molecule' research activities. As a consequence of this, the Formulation Development department was closed down. Hence, the professional sparring with the company's formulation development employees vanished together with research objectives/issues that were supposed to partly be provided from the company's development projects. Despite these changes, it was made feasible to continue the project with a similar scope now anchored in the company's Analytical Development department and with the role of the university departments maintained.

With the author's background as a formulation scientist this is also a target group of this thesis. From the significant growth and use of near-infrared chemical imaging in pharmaceutical applications it is most likely that more and more formulation scientists will be-

come acquainted with this or similar technologies in the future. However, the intended readers are also process chemists, analytical chemists and scientists working with chemical imaging specifically. In fact, this project is rather interdisciplinary integrating the three individual large scientific areas of pharmaceutical technology, near-infrared spectroscopy and multivariate data analysis (Figure 1).

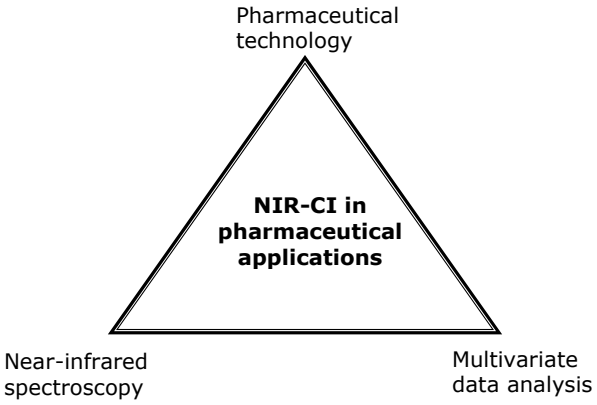


Figure 1 - The interdisciplinary knowledge foundation of this thesis necessary to enable studies of NIR-CI in pharmaceutical applications.

From the initiation of this project, pharmaceutical applications and formulation development studies were intended to make up a much larger part, using NIR-CI as a tool to gain new and increased process and product understanding. However, after purchasing the near-infrared chemical imaging system for the project, it became apparent that more focus had to be laid on the basic foundations of the NIR-CI technology comprised by data acquisition, data processing and image processing. The main part of this project was therefore focused on developing methodologies for these disciplines, which then can be used for many different types of pharmaceutical applications.

This thesis is organised in nine chapters and with the four papers produced during this project (as author and co-author) enclosed in Appendix I-IV.

The thesis will be defended at the Faculty of Pharmaceutical Sciences, University of Copenhagen.

Assessment Committee:

Professor Jukka Rantanen, Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Denmark, (Chairman).

Principal Scientist Lars Hovgaard, Department of Preformulation and Delivery, Novo Nordisk A/S, Denmark.

Dr. Gabriele Reich, Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Germany.

ACKNOWLEDGEMENTS

This PhD project has definitely been a "rich" experience in which I have met many intelligent and inspiring personalities and achieved enormous scientific and personal development. I wish to express my gratitude to all those people who has contributed, supported and helped me in various ways to reach the goal, which has been a tremendous achievement for me.

I would like to thank my university supervisor Associate Professor Claus Cornett[b] and also co-supervisor Associate Professor Jørn Møller-Sonnergaard[b]. I am thankful for their commitment to enter into collaboration on this industrial PhD project. I believe spectroscopic or chemical imaging will develop significantly in the near future and consequently hope to see more PhD or other projects embracing this technology at the Department of Pharmaceutics and Analytical Chemistry.

I am very grateful to the company Novo Nordisk A/S and the people there who gave me the opportunity to do this PhD project. Thanks to Corporate Research Affairs (CORA) for financing and supporting the project all the way through and in particular to Marianne Søndergaard for taking care of all the practical affairs. In this context, I also thank the Danish Agency for Science, Technology and Innovation for partly funding this project.

At Novo Nordisk, I mostly owe my gratitude to Principal Scientist Erik Skibsted[c], who despite all the strategically and organisational changes has "hung in" and stayed on the project all the way through. I realise it would have been difficult to complete the project without his support. At Novo Nordisk, I ended up being in three different departments having five different department managers during my PhD project. Thanks to all the people being involved and for the always warm welcoming and support. A special thank goes to all my former colleagues at Departments 518/261 Product Development/SDF Pharmaceutics. In particular to the people who supported the idea of founding the PhD project and approved financing of the rather expensive NIR-CI instrument.

I am most indebted to Professor Rasmus Bro[d] for his involvement in this project. His incredibly huge knowledge is indisputable and his commitment and investment in me and the project has been outstanding, also considering it was "only" as third-party supervisor. I could simply not have accomplished this project without his help and feel privileged for have been given the opportunity to receive guidance and learn (to learn) from him. Special thanks also go to Post doc's José Manuel Amigo[d] and Thomas Skov[d] for all their help, particular on matlab issues and exchange of ideas. The latter also goes for PhD student Carlos Cairòs[e] and together with José Amigo this collaboration was a boost to my project, where I experienced the fruitful and synergistic outcome of daily working together in a small group on the same scientific subject.

I also wish to show my appreciation to Emeritus Professor Tony Moffat[f] and Dr. Roger Jee[f] for the interest and collaboration at the beginning of the project. Unfortunately, de-

---

[b] Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Denmark.
[c] Novo Nordisk A/S, CMC Analysis & Formulation, Novo Nordisk Park, Måløv, Denmark.
[d] Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Denmark.
[e] Department de Química, Facultat de Ciències, University of Barcelona, Spain.
[f] Department of Pharmaceutical and Biological Chemistry, The School of Pharmacy, University of London, UK.

velopment and changes in the project had it that collaboration faded, but I appreciate all their engagement and inputs.

Last but not least, thanks to all my friends and in particular my Mum and Dad, who have always been there for me, and put up listening to my long talks about the project. Thank you so much all for your patience, understanding and support.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

API	Active Pharmaceutical Ingredient
ATR	Attenuated Total Reflection
AUHC	Area Under Homogeneity Curve
CI	Chemical Imaging
CLMB	Continuous-Level Moving Block
CLS	Classical Least Squares
CPP	Critical Process Parameters
CQA	Critical Quality Attributes
DoE	Design of Experiments
FDA	Food and Drug Administration (United States)
FPA	Focal Plane Arrays
FT	Fourier Transform
ICH	International Conference on Harmonisation
IR	Infrared
MCR-ALS	Multivariate Curve Resolution-Alternating Least Squares
MIR	Mid-Infrared
MIR-CI	Mid-Infrared Chemical Imaging
NIR	Near-Infrared
NIR-CI	Near-Infrared Chemical Imaging
PAT	Process Analytical Technology
PC	Principal Components
PCA	Principal Component Analysis
PLS	Partial Least Squares
QbD	Quality by Design
RMSECV	Root Mean Square Error of Cross-Validation

LIST OF PUBLICATIONS

This thesis is based on the work described in the following papers which will be referred to in the text by the Roman numerals I-IV. The first two papers (I and II) are as first author and the last two papers (III and IV) are closely related work (co-author).

Paper I

C. Ravn, E. Skibsted, R. Bro, Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms – comparing common calibration approaches, *Journal of Pharmaceutical and Biomedical Analysis* **48** (2008), 554-561.

Paper II

C. Ravn, J.M. Amigo, C. Cairos, J.M. Sonnergaard, E. Skibsted, C. Cornett, R. Bro, Application of near-infrared chemical imaging in formulation development of solid dosage forms (Manuscript ready for submission).

Paper III

J.M. Amigo, C. Ravn, Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging, *European Journal of Pharmaceutical Sciences* **37** (2009) 76-82.

Paper IV

J.M. Amigo, C. Ravn, N.B. Gallagher, R. Bro, A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging, *International Journal of Pharmaceutics* (2009), doi:10.1016/j.ijpharm.2009.02.014 (Article in Press, Corrected Proof).

CHAPTER 1

BACKGROUND

1.1 What is near-infrared chemical imaging?

The subject of this thesis is near-infrared chemical imaging (NIR-CI) applied to pharmaceutical solid dosage forms. For readers without prior knowledge of chemical imaging techniques and in particular near-infrared chemical imaging a short introduction to the technology is given here. A more comprehensive and detailed description of the NIR-CI principle and literature are presented in Chapter 4.

NIR-CI is the combination of conventional near-infrared (NIR) spectroscopy with digital imaging. Spectral and spatial information is recorded simultaneously providing a NIR spectrum in each pixel of the sample image. The result is a three dimensional data set, known as a hyperspectral data cube (Figure 2). The x and y axis represent the spatial information and z is the wavelength axis in the spectral measurement. The hyperspectral data cube is translated into a set of, typical chemical, images providing distribution information of each chemical component from the sample surface.

The unique feature of NIR-CI is that in addition to identification and quantification of the chemical components in the sample, their spatial distribution and cluster size and form within the sample is simultaneously revealed. Hence, chemical images visualises the internal structure and elucidates the distribution and cluster size of each chemical component within the sample (Figure 2, left), which can be related to the quality of the product.

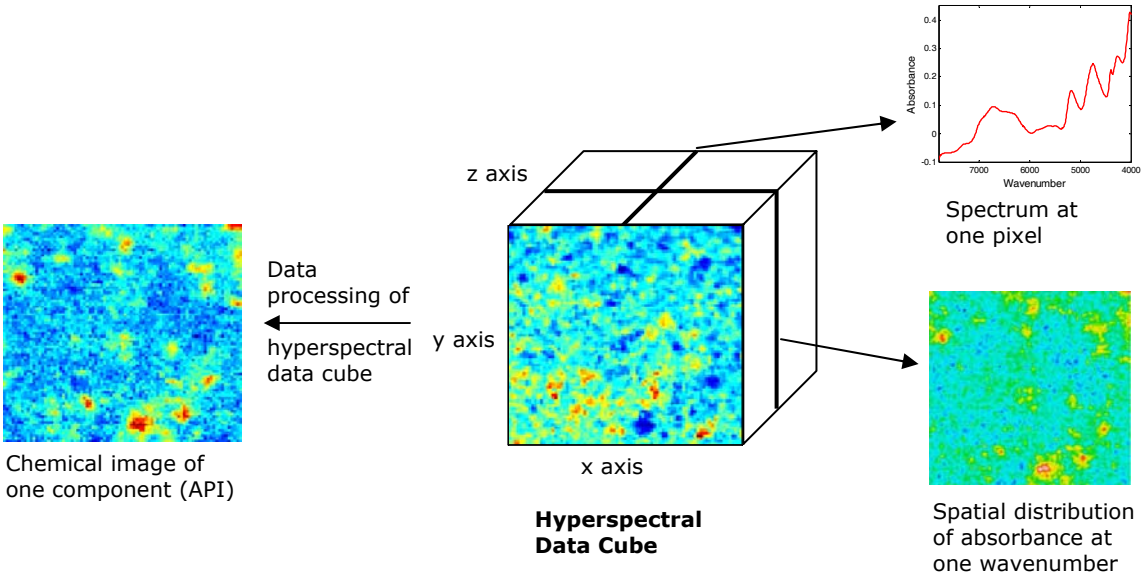


Figure 2 - Three-dimensional hyperspectral NIR data cube recorded from a NIR chemical imaging data acquisition. The information directly obtainable from the hyperspectral data cube is displayed. *Right*: Localised NIR spectra associated with each pixel and images associated with each NIR wavenumber. *Left*: A chemical image of the active pharmaceutical ingredient (API) generated from data processing of the data cube.

1.2 Introduction to project

Regulatory aspect

The pharmaceutical industry is a highly regulated industry. Manufacturing processes are traditionally treated as frozen and based on time-defined end points. Changes to processes are managed through regulatory submissions. The philosophy 'if it isn't broke, don't fix it' has been dominating and even 'broken' processes with frequent batch rejections or re-processing's have not been fixed. The reason for the latter is that the uncertain consequences from regulatory submissions of process changes as well as the high price of submitting post approval changes overpower the possible benefits from optimising the manufacturing processes. The lack of process control and scientific innovation in pharmaceutical manufacturing science led the United States Food and Drug Administration (FDA) to publish a guidance document about the use of process analytical technology (PAT)[1] (Section 4.7). The concept of PAT is to build in quality by design (QbD) instead of merely passively end-product testing to evaluate quality. The guidance promotes the use of different tools to gain process understanding and process control. The tools suggested in the guidance document are:

- ✧ Multivariate tools for design, data acquisition and analysis
- ✧ Process analyzers
- ✧ Process control tools
- ✧ Continuous improvement and knowledge management tools

Near-infrared chemical imaging is one type of novel analytical technology promoted by PAT to create increased process understanding and also to be used for process control.[2-3]

Pharmaceutical formulation development

Pharmaceutical formulation development (Chapter 2) is basically the scientific work of designing and developing a suitable dosage form containing the active pharmaceutical ingredient (API) taking into consideration the physical, chemical and pharmacological properties of the API, administration route, manufacturing process and the patient's use of the dosage form. The aim is to design a quality product and its manufacturing process to consistently deliver the intended performance of the product (ICH Q8).[4]

In formulation development numerous experiments are performed to select the appropriate dosage form, optimal manufacturing process, composition, processing steps and process parameters. It is studied how formulation and process factors affect product quality and performance. The goal is to identify and be able to monitor the critical quality attributes (CQA) and the critical process parameters (CPP), which is consistent with the concept of PAT.

Tablets are the most common pharmaceutical dosage form produced today and were also the first priority dosage form in the former Formulation Development department at the company Novo Nordisk A/S. Therefore tablets were chosen as the dosage form to be analysed in this work. A tablet is a solid dosage form where a mixture of the active

pharmaceutical ingredient(s) and excipients, usually in powder form, are compressed into a solid matrix. The excipients are inactive substances but each plays an important role and function for the tablet performance. The tablet matrix is hence carefully designed in development of new tablet formulations and the functional performance may vary depending on the concentration and distribution of all components within the final matrix. Probably the most significant factor in determining the quality of the formulated tablet is the structure of the matrix that evolves during the manufacturing process.[5] It is therefore a paradox, that when it is time to assess the quality of a tablet the matrix is destroyed by some kind of sample preparation. The conventional analytical techniques used to analyse the identified quality attributes of tablets are destructive methods, focuses on the API component and only yield bulk content information. No information on the excipients and no insight into the structure of the matrix are obtained. Different spectroscopic techniques, in particular near-infrared (NIR) spectroscopy, are now also widely used in pharmaceutical analysis. NIR spectroscopy is a non-destructive method but also only provides bulk information, i.e. determines the gross composition of sample.[2,6-8]

Product and process-intermediate testing

With the conventional analytical techniques available and also with pharmaceutical formulations becoming more complex and advanced the challenge is to fully characterise the processes and products. New analytical technologies are needed to determine the impact of process and ingredient changes and to identify and measure the critical quality attributes (CQA) and critical process parameters (CPP) to gain the process understanding that will enable process and product control. NIR-CI provides increased understanding of the solid dosage form by a technology that generates other types of information compared to traditional analysis techniques, which only provide bulk information of the analysed dosage form.

NIR-CI is a non-destructive analytical technique, requires almost no sample preparation and can be used to determine multiple quality attributes not only for the API but also the excipients. The visualisation of the spatial distribution of components in a sample provides information that can not be obtained by any conventional analytical techniques.

Scope and aim of thesis

Some of the major key-drivers for interest, implementation and further development of the NIR-CI technology in the pharmaceutical industry are:

- *Compositional structure of sample matrix.* The quality, i.e. the physical, chemical and biopharmaceutical properties, of pharmaceutical solid dosage forms is a function of the chemical composition and spatial distribution of all the components in the final matrix.
- *FDA's PAT initiative.*[1] FDA encourages manufacturers to use new technologies that increase the scientific understanding of how formulation and manufacturing process factors affect product quality and performance.

- *Trouble-shooting.* The non-destructive nature and unique component size and distribution information obtainable by NIR-CI increases the likelihood of not only identifying formulation or manufacturing problems but also explaining their cause.
- *Counterfeit medicine.* It is estimated that counterfeit medicine make up more than 10 % of the world's medicine market.[9] This is likely to increase with increasing trade of medicine via the Internet. Counterfeit medicine is a threat to the health of patients and the pharmaceutical industry's reputation and sales. NIR-CI is a rapid method for detecting and comparing counterfeit products without any sample preparation.

The aim of this thesis is to demonstrate how NIR chemical imaging can be used in formulation development of solid dosage forms, primarily tablets.

Through application of NIR-CI in a formulation development experiment it is shown how the unique information obtainable by NIR-CI can lead to enhanced process and product understanding. In addition to a thorough NIR-CI literature review on pharmaceutical applications this should provide inspiration to other relevant formulation development studies where NIR-CI can add value and increase understanding.

However, in order to fully exploit the potential of NIR-CI it is essential to have a thorough understanding of how to properly measure, analyse and interpret NIR-CI data and to understand the limitations of this technology. The NIR-CI technology is still maturing and there is a need to evaluate the data processing methods for their ability to generate accurate chemical images of the pharmaceutical samples. Development and evaluation of data processing methods and extracting essential information from the generated chemical images were also an important objective of this thesis work.

This thesis can be read as an introduction or guidance to near-infrared chemical imaging and its application to pharmaceutical solid dosage forms. The focus is on formulation development of tablets and the concept of this work is presented in **Chapter 2**. Although all experimental work has been NIR-CI studies of tablets, the presented data analysis strategies are applicable to other chemical imaging technologies and can also be used to analyse other solid dosage forms than tablets (powders, capsules, microspheres, extrudates, granules etc.). Near-infrared (NIR) spectroscopy is the basis of the NIR-CI technology and a short introduction to theory on NIR spectroscopy and the diffuse reflectance measuring mode is given in **Chapter 3**. The principles and instrumentation of NIR-CI is presented in **Chapter 4**. This chapter also includes sections describing other chemical imaging technologies (MIR-CI, Raman-CI), the role of NIR-CI in the PAT concept, and a thorough literature review of NIR-CI used in pharmaceutical applications.

A full NIR-CI experiment can be divided into data acquisition, data processing and image processing. Each of these three overall analytical steps is described in **Chapter 5**, **6** and **7**, respectively. The data processing described in **Chapter 6** is a major chapter presenting several approaches on how to analyse hyperspectral raw data obtained from NIR-CI data acquisition. The described data processing approaches will provide a toolbox for analysing most NIR-CI data of pharmaceutical solid dosage forms. **Chapter 6** also contains a stepwise guidance on how to analyse NIR-CI data and the chapter is completed with a comparison and summation of the characteristics of the presented methods. The

data processing chapter can thus be used as a guidance or tutorial on how to analyse hyperspectral NIR data but will also be applicable to other chemical imaging techniques (Raman, MIR). **Chapter 7** on image processing describes developed methods to extract useful information from the generated chemical images. The application of NIR-CI in a development study of a tablet formulation produced by a direct compression process is presented in **Chapter 8**. In this chapter there are also other examples of NIR-CI applied to pharmaceutical tablets, which contributes to understanding the limitations and added value of using the NIR-CI technology. Finally, concluding remarks and suggestions for future research are given in **Chapter 9**.

During the project two publications (**Paper I** and **II**) have been produced. Paper I is accepted and published. Paper II is in a final manuscript form ready for submission. Further, two publications of closely related work are enclosed (**Paper III** and **IV**). The four papers are enclosed as Appendix I-IV and will be referred to in the text by the Roman numerals (I-IV). Some of the results from these papers will be integrated in the following chapters of the thesis.

1.3 Terminology

The combination of vibrational spectroscopy with a microscope to obtain spatially resolved spectra was already introduced back in 1949.[10] It then lasted almost five decades before sufficient advancement in computer power, detector elements and optics made the technology commercially available. The technology and the application knowledge are still maturing and the terminology for this multi-disciplinary technology is not always clear and consistent. It is therefore important to clarify the terminology used in this thesis to avoid any misconceptions.

The overall term for combining vibrational spectroscopy (IR, NIR or Raman) with digital imaging is denoted chemical imaging (CI) in this thesis. Specifically when using near-infrared spectroscopy the term *near-infrared chemical imaging (NIR-CI)* is used, which is the only chemical imaging technique used for experimental studies in this thesis work. Other terms (synonyms) often met in the literature for this technology are: spectral or spectroscopic imaging, vibrational chemical imaging, chemical mapping and imaging, hyperspectral imaging, vibrational microscopy, vibrational microspectroscopy or microspectroscopy.

The full analytical work of a NIR-CI experiment in this thesis is divided into three overall steps defined as:

- **Data acquisition.** Includes sample preparation, instrumental settings and basic spectral transformation. The raw data output from a NIR-CI measurement is organised in a three-dimensional data structure with two spatial axes and one wavelength axis, also called a hyperspectral data cube.
- **Data processing.** The processing of the hyperspectral data cube into a, typically chemical, image by univariate or multivariate image analysis methods. This part includes wavelength selection, spectral pre-processing and the subsequent data analy-

sis to generate images visualising the distribution of each of the components within the imaged sample.

- **Image processing.** The processing of the generated chemical images into relevant and 'useful' information that will qualitatively or quantitatively describe the properties of a sample in relation to the problem investigated. This extracted information could e.g. be a total concentration, a measure of the distribution of the concentration of the active ingredient or the domain sizes of components within the sample.

The term *hyperspectral data cube, hyperspectral image* and *hypercube* are used interchangeably and denote the raw data output from the data acquisition. It is a three dimensional data structure containing two spatial axes and one wavenumber axis (Section 1.1 and 4.1).

A *chemical image* is a visual representation of the spatial distribution of a chemical component in the imaged sample. It is generated from the mathematical processing of a hyperspectral data cube. However, it is important to note that '*chemical image'* does not cover the representation of images generated from all processing's of hyperspectral data cubes, as for example, principal component analysis (PCA) will generate score images which should be denoted as such.

A *component* relates to a chemical component or compound, e.g. the active pharmaceutical ingredient (API) or any excipients used for a tablet composition. The term *component* is primarily used but *compound*, *ingredient* or *analyte* will have the same meaning. It should not be confused with the term *principal component* from multivariate data analysis. The *principal component* (PC) will be specified as such when used.

There are two types of resolution in chemical imaging. The *spectral resolution* refers to the number of wavelength data points in a measured spectrum. The *spatial resolution* refers to the pixel size.

CHAPTER 2

FORMULATION DEVELOPMENT OF TABLETS

2.1 Introduction

Medicine is seldom administered as the pure active pharmaceutical ingredient (API) alone. It is primarily given in a mixture with appropriate additives (excipients) to provide a suitable dosage form for administration, also called the formulated dosage form, drug product or just the formulation. The design of the formulation and the process that results in the most desirable dosage form is the central concept of formulation development. The type and function of excipients used in formulations are varied and each plays an important and specific role for the quality and functionality of the final dosage form.

> *Tablet formulation and design:*
>
> Tablet formulation and design may be described as the process whereby the formulator ensures that the correct *amount* of drug in the right *form* is delivered at or over the proper *time* at the proper *rate* and in the desired *location*, while having its chemical *integrity* protected to that point.[11]

Pharmaceutical development is the term that encompasses all aspects in developing a new drug product for commercialisation, i.e. from drug substance development to selecting packaging material for the drug product. The aim is to design a quality product and its manufacturing process to consistently deliver the intended performance of the product. Formulation development specifically concerns drug product and process development, which generally means designing the dosage form composition and manufacturing process. The International Conference on Harmonization (ICH) lays down the regulatory foundation for this work in the guideline document ”Pharmaceutical Development Q8”.[4]

Formulation development is an essential discipline in a large part of the development phases of a new medicine. It is used in pre-clinical studies for dosing animals, early safety studies, clinical trials, and eventually for developing the commercial formulation. Formulation development is crucial for the quality of the new medicine as the intended therapeutic performance may be compromised without a suitable designed and robust formulation.

2.2 Solid dosage forms - Tablets

Solid dosage forms for oral use are by far the most prevalent pharmaceutical formulation type on the worldwide pharmaceutical market comprising around 60 %.[12] Most solid dosage forms are produced as tablets. A tablet is a compressed unit of a powder mixture of the active pharmaceutical ingredient (API) and a number of excipients. Extracts from the monograph for tablets in the European Pharmacopoeia (6^{th} Edition 2008)[12] define tablets in the following way:

- ✧ Tablets are solid preparations each containing a single dose of one or more active substances. They are obtained by compressing uniform volumes of particles or by another suitable manufacturing technique….
- ✧ Tablets are intended for oral administration.
- ✧ The particles consist of one or more active substances with or without excipients such as diluents, binders, disintegrating agents, glidants, lubricants, substances capable of modifying the behaviour of the preparation in the digestive tract, colouring matter authorised by the competent authority and flavouring substances.
- ✧ Tablets are usually straight, circular solid cylinders, the end surfaces of which are flat or convex and the edges of which may be bevelled. They may have break-marks and may bear a symbol or other markings. Tablets may be coated.

The popularity of tablets as dosage form is related to some of its major advantages compared to other dosage forms:

- High accuracy in correct dosing of API
- Better drug stability
- Easier and economically beneficial manufacturing process
- Easy packaging process and distribution to customers
- Convenient dosage form to handle and administer for the users

The two most common processes for manufacturing tablets are by *direct compression* or *wet-granulation*. The most simple and often preferred process for manufacturing tablets is simply mixing the API with excipients and compressing the powder mixture directly into tablets (direct compression). Tablets by direct compression are the only formulation type and process investigated throughout this thesis work. However, the application of near-infrared chemical imaging (NIR-CI) demonstrated on tablets by direct compression can also be applied to other types of solid dosage forms using other manufacturing processes.[3,5,14-20] For example, the use of NIR-CI to characterise the intermediate and final product in a wet granulation process is apparent. Wet granulation is the process where primary powder particles are made to adhere to form larger multi-particle aggregates called granules in which the original particles can still be identified.[21] The granules are produced by adding an aqueous binder-solution to the powder mixture through mechanical agitation in granulator equipment. Prior to the wet granulation process dry mixing of API and different excipients are often performed to achieve a uniform distribution of each ingredient throughout the mix and later in the final product. After the granulation process the granules are typically dried, sieved, mixed with other excipients and finally compressed into tablets. There are various reasons for selecting a wet-granulation process, for example to provide better powder-flow properties, improve compaction characteristics, prevent segregation effects of formulation ingredients, better control of drug content uniformity at low drug concentrations, reduce powder volume and dust from toxic pow-

ders etc.[21,22] It is important to obtain granules of well-defined size and with a uniform and consistent API content. NIR-CI is an obvious and valuable technology to characterise both single granules (e.g. after sectioning) for their API distribution but also for characterising the granule distribution in the final tablets and thus homogeneity of the product.

2.3 Preformulation

Prior to the development work of the formulation, the physical and chemical properties of the active pharmaceutical ingredient (API) and other derived properties of the API powder are determined in *preformulation*. This characterisation of the API is essential for the formulation development work as it determines many of the experiments and approaches to be performed. Some of the most important preformulation characterisations of the API for the formulation development work are:

- Solubility
- Density
- Polymorphism, solvates/hydrates
- Hygroscopicity
- Particle size (distribution) and particle morphology
- Compatibility with excipients
- Powder flow properties
- Compression properties

2.4 Formulation development of tablets

With the knowledge obtained from the preformulation studies the aim of (tablet) formulation development is now to design and develop a robust process and product composition that consistently will produce a final product with the desired functional characteristics that fulfils the intended therapeutic performance. The essential aspects of formulation development and manufacturing of tablets are illustrated in Figure 3 (based on direct compression).

The goal is as it appears to manufacture robust tablets that consistently fulfil a number of quality attributes. After oral administration the tablets must disintegrate and release the API to make it available for absorption into the systemic blood circulation. The general quality attributes that should be fulfilled for a tablet formulation are listed below together with the European Pharmacopoeia tests (in brackets) applicable to assess the quality attributes:

- Tablets must be mechanical resistant to withstand attrition or breakage from handling during and after manufacturing.
  (Resistance to crushing of tablets[23]; Friability of uncoated tablets[24]).
- The tablets must have a consistent and uniform mass and content of API.
  (Uniformity of dosage units[25]).

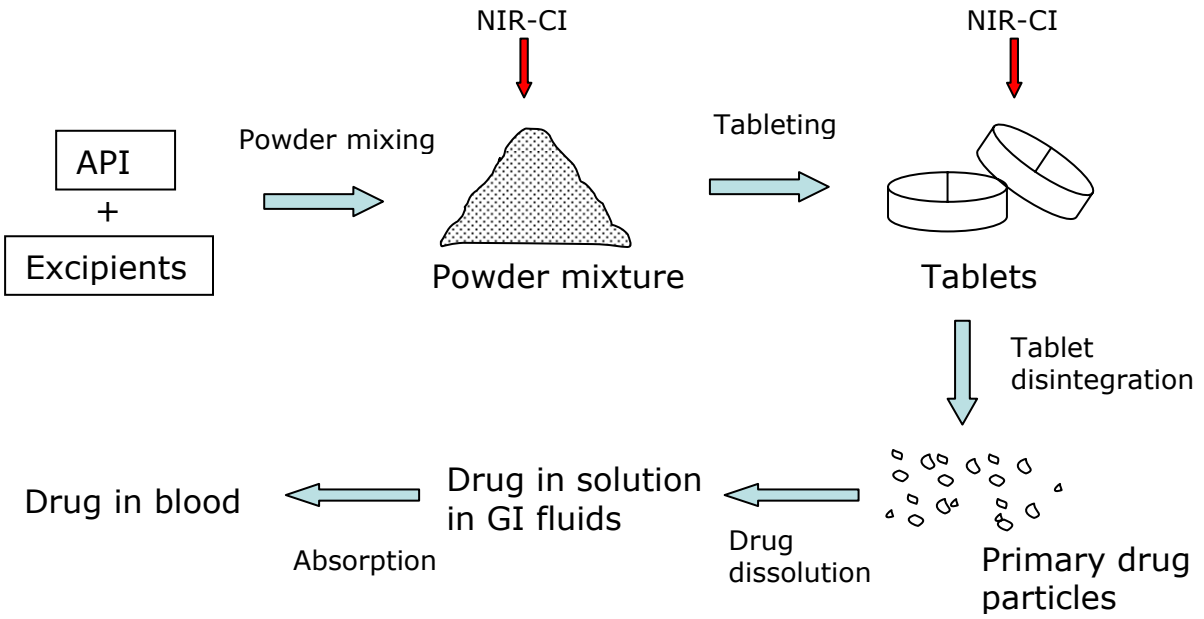


Figure 3 - The direct compression process for tablet manufacturing and fate of the tablet after oral intake. Possible application points for NIR chemical imaging are indicated. GI = Gastro-Intestinal.

- The API should be released from the tablet in a controlled and reproducible way. The dissolution test does not assure satisfactory bioavailability but is the available approach to evaluate this.
  (Disintegration of tablets and capsules[26]; Dissolution test for solid dosage forms[27]).
- The tablet should be chemically, physically and microbiologically stable during the lifetime of the product.
- The appearance (size, shape, colour, and logo) of the tablet should be consistent.

Compliance with all these and other case-by-case additional quality attributes can only be obtained with a carefully designed formulation and manufacturing process. This is the goal of formulation development work and experimental formulation studies are performed to select the appropriate dosage form and optimal manufacturing process, composition, processing steps and parameter settings. The selection of excipients and their concentration in the formulation is an important part of this work. Their role is to ensure that the manufacturing process runs satisfactorily and tablets of the specified quality are obtained. The functions of excipients are diverse and different types such as fillers/diluents, disintegrants, binders, glidants, lubricants etc. are used. The important role of excipients can e.g. be illustrated for the lubricant magnesium stearate. Due to its hydrophobic nature it may retard the tablet disintegration and dissolution if it is present at a too high concentration or mixed too well with the other ingredients.

In the development work of a new formulation studies are performed to identify the critical quality attributes (CQA), the critical process parameters (CPP) and determine the optimal process parameters for manufacturing the formulation. The types of formulation

experiments to perform are highly depending on the formulation type, manufacturing procedure and the unit process steps. It is therefore not possible to make a general comprehensive list of experiments to perform. However, below follows a list of some of the typical experiments that could be performed in developing a tablet formulation by direct compression:

- Selecting the optimal excipients and composition
    - Compatibility studies
    - Screening with different types, quantities and ratios
    - Evaluate compactability, flowability, lubrication effect etc.
- Sieving studies to determine optimal particle sizes
- Mixing studies:
    - Mixing principle
    - Mixing effect (rotation speed and time)
    - Mixing order of ingredients
- Compression force and speed
- Stability studies

Many different product and process parameters are typically involved and varied in each experiment; hence design of experiments (DoE) is often employed in these formulation studies. Depending on the type and objective of each formulation study, powder samples and/or tablets are collected and analysed for the different quality attributes described earlier. Existing analytical techniques such as high performance liquid chromatography (HPLC), ultraviolet spectroscopy (UV), mass spectrometry (MS), and the specific quality tests such as dissolution, resistance to crushing, disintegration etc. are used to determine identity, strength, purity and quality of the tablets. These tests are all destructive, time consuming and only provide bulk information about the sample and are primarily focused on analysing the API content. For example, the content uniformity testing will provide information on the variation in content between a number of dosage forms – but only for the API. It provides no information about the size or distribution of API or the excipients within each tablet. Dissolution testing examines the manner and duration of API release from the tablet formulation but provides no insight into the cause of the profiles obtained. NIR spectroscopy is a non-destructive method that is becoming a traditional analytical technique. However, this method can also only provide the bulk content relationship of components but provides no information on their spatial distribution in the sample.

Investigating the homogeneity in pharmaceutical powder blends or tablets is always a main concern in formulation development and is mainly related to a uniform distribution of the active pharmaceutical ingredient (API) throughout the blend/tablets. There is no definitive definition of homogeneity of a powder blend or tablet batch. Homogeneity is indirectly determined or assessed by calculating an acceptance value based on the mean and/or standard deviation from a number of samples and comparing the results with predetermined acceptance criteria. The pharmacopoeia's testing performed to evaluate the

homogeneity of the final dosage form is covered by "Uniformity of dosage units".[25,28] The content of API in 10 or 30 individual dosage units are analysed and the overall batch homogeneity is assessed acceptable when a calculated acceptance value meet the requirements. As the content uniformity of solid dosage forms is closely related to the uniformity of the powder blend, it is also important to determine blend uniformity. This is the basis for the FDA Guidance for Industry document describing how sampling and assessment of uniformity in powder blends and finished dosage units can be conducted to comply with good manufacturing practice requirements.[29] However, it should be noted that the powder blend uniformity may be affected by the subsequent steps before the final tablets are obtained, hence homogeneity is best determined as late in the manufacturing process as possible, i.e. determining content uniformity of the final dosage form.

It should be evident by now that a tablet formulation is a carefully designed matrix and the functional performance may vary depending on how the components are distributed within the final tablet. The structure of the tablet matrix that evolves during the manufacturing process is thus an important quality factor. However, when it is time to assess the quality of the tablets matrix it is destroyed by e.g. dissolving the sample in a solvent. All the insight into the internal structure and spatial distribution of components relative to each other in the final product matrix is lost. Information that is invaluable and can elucidate intra-tablet variations helping the formulator to see how the process actually performed. While the traditional analytical techniques cannot provide this important spatial information, it is possible by chemical imaging techniques such as e.g. near-infrared chemical imaging (NIR-CI).

Moreover, there are trends in formulation development science where traditional analytical techniques may fall short on fully characterising the product quality and NIR-CI could be useful:

- More potent API's, i.e. it is crucial and a challenge to obtain and maintain content uniformity as the content per tablet decreases with more potent API's.

- The increasing number of poorly soluble drug substances being developed or desire for more complex drug release profiles call for the development of more advanced tablet formulations. Examples are layered tablets, barrier layers, encapsulated drug molecules, microspheres of selective drug release, osmotic pumps etc. Common for these highly engineered drug delivery systems is that the physical structure is as critical as the chemical composition.[2]

CHAPTER 3

NEAR-INFRARED SPECTROSCOPY AND DIFFUSE REFLECTANCE MODE

3.1 NIR spectroscopy

NIR chemical imaging builds upon the well established near-infrared (NIR) spectroscopy technology. The illumination in the NIR-CI instrumentation used for the data acquisition is provided by near-infrared radiation. The near-infrared region is located adjacent to the mid-infrared (MIR) region starting at 4000 cm^{-1} and extends up to the visible region at 12800 cm^{-1} (wavelength range 2500 nm to 780 nm). Near-infrared belongs together with mid-infrared (MIR, 4000-400 cm^{-1}) and Raman (4000-50 cm^{-1}) to the vibrational spectroscopy techniques that utilises absorption bands in the spectra emerging as a consequence of molecular vibrations in the chemical compounds. The location of the NIR and MIR regions in the electromagnetic spectrum is shown in Figure 4.

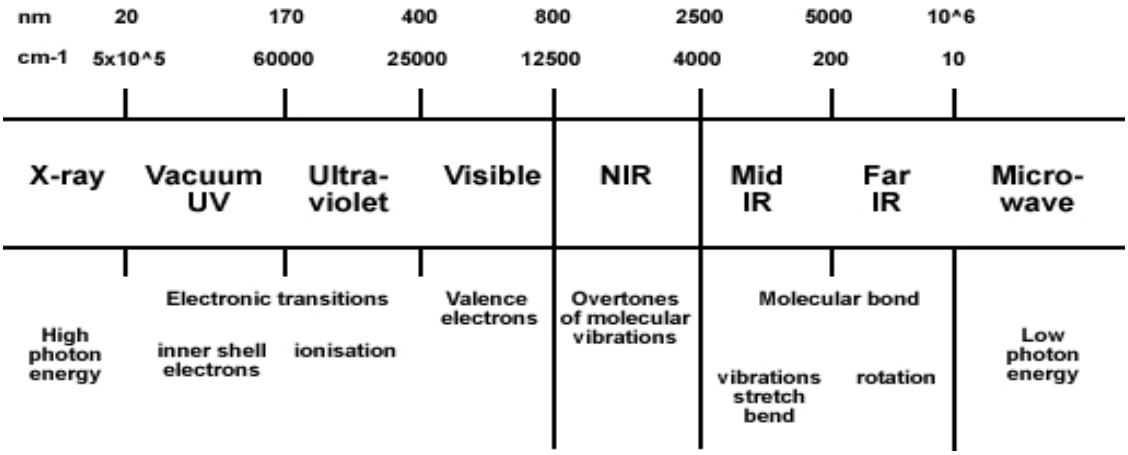


Figure 4 - Electromagnetic spectrum.

The absorptions in the NIR region are primarily due to overtones and combinations of the fundamental molecular vibrations from the mid-infrared region between X-H bonds (X = O,N,C,S). The absorption peaks are 10-100 times weaker than their corresponding fundamental infrared absorption bands and characterised by very broad and overlapped bands. Unlike IR spectroscopy, absorption bands in NIR spectroscopy are not well suited for identification purposes and the NIR spectra for compounds of chemical resemblance may be very similar in structure and difficult to resolve. However, the development of chemometrics and increased computer power has made it possible to interpret and extract the unique information from the NIR spectra.

3.2 Diffuse reflectance spectroscopy

Depending on the sample properties NIR spectra can be obtained in three different modes: transmission, diffuse reflection or transflection. Only diffuse reflectance measurements on solid samples have been conducted throughout this project and only this mode will therefore shortly be introduced further here.

In diffuse reflectance spectroscopy, the illumination source and detector are located on the same side of the sample. It measures the amount of light reflected from the sample surface, which contains a specular component and a diffuse component.[6] The specular (speculum (lat.) = mirror) component is the light that is just reflected from the sample surface rather than absorbed and contains little or no information about the sample. The diffuse component is the light reflected after interaction with the sample. The diffuse reflected light carries information about the sample materials at various depths and locations in the sample. It is thus the diffuse component that contains information on the composition that is of interest. The contribution from the specular component is minimised by appropriate position of the detectors relative to the sample and by preprocessing the collected spectra.[6] In practice, relative reflectance (R), i.e. ratio of the intensity of light reflected by a sample and a standard ($R = R_{sample}/R_{standard}$), is measured.

Over the last two decades, NIR spectroscopy has occupied an increasing role in pharmaceutical analysis especially along with the PAT concept. NIR spectroscopy is fast, non-destructive, flexible, non-invasive, requires little or no sample preparation, and the technique provides a lot of information about the measured sample. It is widely used for raw material characterisation, process monitoring, and product quality control. Numerous publications on the fundamental NIR spectroscopy theory and its application to pharmaceutical samples can be found in the literature, and good overviews and reviews have also been published.[3,6-8,30-32]

Most of the beneficial characteristics and processing techniques of conventional NIR spectroscopy are also valid when it is utilised for NIR chemical imaging.

CHAPTER 4

PRINCIPLES OF NEAR-INFRARED CHEMICAL IMAGING AND INSTRUMENTATION

4.1 Principles of NIR chemical imaging

In conventional NIR spectroscopy a bulk NIR spectrum is measured that reflects an average composition of the measured sample. NIR-CI adds spatial information to the spectral information by combining conventional NIR spectroscopy with digital imaging. In NIR-CI the spectral and spatial information is recorded simultaneously. The result is creation of a three dimensional spectral image, also termed a *hyperspectral data cube* or *hyperspectral image*, where two dimensions provide the spatial information (x and y) and the third dimension provides spectral information (z) (Figure 5).

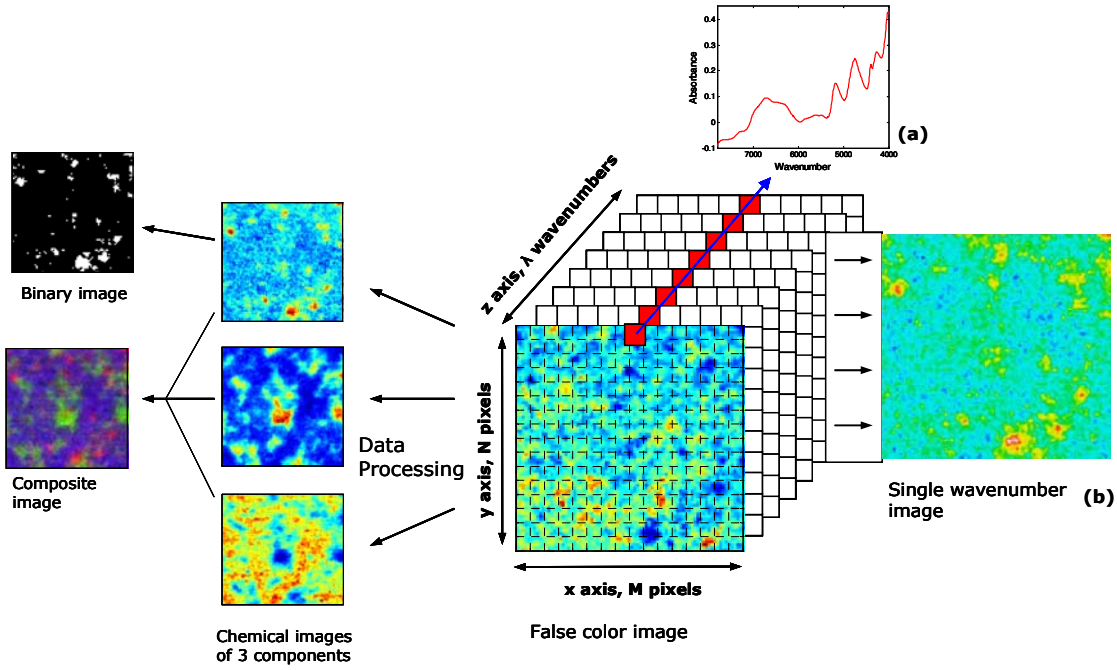


Figure 5 - Three-dimensional hyperspectral data cube ($M \times N \times \lambda$) shown as false color image and the information that may be obtained directly from the data cube (single pixel spectrum (a) and single wavenumber image (b)) and through data processing (chemical, binary and composite images, all left).

The hyperspectral data cube can either be viewed as an array of spectra (one for each pixel) from which images can be created or as a row of images from which spectra can be extracted. Selecting a single pixel (xy-coordinate) through the z-plane will show the spectrum recorded at this particular spatial location, which provides the spectral signature of chemical components present in that exact part of the sample (Figure 5a). Selecting an image plane (xy-plane) at a specific wavenumber (z-value) will show the intensity

values for all pixels at that wavenumber also called a single wavenumber image. The single wavenumber image contains spatial distribution information about a chemical component with a distinct spectral characteristic at that particular wavenumber (Figure 5b). This is the simplest image generation algorithm.

The hyperspectral data cube is the raw data outcome from a NIR-CI data acquisition. It is often visualised as an image using either a grey scale or colour scale to represent intensity, for example the average absorbance intensity value for each pixel (spectrum). Such false colour image will not in itself reveal anything about the chemistry in the sample. The NIR spectra can be regarded as chemical fingerprints and the goal is to translate the spectral signature from each pixel into chemical information. There are a number of univariate and multivariate analytical tools available for this purpose and some of these approaches are dealt with in more details in Chapter 6. The result from the data processing of a hyperspectral data cube is a set of, typically chemical, images providing information on the composition, size of clusters, and distribution of each component within the sample matrix.

The unique feature of NIR-CI is that in addition to identification and quantification of the chemical components, their size and spatial distribution within the sample is also revealed. NIR-CI will also measure samples non-destructively so they are preserved for further testing.

4.2 NIR-CI instrumentation

The instrumentation for acquisition of hyperspectral NIR data is best described as the coupling of a microscope with a NIR spectrometer. There are two main types of NIR-CI systems to obtain the raw data hyperspectral data cube:

1) Line mapping or pushbroom system;

2) Global imaging or staring imager system.

Only the line mapping system has been used in the experimental studies in this thesis but as an introduction to the NIR-CI technology and to get acquainted with the principles reported in literature, general knowledge about both principles are needed and thus introduced below.

Using the term *chemical imaging* for a line mapping system has been found inappropriate as only global imaging and not line mapping is considered a real imaging technique.[16] Despite the differences in data acquisition the output result for both techniques is a *hyperspectral data cube* and the same *chemical image* will be obtained for the same sample using the same instrument settings and data processing method. Therefore, no distinction is made between line mapping and global imaging in this thesis – both are classified as chemical imaging systems.

LINE MAPPING

The line mapping principle is based on a step-and-acquire acquisition mode. The sample area and pixel size are initially defined and this determines the number of pixels in the image. For each acquisition diffuse reflectance spectra are collected from a line of, typical

16, adjacent points (pixels). With a very accurate positioning sample stage the sample is moved and another line of spectra are obtained. In this way a grid of spectral information is build up from the lines of spectra until spectra from all pixels in the defined sample area are obtained to constitute the hyperspectral data cube.

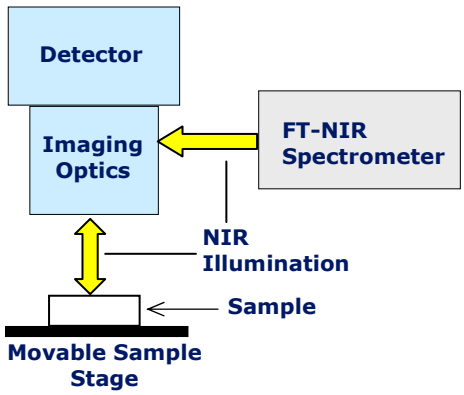


Figure 6 - Schematic presentation of line mapping instrumentation.

A schematic presentation of the line mapping instrumentation is shown in Figure 6. The sample is positioned on a motorised moving sample stage. A Fourier Transform (FT) NIR spectrometer is employed as wavelength filter and provides the light source (tungsten lamp). The radiation is focused onto the sample surface using a Cassegrain objective lens and is after interaction with the sample focused through a set of apertures onto the line of detectors.

GLOBAL IMAGING

The global imaging system contains - unlike line mapping - no moving parts and generates images at each selected wavenumber rather than collecting NIR spectra directly from each pixel. Global imaging measures the NIR absorption intensity values in each pixel of the defined sample area at one particular wavelength at a time. The imaging technique uses two-dimensional focal plane arrays (FPA), which are cameras composed of many thousands individual detector elements (pixels). The number of pixels in an image is thus fixed for global imaging systems and the pixel size and hence sample area are defined by the magnification optics (e.g. 40 µm/pixel objective leading to a 12.8 × 10.2 mm image using a 320×256 pixels FPA). The third dimension of 'spectral' information is obtained by changing the wavelength of light using tuneable filters, e.g. liquid crystal or acoustic optical tuneable filters. The NIR absorption intensity values are measured in every pixel of the FPA at each individual wavelength, which is sequentially changed by the tuneable filter. Images at each individual wavelength are in this way built up like a deck of cards to generate the hyperspectral data cube.

A schematic presentation of the global imaging instrumentation is presented in Figure 7. The NIR light is not focused but rather defocused onto the sample to cover a larger area. After interaction with the sample the reflected light is collected through imaging optics and passed through the tuneable filter before an image of the sample at the selected wavelength is recorded on the focal plane array detector.

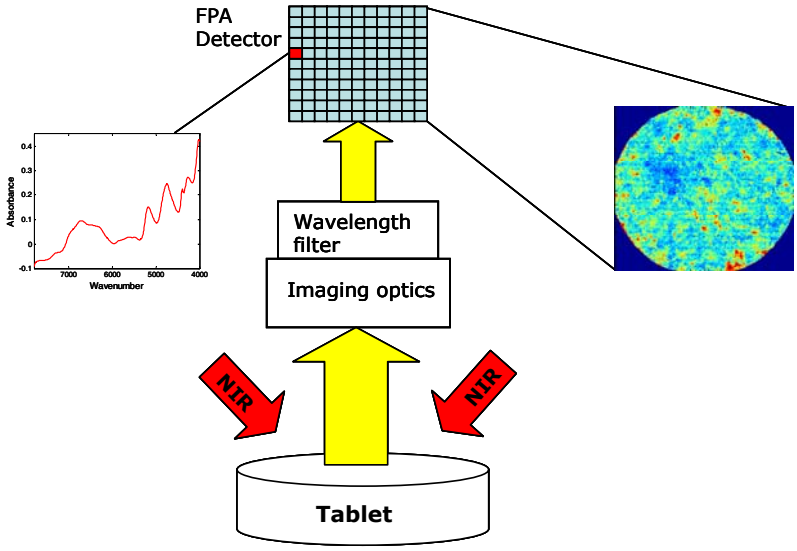


Figure 7 – Schematic presentation of a global imaging system.

4.3 NIR line mapping versus global imaging

Irrespective of whether a line mapping or global imaging system is used for a NIR-CI experiment the output results are the same, i.e. obtaining a hyperspectral data cube. There are nonetheless some important differences between the two systems that may decide whether a NIR-CI line mapping system or global imaging system should be purchased or alternatively used (should one be so lucky to have both types of instruments available).

The line mapping system is based on a Fourier-Transform (FT) NIR spectrometer which is known to provide high wavelength accuracy. The spectral resolution is generally better for FT-NIR based mapping techniques providing higher quality of the spectral data. Mapping systems typically have spectral resolution down to 1-4 cm^{-1} compared to 8-10 cm^{-1} for global imaging systems.[16] The sampling scheme utilised in line mapping illuminates the line of pixels directly in a staggered pattern, which reduces the inter-pixel mixing effects.[33] For line mapping there is also no correction of 'dark' detector response as for global imaging systems.[15] However, line mapping systems are not so flexible for irregular samples geometries and large sample areas, but do have the option to acquire data from several areas and quilt them together to form a larger image.

The major advantages of global imaging are the larger field of view, speed of data acquisition and ruggedness. By changing the magnification objectives much larger areas can readily be analysed by global imaging compared to line mapping, and global imaging has for example been used to analyse whole blister packs[15] and for high-throughput analysis of multiple tablets simultaneously.[35] Due to the principle of obtaining images at selected wavelengths and using fast tuneable filters, data acquisition time for global imaging is not so dependent on the area analysed as in line mapping. The tuneable filters can also be used to select a subset of available wavelengths and thereby only analysing absorp-

tion bands of useful information. This will make the data acquisition time even shorter. However, narrowing the spectral range and thereby reducing the acquisition time is also possible for FT-NIR based line mapping instruments. In general, data acquisition time for global imaging systems is a few minutes compared to 10 to 40 minutes for line mapping systems depending on the selected area, pixel size, spectral resolution and number of scans per wavenumber.

The illumination source in a global imaging system is usually provided by four quartz-halogen lamps. Some users of global imaging have experienced significant heat generation from the lamps and this can potentially be a problem for heat-labile compounds or compounds with a low melting point as e.g. for some polymers (unpublished observations).

The line mapping uses a moving sample stage to move the sample during data acquisition and raster scan the defined sample area. Global imaging systems have no moving parts. This makes global imaging more robust for manufacturing environments and thus at-line or on-line PAT applications.

The cost is often an important factor when purchasing analytical equipment. The global imaging system tends to be the more expensive of the NIR chemical imaging systems, also considering that different magnification objectives are sold separately at additional prices. Should both instrumentation types be available the choice to use will probably depend on the type of application. The line mapping may be most suitable for initial interrogation of a new or unknown formulation, for research and problem-solving experiments or in other cases where most spectral information from the sample is needed. The global imaging system is preferable when larger areas or multiple samples are analysed, when fast data acquisition is a requirement, or when the analytical environment requires the most rugged and robust NIR-CI instrument.

4.4 Chemical imaging technologies (NIR-, MIR- and Raman-CI)

NIR chemical imaging applied to pharmaceutical tablets is the main theme of this thesis and NIR-CI is the only chemical imaging technique used in the experimental work. However, the possibility of utilising other chemical imaging technologies (MIR and Raman) in pharmaceutical applications also exists. A short comparison between MIR-, NIR- and Raman chemical imaging for pharmaceutical applications will shortly be presented here.

The nature of the analysed sample is important to consider when deciding on the most suitable chemical imaging technologies for a given analytical application. Most ingredients for pharmaceutical tablets (API and excipients, Chapter 2) are white to off-white powders, which is why traditional microscopic analysis of a compressed tablet is unsuitable.

Most pharmaceutical API's are organic compounds appearing as crystalline powder. The API is often a salt of a specified polymorphic form that need to be distinguished from other related API forms (free acid or base, other salt or polymorphic forms, hydrate or solvate).[16] NIR spectroscopy involves the overtone and combination absorptions of the fundamental MIR absorptions from X-H bonds (Chapter 3), however in MIR spectroscopy the most intense absorption will come from polar bonds (C-O, C=O, Si-O etc.). Raman spectroscopy is a spectrally rich scattering technique and is often better for identification

of different API polymorphic forms as it exhibit strong spectral differences between these.[16] Furthermore, Raman spectroscopy is complementary to MIR and NIR as it is better at detecting non-polar symmetrical bonds (e.g. C=C, C-C, S-S).

The pharmaceutical excipients are varied in nature ranging from hydrophilic to hydrophobic, crystalline to amorphous and organic to inorganic chemical compounds. A large part of the frequently used excipients are carbohydrates (sugars, celluloses and starches). They are characterised by their many X-H bonds and are quite similar in chemical structure although they exhibit widely different functions in a tablet. For this purpose, NIR-CI has an advantage, as both MIR and Raman do not discriminate mixed carbohydrates as well as NIR.[36-37] For inorganic excipients that have no NIR spectrum, e.g. the commonly used calcium phosphate, Raman is superior. An ideal solution for analysing a pharmaceutical matrix composed of organic and inorganic compounds (and maybe also non-polar chemical bonds) would therefore be to analyse the sample by both NIR and Raman chemical imaging. The combination of these two techniques to analyse a formulation containing some of the different types of chemical compounds mentioned has been successfully demonstrated.[36]

To illustrate the spectral appearance of these different excipients, the diffuse reflectance NIR spectra from a NIR-CI analysis of several commonly used pharmaceutical excipients and also of an API compound are displayed (Figure 8). The spectra are the mean spectrum of the hyperspectral data cube obtained from NIR-CI analysis of each pure component material.

A common feature for the MIR, NIR and Raman chemical imaging techniques are that they are surface analysis techniques, i.e. based on radiation with relatively limited penetration depths into the sample.

Due to strong absorption and virtually no combination bands in the MIR spectral range this offers higher specificity and sensitivity compared to NIR. The consequence of this for MIR-CI applications is however, that some kind of sample preparation (sample dilution) to limit the amount of sample material that interacts with the light is required. To avoid such sample dilution MIR-CI is usually employed utilising the attenuated total reflectance (ATR) technique to analyze tablet surfaces.[15,19,34] Close contact of the sample with the surface of the ATR crystal used for this technique is required. The close contact with the crystal surface is best obtained with liquids and extremely flat or deformable sample surfaces. When analysing solid dosage forms, pressure needs to be applied, which may damage the sample or the crystal.[34]

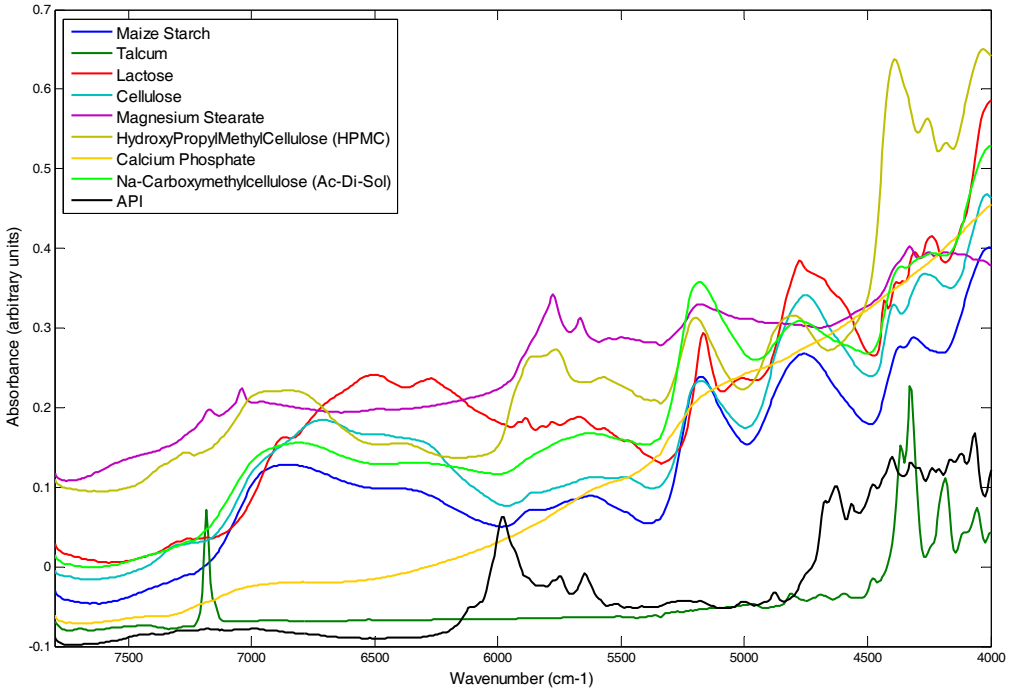


Figure 8 - NIR diffuse reflectance spectra from a NIR-CI analysis of common pharmaceutical tablet excipients and one API compound. The spectra are the mean spectrum of the hyperspectral data cube obtained from NIR-CI analysis of each pure component material.

Compared to MIR-CI both NIR and Raman chemical imaging work well in diffuse reflection and are suitable for surface analysis of opaque or thick materials such as tablets. NIR and Raman chemical imaging requires little or no sample preparation. Raman chemical imaging provides the best spatial resolution being as low as 1-2 µm (typical 2-5 µm) compared to about 5 µm (typical 6-25 µm) for NIR-CI.[16,36-37] The disadvantage for Raman-CI is longer data acquisition time and limitation of the sample area that can be measured. Due to the focusing nature of the laser beam used in Raman, in particular at high resolution, sample heating/burning can also be a problem.[34,38] And finally, as for all Raman based techniques fluorescence may interfere and overwhelm the Raman signal.

So, primarily due to speed of data acquisition, ease of operation, ability to most effectively map/discriminate pharmaceutical components, and instrument price, NIR-CI was the chemical imaging technique purchased and used for exploring the pharmaceutical tablets in this project. The challenge, as for all NIR techniques, lies in the interpretation and chemical information extraction from the broad and overlapped absorption bands but multivariate data analysis will facilitate in mitigating this (Chapter 6).

Finally, it should also be mentioned that new imaging technologies are emerging that may complement or compete with NIR chemical imaging. Some of these are terahertz imaging,[39] time-of-flight secondary-ion mass spectrometry (TOF-SIMS),[40] NMR-imaging,[41] and different tomographic imaging techniques.[42]

4.5 Measurement variables in NIR-CI data acquisition

The principles of NIR diffuse reflectance in NIR-CI measurements were described in Chapter 3, the principles of data collection and instrumentations were discussed above, and the aspects of data acquisition are presented in the next chapter. In relation to these aspects it is important to consider what the NIR-CI instrument is actually measuring taking both the sample and measurement characteristics into account. This section will present and discuss some of the measurement variables and their impacts that are important to be aware of. Although there may be no easy solution or final answers to these measurement variables, it is important to be familiar with them for conducting a proper NIR-CI experiment. For example, in order to take care and be critical in the interpretation of NIR-CI results.

When the NIR radiation - appearing as millions of photons – reaches the sample surface, it may be immediate reflected (specular reflectance, Chapter 3) or enter the sample. Some of the photons entering the sample will be lost completely by absorption in the sample material. Other photons will pass through the entire sample or be scattered/reflected to a degree that will not be captured and detected by the instrument. The photons of interest will penetrate into the sample and be partially absorbed by every sample material/particle it interacts with. These photons are reflected to various degrees and may eventually return to the sample surface from where they can be detected. The spectral signature of this diffuse reflected NIR radiation will contain information about the sample components at various depths and locations in the sample. The light penetration into the sample and fate of the reflected light will depend on various measurement variables such as particle size, particle shape, NIR radiation wavelength, and the optical and chemical properties of sample components.[38,43-45]

A model pharmaceutical tablet is schematically shown to illustrate some of these measurement variables (Figure 9).

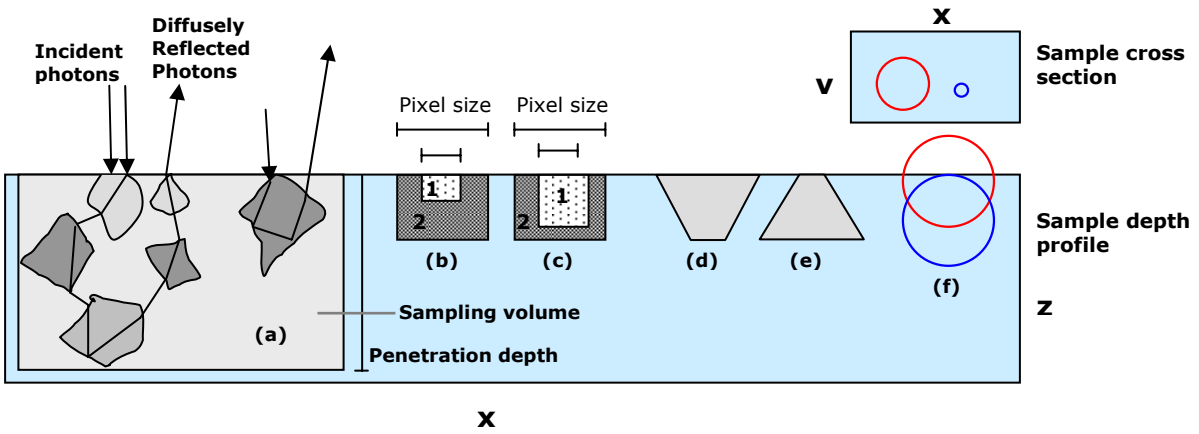


Figure 9 - Model sample of a pharmaceutical solid-mixture showing different hypothetical examples (a-f) used to explain important measurement variables of NIR-CI data acquisition.

In Figure 9a it is illustrated how the NIR radiation are reflected at particle/particle interfaces and penetrate into the sample at various depths and locations. The grey-shaded

area illustrates the penetration depth and sampling volume that photons entering this area will have interacted with. The maximum physical penetration depth is often translated into an *information depth*, which is the depth of sample contributing to the measured reflected radiation, and a sample volume can be calculated from this information depth.[45] The information depth is wavelength dependent and will be discussed further in Section 6.3.

It seems immediate apparent that any spatial resolution will be lost by this substantial reflection pattern. The low content illustration of component 1 in Figure 9b would then be difficult to detect because it would be overwhelmed by the signal of the much higher amount of component 2 adjacent to or below component 1. However, it seems that the significant penetration depth and reflection merely add a uniform baseline to the signal generating mixed spectra. The mixed spectra will have sufficient absorption information from the surface component/particles to extract this information by utilising appropriate data processing.[38] The resulting mixed spectra are proven by NIR spectra from the hyperspectral data cube of a real pharmaceutical tablet in Figure 13 (Section 5.1) and Figure 38a (Section 8.1). The sample behind the spectra in Figure 38a is predominantly microcrystalline cellulose (95 % w/w) with low content (5 % w/w) of fine API particles uniformly distributed, i.e. comparable to Figure 9b. The spectral API peaks (around 5980 cm^{-1}) from 'hot spots' of API are shown to be quite weak. The other example in Figure 38b shows the same formulation in overall batch composition but with much larger API particle sizes appearing as large clumps and unevenly distributed in the tablet, i.e. comparable to Figure 9c. The spectra from API 'hot spots' now show much stronger spectral API peaks and resemble much more the spectrum of pure API component (Figure 38c). Thus, despite the complex and unpredictable interactions between photons and sample material, the information obtained (after data processing) from any point on the sample surface will primarily come from the sample surface.[38,44-45] Both real case examples described above are sought illustrated in the model tablet (Figure 9b and 9c).

The random reflection of NIR radiation discussed above could also influence the precision of the method. It is crucial that repeated measurements of the exact same sample and the exact same area using identical analytical conditions will provide the same results. Data from such an experiment are presented in Section 8.2.

The model samples in Figure 9b and 9c also illustrate the impact of spatial resolution (pixel size). Component 1 would easily be determined in both examples (b and c) using the small pixel size. The difference will only be that the signal and hence the concentration will be stronger/higher for the sample in Figure 9c compared to Figure 9b. Using the large pixel size will relatively provide a much smaller contribution from component 1. It would thus make it more difficult to detect component 1. This could e.g. mean that component 1 would not be included for this pixel when creating a binary image, of course depending on the threshold value selected (Section 8.1).

Pharmaceutical tablets are as before mentioned usually composed of a number of components having different physical and chemical properties. The chemical images of irregular shaped components will indeed depend on what part of the particle is presented

at the sample surface (Figure 9d-9e). This will also be the case for more regular shaped particles. As an example, consider a sample composed of similar sized spherical particles (Figure 9f). Although having the exact same form and size, the interpretation of a chemical image obtained from a cross-section of the sample may be that the sample is composed of different sized particles (see xy-plane 'Sample cross section', Figure 9f). However, what the results just illustrate is spherical domains sectioned at different depths. This clearly demonstrates that care should always be taken in evaluation of particle/domain size results as absolute determinations.[44,46] A possible solution to this issue is to measure several samples of a product or measure a number of cross sections of each sample. An example of the latter is demonstrated in Section 8.2.

The optical and chemical properties of the sample components will also impact the interaction pattern with the NIR radiation. Some components may have a strong absorption, some may not absorb at all, and some may reflect most or all radiation. It will require some well-designed, systematically NIR-CI experiments of well-defined model samples to explain and understand the impact of these different component properties (cf. further suggestions Section 9.2).

The aspects that have been discussed in this section raise rather than answer questions for which there currently is no easy answer or solution. It addresses some of the limitations and uncertainties on characterising today's chemical imaging analyses of pharmaceutical solid-mixtures. The simplified hypothetical illustrations in Figure 9 serve to facilitate the perception of some of these issues but the examples are obviously not exhaustive. The model figure is however useful to have in mind when evaluating chemical imaging data but may also serve as a basis for designing new experimental studies that may address some of the unsolved aspects. Such new studies are important for further development and better understanding of chemical imaging technologies and suggestions are given in Section 9.2 (Future research).

4.6 Review of pharmaceutical applications using NIR-CI

The potential of NIR-CI has been explored through many applications in both biological, chemical, pharmaceutical and food industries. The focus of this thesis work is on the use of NIR-CI in pharmaceutical applications and only reported applications from this area are reviewed here.

The application of NIR-CI to pharmaceutical samples is still relatively new and not nearby as widely reported as for NIR spectroscopy. Due to advances in detectors, optics and computers commercial NIR-CI instrumentation became available around a decade ago and the pharmaceutical industry slowly started exploring this new technology. However, data acquisition was still rather slow compared to today's instrumentations and some of the first articles report acquisition time on half a day or even days.[47-48]

The majority of the early NIR-CI literature in pharmaceutical analysis describes the general principle of this new technology and its potential use. The early applications include root-cause analysis of manufacturing problems, product development, quality assurance and quality control but are mostly feasibility studies on relatively simple model systems or examples based on a single pharmaceutical sample.[2,5,15,18,47,49-50,52] The pharmaceutical

NIR-CI research later moved into developing methods to optimise the analysis of hyperspectral NIR images and investigating the factors affecting NIR chemical images of solid dosage forms.[36,38,43,45,53-55] In the past few years the number of pharmaceutical applications using NIR-CI has increased significantly and recently NIR-CI is seen integrated in formulation development,[14,56-57] used for mechanistic understanding of pharmaceutical processes[58-60] and three reviews have also been published.[8,34,61] Additionally, the NIR-CI technology primarily with focus on pharmaceutical applications is treated more extensively in a few book chapters.[15-16,33,48]

The primary focus of this thesis is on the use of NIR-CI in formulation development of solid dosage forms and recently two papers were published on that particular objective. Both development cases concerned evaluation of the uniform distribution of components in the formulated product using NIR-CI. Hilden et al.[56] examined whether the difference in particle size of an extra-granular component affected the uniformity of its distribution within finished tablets. Chemical images of the tablets were generated by principal component analysis (PCA) and transformed into binary images from which particles size statistics were determined and used for evaluation. Gendrin et al.[14] used NIR-CI to evaluate the homogeneity of two types of process intermediates obtained with different process parameters from an extrusion process. Both univariate (single wavelength selection) and multivariate (classical least squares) analysis were used to generate chemical images that were evaluated visually and by histogram statistics. Both studies demonstrate NIR-CI as a valuable tool for supporting formulation development studies and optimising products and processes.

With the increasing number of studies on the use of NIR-CI in pharmaceutical applications the diversity of the investigations are also increasing. Table 1 presents a categorisation of the different types of pharmaceutical applications published on the use of NIR-CI. Together with the generated list of data analysis methods used in these studies (Table 2, Section 6.2) this provides a comprehensive overview of published literature on the use of NIR-CI in pharmaceutical applications. The type of NIR-CI instrumentation used for each study can also be read from Table 1.

The recent 1-2 years of steadily increasing number of publications on the use of NIR-CI in pharmaceutical applications reveal that more and more research groups from industry and academia are exploring and developing the capability and potential use of NIR-CI for analysing pharmaceutical samples. Thus, with this increasing growth NIR-CI will likely play a more and more dominant role in pharmaceutical analysis; maybe even evolve in a similar fashion as for NIR spectroscopy.

Pharmaceutical Application	Reference	
Compositional information or Quality assessment	3, 5, **8**, 15, **15**, *17*, 20, 33, *36*, **37**, 50, **53**, 62, *71*, 72, 73	*Paper I,III*
Fundamental NIR-CI technology; Data processing or image processing development	15, **15**, **38**, 43, 44, **45**, **53**, 63, 64, 65, *71*, 72, 73, 74, 75, 76, 77	*Paper I-IV*
Powder blend and tablet uniformity, (content uniformity)	3, 15, 18, 33, (35), 49, **51**, 59, 64, (65), 66, 77	*Paper II,III*
Process information – Trouble-shooting	15, **15**, **37**, **51**, **53**	
Formulation development	14, **15**, 56, 57	*Paper II*
Mechanistic process understanding	58, 59, 60	
Particle/domain size determination	**15**, 33, **53**, 56, 58, 59, 69, *71*	*Paper II*
Counterfeit product detection	15, 62, 67, 68	
Coating studies	5, 69, 70	
High throughput analysis	15, 35, 57	
Trace contaminant, polymorph	15, **16**	

Table 1 - A comprehensive list of pharmaceutical applications using NIR-CI. The chemical imaging systems used/described in the studies are: Normal font = Global imaging; *Italic* = Line mapping; Bold = Both systems.

The increased interest on the use of NIR-CI in pharmaceutical applications is also been experienced for the other chemical imaging principles, mid-infrared chemical imaging[19,78-82] and Raman chemical imaging.[17,44,46,71,83] Applications reported for MIR-CI and Raman-CI will not be presented further here, but readers are encouraged to look for more information in the supplied references, which provide a fine introduction on the use of MIR-CI and Raman-CI in pharmaceutical applications. However, one interesting study should shortly be mentioned though. The study is special as it combines the use of Raman-CI and NIR-CI to study pharmaceutical formulations.[36] The complimentary nature of NIR and Raman spectroscopies (section 4.4) provides the opportunity for a more complete chemical imaging analysis of pharmaceutical formulations. Clarke et al.[36] analysed solid dosage formulations by separate FT-NIR and Raman chemical imaging instruments assuring to analyse the exact same area of each sample. This 'chemical image fusion' method showed how some formulation ingredients were better mapped by Raman-CI (the API and an inorganic excipient) and some by NIR-CI (a diluent, disintegrant and lubricant). The most appropriate chemical image of each ingredient were selected from the two techniques and thereafter merged to generate a final composite image. This provided an improved evaluation of the pharmaceutical formulations compared to using only one of the chemical imaging technologies. The complementary benefits of NIR-CI and Raman-CI were demonstrated, but it probably requires a commercially available instrument cov-

ering both techniques in one before it will find wider pharmaceutical applications. Because if the synergy effect of this chemical image fusion technique is so advantageous, then how come that, to the author's best knowledge, no other work has been published using this fusion technique since it was introduced in 2001?

4.7 Process analytical technology and NIR chemical imaging

The Process Analytical Technology (PAT) initiative driven by the United States Food and Drug Administration (FDA) was shortly mentioned in the Introduction to the project (Section 1.2). It is further presented here as it plays a key role in promoting the interest and implementation of NIR chemical imaging in pharmaceutical industry and academia.

FDA initiated the PAT initiative as the pharmaceutical manufacturing compared to other industries was characterised by lack of innovation and hesitation to introduce new technologies to improve the manufacturing process. In general the performance of the pharmaceutical industry was (and is) characterised by high manufacturing costs and poor efficiency. The predominant reason cited is the rigid regulatory system. Despite frequent manufacturing problems, batch rejections, reworks, recalls etc., the uncertainties of facing regulatory changes and for example losing marketing authorisation has led many pharmaceutical manufacturers to leave (not very robust) processes unchanged.

In September 2004, a final version of the PAT guidance document was released by the FDA[1]. The world wide regulatory framework and implementation of PAT is facilitated by guidelines from the International Conference on Harmonization (ICH). The concept of PAT is now adapted in the ICH guidelines of Pharmaceutical Development (ICH Q8)[4] and Quality Risk Management (ICH Q9).[84]

> *FDA considers PAT to be:*
>
> …a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e. during processing) of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality.[1]

Briefly, the concept of PAT is quality by design (QbD); i.e. instead of merely passively testing the quality of the manufacturing process steps and end products the quality should be built-in or should be by design.

The goal of PAT is to design and develop well-understood processes that will consistently ensure a predefined quality at the end of the manufacturing process. Through an in-depth characterisation and understanding of the pharmaceutical ingredients, the processes and their relationship, the FDA will allow unit operations to be adjusted along the manufacturing process to maintain an end product of consistent quality. The manufacturing process is controlled through timely measurements and the ultimate goal is to release the product immediately after the last manufacturing step, i.e. real time release. This is indeed in contrast to the traditional manufacturing paradigm with time-defined end points for processes and extensive end product testing to release the products.

An important element of the PAT initiative is promoting technologies that can provide increased process and product understanding and identify and monitor all the critical quality attributes (CQA) and critical process parameters (CPP). The role of NIR-CI in PAT lies currently in the *identification* of CQA/CPP and root cause analysis. In order to adequately *monitor* CPP there is a need to develop NIR-CI hardware and software. NIR-CI may be useful for some at-line applications but development in instrument design, acquisition speed, flexibility, ruggedness etc. are needed for NIR-CI to be a complete PAT tool and applicable for on-line and in-line measurements.

The PAT guidance recommends PAT principles and tools to be introduced during the development phase. The mechanistic understanding of how formulation and process factors affect product performance is established in early development and used later for setting the product and process specifications. It is here, that NIR-CI currently can play a major role in PAT.

The FDA itself has shown a great interest in NIR chemical imaging and conducted different regulatory research to evaluate the technology and its use for pharmaceutical applications.[3,20,49,62-63] The PAT framework is often mentioned in literature reports on NIR-CI in pharmaceutical applications and there are also a few of them specifically describing NIR-CI in relation to PAT.[2,15,61,85]

CHAPTER 5

DATA ACQUISITION AND RESEARCH INSTRUMENTATION

5.1 Data acquisition

Data acquisition is the first of three steps in a full NIR-CI analysis and is the fundament of a successful experiment. The quality of the hyperspectral raw data is depending on the data acquisition and no data processing methods are able to correct for poor quality data and provide accurate images of the distribution of sample components if the data acquisition is not performed properly. However, one of the challenges for the NIR-CI technology lies in how to confirm the best or most appropriate data acquisition. This is a difficult task as no calibration standards (e.g. a 'NIR-CI calibrated tablet') exist to identify these settings.

> *Definition of Data Acquisition in this thesis:*
>
> Includes sample preparation, instrumental settings and basic spectral transformation. The raw data output from a NIR-CI measurement is organised in a three-dimensional data structure with two spatial axes and one wavelength axis, also called a hyperspectral data cube.

SAMPLE PREPARATION

One of the advantages of NIR-CI is that almost no sample preparation is needed prior to the data acquisition. Due to the focusing nature of the line mapping system flat sample surfaces are preferable. For tablets having an inherent flat surface from using a flat punch-set no sample preparation by sectioning is necessary. For convex tablets or when it is desirable to explore the internal sample matrix a microtomer (e.g. Leica EM Trim, Leica Mikrosysteme GmbH, Vienna, Austria, Figure 10) can be used to mill flat the sample surface.

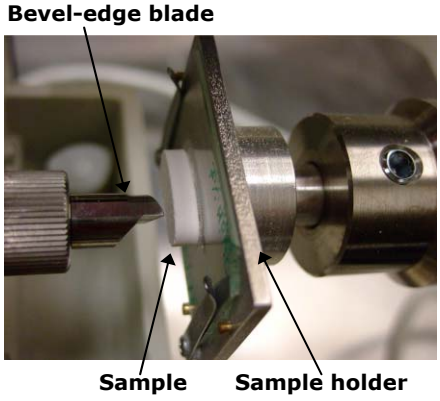


Figure 10 - Microtomer (Leica EM Trim) used to section tablets for obtaining a flat sample surface.

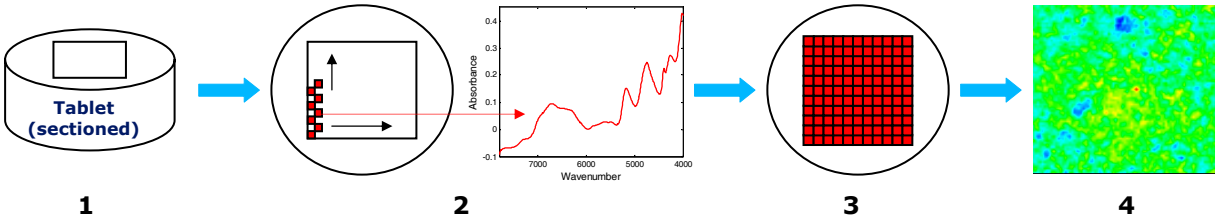
The microtoming milling technique is however not without issues as some component materials may smear or be displaced during this sample preparation.[38] For the microtomer milling process the tablet is glued onto a microscope slide using a cyanoacrylate adhesive. The milling is done with respect to the microscope plane, which has the advantage that the flat tablet surface automatically will be parallel with the image plane.

INSTRUMENTAL SETTINGS AND RESEARCH INSTRUMENTATION

The instrumental settings to consider depend on the type of NIR-CI system being used. For the line mapping system used for this thesis work the instrumental settings comprise area to be measured, spatial resolution (pixel size), spectral resolution, number of spectral scans and wavenumber range selection.

The hyperspectral NIR data cubes acquired in all experiments throughout this thesis was collected using a Spectrum Spotlight 350 FT-NIR line mapping system from Perkin Elmer (Seer Green, UK, Figure 11a).

(a)



(b)

Figure 11 - (a) Photograph of the Spectrum Spotlight line mapping instrument (Perkin Elmer, UK). (b) Schematic illustration of the steps in line mapping data acquisition. (1) The tablet with a flat surface is positioned on the sample stage and sample area defined. (2) Spectra are obtained on a linear detector for each pixel size in the defined area. (3) With sample stage movement a grid of spectral information is collected. (4) A hyperspectral data cube is finally obtained displayed as a false colour image.

Prior to each analysis a background reference was measured (cf. sub-section *Spectral correction* below). The samples were in general fixed either onto a microscope slide as described above or in an appropriately sized rubber-ring holder, which were then mounted on the XYZ motorised stage. Each sample was moved across the field of view of a 16-element mercury-cadmium-telluride (MCT) linear detector array. The pixel size generally used was 25 x 25 µm (alternatively 6.25 x 6.25 µm) so the 2D image was built up line by line in steps of 400 µm x 25 µm. The third dimension of NIR diffuse reflectance spectra were each the average of 8 scans from the wavenumber region 7800-4000 cm^{-1} using a 16 cm^{-1} spectral resolution (= 476 wavenumber channels). The Spotlight instrument and a schematic presentation of the line mapping data acquisition are both displayed in Figure 11.

SPECTRAL TRANSFORMATION

Spectral correction

As the spectral responses obtained from a NIR-CI measurement contain information from both the sample and the instrument it is necessary to correct for the instrument response by an initial background measurement. This is performed measuring the intensity of light (I_b) diffusively reflected from of a high-reflectance standard Spectralon™ (Labsphere, Inc., North Sutton, NH). The Spectralon™ is a white ceramic disc made of polytetrafluoroethylene (PTFE) that reflects almost 100 % light at each wavenumber. The intensity of light reflected from the sample is then measured (I_s). The final NIR diffuse reflectance raw data (R) from the data acquisition is then calculated as the ratio between the sample and the background measurements ($R = I_s/I_b$). The diffuse reflectance data (R) is organized in a three-dimensional structure (the hyperspectral data cube).

Conversion to absorbance

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

$$A = \log_{10}(1/R) = \varepsilon \times \ell \times c, \qquad [1]$$

where ℓ represents the path length of the sample and ε is the molar absorptivity that is specific to each analyte and a function of wavenumber.

Unfold 3D hyperspectral data cube

The generated three-dimensional hyperspectral data cube can be analysed by both ordinary two-way and three-way methods but the two-way methods have been found most suitable for this type of data.[86] To make hyperspectral image data amenable for two-way methods it is necessary to unfold the 3D hyperspectral data cube to a 2D matrix, in which each row is a spectrum from one of the pixels (Figure 12).

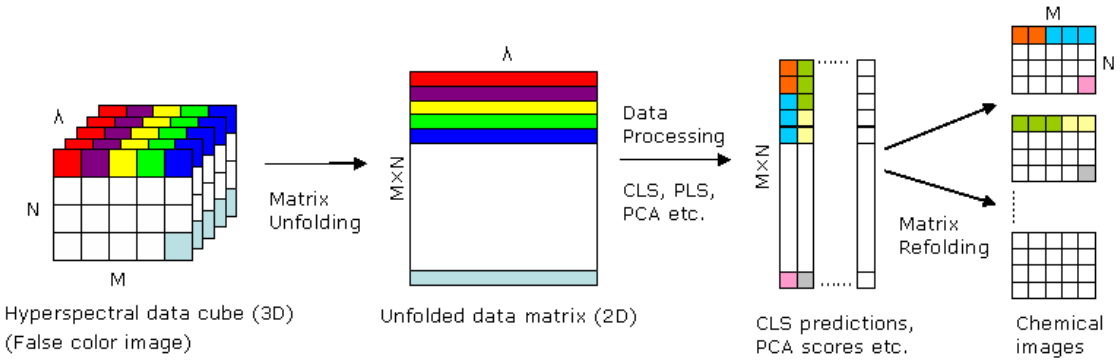
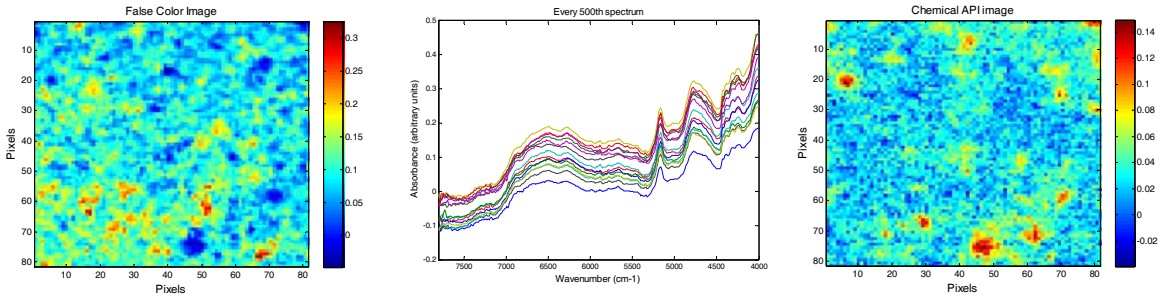


Figure 12 - Schematic description of 3D hyperspectral data cube unfolding for data processing approaches (CLS, PCA, PLS etc.) and subsequent matrix refolding.

Once all data acquisition and data processing has been performed the resulting 2D matrix of e.g. CLS or PLS prediction values, PCA score values etc. is refolded to retain the pixel location from the corresponding spectrum and generate the chemical images (Figure 12). To illustrate the actual output from a NIR-CI data acquisition (before unfolding) and comparing it to the final result from data processing, a typical example of a hyperspectral data cube, its underlying spectra and a resultant chemical image are shown in Figure 13.



(a) False Color Image **(b) Spectra from hypercube** **(c) Chemical image of API**

Figure 13 - (a) Illustration of a typical hyperspectral NIR data cube of a five-component tablet formulation shown as a false colour image; (b) its underlying spectra (every 500th spectrum); and (c) the resultant chemical image of API generated from data processing using classical least squares.

5.2 Other research instrumentation

For some of the NIR-CI experiments samples were compared with conventional NIR spectroscopy analysis (II). A Bruker FT-NIR Multi Purpose Analyzer instrument (Bruker Optics, Ettlingen, Germany) equipped with a tablet auto-sampler and capability of analysing transmission spectra of whole tablets was utilised for this purpose. Transmission NIR spectra were measured through both sides of tablets and an average spectrum was calculated for each tablet. A spectral resolution of 16 cm^{-1} and 64 scans per spectrum were used. The wavenumber region from 12500 cm^{-1} to 5800 cm^{-1} was measured. Air was used to record a background spectrum.

CHAPTER 6

DATA PROCESSING – ANALYSING THE HYPERSPECTRAL DATA CUBE

6.1 Introduction

Processing of the hyperspectral data cube concerns analysis of the spectral data contained within the data cube resulting in one or more images reflecting the spatial distribution of components within the imaged sample. Often, although this depends on the data processing method, the chemical content of the components in the imaged sample is estimated resulting in concentration or chemical images.

Data processing is the second of the three overall steps in the full analytical work of a NIR-CI experiment as defined in this project.

> *Definition of data processing in this thesis:*
>
> The processing of the hyperspectral data cube into a, typically chemical, image by univariate or multivariate image analysis methods. This part includes wavelength selection, spectral pre-processing and the subsequent data analysis to generate images visualising the distribution of each of the components within the imaged sample.

Data acquisition is described in Chapter 5, data processing is described in this chapter, and the image processing will be described in Chapter 7. Extracting all the elements from each of these three overall steps, a full NIR-CI analysis can be divided into the following steps:

1) Sample preparation
2) Instrumental raw data acquisition (hyperspectral data cube)
3) Convert reflectance data to absorbance
4) Unfold 3D data cube to a 2D matrix
5) Wavenumber selection
6) Pre-processing spectra
7) Perform data analysis on two-way data to generate images reflecting the spatial distribution of chemical components in the imaged sample (e.g. chemical images)
8) Extract information from generated images by image processing tools

As defined above data processing comprise step 5-7 and is initiated by wavenumber selection and pre-processing spectra. The data analysis method is then applied to extract, usually chemical, information from the hyperspectral data cube. A univariate approach

can be applied by simply selecting a single image plane (cf. Figure 5). However, this approach only uses a small amount of the immense information contained in the hyperspectral image cube and will often fall short of determining the accurate spatial information for more complex samples. Since a NIR-CI experiment also easily generates more than 10000 spectra and 4 millions data points, data are best suited for multivariate data analysis methods. Some of the most commonly used multivariate methods are described and discussed in this chapter.

To perform data processing of the hyperspectral data cubes various commercial software's or free open source programs for download are available. The commercial software's are often purchased with the instrument hardware (e.g. ISys by Malvern; Hyperview by Perkin Elmer; CI Xpert™ by ChemImage; OPUS/3D by Bruker). Image processing programs freely available for download are for example ImageJ[87] and JIMIA.[88] This type of ready-to-use data and image processing programs may be easy accessible, intuitive, fast and convenient to use particularly with no prior knowledge of multivariate image analysis. However, there is often a big difference in how well developed and supported these programs are and how many features they contain. Thus, limitations in processing the hyperspectral data are often met and the programs are often inflexible turnkey solutions that perform "black-box" analysis, where it is impossible to follow and evaluate the effect on the data for each step applied during data processing. As an example, the software enclosed with the line mapping instrument used for all experiments in this thesis work was evaluated and assessed to be inadequate in several aspects (however, it was also only promoted as experimental software). All raw reflectance data (R) produced for this project was therefore imported and processed in the Matlab® programming language[89] and some of its additional toolboxes. This provides almost unlimited opportunities, large flexibility and provides a much more comprehensive understanding of each data processing step.

The high importance of reliable data processing methods was realised during this project. The data processing methods applied to hyperspectral data cubes are by various well-defined manners able to generate images reflecting the composition of a sample. However, the real challenge is to select the method that generates the 'best' or most accurate images. Inadequate methods will produce inaccurate chemical images that will lead to erroneous conclusions drawn not only from the immediate apparent visualisation of the components distribution but also from the subsequent image processing analysis. Unfortunately, there is no universal data processing method that is superior for all hyperspectral data cubes. And when different univariate and multivariate methods are applied to the same pharmaceutical samples the results are often difficult to compare. The images may visually look similar but can actually provide different chemical information. Hence, there is a need for objective criteria such as e.g. a 'NIR-CI calibrated tablet' (in analogy to a certified reference material) to assess what the 'best' or most accurate image is. This is not currently available (although ASTM are working on the subject[16]), so the choice of proper analysis therefore mostly depends on the type of data set, the purpose of analysis and experience. Yet, there is one common way to compare and evaluate different data processing methods. This requires that a calibration data set is available. Regression analysis can then be performed for each method and the prediction error,

root mean square error of prediction (RMSEP) or root mean square error of cross-validation (RMSECV), can be estimated. The method providing the lowest prediction error will be considered to generate the most accurate chemical images. This evaluation approach is the main subject of paper I, where three different common calibration approaches were compared.

In general, it was realised that to further develop and mature the NIR-CI technology it was necessary to:

1) describe each data processing method in adequate details;
2) evaluate the reliability and accuracy of the chemical images produced;
3) discuss how to interpret the obtained results in relation to the experimental data and the fundamental of the data analysis method to fully understand the relationship.

The work by Sasic[71] is an excellent example of such a thorough work with detailed discussions of the PCA results and score images produced from Raman and NIR chemical images of both minor and major component in common pharmaceutical tablets. During this thesis work the classical least squares (CLS) method was found to be fast, easy to use and relatively accurate. CLS therefore became the preferable data analysis method, especially when only spectra from the pure components were available. Hence, CLS was employed in all the papers related to this thesis work (I-IV). A more detailed discussion is therefore given on particular CLS, including how to interpret the results and the CLS generated chemical images (Section 6.3).

6.2 Data analysis method used for NIR-CI studies in pharmaceutical applications

In most of the NIR-CI literature analysing hyperspectral data cubes of pharmaceutical samples the focus has more been on the actual application than a critical evaluation of the data processing methods and their appropriateness and accuracy. Table 2 is created from a thorough examination of the NIR-CI literature in pharmaceutical applications and provides an overview of the data analysis method used in these studies. Many of the early NIR-CI studies used univariate methods or multivariate turnkey approaches from commercial software. As it was realised that data processing is crucial for successful NIR-CI experiments, studies began to focus more on developing, evaluating and thoroughly discussing data processing methods for their reliability and ability to obtain the most meaningful and accurate images. For example, Clarke[53] presented and discussed a principal component analysis (PCA) and partial least squares analysis (PLS) method for analysing NIR-CI data and how to extract process related information from these. Two other studies, one comparing five different data processing approaches[64] and one comparing CLS and PLS regression[65] were also contributing to the data processing development work. Paper I (see also section 6.4) is also contributing to this work. Three common calibration approaches (single wavenumber, CLS and PLS1) were compared by their ability to generate accurate chemical images. A calibration data set of tablets with five components was used for analysis. Chemical images of the API and the two major excipients in

the formulation were produced. The accuracy of the generated chemical images was evaluated by the concentration prediction ability.

Data Analysis Method	Reference	
Single wavenumber	14, 15, 17, 18, 33, 35, 38, 49, 60, 64, 69	Paper I
Peak-height ratio, peak area	33, 36, 37, 64	
Correlation coefficient	33, 36, 64, 72	
Principal Component Analysis (**PCA**)	5, 15, 16, 33, 37, 38, 44, 53, 56, 59, 67, 71, 72, 73	
Multivariate Curve Resolution (**MCR**)	72, 74, 75, 76	Paper III
Classical Least Squares (**CLS**)	14, 16, 65, 72	Paper I-IV
Partial Least Squares (**PLS**, pure spectra), also named PLS-Classification	15, 16, 20, 33, 37, 38, 49, 50, 53, 57, 58, 59, 60, 62, 63, 66, 67	Paper IV
Partial Least Squares (**PLS**, calibration set)	16, 64, 65, 77	Paper I

Table 2 - Data analysis methods theoretically described or used for analysing hyperspectral NIR data cubes in pharmaceutical applications.

6.3 Data processing of NIR-CI data

Prior to applying the actual data analysis method that generates the chemical images the spectral range and pre-processing methods must be selected. These two steps are included in the definition of data processing and described below prior to the actual data analysis methods. The efficiency of pre-processing method or spectral range selection can only be evaluated in conjunction with the data analysis methods by using a calibration data set and select best performance by the lowest prediction error obtained.

WAVENUMBER SELECTION

NIR-CI data is usually collected in the spectral region 7800-4000 cm^{-1} (1280-2500 nm). The cut off at 7800 cm^{-1} (the NIR region normally spans to 12800 cm^{-1}) is usually performed because the impact from noise becomes too large at higher wavenumbers where the spectral absorption is low.

Multivariate methods often excel above univariate methods because of the possibility to include the entire measured wavenumber range. Nevertheless, the precision of a multivariate method can, in some cases, be improved by a proper variable, i.e. wavenumber, selection. As explained, this requires a calibration data set to evaluate and select the spectral range that provides calibration models with the lowest prediction error. Table 3 in the next subsection on pre-processing shows a few results from such a wavenumber

range evaluation for the CLS method. This is just for illustrating that prediction results can be improved by appropriate spectral range selection.

For partial least squares regression (PLS) a variety of variable selection methods exist such as for example variable importance in the projection (VIP)[90-91] and iPLS[92]. When no such methods or calibration data set are available to confirm the most suitable variables, the selection must be decided from experience and visual inspection of the spectral raw data. The general approach used for the CLS analysis has been to remove the noisy wavenumber ends and reduce the spectral range from 7800-4000 cm^{-1} to e.g. 7500-4200 cm^{-1}.

Another important impact of wavenumber selection is the consideration of depth of penetration of the NIR light into the sample. A typical average spectrum from a hyperspectral data cube obtained from a NIR-CI analysis of a pharmaceutical tablet shows that the absorption increases as the wavenumber decreases (Figure 14). This means that the NIR light cannot penetrate as deeply as it can at higher wavenumbers. Clarke et al[45] confirmed this and found an exponential relationship of NIR light penetration with wavenumber (Figure 14). The information depth, which is the depth of sample contributing to the measured reflected radiation, was found to increase exponentially with increasing wavenumber. From 9090-4000 cm^{-1} (1100-2500 nm) the information depth varied from 777 µm to 109 µm. However, it was also determined that half of the contribution to any obtained reflectance spectrum will come from the top surface (from 180 µm to 25 µm for the spectral range 1100-2500 nm).[45] The consequence of this relationship actually makes it possible to actively adjust the penetration depth by appropriate wavenumber selection.

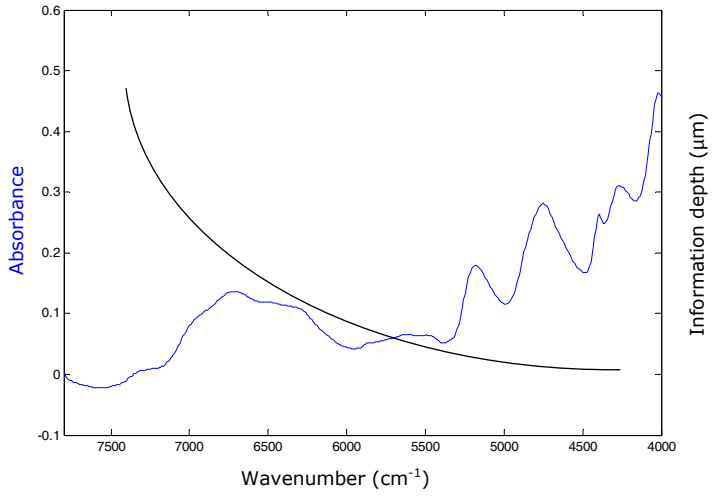


Figure 14 - Schematic illustration of a typical average spectrum from a hyperspectral NIR data cube and the exponential relationship between information depth (bold) and wavenumber determined by Clarke et al.[45]

SPECTRAL PRE-PROCESSING

The raw NIR diffuse reflectance spectra obtained from a NIR-CI measurement contain both chemical and non-chemical information about the solid sample.[15,93] The source of non-chemical information may originate from the sample (e.g. uneven sample surface or differences in sample density) and/or the instrumentation (e.g. changes in lamp intensity, detector response or other instrumental noise). The resulting effects are typically observed as spectral baseline offsets or a sloping baseline.

As it is usually the chemical information that is of interest in NIR-CI experiments the non-chemical biases are sought removed by different pre-processing techniques. These pre-processing techniques are routinely used in conventional NIR spectroscopy on smaller sets of spectra but can also be applied to hyperspectral data cubes of thousands of spectra. The most common pre-processing methods used in NIR-CI experiments on pharmaceutical solid dosage forms are first and second Savitzky-Golay derivative transformation[94], standard normale variate (SNV)[95], multiplicative scatter correction (MSC)[96], baseline corrections or a combination hereof. Note, that Savitzky-Golay derivative transforms imply choosing derivative order, filter width and polynomial order. For example, a first derivative transform with a nine-point filter width and polynomial order three is here denoted "first (or 1st) derivative (9/3)".

Careful selection of spectral pre-processing method(s) can improve the final results. However, it is difficult to confirm the best pre-processing unless a calibration data set is available to estimate the prediction error for this evaluation. Using this latter approach to compare the effect of using different pre-processing treatments in NIR-CI data processing have only been investigated in a few studies (I).[54,65] The pre-processing method evaluation was only a secondary part of paper I. In that study, results from two pre-processing methods were shown for each of three data analysis method applied to a calibration data set of hyperspectral NIR data from a conventional pharmaceutical tablet formulation. These two pre-processing methods were selected as the two best of a range of different pre-processing methods evaluated for their estimated prediction error results (RMSECV) from a regression analysis. All these results were not included in the study (I), but Table 3 below presents the results that formed the basis for selecting the two best performing pre-processing methods for the CLS method (the two best pre-processings selected are highlighted by grey-shaded areas). Note, that only the fourth lowest prediction error value was obtained for API using '1st derivative (9/3)', but this pre-processing was nevertheless chosen as a compromise due to its better prediction results for the two other components, cellulose and lactose (compared to particularly the competing pre-processings 'SNV + 1st derivative (9/3)' and 'SNV + 2nd derivative (9/3)').

CLS analysis RMSECV-values (9 calibration spectra)	API	Cellulose	Lactose
	Wavenumber range: 4200-7500 cm^{-1}		
1st derivative (9/2)	0.59	1.27	1.79
1st derivative (11/2)	0.65	1.27	1.78
1st derivative (9/3)	0.40	1.16	1.82
SNV + 1st derivative (9/3)	0.30	1.39	2.93
1st derivative (11//3)	0.45	1.23	1.87
1st derivative (9/3) + mean center	0.49	2.71	7.93
1st derivative (15/3)	0.55	1.28	1.80
MSC	1.96	2.21	3.06
MSC + mean center	1.48	2.74	7.82
SNV + 2nd derivative (9/3)	0.31	1.63	2.85
2nd derivative (15/3)	0.43	1.47	1.85
2nd derivative (9/3) + mean center	0.42	2.22	8.35
2nd derivative (9/3) (4200-7500 cm^{-1})	0.23	1.34	1.94
2nd derivative (9/3) (4000-7800 cm^{-1})	0.65	1.84	2.19
2nd derivative (9/3) (4800-6500 cm^{-1})	0.48	1.86	1.84

Table 3 - The effect of using different spectral pre-processing treatments in data processing of hyperspectral data cubes. Prediction results (RMSECV) are shown for three different components in a conventional five-component pharmaceutical tablet formulation from CLS regression analysis of a calibration data set containing nine samples. The effect on prediction results for selecting different wavenumber ranges are also shown for '2nd derivative (9/3)' pre-processing. The two pre-processing methods selected for further calculation (cf. Paper I) is highlighted in the grey-shaded areas.

DATA ANALYSIS METHODS OF HYPERSPECTRAL DATA CUBES

Some of the most commonly used univariate and multivariate analysis methods for hyperspectral image analysis and methods applied in this thesis work are presented in this subsection. The presented methods will manage to satisfactorily and reliably analyse most hyperspectral NIR data from pharmaceutical applications. Paper I contributes to the knowledge foundation being built for developing reliable data processing methods and these results are presented in Section 6.4.

Univariate method

One univariate method is simply to select an image plane from the hyperspectral data cube (cf. Figure 5). The images produced are thus the absorbance intensity values in each pixel at the specific wavenumber selected. This is the easiest and fastest way of analysing hyperspectral data cubes. For this method to be successful and provide reliable chemical images of the component of interest, it is necessary to select a unique, distinct absorption band with little spectral overlap from the other components in the sample. This may work for simple samples, but it is often difficult to find distinct, characteristic absorption bands for all components in more complex pharmaceutical samples. Often the API has one or more distinct, well resolved bands but many excipients – e.g. the fre-

quently used carbohydrates – have broad peaks and often very similar spectra. This was well illustrated by the single wavenumber chemical images of the API and two major carbohydrate excipients produced of the 5-component pharmaceutical tablets in paper I.
Fairly accurate single wavenumber chemical images of API but not of the two carbohydrate excipients were obtained. The distribution information of the two excipients was also not very well resolved, which confirm the high RMSECV results.

The distinct absorption band for each compound in a sample is best selected from the NIR spectra of the pure components if available. For proper wavenumber selection it is often necessary to pre-process these spectra to resolve and highlight subtle spectral peaks. The appropriate pre-processing method to use is sample dependent and should be determined case-by-case. However, second order derivative transform is often suitable for the task and advantageously compared to e.g. first derivative spectra, as the absorption peaks for second derivative spectra will appear at the same position as for the original peaks.

Multivariate methods

The univariate method suffers from background variation of the sample components not of interest and other interferences. This can be overcome by using multivariate techniques. The advantage of multivariate analysis is that the entire spectrum can be used and interferences can be separated from the information of interest.

Principal Component Analysis (PCA)

Principal component analysis (PCA) is probably the most used method in multivariate image analysis.[97-99] Table 2 also indicates that it has been frequently used for analysing hyperspectral data cubes in pharmaceutical applications. However, recently publications have pointed out drawbacks and limitations of using PCA for analysing hyperspectral data cubes of pharmaceutical samples.[71,72,74.]

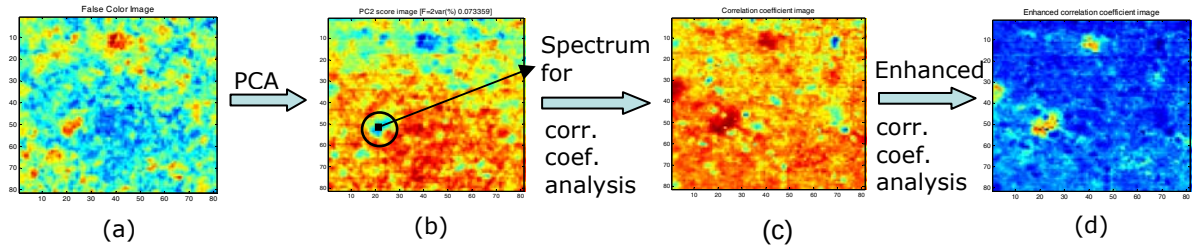
PCA is a multivariate data compression method that seeks to extract the major systematic variation in a data matrix and reduce it to a few variables, called principal components (or PCs). The principal components are new uncorrelated linear combinations of the original variables. The first principal component (PC1) describes the largest variation in the data set, the second PC the second largest variation and so forth. A loading is produced for each PC. Loadings provide information about the relation between the original variables and the PCs. In PCA models of spectral data this means that the loadings look similar to spectral features and even so, similar to the pure spectrum of a compound if a principal component mainly explains the variation for this pure compound. A score value is also assigned each sample (spectrum) for each PC based on the relative contribution of each PC. With a score value for each pixel spectrum, score images for each PC are produced as the final result from a PCA analysis of hyperspectral data.

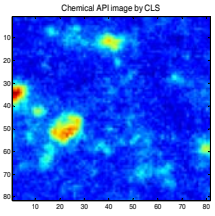
The primary drawback reported for PCA on NIR-CI data is that a principal component score image can not be unambiguously assigned to a particular chemical compound because it is difficult to associate the loadings with the pure compound spectra. Care should

therefore be taken to draw conclusions on the distribution of a compound from a PC score image. Another problem is that the numbers of PCs are often not in accordance with the number of chemical compounds in the sample.[16,71,72,74]

Despite the limitations reported for PCA it still has a role to play in analysing hyperspectral data cubes of pharmaceutical samples. Due to its unsupervised nature it can be used without any prior knowledge of the sample and do not require any reference samples (e.g. pure component spectra or a calibration data set). It is thus suitable for initial exploratory analysis of a sample, especially samples of unknown compositions, and may for example reveal the presence of un-intended products (polymorphs, hydrates, contaminants etc.) that will not be captured by supervised methods.

Acknowledging the apparent limitations of PCA, a method has been developed to optimise PCA for its exploratory use.[73] The method is based upon an initial PCA analysis, but takes it a step further. It combines PCA results with correlation coefficient calculations between one single spectrum selected from the image and all other image spectra. Furthermore, an enhanced contrast function is introduced that aims at enhancing the chemical and spatial information from the correlation coefficient image. An example is included to illustrate the method (Figure 15). The tablet sample used in this example is a five-component tablet formulation also used for other NIR-CI experiments in this thesis (cf. Table 4, Batch 9). In fact, it was the exact same sample analysed by single wavenumber analysis, CLS and PLS1 in (I), which therefore can be used for comparison (Figure 4 in Paper I).





(e)

Figure 15 - Schematic presentation of a traditional principal component analysis combined with correlation coefficient calculations of hyperspectral image data of a five-component tablet sample.[73] The false colour image of the hyperspectral data cube (a) is first analysed by PCA. Secondly, a single spectrum [(x,y) = (52,20); α=0.1] from the PCA score image (b) is used for initial correlation coefficient analysis (c). An enhanced autocorrelation coefficient routine is used to generate the final enhanced coefficient image (d). The CLS generated image of same tablet sample is shown for comparison (e, using 1st derivative (9/3)).

A conventional PCA of the hyperspectral NIR data cube is initially performed (Figure 15a). Any of the generated principal component score images can subsequently be used to select a single pixel spectrum (Figure 15b). This spectrum is used for calculating the correlation coefficient with all other spectra from the hyperspectral image and thus generating an image based on correlation coefficients (Figure 15c). However, a disadvantage work-

ing with correlation coefficients in hyperspectral image analysis is that many of the spectra in the image are highly correlated with each other, making the interpretation difficult. An enhancement function was therefore introduced to increase the discrimination between the correlation coefficients.[73] The final results show the most similar pixels to the first selected pixel in an image based on the enhanced correlation coefficient (Figure 15d). For comparison the chemical image of the same sample analysed by classical least squares (CLS, see next subsection) is also displayed (Figure 15e).

The major advantage of the presented PCA-based method is that it requires no prior knowledge of the composition of the sample and the input for the analysis is just the hyperspectral data cube from a NIR-CI data acquisition. There are some challenges associated with this method (e.g. how to select appropriate discrimination power by adjusting the α-parameter) but these will not be discussed further here.[73] Despite the challenges, this method is found very useful for an initial exploratory and scrutinising study of a new and unknown tablet formulation.

Classical Least Squares (CLS)

The classical least squares (CLS) model is most often used in spectroscopy and is also directly applicable to hyperspectral data analysis.[80,100-101] CLS was therefore an obvious model to explore in this project and is very useful because a single model is used to estimate quantitative and easy interpretable chemical information for multiple components. CLS is fast and easy to use and major advantages are that CLS only needs the pure component spectra to estimate concentrations and that it generates chemical concentration images rather than abstract score images.[101] The latter is also very useful for the subsequent image processing when extracting essential information from the generated chemical images.

The characteristic mixture absorbance spectra from a hyperspectral NIR data cube can be viewed as the weighted sum of absorbance of each pure component spectrum constituting the sample plus experimental noise. The linear relationship between absorbance and concentration (equation [1]) is utilised in CLS and the CLS model can be considered a multi-component Beer-Lambert's law given:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \qquad [2]$$

where **D** is an $(M \times N) \times \lambda$ matrix of unfolded measured spectra with $M \times N$ spectra (pixels) and λ wavenumbers (variables), **S** is a $k \times \lambda$ matrix of k pure component spectra, **E** is a $(M \times N) \times \lambda$ matrix of residuals and **C** is the resulting $(M \times N) \times k$ matrix of mixing coefficients (concentration) when applying the CLS model (Figure 16).

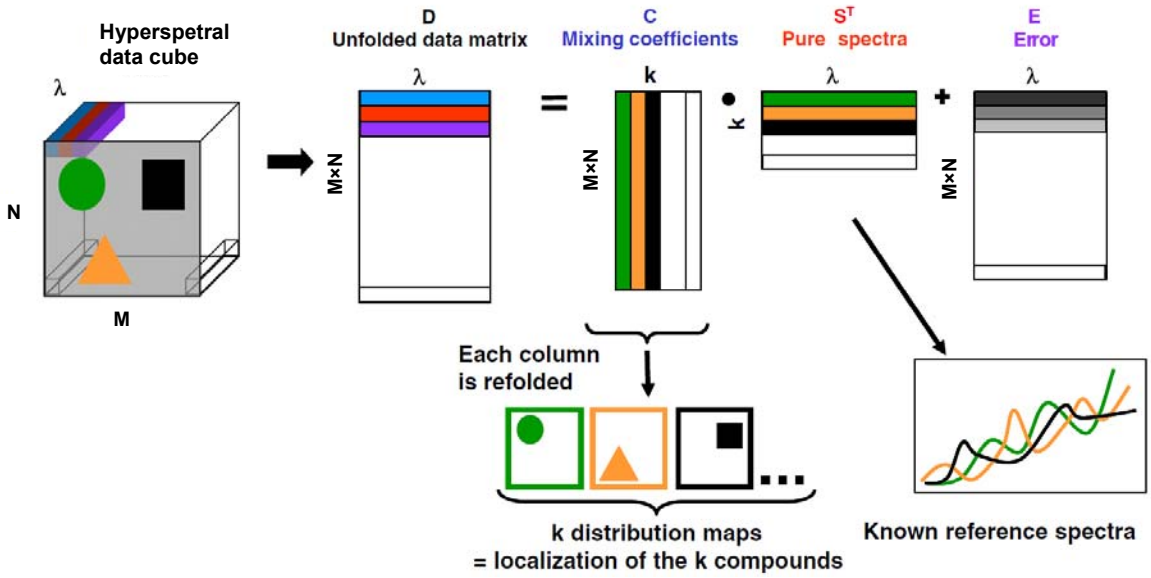


Figure 16 - Schematic presentation of analysing a hyperspectral data cube using the CLS algorithm (Reprinted slightly modified from Gendrin et al.[14] with permission from Elsevier).

Basically, the principle of CLS when analysing all the unknown pixel spectra ($M \times N$) is to find the linear combination of the pure compound spectra that gives the smallest (sum-squared) residual spectrum. **D** is given by the NIR-CI measurements, **S** is measured or known in advance and **C** is then estimated by minimising the sum of squares of the elements in **E** by:

$$\hat{\mathbf{C}} = \mathbf{DS}(\mathbf{S}^T\mathbf{S})^{-1} \qquad [3]$$

Refolding each column of the resulting $(M \times N) \times k$ matrix of estimated concentrations $\hat{\mathbf{C}}$ back to the original image size $M \times N$ will generate k chemical images with an estimated concentration of the k component in each pixel as shown in Figure 16.

As it appears, a prerequisite for using CLS is that all the spectra of the pure components constituting the sample are available. This is fortunately almost always the case in developing a new tablet formulation. One potential problem using CLS is that signals in **D** from sources not included in **S** can seriously bias the estimates of **C**. Further, obtaining good estimates of **S** can be difficult as the pure spectra from a pure component sample may not sufficiently represent the spectra of the pure components in the process mixture due to matrix effects, porosity, instrument differences when collecting data etc.[101]

As it appears from Table 2, the useful role of CLS has been explored for analysing hyperspectral NIR data of pharmaceutical samples.

How to interpret CLS images and prediction values

As described above, CLS is very useful because it estimates concentrations from spectra.[101] But how should these estimated CLS concentration predictions actually be interpreted? In many applications using CLS (cf. Table 2) the pure spectra, **S**, come from images of pure components and the average spectrum of such an image is then taken to be concentration 1 (100%). Thus, when predicting the concentrations of components in a

mixture image, these concentrations are then really fractions rather than absolute concentrations. Moreover, it is quite likely that the density of a pure component is different in the pure component compacts compared to the mixture in tablets. Hence, the sum of the concentrations found from the above CLS model will not necessarily add to 100%.

This is in itself not problematic, as long as it is understood that a CLS prediction of, say 40% component A, does not necessarily imply that the sample contains 40% of component A in the sample. It means that there is 40% of what was in the pure component sample. Mostly, though, the CLS predictions may be converted to something that can be considered the mean percentage of a component in the actual sample. There are several different ways of transforming the initial predictions into predictions of fractions in terms of the actual sample. These methods have different properties and different prerequisites as will be shown below.

The hyperspectral NIR data cube of a five-component pharmaceutical tablet was analysed by CLS (Figure 17). The uncorrelated CLS prediction results are displayed in Figure 17a. Two different ways of scaling these CLS prediction results are shown in Figure 17b and 17c. Scaling each individual pixel prediction value to the sum of the prediction values in each pixel will provide the results shown in Figure 17b. All pixel value will sum to a total concentration of 100 %. However, scaling each individual pixel prediction value to the sum of the mean concentrations for all components in the entire chemical image (0.8737 in the example in Figure 17a) will provide the results depicted in Figure 17c. The mean of all component predictions will now sum to 100 % but the variation will be similar to the uncorrelated results.

The above considerations on interpreting and scaling CLS prediction results are only preliminaries. Further systematic studies are needed to fully understand how best to evaluate and handle CLS prediction results. Such studies would preferable include chemical reference analysis method (Section 9.2).

It is important to be aware of the different factors that may affect the accuracy of the CLS prediction results. For example, it is important to assure that the sample materials used to generate pure spectra are as representative as possible of the component material in the final dosage form. The pure spectra are typically generated from compacts produced by direct compression of the pure component material. The likelihood that pure component material approximates sample material is therefore higher for a direct compression process than e.g. a wet-granulation process. The sample material in wet-granulation may change form and size during the many different process steps and cause differences in the spectral response between pure component and sample materials. For a direct compression process, the largest contribution to any spectral differences between the pure component compacts and the final tablets probably originate from density/porosity differences. An example to illustrate this impact is presented below. In a small pilot study, spectra from pure component compacts produced by high and low compression force were obtained by NIR-CI analysis. A sample from a pharmaceutical powder blend compressed into a compact on a hydraulic tablet press (similar to producing the pure component compacts) was also analysed by NIR-CI. The powder compact

was from a three-component formulation composed of 10 % API, 45 % cellulose and 45 % lactose (all in % w/w).

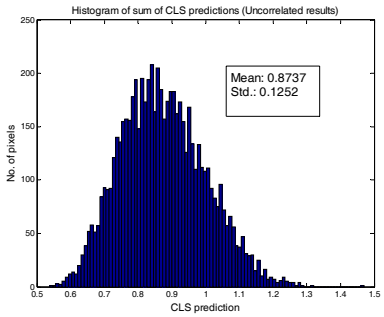
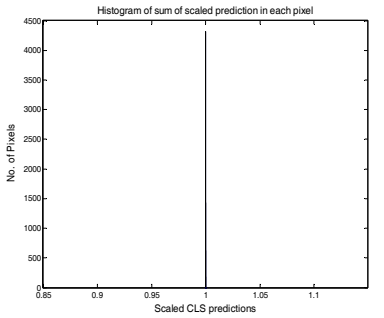
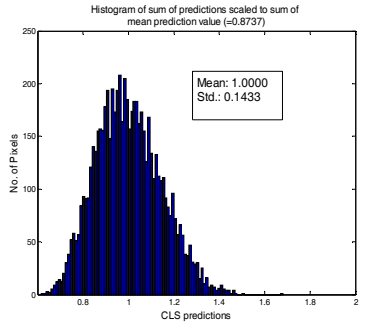
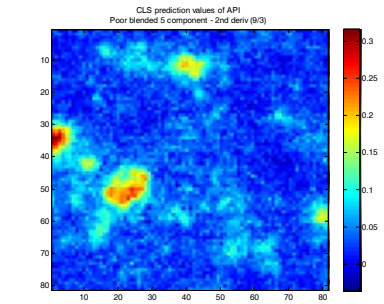
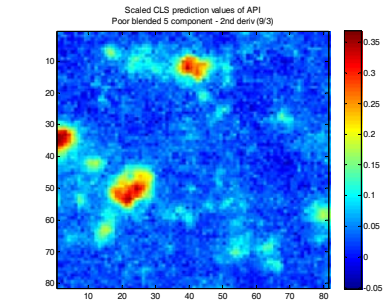
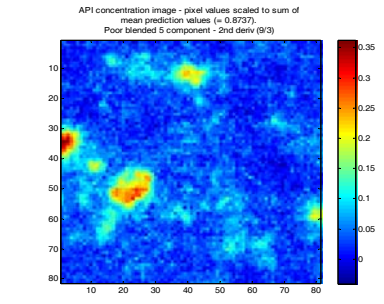
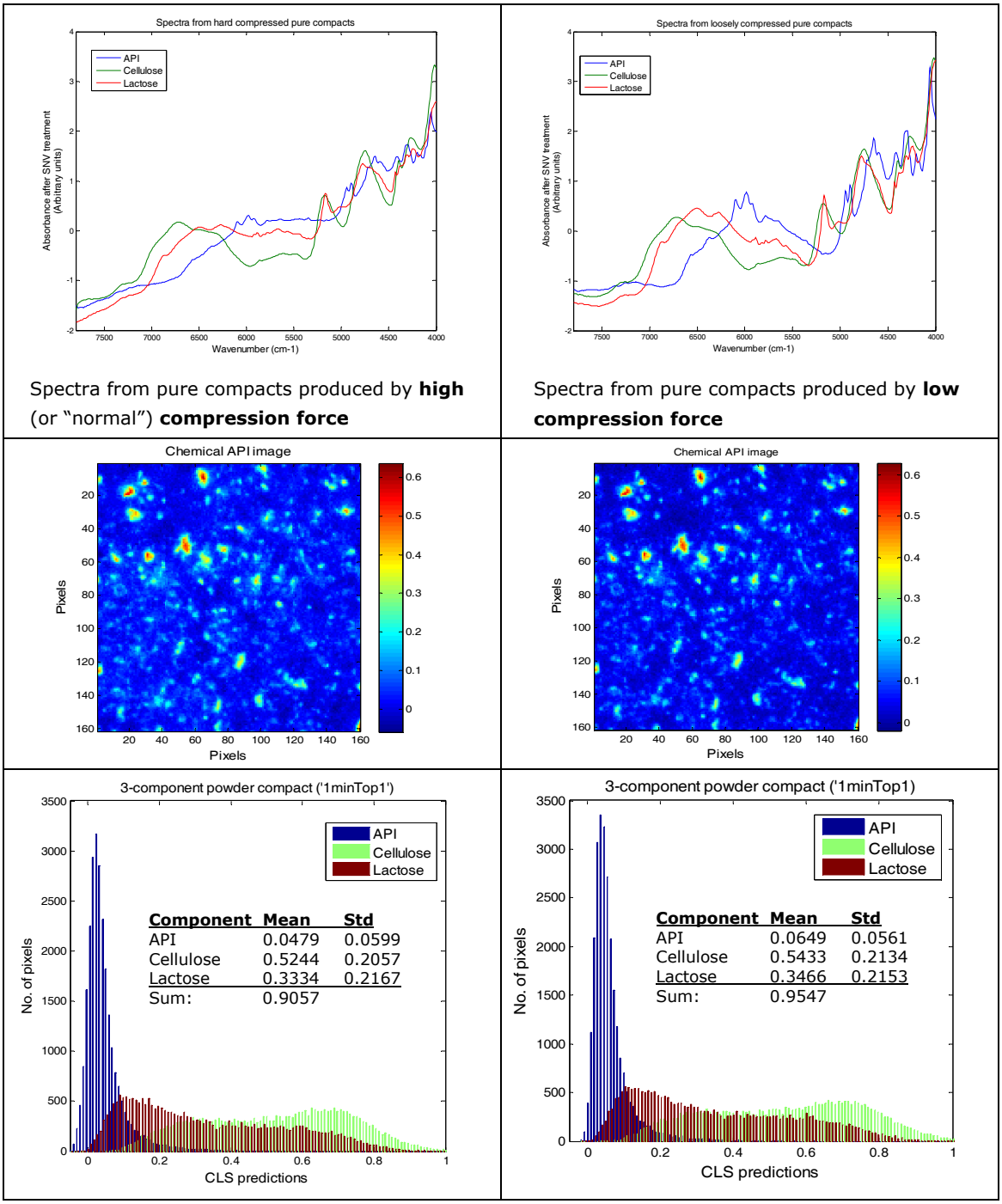
(a)	(b)	(c)
Uncorrected CLS results:	*Scaling each individual pixel prediction value to its pixel sum value:*	*Scaling each individual pixel prediction value to the sum of mean analyte concentrations (0.8737):*
0.0416 (API)	0.0482 (API)	0.0476 (API)
0.0229 (Mg. Stearate)	0.0260 (Mg. Stearate)	0.0262 (Mg. Stearate)
0.2921 (Cellulose)	0.3278 (Cellulose)	0.3344 (Cellulose)
0.4834 (Lactose)	0.5595 (Lactose)	0.5532 (Lactose)
0.0337 (Talc)	0.0384 (Talc)	0.0386 (Talc)
0.8737 Sum	0.9999 Sum	1.0000 Sum
		
Mean: 0.8737	Mean: 1.0000	Mean: 1.0000
		

Figure 17 - Different ways of scaling the concentration predictions from a CLS analysis of a hyperspectral NIR data cube of a five-component pharmaceutical tablet formulation. (a) Uncorrelated or not-scaled CLS predictions; (b) Scaling each individual pixel prediction value to its pixel sum value; (c) Scaling each individual pixel prediction value to the sum of the mean component concentrations (0.8737 in this example).

The spectra from the pure component compacts produced by high pressure and the results from a CLS analysis of the three-component powder compact using these pure spectra are shown in Figure 18 (left column). The corresponding CLS results for the exact same sample data using spectra from the pure component compacts produced by low pressure are shown in Figure 18 (right column).



Spectra from pure compacts produced by **high** (or "normal") **compression force**

Spectra from pure compacts produced by **low compression force**

Figure 18 - CLS analysis of a three-component powder compact sample. Results were obtained using spectra from hard/normal compressed pure component compacts (left column) and using spectra from loosely compressed pure component compacts (right column). The results shown are in row order: pure component spectra, chemical API image by CLS and histogram statistics (i.e. mean CLS predictions) for the three components.

The results from this pilot study were far too scarce to be conclusive, but indicate that the density/porosity of the pure component compacts may affect CLS prediction results. The results indicate that using pure spectra from loosely compressed compacts yield higher content predictions, in particular for API in this example (35 % increase, 0.0479 vs. 0.0649). Two other similar powder compact samples (replicates) were analysed in a simi-

lar manner and provided similar results (not shown here, but the API increase for these two samples were 21 % and 24 %, respectively). Whether these results are generally valid and apply to all components in the samples must be verified by detailed designed studies and is a task for future research (Section 9.2).

Partial Least Squares (PLS)

PLS is a common and well-described multivariate regression method used to build quantitative calibration models.[100,102] It is a regression method used to relate two data matrices **X** (spectra) and **Y** (reference values) with each other. In PLS model building of hyperspectral images, **X** is the spectra or mean spectrum of each hyperspectral data cube and **Y** the concentration values of one (PLS1) or more (PLS2) of the components in the sample. The calibration data set required for PLS is composed of several samples spanning an appropriate concentration range to build a model for new predictions. For PLS analysis only components that have been varied in concentration in the calibration data set are modelled. This is in contrast to CLS analysis where all components are (and should be) modelled.

In this project, only PLS1 models for each component that varies in concentration have been built as described above. The models are used by applying them to all single-pixel spectra in a hyperspectral image ($M \times N \times \lambda$) of an unknown sample composed of same components as the calibration data set. The resulting chemical image for each component consists of concentration prediction values of that component for each pixel. The mean of all pixel score values is an estimate of the concentration in the analysed sample.

PLS has in general been superior, i.e. the most accurate, when compared with other data processing methods for analysing hyperspectral NIR data cubes of pharmaceutical samples (I).[64-65] This means that PLS can be assumed to provide the most accurate chemical images. The major limitation for not mainly using PLS is that it is time-consuming to produce and measure a calibration data set. Further, it may not always be possible to obtain a calibration set, for example, in the early development of a pharmaceutical formulation due to the often limited amount of API available.

In PLS there is also the challenge to obtain the most accurate Y matrix (reference values). Using the concentrations from the theoretical composition does not take the batch variation into account. This can be included by analysing batches by conventional analytical techniques (e.g. HPLC); however, this usually only provides the API concentration as methods are normally not developed for excipients. A more accurate approach would be to obtain the reference values for each sample (tablet). I.e. first analyse the tablet by NIR-CI and then by a reference method. Or if sample preparation to obtain a flat tablet surface is necessary for the NIR-CI analysis, then first analyse the tablet by a non-destructive reference method and then by NIR-CI. The latter approach requiring a non-destructive analytical method could e.g. be provided by NIR spectroscopy analysis (also for excipients, but again only if a reference method for them exists). Moreover, an approximation is made when the measured reference value is determined for the entire tablet but only a smaller part of the tablet is analysed by NIR-CI (sampling issue). Ide-

ally, the Y reference values should be obtained by analysing the component concentration in each pixel, which is obviously impossible.

As appears from Table 2 many reported studies of NIR-CI in pharmaceutical applications use an alternative PLS method (in Table 2 denoted 'PLS2 (pure spectra)'). This method is based on PLS-Discriminant Analysis (PLS-DA) in conjunction with pure spectra and is commonly denoted PLS-Classification (PLS-Class). This PLS-Class model is build from a calibration matrix (**X)** of pure component spectra and a dummy matrix (**Y**) as target matrix, where '1' and '0' indicates belonging or not belonging, respectively, to each pure component spectrum.[15-16]

However, it can be mathematically shown that the PLS-Class method is actually doing CLS and that there are advantages of using CLS rather than PLS-Class (IV). The equivalence of PLS-Class and CLS is also illustrated by an example (Figure 19). The hyperspectral data from a NIR-CI analysis of a five component conventional pharmaceutical tablet was analysed by PLS-Class and CLS. The chemical images of API were generated and their predictions plotted against each other (Figure 19). From this plot it is also evident that PLS-Class and CLS provide the exact same results. Identical relationships were found for the other four components in the tablet.

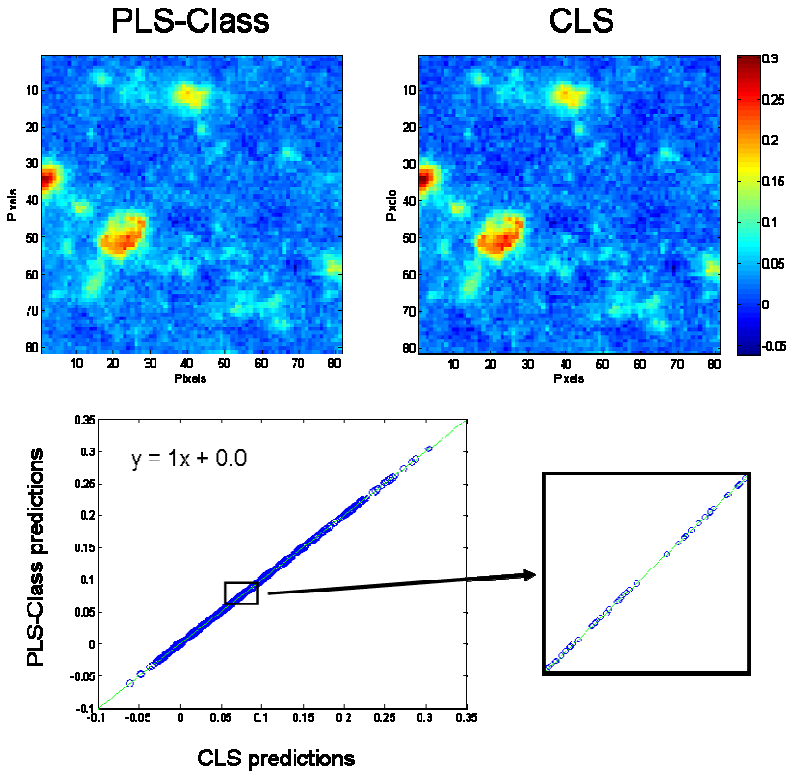


Figure 19 - Top: Chemical images of API predicted by PLS-Class and CLS methods. Bottom: CLS predictions versus PLS-Class predictions of API.

Multivariate Curve Resolution (MCR) – Alternating Least Squares (ALS)

During the work of developing robust and reliable data processing methods for analysing hyperspectral NIR data cubes focus was on developing methods without a time-consuming calibration stage. Multivariate curve resolution (MCR) models were in this context evaluated for their potential to provide reliable quantitative and spatial information of all the five components in a conventional tablet formulation (Paper III). Recently, increased interest has been shown for employment of multivariate curve resolution techniques to study hyperspectral NIR data of pharmaceutical tablets.[72,74-76]

The basis of MCR is very similar to CLS and they can both be described by the bi-linear model $\mathbf{D} = \mathbf{CS}^{\mathbf{T}} + \mathbf{E}$ (cf. equation [2]). CLS uses spectra from the pure components to create **S**, whereas MCR is a class of methods for estimating **S** from the data, i.e. the spectra from pure components are not a prerequisite.[72,76,101,103] The advantage of MCR in this relation is that matrix effects are included in the estimated spectra, which eliminates the issue of obtaining representative pure component sample material to generate pure component spectra as earlier discussed for CLS. In MCR the estimation of **C** and **S** matrices are performed iteratively and the algorithm alternating least squares (ALS) is often used for this purpose (MCR-ALS). One drawback of MCR-ALS is related to difficulties in estimating **S** (due to the phenomena of 'lack of selectivity' or 'ambiguities') and an augmented version of MCR-ALS has been developed to overcome this.[72,76,103] A detailed description of the two MCR-ALS method and their differences and limitations will not be given here but can be found in Paper III and the supplied references.

MCR-ALS and augmented MCR-ALS were compared with CLS in the study of NIR-CI data for providing quantitative and spatial information of all ingredients in a five component tablet formulation (III). The conventional tablet formulation was composed of components of both low and high concentration (minor and major components, respectively). The tablets analysed were the six tablets (replicates) of batch 9 also analysed for the common calibration approaches in Paper I (cf. Table 4, Section 6.4). Unlike that calibration study (I), all five minor and major components in the formulation were included for calculation in this study (III).

The analysis of the tablet samples in (III) demonstrated CLS and in particular augmented MCR-ALS to provide reliable semi-quantifications of the major components in the formulation. For the minor components only augmented MCR-ALS provided reliable results. The MCR-ALS techniques were found useful to analyse NIR-CI data of pharmaceutical tablets without the need of a calibration stage and the potential applications are for analysing samples where no prior information about them are available.

6.4 Comparing common calibration approaches (I)

As emphasised before NIR-CI provides information about what chemical components are available in the sample, how much of each is present, and most importantly and unique for this technology, their spatial location. The quantitative determination of the components in the sample is valuable information, in particular because it does not only quantify the API but also the excipients, which are almost never quantified by conventional analytical methods. However, the quantification in itself is not the goal using NIR-CI. Other conventional analytical techniques are better and more accurate for doing this. Further, the NIR-CI quantitative determination will always be an approximation, as it will only measure a part of each sample. Nevertheless, accurate quantification using NIR-CI is important and a desirable goal, because it assures reliable results and generation of accurate chemical images. Accurate chemical images are important for sample evaluation and in particular for the image processing, where the essential and process related information from the images are extracted.

A study (I) was performed comparing three of the data analysis methods presented earlier in this chapter; namely (1) single wavenumber, (2) classical least squares (CLS), and (3) partial least squares (PLS1). The methods were compared by their ability to generate accurate chemical images, which was evaluated by their concentration prediction ability. Two different pre-processing methods were evaluated for each method. The approach was to analyse a calibration data set by NIR-CI and evaluate the most accurate method as the method providing the lowest prediction error (RMSECV) from regression analysis. The calibration set was produced from nine batches of a five-component direct compression tablet formulation only varying the concentrations for components of high content in the tablet formulation (Table 4). Six tablets (replicates) were analysed by NIR-CI for each batch and chemical images and the predicted concentrations were computed for API, cellulose and lactose for each of the three data processing methods. Regression analysis was performed from the nine predicted concentrations (mean of 6 replicates) versus the nominal concentrations.

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

Table 4 - Composition (% w/w) of the 5-compound pharmaceutical tablet formulations constituting the 9 calibration batches. *Particle size measures (µm) of the cumulative volume distribution given by the 10%, 50% and 90% percentiles (D[v,0.1]/D[v,0.5]/D[v,0.9]) obtained using a Malvern Mastersizer 2000 laser diffraction system.

The results from the regression analyses are shown in Table 5. PLS1 proved to be the most accurate method for all three components, which means that PLS1 can be assumed

to generate the most accurate chemical images. However, CLS proved to be an excellent alternative to PLS1 providing only slightly less accurate concentration predictions.

	Single wavenumber			**CLS**			**PLS1**		
Components	API	Cellulose	Lactose	API	Cellulose	Lactose	API	Cellulose	Lactose
Preprocessing	**2nd derivative (15/3)**			**1st derivative (9/3)**			**1st derivative (9/3) + mean center**		
Wavenumber region (cm^{-1})	5984	4280	5168	4200-7500			4256-4744; 5840-6176 (105 var.)	4088-4488 (52 var.)	4248-4520; 5008-5400 (85 var.)
Corr. coef. R^2	0.98	0.95	0.92	0.98	0.98	0.96	0.99	0.98	0.99
Slope	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	0.99
RMSEC	0.29	1.560	2.03	0.30	0.93	1.50	0.05	0.24	0.32
RMSECV	0.38	1.94	2.46	0.40	1.16	1.82	0.18	0.70	0.62
Preprocessing	**SNV + 2nd derivative (9/3)**			**2nd derivative (9/3)**			**2nd derivative (9/3) + mean center**		
Wavenumber region (cm^{-1})	5984	4280	5168	4200-7500			4304-4792; 5816-6200 (112 var.)	4104-4496 (50 var.)	4264-4536; 5016-5400 (86 var.)
Corr.coef. R^2	0.96	0.97	0.86	0.99	0.97	0.95	0.98	0.98	0.97
Slope	1.00	1.00	1.00	1.00	1.00	1.00	1.04	1.04	0.92
RMSEC	0.44	1.53	2.59	0.17	1.11	1.60	0.07	0.33	0.35
RMSECV	0.54	1.97	3.38	0.23	1.34	1.94	0.23	0.87	0.88

Table 5 - Results from concentration predictions of API, cellulose and lactose for the single wavenumber method, CLS and PLS1. For each calibration method results are presented for the two preprocessing treatments showing best predictions. The grey-shaded area points out the best predictions obtained.

6.5 Comparing and selecting data processing approaches

A range of different data processing methods to analyse hyperspectral data cubes from NIR-CI experiments have been described. Some of the methods have been evaluated for their applicability and accuracy, and interpretation of the results has been discussed.

The presented methods provide a varied and comprehensive 'toolbox' of data processing approaches that can be used to analyse most types of NIR-CI data from pharmaceutical applications. The methods may also be applied to hyperspectral data of pharmaceutical samples obtained from other chemical imaging techniques. An overview of the presented methods and their characteristics are summarised in Table 6.

The most suitable technique to choose for a certain application will likely depend on the purpose of the analysis and the type of data set. The single wavenumber method and PCA are for example most suitable for initial exploratory analysis. For quantitative analysis and more accurate distribution information calibration PLS has proven to be the superior method to use. However, PLS requires a calibration data set that is time consuming to produce and measure and also requires that sufficient amount of component material is available. When there is no calibration data set available, CLS and augmented MCR-ALS are excellent alternatives for quantitative purposes. CLS was during this thesis work experienced as fast, easy to use, easy to interpret, and providing reliable results. It thus

became the choice of method when no calibration data set but only the pure component spectra were available.

Data analysis methods	Samples required for method	Characteristics of method and preferred use
Single wavenumber Univariate	NIR spectra of pure components (to identify distinct absorption peaks)	Qualitative use. Easy, seemingly intuitive and fast. Exploratory analysis, valid for simple sample matrices. Distinct spectral band for compounds required.
PCA Multivariate Unsupervised	No prior knowledge of sample required	Qualitative use. Exploratory analysis, valid for complex matrices. Provides principal component score images.
CLS Multivariate Supervised	NIR spectra of pure components	Quantitative use. Relatively accurate, fast and easy. Requires only pure component spectra. Assumes Beer's law valid. Provides concentration/chemical images. Easy interpretable from a chemical perspective.
MCR/MCR-ALS Multivariate Supervised	No prior information about samples needed. Can estimate spectra from sample data.	Quantitative use. Relatively accurate but not as well-defined, fast and easy as CLS. Initial estimates of **S** using pure component spectra are often preferable. Assumes Beer's law valid. Provides concentration/chemical images.
PLS Multivariate Supervised	Calibration data set of known compositions	Quantitative use. Accurate, robust predictive precision. Requires full calibration data set. Provides PLS scores or predicted concentration images.

Table 6 - Comparison of data analysis methods suitable for analysing hyperspectral NIR data cubes from most types of pharmaceutical applications and products.

CHAPTER 7

IMAGE PROCESSING

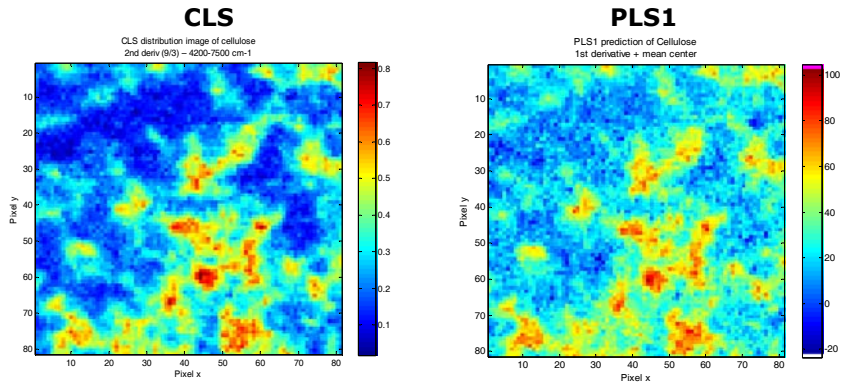
7.1 Introduction

Image processing is the third and last step of a full NIR-CI experiment as defined in this thesis (Section 1.3). Image processing also constitute a very important part for a successful NIR-CI experiment, as this is the step where the useful and relevant process and product related information contained in the generated images is extracted and revealed. Obviously, it is important to extract the relevant and correct information contained in the generated chemical images for a successful NIR-CI experiment.

> *Definition of Image Processing in this thesis:*
>
> The processing of the generated chemical images into relevant and 'useful' information that will qualitatively or quantitatively describe the properties of a sample in relation to the problem investigated. This could e.g. be a total concentration, a measure of the distribution of the concentration of the active ingredient or the domain sizes of components within the sample.

There is a need to obtain objective and quantitative measures from the images derived from data processing. The chemical images can visually be very convincing but care should be taken on the qualitative interpretation of these images. It may be impossible to discriminate small differences in chemical images by visual examination and sometimes the images can even be 'misleading'. An example of the latter is shown by comparing two chemical images of cellulose of the exact same sample generated from CLS and PLS1 data processing in the study described in (I) and Section 6.4 (Figure 20). The visual distribution information obtained is quite similar but the CLS image appears easier to interpret due to the higher contrast. However, the best cellulose concentration predictions (lowest RMSECV), and thus presumable the most accurate chemical image, was obtained by PLS1.

Other reasons for obtaining objective and quantitative information from chemical images are that such numerical measures are necessary for the employment of the NIR-CI technology in quality control (QC) and in order to be a complete PAT tool, e.g. by enabling real-time monitoring of critical attributes.



RMSECV: 1.34%

RMSECV: 0.70%

Figure 20 - Chemical images of cellulose (20.0 % w/w nominal) from a five-component tablet formulation generated by CLS (left) and PLS1 (right). The CLS image provides higher contrast and is easier to visually interpret, but the most accurate content prediction was obtained by PLS1 (RMSECV=0.70 % for PLS1 versus RMSECV=1.34 % for CLS).

Despite the importance of image processing this is the step of the three overall NIR-CI analytical steps that has been paid the least attention to in published papers so far.[53] This is probably partly because it is not so straightforward, and partly because focus at first has been on developing accurate and reliable methods for data acquisition and data processing. The latter makes sense from the perspective that it is important to first "lay down the foundation". If the spectral quality from the data acquisition is poor, no multivariate image data analysis method is able to compensate for this and still generate accurate results. If the data processing method is sub-optimal, inaccurate chemical images will be generated that will lead to erroneous conclusions in the subsequent image processing analysis. And finally, even when an accurate chemical image is generated, poor image processing methods may extract the wrong product or process-related information from the images. It is therefore imperative that each of the three overall steps is thoroughly investigated and their strengths and limitations known. In the present chapter different image processing methods are presented that will help to further develop the value of the NIR-CI technology.

As the unique information obtained from chemical imaging techniques primarily are the spatial distribution (which is closely related to sample homogeneity) and particle/domain sizes, it is obvious to develop image processing tools extracting and determining quantitative and objective measures of these characteristics. Such tools and measures are presented in the following.

7.2 Histogram statistics

The histogram representation of a chemical image statistically characterise the intensity distribution across the imaged sample. The histogram plot displays the obtained image intensity values (e.g. the CLS prediction values) on the abscissa and the number of individual pixels for each of these intensity values on the ordinate. The *mean* represents the

mean value of all pixel predictions of a particular component over the whole image, providing an estimate of how much of the component is present in the sample. The *standard deviation* is a measure of the variation of image prediction values around the mean and provides a quantitative estimate of the overall sample homogeneity at the particular spatial resolution utilised. Sometimes *skewness* and *kurtosis* are also used to characterise the histogram distribution. Skewness is a measure of the asymmetry or tailing of the histogram, whereas kurtosis gives information on the peak shape of the histogram.[15,34] In general, the histogram statistics from NIR-CI experiments have been utilised to study the spatial and chemical homogeneity of pharmaceutical samples.[14,15,17,33,49,59,64,66] This homogeneity concept from histogram statistics introduced for NIR-CI analysis is different from the traditional homogeneity concept described earlier in Section 2.4. The thousands of individual measurements obtained from microscopic pixel locations of each sample provide data for intra-sample statistical analysis. This homogeneity is thus pertaining to each sample/tablet, i.e. intra-tablet homogeneity, more than the overall batch as in conventional homogeneity testing.

An example of histogram plots and statistics are displayed in Figure 21. The two chemical images of API having same nominal concentration represent a homogenous (top) and heterogeneous (bottom) tablet, respectively.

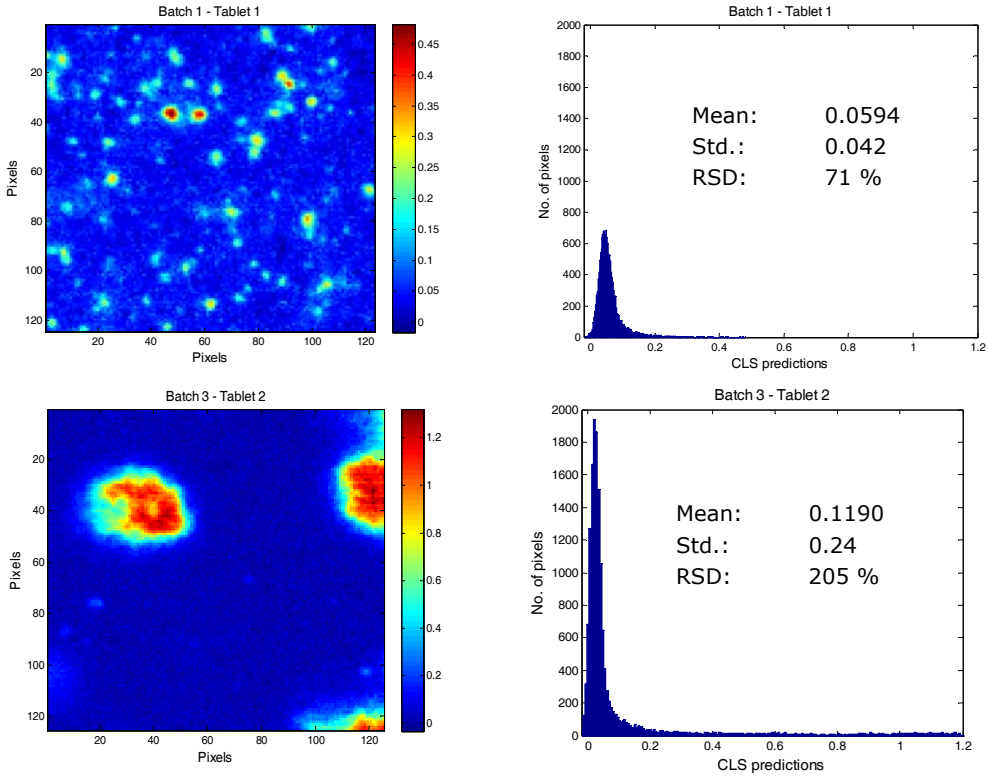


Figure 21 - Chemical API images generated by CLS and their corresponding histogram plots and statistics. The two images represent a homogenous tablet (top) and inhomogenous tablet (bottom), respectively.

Histogram statistics of chemical images can be useful for some applications but may not always be a suitable and complete measure to determine homogeneity or blend uniformity in pharmaceutical samples. This homogeneity evaluation is based on the inherent spa-

tial resolution, i.e. original pixel size, and basically determines the distribution of these single-pixel concentration predictions. The limitation of this principle is illustrated in the simulated images of two-component pharmaceutical samples in Figure 22. The first two samples in the upper row (*a* and *c*) are considered homogenous and the first two samples in the lower row (*b* and *d*) in-homogenous. The difference between the left column samples (*a* and *b*) and the middle column samples (*c* and *d*) samples is the spatial resolution of the images. The calculated standard deviation between pixel values for the left images with high spatial resolution would be identical although their blend uniformity is clearly different. However, the standard deviation calculated using the low pixel resolu uniformity differenc tion shown in the middle column images (*c* and *d*) would clearly demonstrate the blend es between the left and right sample. This is further illustrated in Figure 22*e* and 22*f*, where the homogenous pixels (low standard deviation) for image (*e*) are broken down to smaller pixels in image (*f*) expressing another blend uniformity (higher standard deviation) for the same sample.

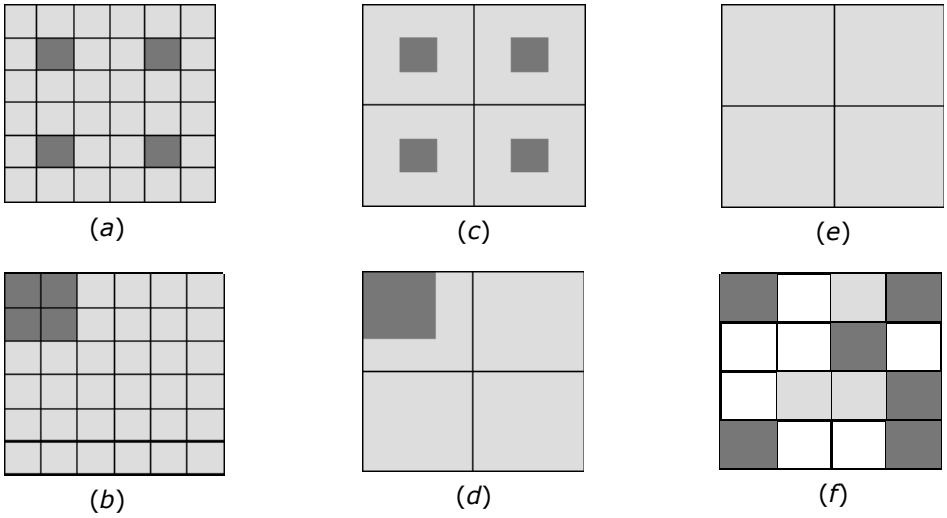


Figure 22 - Simulated chemical images. The two image samples (*a*) and (*c*) and the two image samples (*b*) and (*d*) are identical regarding distribution of dark grey component but differ in spatial resolution (36 and 4 pixels, respectively). The standard deviation calculated over the intensity values in each pixel of the images having the same spatial resolution will reveal identical values for (*a*) and (*b*) but differ between (*c*) and (*d*). The images (*e*) and (*f*) illustrate how a homogenous sample in image (*e*) may look different and more in-homogenous using a higher spatial resolution (smaller pixels) in image (*f*). This illustrates the influence of the spatial resolution on histogram statistics of images and how it may not always be able to distinguish between samples of different homogeneity grade interrogating at only one spatial resolution.

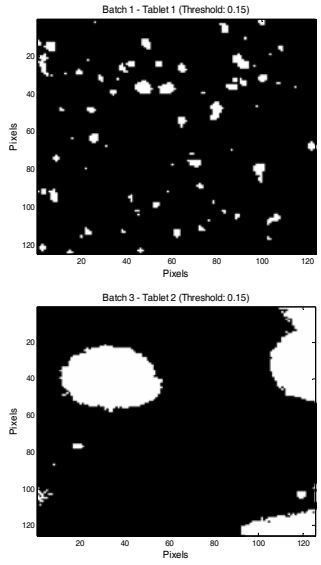
It was based on this realisation that the homogeneity concept of calculating the standard deviation of prediction values in a chemical image was extended to be a function of the spatial resolution (scale of scrutiny). This concept is presented in Section 7.4.

7.3 Binary images and domain statistics

BINARY IMAGES

The creation of binary images (image binarisation) aims at separating the component of interest from the other components in the image. The binary images make component distribution easier to visually interpret, but binary images are also used as basis for generating overlay or composite images and for measuring the number and size of particles or domains within a sample.

A binary image of an individual component can be generated from a chemical image by setting a threshold value from the colour-scale intensity bar on the chemical image or from the histogram plot. All pixel values above the threshold value are then given the value '1' (and e.g. coloured white) and all pixel values below the threshold value are given the value '0' (and coloured black). The binary image can now be generated as a black-white (or other two-coloured) image. Figure 23 displays the binary images of API generated from the chemical images shown in Figure 21.



No. of domains:	71
Mean equivalent diameter:	72 µm
Max. equivalent diameter:	178 µm
Min. equivalent diameter:	28 µm
Area covered:	4.0 %

No. of domains:	11
Mean equivalent diameter:	262 µm
Max. equivalent diameter:	976 µm
Min. equivalent diameter:	28 µm
Area covered:	15.7 %

Figure 23 - Binary images of the chemical API images in Figure 21. The threshold value 0.15 is used to create both binary images. The calculated domain size statistics (see later in this section) are also shown for each binary image.

THRESHOLD SELECTION

Selecting the appropriate threshold value is a real challenge in creating binary images. To the author's best knowledge no objective method has been reported to select a threshold in chemical imaging of pharmaceutical products. Most of the NIR-CI studies of pharmaceutical solid dosage forms that have demonstrated the value of generating binary images, primarily for the purpose of calculating particle/domain size statistics, have applied user-defined threshold values.[16,33,46,51,53,56,58-59,71] Otsu[104] has introduced a general, objective method of automatic image threshold selection that minimises the intra-class vari-

ance between pixels. It determines an individual threshold value for each image. However, Otsu's method of image threshold selection was not found useful for any of the samples analysed in this project and was also reported insufficient in another study.[71]

As in other areas of NIR chemical imaging analysis the difficult task is to confirm the selection of the most appropriate threshold value. It would require some kind of 'NIR-CI calibrated tablet' for each product, which is impossible to obtain. With the lack of such standards, the best way seems currently to determine the threshold value as a function of the sample and problem investigated. Although this means subjectively selecting the threshold value it may be possible to select one fixed threshold value for each component appropriate for all samples of each formulation. Such consistent *formulation-specific* threshold values would probably be the most suitable approach for comparing different samples of the same formulation, although it may not be based on the most 'correct' threshold values. Another possible approach is a *method-specific* threshold value selection, where the threshold value is selected using the same approach/method for each sample. This could for example be determining the threshold value as the mean value plus one, two or three standard deviations for each generated component image.

Another method-specific threshold value selection is suggested and used for the study in Paper (II) and Section 8.1. In this study, chemical API images were generated using CLS. The threshold value to generate binary API images was determined from CLS analysis of the hyperspectral data cube of the pure API compact using only the mean API spectrum from this sample. The histogram plot of the resulting chemical API image (of the pure API compact) showed a symmetrical distribution around the mean of 1.0 (Figure 24).

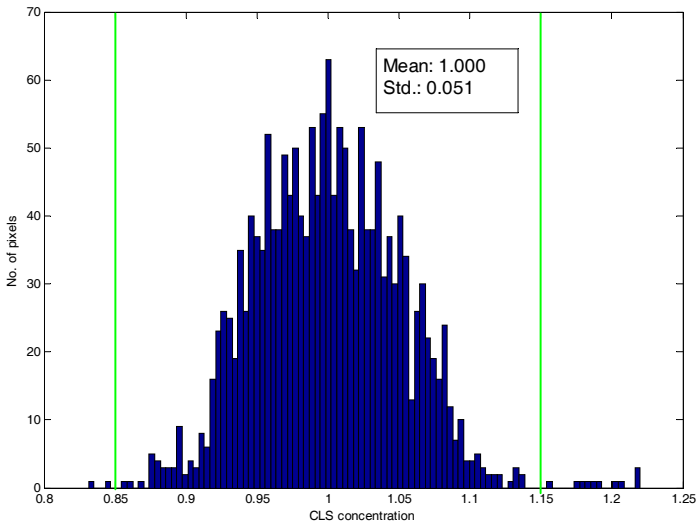


Figure 24 - Histogram representation of the chemical API image from analysing a pure API compact by CLS using only the mean API spectrum of the hyperspectral data from the pure API compact. The threshold limit is set as three time the standard deviation from this analysis (= 0.15, vertical lines).

The spread of the data around the mean was interpreted as an estimate of the methods precision (repeatability) assuming a uniform pure API compact (i.e. assuming API concentration 1.0 in each pixel). The threshold value was hence determined as equal to three times the standard deviation (Std. = 0.051, Figure 24) from the CLS analysis of the pure API image; resulting in a threshold value of 0.15.

The interpretation of threshold values selected by this approach is that a concentration value equal to or below the threshold value in a chemical image could within statistical probability (three standard deviations) just as well be regarded as zero as determined by the 'precision' analysis of the pure component image. So in order to statistically confirm the component in a chemical image the concentration/intensity values must be above the threshold value. Whether this threshold selection method is general applicable or just useful for some samples needs to be studied further on different types of pharmaceutical samples. The advantage of the method is that it determines a threshold value in an objective and well-defined manner.

DOMAIN SIZE STATISTICS

To calculate particle/domain size statistics within an imaged sample it is necessary first to create a binary image as described above. Images can often not discriminate between one large particle, agglomerated particles or closely spaced but separate particles. It is therefore more the domain size statistics rather than the particle size statistics that is determined and the term 'domain statistics' is therefore used. Further, this discrimination often becomes more difficult at higher component concentrations.[38] Although the calculation of the domain size statistics of components is an approximation of the true solid particle size statistics in the imaged sample, the domain statistics are useful to determine if any changes in domain size or form happen during the manufacturing process and to compare different formulations and processes.

In formulation development of solid dosage forms, the effect of API particle size is usually evaluated for its influence on the process and product performance. NIR-CI and domain size statistics is a powerful combination to study the effect of varying API particle size on its distribution within the formulation.[16,58] This is demonstrated in this project for a two-component formulation produced by direct compression (Section 8.1 and II). Calculation of domain size statistics will also be very useful for wet-granulation formulations, where it is important to control and hence measure the size and form of individual granules and of the granules in the final tablet formulation, but also determine the API distribution within each granule.[12,16,22]

The useful role of image binarisation and calculating domain sizes has also been demonstrated for various trouble-shooting investigations.[16,37,51,53] In these studies variation in domain sizes of different components (disintegrant, diluent, binder) were found to cause changes in the product quality (dissolution, content uniformity) or problems in the manufacturing process (poor powder flow, adherence to tablet tooling).

The actual domain size statistics in this project is calculated using routines developed in Matlab® 7.5.0 (R2007b) software[89] and functions from the Image Processing Toolbox 6.0

(Mathworks).[105] Many different properties can be obtained but the following has been extracted and focused on in this project (Table 7):

DOMAIN STATISTIC PROPERTY	EXPLANATION OF PROPERTY
Number of domains	The number of separate domains registered in the binary image
Mean Equivalent Diameter (µm) (Std., Min. and Max.)	Domains appear in many different shapes. The Equivalent Diameter represents the diameter of a circle with the same area as the region of the domain of interest. The mean is calculated for the equivalent diameter of all domains. The standard deviation, minimum and maximum equivalent diameters can also be determined from this calculation.
Area Covered (%)	The area covered is the number of pixels of the calculated component (e.g. white pixels) in percent of the total number of image pixels. The percent area covered thus gives a rough estimate of the component concentration.

Table 7 - The domain size statistics properties determined in this thesis work.

The results from a domain size statistics calculation is shown in Figure 23 but the utilisation of these statistics are also demonstrated in the pharmaceutical application discussed in Section 8.1.

7.4 Measures of homogeneity

CONVENTIONAL HOMOGENEITY TESTING

As earlier described (Section 2.4) the conventional homogeneity testing determines the chemical homogeneity, i.e. the bulk content relationship of components, but provides no information on the spatial homogeneity. The conventional homogeneity testing can thus be considered an *inter-tablet* homogeneity versus the *intra-tablet* homogeneity provided by NIR-CI analysis. NIR-CI provides information on the spatial homogeneity and the standard deviation of the histogram plot of a chemical image is an objective measure of the variation of image data about the mean indicating the degree of homogeneity (Section 7.2). However, as mentioned earlier determining sample homogeneity in this way is dependent on the spatial resolution used, and a sample may be considered chemically homogenous at one spatial resolution and heterogeneous at another (Figure 22). This concept was therefore further developed to be a function of the spatial resolution (scale of scrutiny).

MACROPIXEL ANALYSIS METHOD

Instead of only using the predicted pixel values from the original spatial resolution (pixel size) defined by the data acquisition, the refined method for sample 'homogeneity/heterogeneity' presented samples the chemical image by many different spatial resolution's (called 'sampling squares' or 'macropixels'). These macropixels are a square cluster of pixels and varies in size from the original pixel size up to the size of the entire chemical image. For example, a chemical API image generated by CLS and consisting of 30 × 30 pixels could be sampled by a macropixel of 10 × 10 pixels (Figure 25). The chemical image is sampled by all the 10 x 10 macropixels it is possible to position in the

chemical image by moving the macropixel systematically one original pixel size step each time. This sums to 441 macropixels sampling the image in Figure 25. A measure of the homogeneity could then be calculated as the standard deviation of the mean CLS prediction obtained for each macropixel, i.e. the standard deviation of the 441 mean CLS predicted API concentrations from the 441 macropixels in the example shown. For practical reasons it is preferable to obtain a square chemical image (e.g. by truncation). The number of sampling squares for a certain size can then be calculated from:

$$\text{Total number of macropixels} = (\text{Image pixel size} - \text{macropixel size} + 1)^2 \quad [4]$$

[For the example in Figure 25: $(30 - 10 + 1)^2 = 441$ macropixels]

This new NIR-CI homogeneity approach ensures a much more extensive and detailed sampling of each image, which is thought to provide a better and more comprehensive basis for evaluating the homogeneity of pharmaceutical samples.

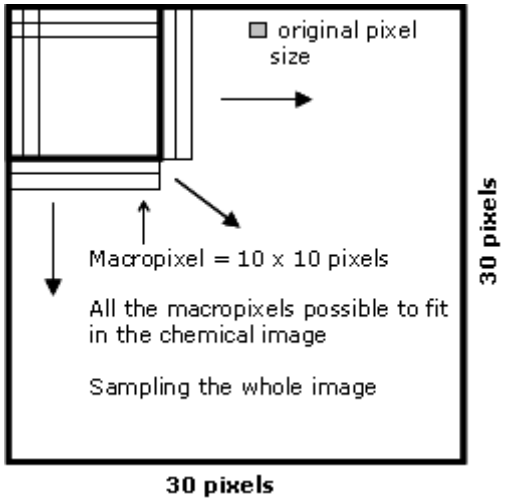


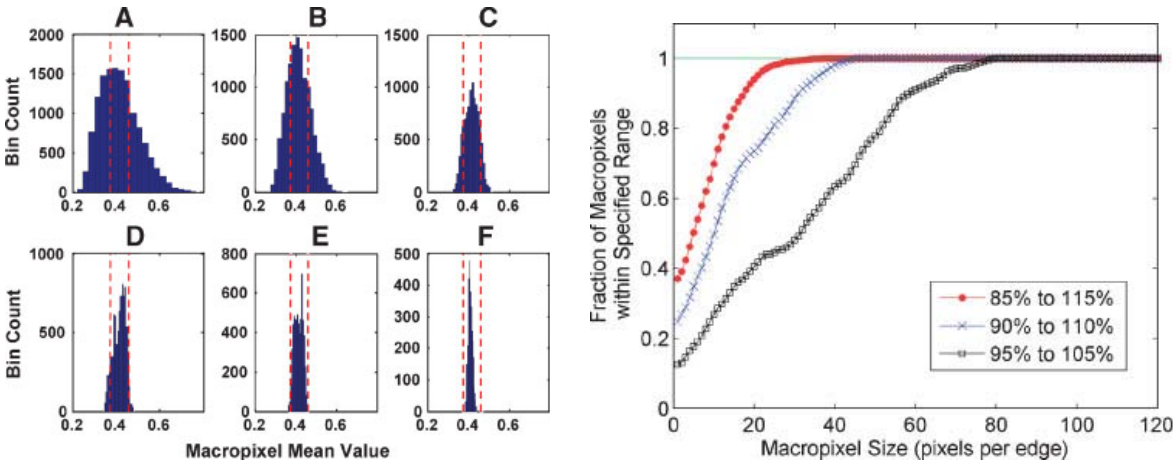
Figure 25 - Illustration of the sampling principle for the novel NIR-CI homogeneity measure approach. The entire chemical image is sampled by all the macropixels it is possible to position in the chemical image. The relative standard deviation of the CLS prediction values for each macropixel is calculated and plotted as a function of the squared macropixel size. This generates a so-called *homogeneity curve*.

This methodology of extensively sampling a chemical image by all possible macropixels ('macropixel analysis') was first introduced by Hamad et al.[63] and the method was specifically referred to as 'Continuous-Level Moving Block' (CLMB). The method introduced by Hamad et al.[63] for evaluating the CLMB results is described in the following, but a different evaluation approach has also been developed and is presented ('homogeneity curves').

EVALUATION OF MACROPIXEL ANALYSIS

Mean or standard deviation

Each macropixel value calculated in the Continuous-Level Moving Block (CLMB) approach is the mean value of all the original pixel predictions contained within each macropixel. In the example above (Figure 25) a macropixel value is thus the average of 10 × 10 = 100 original pixel prediction values. A histogram plot can be generated for each macropixel size similar to the histogram statistics described earlier in this chapter (Section 7.2). For each macropixel size Hamad et al.[63] calculate the fractions of macropixel values within a 90-110 % specified range of the mean value for the complete set of macropixels. This approach is illustrated in Figure 26. The concept is then to choose the minimum macropixel size that provides all its macropixel values within the 90-110 % range of the macropixel mean value of the image (macropixel size 43 in Figure 26b).



(a) **(b)**

Figure 26 - (a) Histograms resulting from CLMB analysis of different macropixel sizes. The dashed vertical lines represent the specified range from 90-110% of the mean image values. (b) Plot of fractions of macropixel values within specified ranges as a function of macropixel size. The goal is to find the macropixel size with all macropixel values within the specified range (i.e. fraction = 1.0, e.g. macropixel 43 for the 90-110% range). (Reprinted from Hamad et al.[63] with permission of Wiley-Liss, Inc. a subsidiary of John Wiley & Sons, Inc).

In this project (II and Section 8.1), the standard deviation of the set of macropixel concentration prediction values for each macropixel size was calculated. The standard deviation is plotted against the squared macropixel size (e.g. 10^2 = 100 for the example in Figure 25), which will generate a so-called *homogeneity curve*. An example of this homogeneity curve is shown in Figure 27 for a chemical API image of CLS predictions for a two-component tablet formulation.

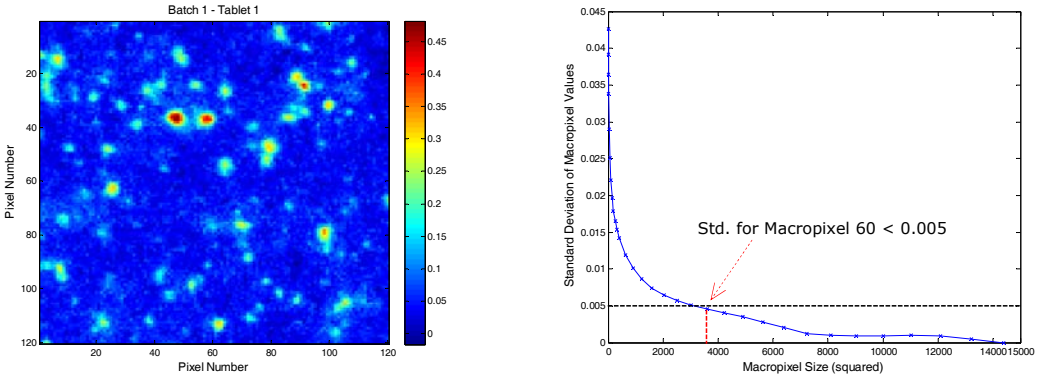


Figure 27 - CLS predicted chemical API image of a two-component tablet (left) and its corresponding 'homogeneity curve' showing standard deviation of macropixel values as a function of squared macropixel size.

Homogeneity curves

Visual evaluation and 'univariate' approach

Visual evaluation of the shape and position of the homogeneity curves would in many cases be sufficient for the purpose of comparing the homogeneity of different products. The goal for a homogenous sample is low standard deviations between macropixels, i.e. the lower-lying a homogeneity curve the more homogenous a sample. However, for not very distinct samples with closely spaced homogeneity curves, it may not be so easy to visually evaluate and compare their homogeneity. It would therefore also for this chemical image evaluation method be useful to move beyond the subjective and qualitative analysis and determine a numerical value as a quantitative measure of this homogeneity parameter.

An option to objectively and numerical assess homogeneity from these curves would simply be to select a macropixel size and a standard deviation as a specification limit for which the homogeneity curve values would need to be below in order to be assessed homogenous. This univariate approach is actually similar to the standard deviation of the histogram statistics using the original spatial resolution. However, the advantage for the macropixel method is that the intra-tablet homogeneity can be evaluated at any suitable level of scrutiny (i.e. macropixel size). For example, assume the acceptance limit for homogeneity of the product in Figure 27 is a standard deviation of macropixel values below 0.005 at macropixel size 60. As illustrated the standard deviation value for a macropixel value of 60 (= 0.0046) is below the acceptance criterion of 0.005 and the sample thus fulfils the homogeneity specification.

This approach is very much in line with today's conventional homogeneity concept in the way that a blend or a tablet batch is considered homogenous if the content uniformity results comply with the acceptance criteria determined by its purpose of use. The appropriate macropixel size to select for evaluation will depend on the degree of homogeneity (quality) needed, which depends on the purpose of use for each product. The challenge will be to select the appropriate macropixel size and standard deviation value as specifi-

cation limit for each product. However, strictly selecting such an acceptance limits is probably most useful for routine quality control purposes, whereas in formulation development it would be sufficient to compare the standard deviations between products at a certain sampling square without necessarily setting a specification limit.

Area Under Homogeneity Curve

As the goal is to obtain as low standard deviations between macropixels as possible an obvious evaluation approach is simply to determine the area under each homogeneity curve. This method has the advantage that it includes all the information contained in the homogeneity curves. An application using this approach is presented in Section 8.1 (II).

Deviation from Ideal Mixing

A very interesting aspect of evaluating the chemical images by the macropixel analysis approach using the calculated standard deviations (and thus 'homogeneity curves') is that it can be related to and possible explained by the theory of mixing particles.[12,106-109]

> *Definition of Mixing:*
>
> Mixing may be defined as a unit operation that aims to treat two or more components, initially in an unmixed or partially mixed state, so that each unit (particle, molecule, etc.) of the components lies as nearly as possible in contact with a unit of each of the other components.[12]

Tablets are produced as individual dose units from a powder mixture as a result of a mixing process. However, each tablet can also be considered a 'mixture' when divided into small sub-samples (pixels). For illustration and better understanding purposes, a powder mixture of two components of equal quantity and identical particle size, shape and density are considered (Figure 28).

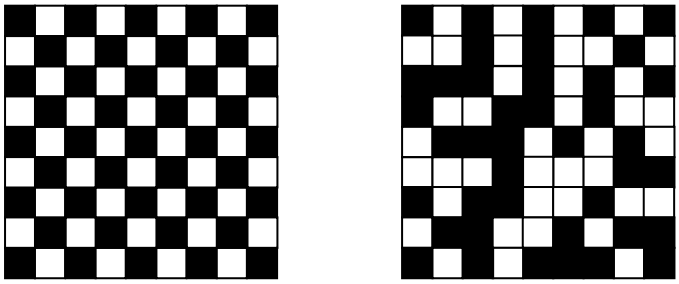


Figure 28 - Illustration of two states of binary powder mixtures: (a) Perfect mix; (b) Random mix. There are an equal proportion of identical sized black and white components in each mixture.

If a binary mixture according to the definition of mixing above is achieved, it produces a theoretical ideal or *perfect mix* (Figure 28a). The chance of a perfect mix occurring is practical impossible, and also unnecessary and sometimes undesirable. In practice, the best type of mix obtainable is a *random mix*, where the probability of finding a particle of a given component is the same at all positions in the mixture and equal to the proportion of this component in the entire mixture (Figure 28b).[12,107-108]

In formulation development and manufacturing of tablets the goal is to produce individual dose units (tablets) from the powder mixture with minimal variation in the composition. This is sought obtainable by selecting the appropriate particle sizes, mixing process etc. There will always be some variation in composition of samples taken from a random mixture. For samples taken from a random mixture the standard deviation can be calculated as:[107]

$$\sigma = \sqrt{\frac{P_A \cdot P_B}{n}} = k \cdot \sqrt{\frac{1}{n}} \qquad , where\ k = \sqrt{P_A \cdot P_B} \qquad [5]$$

P_A and P_B are proportions of the two components A and B and n is the number of particles in the scale of scrutiny sample. As obtaining the random mixture is the goal, i.e. the optimal mixture, the standard deviation (σ) in Equation 5 expresses the ideal degree of mixing. It appears from the equation that the standard deviation can be reduced by either increasing the number of particles in the scale of scrutiny sample by decreasing the particle size, or by increasing the scale of scrutiny sample. Equation 5 is based on particles of identical size, shape and density, which is never the case in practice of pharmaceutical powder mixing. For multi-sized particle mixtures, the ideal degree of mixing for a binary mixture can be expressed as:

$$\sigma = \sqrt{\frac{P_A}{P_B} \frac{P_B \cdot \overline{w}_A + P_A \cdot \overline{w}_B}{W}} \qquad [6]$$

where W is the mass of the sample, $\overline{w}_A$ is the mean particle mass of component A and $\overline{w}_B$ is the mean particle mass of component B. Equation 6 can be approximated powder mixtures of several components by considering them pseudo-binary-mixtures of component A (e.g. API) and components B (all the excipients).

The principles described above can be applied to the sampling of a chemical image. The chemical image represents the entire powder mixture and the macropixels represents the scale of scrutiny. In Equation 5 the number of particles (n) in a sample is proportional to the sample size. The macropixels can thus substitute n in Equation 5 and a plot of the ideal homogeneity curves can then be obtained for every two-component mixture. The ideal homogeneity curve for the sample in Figure 27 (P_A= 0.05 and P_B = 0.95, k=0.22) is depicted in Figure 29. Acceptance limits can be defined around the 'ideal' homogeneity curve for which homogeneity curves of samples need to be within in order to be assessed suitable homogenous. The upper and lower ± 25 % limits are shown in Figure 29 (red curves) and it is seen that the sample from Figure 27 according to this is quite far from being homogenous (green curve). However, this calculation was not taking particle size

or density into account and there can be many other reasons for the homogeneity curve position of this sample.

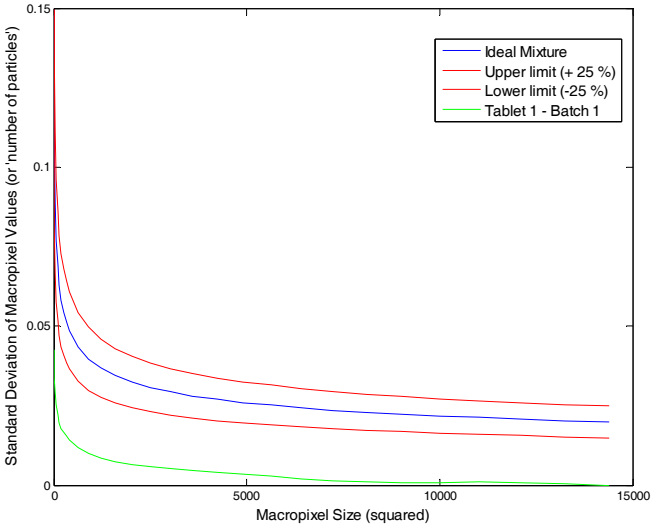


Figure 29 - The ideal homogeneity curve for a 5:95 binary mixture (blue). The ± 25 % upper and lower limit homogeneity curves with regard to the ideal homogeneity curve are shown in red. The homogeneity curve for the sample in Figure 27 is shown in green.

The need to investigate the behaviour and characteristics of homogeneity curves occurring from different pharmaceutical samples is evident. Unfortunately, time did not allow enough exploration of this, but some of the work that has been done and is on-going is shortly presented here. The work is related to investigate the relationships between chemical images and the shape and position of their homogeneity curves. Such fundamental studies are preferable performed using simulated chemical images to standardise and control the concentration and homogeneity ('randomness') of the images. The initial work of these studies is presented in Figure 30, Figure 31 and Figure 32. In Figure 30 the concentration is identical in all three images but the particle size (and 'homogeneity') is varied. In Figure 31 the concentration and the particle size are identical for all three images but the distribution (or homogeneity/randomness) is varied. Finally, in Figure 32 the concentration is varied in the four images but the particle size is identical and the distribution ('homogeneity/randomness') is attempted to be maintained.

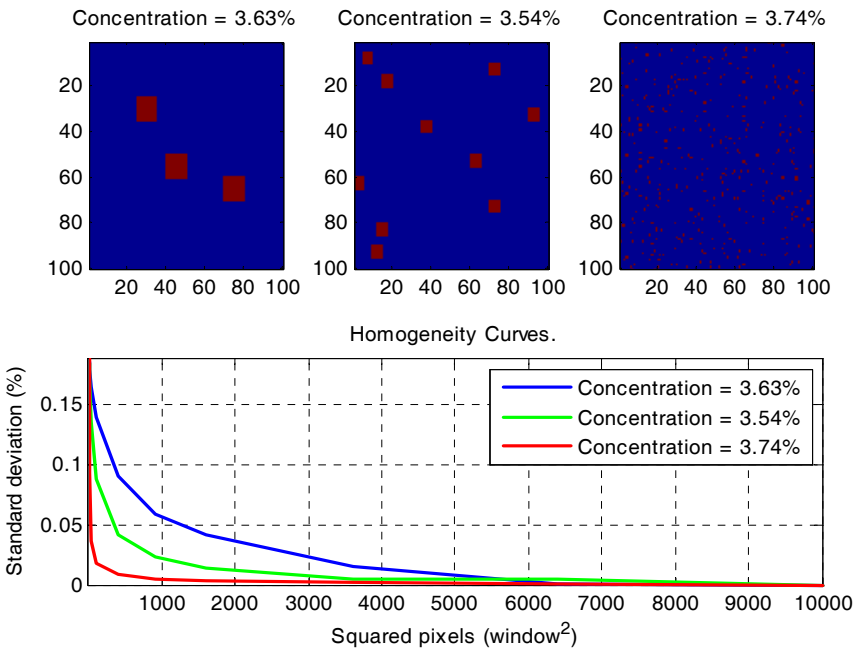


Figure 30 - The concentration in the images is identical but the particle sizes and homogeneity are varied.

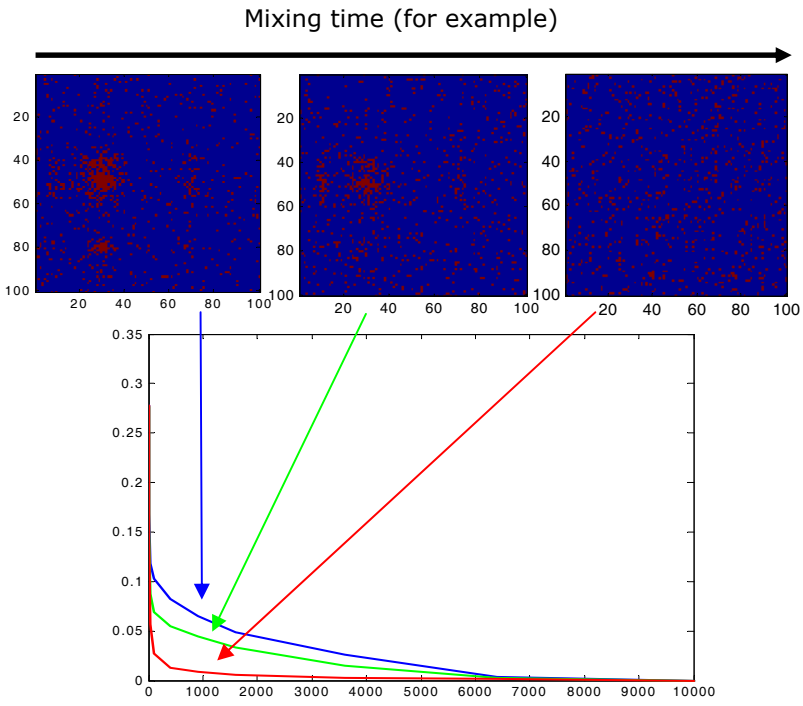


Figure 31 - The concentration and particle sizes are identical in the images but the homogeneity (or randomness) is varied.

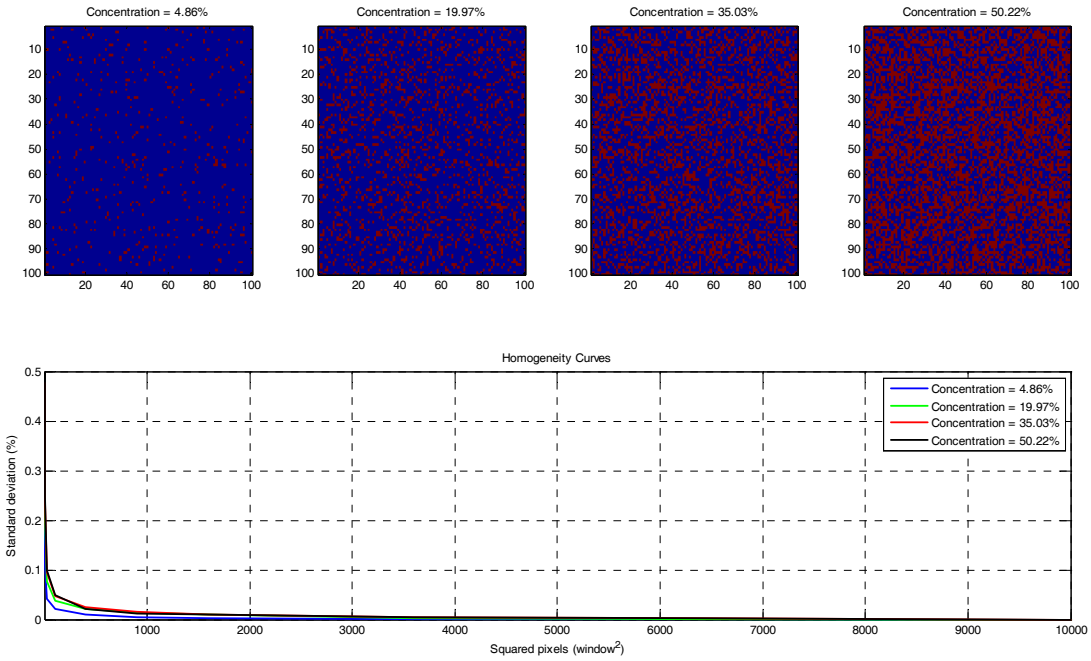


Figure 32 - The concentration in the four simulated chemical images is varied but the particle size and distribution ('homogeneity') are maintained.

The results from Figure 30-32 suggest that the homogeneity curves may contain more structural information on for example particle size and the distribution of a component (homogeneity or randomness) in the samples, than is revealed in any of the earlier mentioned evaluation methods in this chapter. This information and possible relationships may be very relevant information and hence needs to be investigated further (cf. suggestions Section 9.2).

7.5 Other image processing techniques

MASKING TECHNIQUE[g]

NIR-CI analysis of larger sample areas is significantly faster using global imaging compared to line mapping technique (Section 4.3). In most NIR-CI analysis using line-mapping only a selected rectangular area of the tablet surface is analysed. This obviously contributes to the sampling issue discussion of how representative this area is of the tablet surface (more on this in Section 8.2). This particular sampling issue can however be eliminated by analysing the entire tablet surface area. Tablet formulations are however not produced in a rectangular form and this will therefore yield an area with pixels not belonging to the sample. These pixels must be excluded in the data processing. A matlab-routine was developed for such an analysis masking the pixels not belonging to the sample. An example of this is shown in Figure 33, where the entire tablet surface of a 6 mm diameter tablet is analysed by CLS and compared to a similar analysis of a 3 × 3 mm area of the same tablet surface.

[g] This is a freeware function ('BACRA') available from http://www.models.life.ku.dk/users/jose_manuel_amigo/index.htm

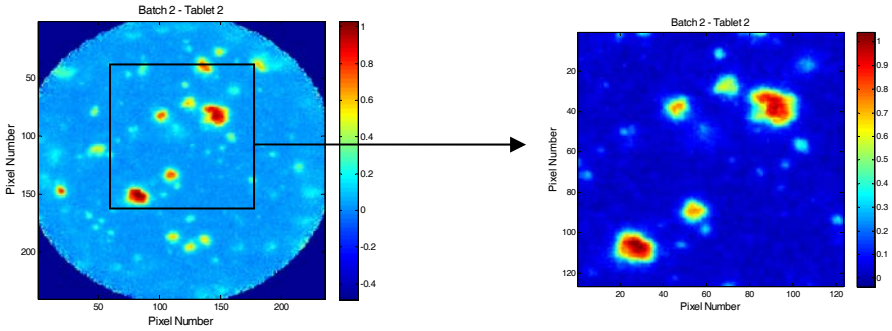


Figure 33 - Chemical API image by Classical Least Squares (CLS) of the entire 6 mm diameter tablet surface (left), and a 3×3 mm area of the same tablet surface (right). The sample is a two-component tablet formulation containing 5 % w/w API (Batch 2 - Tablet 2, Figure 36, and Paper II).

COMPOSITE IMAGES

A common way in the NIR-CI literature to visualise the distribution of multiple components in a sample simultaneously is to produce composite images.[15,71] Composite images provide a visualisation of the relative distribution of components and are appealing because multiple components are shown together.

A composite image can be produced as a red, green and blue (RGB) image in an M×N×3 format. However, such a composite image is usually generated from binary images, i.e. it involves the difficult threshold value selection of the chemical images discussed in Section 7.3. Some of the problems and approximations needed for producing composite images have been discussed.[44,71] The difficult task of threshold value selection often results in heavy overlaps of components in composite images, which may confuse the interpretation. Sasic[71] describes how re-assignment of overlapped pixels subjectively can be performed, but concludes it will generate arbitrary composite images that may even be misleading.

A RGB composite image based on binary images of API, cellulose and lactose from a five-component tablet formulation is shown as an example (Figure 34d). The three chemical images used to produce the composite image and the threshold values selected for each chemical image are also shown (Figure 34a-34c).

Another way of producing a composite image is to directly overlay the chemical images (Figure 34e). Each of the chemical images of API, cellulose and lactose (Figure 34a-34c) are first transformed into a red, green and blue image, respectively. The colour intensity of each pixel in a red, green or blue image will be proportional to the intensity values from the chemical image it is generated from. The three red, green and blue images representing the three components will finally be overlaid. This composite image will naturally be full of overlaps but the high intensity areas of each component will protrude and thus show the distribution of these areas for each component (Figure 34e).

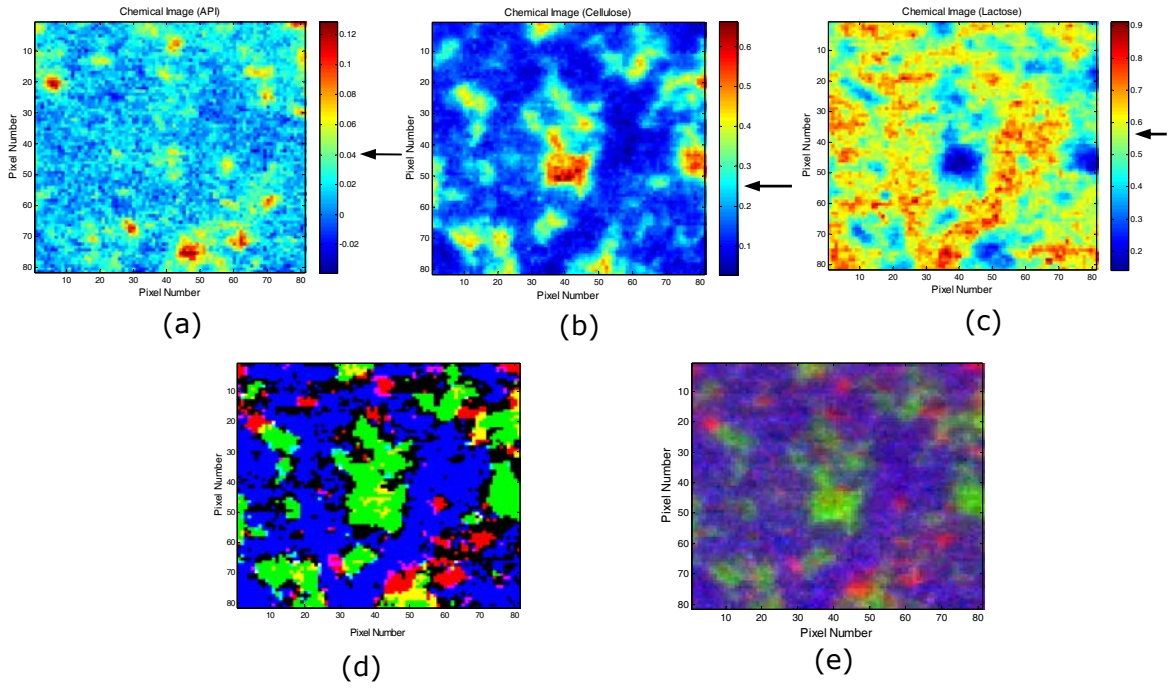


Figure 34 - Chemical images of (a) API, (b) cellulose, and (c) lactose of a five-component tablet formulation. A red-green-blue composite image (d) is produced from binary images of the three chemical images (a-c) generated from threshold-values indicated by the black arrows at the colour-bar for each image. An alternative composite image (e) is produced from directly overlaying the chemical images.

As discussed above, composite images are appealing but not without issues and care should be taken in their interpretation. The individual chemical images are a more visual reliable outcome of the experiments and these chemical images have therefore been selected as the primary image presentation form of NIR-CI analysed samples throughout this project.

CHAPTER 8

NIR-CI IN PHARMACEUTICAL APPLICATIONS

The theoretical framework and method descriptions necessary for performing a complete and successful NIR-CI experiment were presented in the previous chapters. This work is unified in this chapter presenting different studies of NIR-CI applied to pharmaceutical samples utilising one or more of the developed data processing method and image processing tools.

8.1 Effect of different API particle size on tablet homogeneity

The results from this study are the foundation of paper II. This section and paper II can be read separately, but are obviously overlapping and complementary. There will be results presented in this section not included in paper II and some of the work is on the other hand more detailed described in paper II. The study demonstrates how NIR-CI can be integrated in formulation development of tablets. The effect of using different API particle size on its distribution within the tablets is examined.

During formulation development of tablets the effect of API particle size will be evaluated for its influence on the process and final product performance. As revealed from Equation 5 more uniform distribution of API is in general obtained per unit dose (tablet) with increased number of particles, i.e. with finer (smaller) API particle size. However, small particle sizes produce cohesive powders, which have poor flow properties, may cause segregation when mixed with larger particle sized excipients, and will tend to agglomerate. This will require use of a high mixing efficiency. Using a larger API particle size will provide better flow properties and generally be easier to mix. However, it will also be more prone to segregation, especially if the particle sizes of API and excipients are too different. Further, larger API particle size may also be difficult to mix to a high degree of homogeneity if the content of API is low.[12,108] The mentioned issues on particle size and mixing efficiency demonstrate a number of factors to be aware of. It is therefore important to be able to analyse and characterise the effects from these factors and NIR-CI may be an extremely helpful tool in that context.

Three different sieved fractions of API particle size were utilised in this study to produce three different tablet batches by direct compression of a two-component tablet formulation. All other process factors were kept constant. Four tablets from each of the three batches were analysed by NIR-CI line-mapping (Chapter 5), data processed by CLS (Chapter 6) and evaluated by different image processing tools (Chapter 7).

Materials and methods (see also Paper II)

A two-component powder blend consisting of 5 % (w/w) active pharmaceutical ingredient (API) (a proprietary compound) and 95 % (w/w) silicified microcrystalline cellulose (ProSolv SMCC® HD90, JRS Pharma, Germany) was used to produce the three different batches. The three batches were manufactured with the following different API particle size fractions: (1) < 125 µm; (2) 125 – 355 µm; and (3) > 355 µm, in the following referred

to as batch 1, batch 2 and batch 3, respectively. Each blend of 150 g was loaded in a 1 L stainless-steel container (~ 50 % of total container volume) and mixed on a lab-scale Turbula® mixer for 5 minutes at 70 rpm.

Immediately after mixing, the two-component powder blend was tabletted on a single punch Diaf tablet press. The tablets manufactured had a flat surface, diameter of 6 mm and aiming at target weight 100 mg and target hardness 75 N. Pure compound reference samples of the API and microcrystalline cellulose (MCC) were also produced for the classical least squares (CLS) image analysis.

A near-infrared line mapping system (Spectrum Spotlight 350 FT-NIR Microscope, PerkinElmer, UK) was utilised to obtain the hyperspectral raw data. The following instrumental settings were used: Area: 3 x 3 mm; Wavenumber range: 7800 – 4000 cm^{-1}; Pixel size: 25 x 25 µm; Spectral resolution: 16 cm^{-1}; 8 scans to average a single spectrum.

CHEMICAL IMAGES

The immediate generated chemical image of a sample is presented as non-scaled. The non-scaled chemical API image of 'tablet 2-batch 1' is compared to its scaled version (from 0 to 1) in Figure 35.

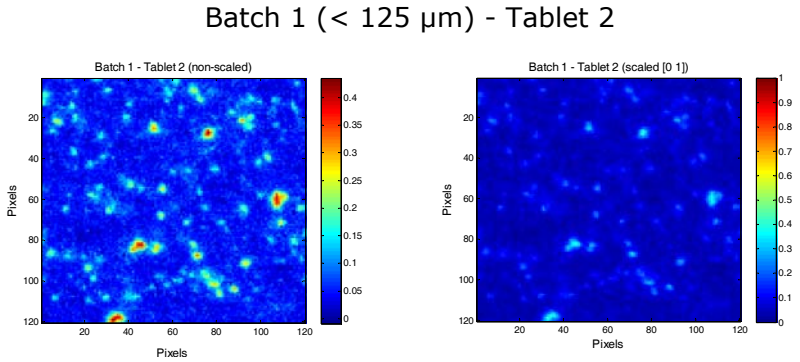


Figure 35 - The chemical API image generated by CLS of tablet 2 from batch 1 as non-scaled (left) and scaled [0 to 1] (right).

Figure 35 illustrates that it appears seemingly easier to visually interpret a chemical image of a single sample in its non-scaled version. However, to compare chemical images of different samples it is necessary to bring to same scale. The chemical images of API generated by CLS of the four tablets from each of the three batches having different ingoing API particle size are shown in Figure 36.

The scaled chemical API images corresponding to the finest API particle size (Figure 36, upper row) display a higher number of smaller domain sizes than observed for the chemical images of API produced from medium particle size and large particle size (Figure 36, medium and lower row). The four images of tablets from batch 3 with API particle sizes above 355 µm only display large API domains for tablet 2 and 4. The reason is that there simply are no (large) API domains in the 3 x 3 mm sampled area of tablet 1 and 3. This is confirmed by the low CLS concentration predictions as appears from Table 8. The table displays the histogram statistics of the chemical images from Figure 36. It confirms the low content of API in tablet 1 and 3 (2.5 % and 3.4 %^{h}) compared to tablet 2 and 4 (12.5 % and 10.2 %) for batch 3.

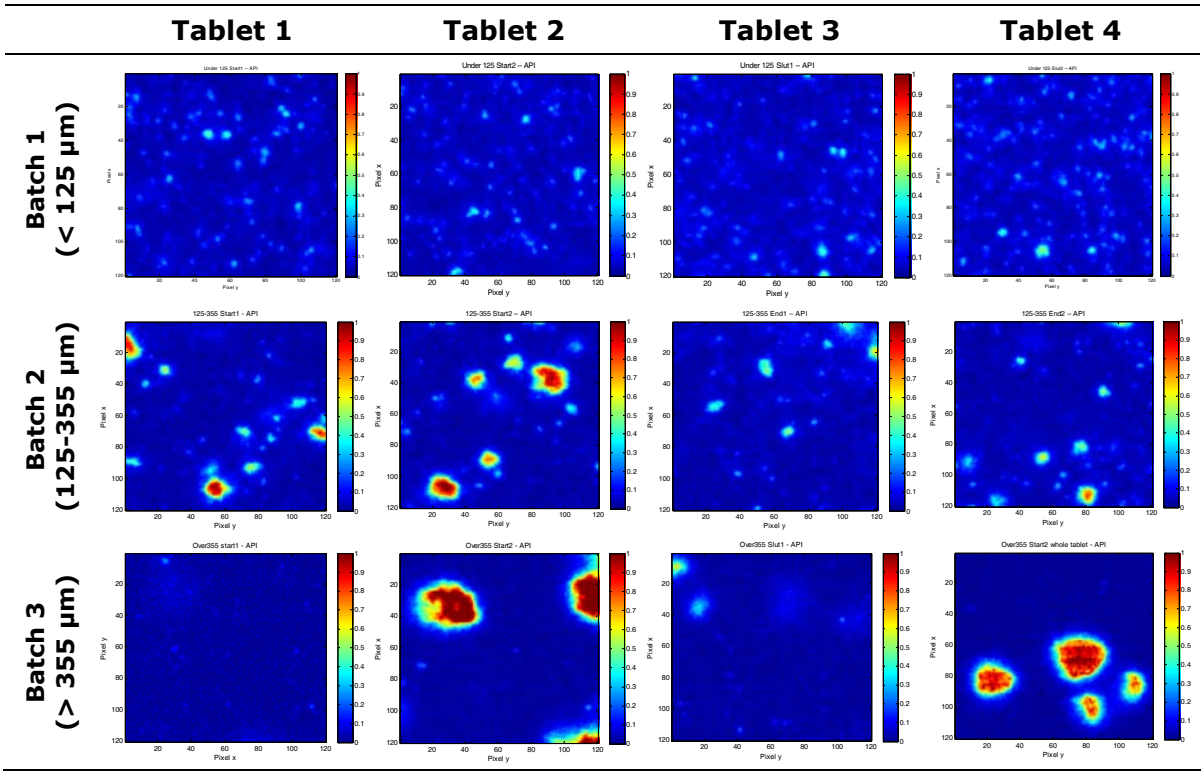


Figure 36 - Chemical images of API generated by the CLS algorithm. The chemical images are of four tablets (3x3 mm) from each of the 3 batches with different API particle sizes: < 125 µm (batch 1, upper row); 125-355 µm (batch 2, middle row); and > 355 µm (batch 3, lower row). All images are scaled from 0 to 1.

	Batch 1 (<125 µm)		Batch 2 (125-355 µm)		Batch 3 (>355 µm)	
	Mean	Std.	Mean	Std.	Mean	Std.
Tablet 1	6.0	4.3	7.2	11.0	2.5	2.0
Tablet 2	6.2	3.9	8.8	14.5	12.5	25.3
Tablet 3	7.1	4.2	5.9	5.8	3.4	4.3
Tablet 4	7.6	4.7	6.4	7.1	10.2	21.2
Mean	6.7		7.1		7.2	
Std-b (between)	0.76		1.3		4.9	
Std-w (within)		4.3		9.6		13.2

Table 8 - Histogram statistics of the CLS generated chemical API images from Figure 36 showing the mean value and standard deviation (Std.) of the API content predictions for each of four tablets within the three batches (CLS predictions × 100)[h]. The standard deviation Std-b is the standard deviation of the four mean values for each batch. The standard deviation Std-w is the mean of the four standard deviations (Std.) for each batch.

[h] The unit '%' for CLS predictions (× 100) means percentage of the pure component compact used for CLS analysis and should not be mistaken by the common perception of '%' as relative content in the actual sample. Mostly, though, the CLS prediction of a component approximates the mean percentage of the component in the actual sample (cf. Section 6.3, CLS).

The objective and quantitative histogram statistics from Table 8 are well in accordance with the qualitative results from the visual evaluation. The mean of the mean values for all four tablets for each batch are quite similar but the standard deviation (Std-b) increases from 0.76 % to 1.3 % and to 4.9 % for batch 1, batch 2 and batch 3, respectively. This reflects the increasing variation in content prediction between the four tablets from each batch and suggests an increasing in-homogeneity of API in the tablets from batch 1 through to batch 3. The standard deviation (Std-w) reflects a weighted mean value of the variation in pixel content predictions within each of the four tablets from each batch. The Std-w increases from batch 1 through to batch 3 (4.3 %, 9.6 % and 13.2 %, respectively). These results also indicate a decreasing uniform distribution of API in the tablets going from batch 1 to batch 3.

The histogram plots of the chemical images of tablet 2 for each of the three batches are shown as representative examples in Figure 37. This is to illustrate the basis of the statistic results in Table 8. An important difference between these histograms is the length and size of the tail of pixel predictions that extends toward the high CLS prediction values. The long tail of pixels for 'batch3-tablet2' represents regions with high API contribution (i.e. high API concentration). The short tail of pixels for 'batch1-tablet2' also represents regions of API but of much lower concentration. These results support what is revealed from the chemical images (Figure 36). However, the histogram statistic information is spatially independent and hence does not reveal anything about the domain size of the API regions.

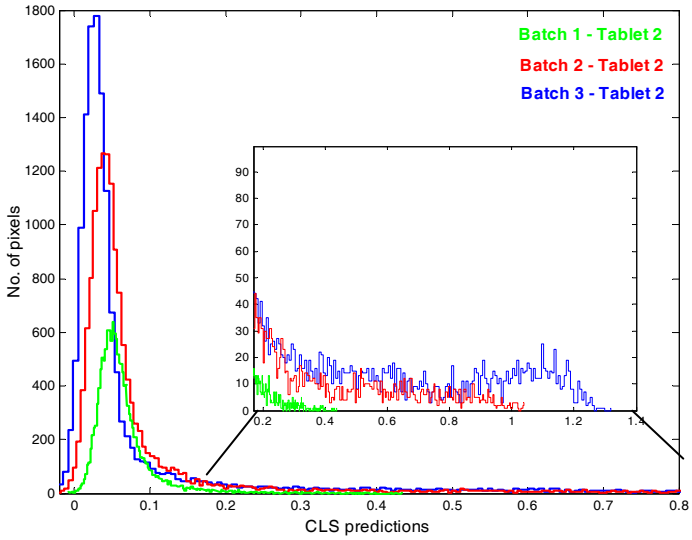


Figure 37 - The histogram plots of tablet 2 for batch 1 (green), batch 2 (red) and batch 3 (blue).

The large and highly concentrated API domains can also be revealed from the measured NIR diffuse reflectance raw spectra (Figure 38). The spectra behind regions in the chemical images with high versus low concentration of API are shown for tablet 2 from batch 1 (Figure 38a) and batch 3 (Figure 38b), respectively. For reference the spectra of the pure components are also shown (Figure 38c). The difference between the spectra is high

(red) and low (blue) concentrated API regions. This difference is more pronounced for the spectra of batch 3 (Figure 38b). Note also the high resemblance of the red spectra for batch 3 and the pure API reference spectrum. This indicates an API rich domain not only contained in the surface at this particular location of the imaged sample but likely spanning the entire information depth (cf. Section 4.5).

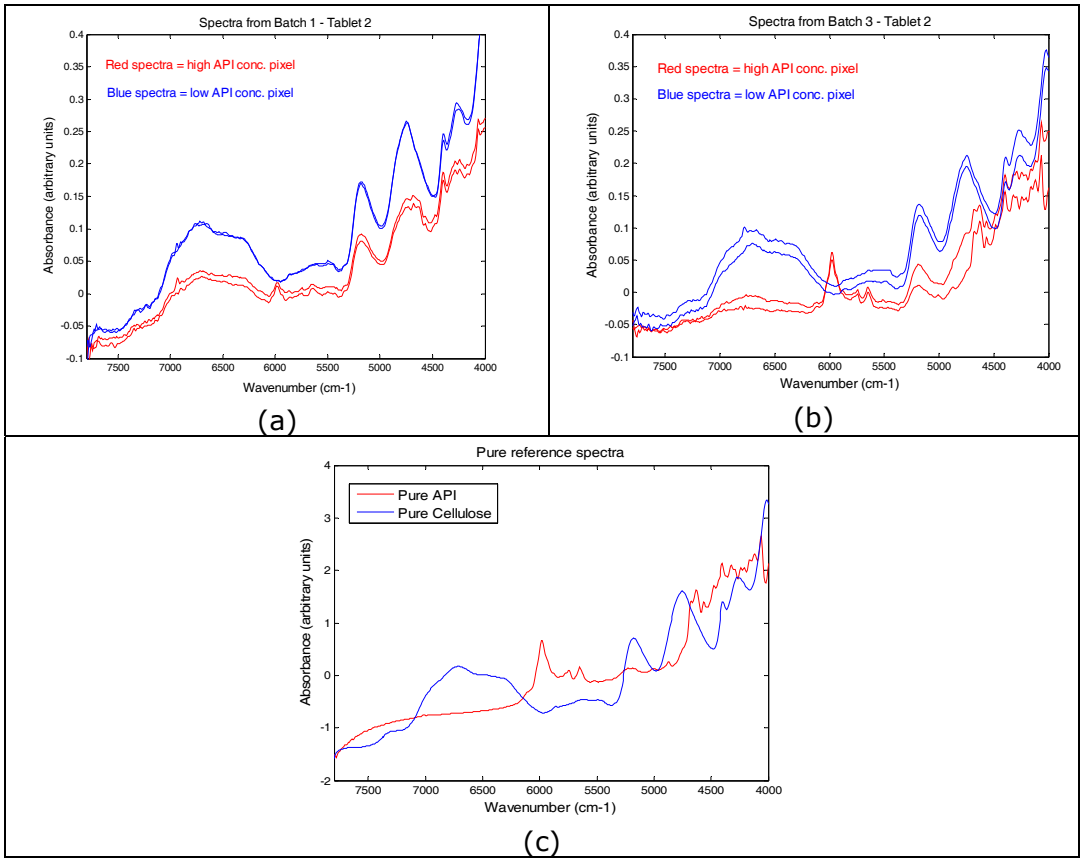


Figure 38 - NIR diffuse reflectance spectra from the hyperspectral data cube of tablet 2 from batch 1 (a) and tablet 2 from batch 3 (b). The red and blue spectra correspond to regions of high and low API concentration, respectively. The pure component reference spectra of API and the excipient (cellulose) are shown for comparison (c).

BINARY IMAGES AND DOMAIN STATISTICS

With the evident visual differences in API domain sizes between the three batches in this study, domain statistics is an obvious image processing tool to compare and evaluate the chemical images. The binary images are generated from the chemical images using a threshold value of 0.15 for all images (Figure 39). The method used to determine this threshold value is described in details in Section 7.3 and (II). From the binary images the effect of different ingoing API particle size on its size and distribution in the final tablet is even more evident compared to the corresponding chemical images in Figure 36.

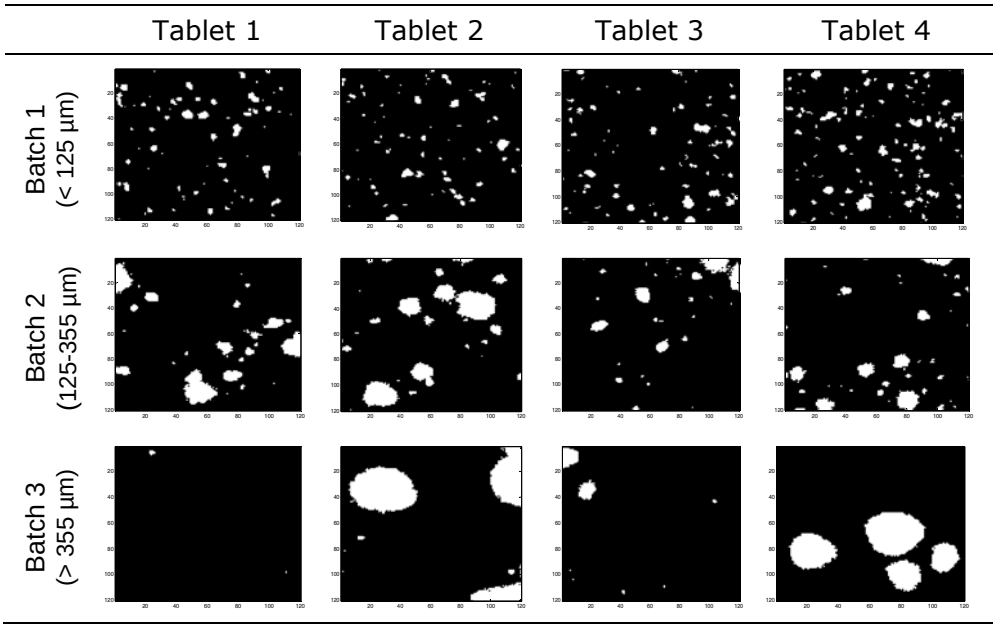


Figure 39 - Binary images generated from the chemical images in Figure 36 using a threshold value of 0.15.

While absolute particle size could not be determined from the binary images (Section 7.3), the API domain size statistics above the spatial resolution (25 x 25 µm) was calculated (Table 9). These results clearly confirm the earlier visual observations with the number of domains decreasing but the size of domains increasing from batch 1 to batch 3. The results for batch 3 are differing even more from the other batches if the two tablets (no. 1 and 3) with hardly any API are excluded (results in brackets, Table 9).

	No. of domains			**Mean equivalent diameter (µm)**			**Area covered (%)**		
	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3
Tablet 1	69	30	2	71	127	69	4.1	8.2	0.1
Tablet 2	68	25	9	66	158	299	3.6	11.0	16.4
Tablet 3	71	31	7	73	95	124	4.5	4.7	1.9
Tablet 4	104	42	4	73	96	659	6.6	6.0	15.9
Average	78	32	6 (7)	71	119	288 (479)	4.7	7.5	8.6 (16.2)
RSD (%)	22	22	57	5	25	92	28	37	102

Table 9 - Domain size statistic results of API based on binary images generated from CLS chemical images of API and a threshold value of 0.15. For batch 3 average results only calculated for tablet 2 and 4 are shown in brackets, i.e. excluding the two tablets with hardly any API.

It should be noted that the mean equivalent diameter for batch 2 (119 µm) and for batch 3 (288 µm) are lower than the actual ingoing API particle size fractions (125-355 µm and > 355 µm, respectively). This is probably due to the often several "unit-sized-pixel" (here 25 x 25 µm) domains appearing in most binary images. Whether these "unit-sized-pixel" particles/domains are representative of the reality in the tablet sample is questionable, but they impact to a lower calculated mean equivalent diameter. There is definitely room

for optimising this domain size statistic methodology. This should include when or how to include/exclude the "unit-sized-pixel" domains but also how to separate contiguous features in a binary image. Different mathematical morphology and image segmentation method are probably needed to refine this method.

As a consequence, the obtained results in Table 9 are only a rough estimate of the domain sizes in the imaged sample. However, such results are still very valuable when comparing samples or determining the relative size changes e.g. as a consequence of changed process parameters.

HOMOGENEITY CURVES

The homogeneity curves (Section 7.4) obtained for the 12 tablet samples in this study (4 tablets from each of the 3 batches) are shown in Figure 40.

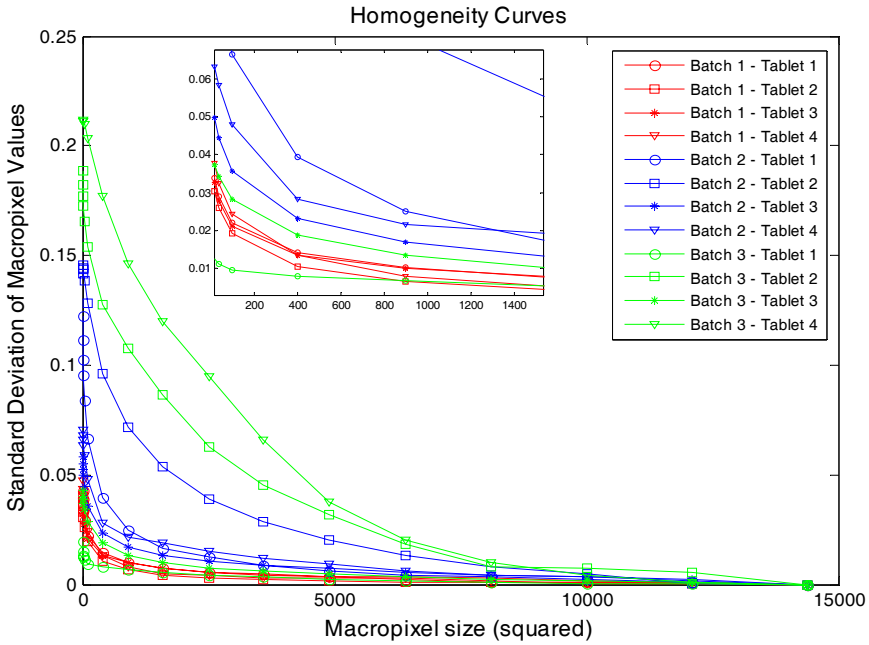


Figure 40 - Homogeneity curves showing standard deviations of CLS predictions versus squared macropixel size. The curves are shown for each of four tablets from each of the three batches. The insert enlarges the area around origin.

Different approaches for evaluating homogeneity curves were discussed in Section 7.4. A visual evaluation and calculation of the area under the homogeneity curves (AUHC) is utilised here.

The homogeneity curves in Figure 40 were generally positioned lower in the graph for tablet samples with smaller API domains more evenly spread out on the tablet surface compared to tablet samples from batch 3 where larger API domains were distributed less evenly throughout the tablet surface. The two curves for tablet 1 and 3 for batch 3 break this pattern but they actually just confirm the functionality of this methodology. As mentioned before, there were hardly any API in these samples and they therefore erroneously appear well blended. These homogeneity curves are therefore found among the homogeneity curves for the more homogenous batch 1 (see insert, Figure 40). Such de-

viating samples would always easily be captured by comparing the homogeneity curves with the low CLS predictions of API (Table 8) or assay values from other conventional analytical methods.

As earlier mentioned it is always desirable for image processing tools to move beyond subjective and qualitative analysis (as for visual evaluation) and determine objective and quantitative measures. The trapezoidal area under the homogeneity curves (AUHC) are determined as such a measure for the chemical API images (Table 10).

AUHC	Batch 1	Batch 2	Batch 3
Tablet 1	49	114	(30)
Tablet 2	30	305	457
Tablet 3	53	99	(67)
Tablet 4	42	122	594

Table 10 - The area under the homogeneity curves (AUHC) from Figure 40 calculated by trapezoidal integration. The results from the deviating tablet 1 and 3 from batch 3 are shown in brackets.

A sample is considered homogenous having low standard deviations between macropixels and thus a lower-lying homogeneity curve and low AUHC-values. It is seen that the AUHC-values are in accordance with the position of homogeneity curves in Figure 40 and also the results from earlier image processing approaches. The AUHC-values for batch 1 are lower than batch 2 that again are lower than batch 3. Tablet 1 and 3 from batch 3 will once again, due to the lack of API in the samples, erroneously be assessed homogenous samples from their low AUHC-values.

CONCLUSION

From the results presented in this study and (II) near-infrared chemical imaging has been demonstrated to be a useful analytical technology for the study of tablets from a formulation development experiment. Domain sizes of API and their spatial distribution within tablets were determined and the homogeneity of each tablet evaluated based on this information. Valuable information, that is not readily obtainable by other analytical techniques. Small ingoing API particle sizes resulted in small and well-distributed API domains in the manufactured tablets. With larger ingoing API particle sizes the API domains increased within the tablets and the API distribution became worse. The results are in line with the findings from a similar Raman chemical imaging study[46] but opposite to an equivalent NIR-CI study.[16] Both of these referenced studies determined the API domain size in tablets using different ingoing API particle size. The opposite results from the NIR-CI study were caused by agglomeration of the smallest API particle size.[16] All these studies emphasise the valuable information and useful role the chemical imaging technologies can play in tablet formulation development experiments.

This study investigated the effect of using different API particle size on the size and distribution of API in the final tablet formulation. However, the main theme was the evaluation of tablets by near-infrared chemical imaging and in particular the evaluation of re-

sults obtained using different image processing tools (Chapter 7). The study could therefore be considered a model-study as the tablets and evaluation approaches used could just as well have been from another formulation development study, e.g. studying the a powder mixing process (Section 2.4). Hence, a general approach for use of NIR-CI in formulation development studies of tablets has been demonstrated.

8.2 Sampling issues

The valuable role of NIR chemical imaging to characterise solid dosage form matrices and gain increased understanding from the unique spatial distribution information of components provided has been demonstrated in this thesis and from the numerous applications published (cf. Table 1). In this project, it has also been shown that accurate component predictions and thus accurate chemical images can be obtained using NIR-CI (Section 6.4 and I).

Despite the valuable information and accurate determinations provided by NIR-CI, the conclusion from most of these studies has been based on chemical images from the sample surface of one or a few tablets. To further develop the NIR-CI technology, e.g. for use in quality assurance (QA) and quality control (QC), there is a need to define representative sampling sizes. This means determining how representative a single chemical image is of a tablet, and further how representative that tablet is of the whole batch. To the author's best knowledge no studies have been published trying to thoroughly investigate these sampling issues.

REPEATABILITY STUDY

The importance of obtaining reliable results from repeated measurements of the exact same sample, the exact same area and under identical conditions was discussed in relation to the large scattering nature of diffuse reflectance measurements of solid-mixtures (Section 4.5). In the small experiment presented here, the repeatability is studied for eight repeated measurements (sample 1-8) on a single tablet. The analysed tablet was collected from batch 1 in the experiment discussed above (section 8.1 and II).

Materials and methods

A single tablet sample from batch 1 (section 8.1, II) was analysed. This tablet formulation was composed of 5 % (w/w) API (a Novo Nordisk proprietary compound) sieved to a particle size < 125 µm and 95 % (w/w) microcrystalline cellulose (ProSolv SMCC® HD90, JRS Pharma, Germany, non-sieved). The tablet was 6 mm in diameter with a flat surface, so no sectioning was necessary. The tablet was fixed onto a microscope slide using cyanoacrylate glue and measured directly on the flat tablet surface.

A near-infrared line mapping system (Spectrum Spotlight 350 FT-NIR Microscope) was utilised to obtain the hyperspectral data cube. An area of approximately 3 × 3 mm was analysed using a pixel size of 25 × 25 µm obtaining an image of exactly 125 × 124 pixels (= 15500 spectra) for each of the eight measurements. Each spectrum was the average of 8 scans from wavelength region 7800-4000 cm^{-1} using a 16 cm^{-1} spectral resolution. The eight measurements (sample 1-8) were conducted in succession analysing the exact same area and using the same background measurement for correction. The collected hyperspectral images were analysed by CLS (Section 6.3) and the mean CLS prediction value of API and the standard deviation were calculated for each chemical image.

All eight CLS generated chemical images of each component appeared visually similar. Hence, only the chemical images of API and microcrystalline cellulose for the samples with the lowest and highest mean API prediction value from the CLS analysis are shown (Figure 41). Table 11 displays the mean and standard deviation of all CLS pixel prediction values for the two components for each of the eight sample measurements.

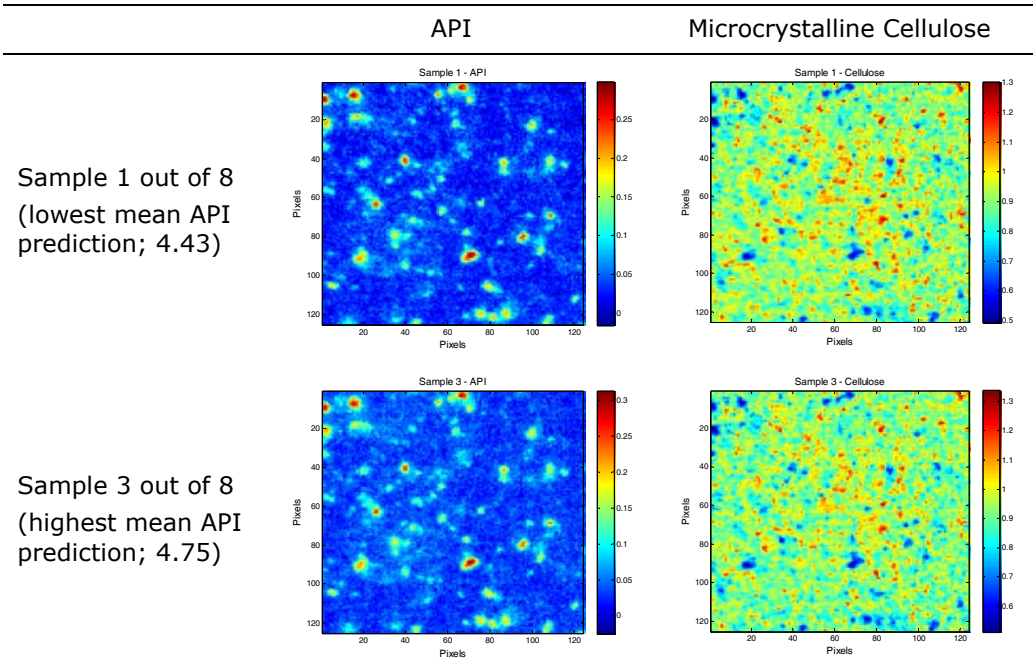


Figure 41 - Chemical images (non-scaled) of API and microcrystalline cellulose for two samples in a repeatability study measuring the exact same tablet and area eight consecutive times. The two samples shown are from the measurements giving the highest and lowest mean API prediction value from the CLS analysis.

	Mean API	Std. API	Mean Cellulose	Std. Cellulose
Sample 1	4.43	3.06	90.95	8.95
Sample 2	4.62	3.03	91.60	8.83
Sample 3	4.75	3.05	92.22	8.99
Sample 4	4.67	3.06	92.49	9.04
Sample 5	4.68	3.07	92.60	8.99
Sample 6	4.63	3.06	92.80	8.97
Sample 7	4.58	3.05	92.79	8.93
Sample 8	4.54	3.06	92.82	8.91
Mean	4.61	3.06	92.28	8.95
RSD (%)	2.1 %	0.4 %	0.7 %	0.7 %

Table 11 - The mean values and standard deviations (Std.), i.e. histogram statistics, of the two components (API and microcrystalline cellulose) from the eight CLS generated chemical images (CLS predictions × 100). The eight samples are from a repeatability study analysing the exact same tablet and same area of the tablet eight consecutive times.

The results in Table 11 prove the high precision (repeatability) of the NIR-CI measurements with the relative standard deviation (RSD) of API and microcrystalline cellulose mean content predictions from CLS analysis only being 2.1 % and 0.7 %, respectively. The results also show that the variation within each measurement is quite similar for all eight determinations (RSD 0.4 % and 0.7 % for API and cellulose, respectively).

The results confirm that despite the complex and deep scattering from the NIR diffuse reflectance measurement in a NIR-CI data acquisition, reliable and reproducible results will be obtained and the spatial resolution maintained. This finding is actually a very important finding to consolidate the validity of chemical images from NIR-CI experiments.

CROSS SECTIONS OF A SINGLE SAMPLE – VOLUMETRIC CHEMICAL IMAGE

It was earlier discussed how care should be taken in the evaluation of chemical images and in particular for absolute determinations of particle/domain sizes of components (Section 4.5, 7.3 and 8.1). The domain size results will depend on the sample depth profile analysed and will generally underestimate domain size of spheres and overestimate the size of cubes.[44] This sampling issue is related to how representative a single chemical image is of a tablet and is investigated by the example shown below.

A single tablet was cross-sectioned 12 times. Each cross-section was analysed by NIR-CI to obtain hyperspectral data cubes from 13 surfaces of the tablet at different sample depths. First, the tablet was analysed by NIR-CI at the sample surface. A layer of approximately 200 µm was then removed by microtoming (Section 5.1) and the newly exposed surface was analysed. This process was repeated until hyperspectral NIR data cubes were obtained from all 13 surfaces (from tablet surface to bottom). Each surface was analysed by identical data acquisition conditions and it was attempted to analyse the exact same area for each sectioned surface according to described approach by Clarke et al.[36] For the best possible illustration purpose, a presumably heterogeneous tablet was selected for this study. This would increase the chances to observe some significant features or differences in the cross-section layers. Further details of sample and data acquisition are given below.

Materials and methods

The single tablet used for analysis was composed of 10 % (w/w) API (a Nordisk proprietary compound) and 90 % (w/w) microcrystalline cellulose (AvicelPH102, FMC Biopolymer, Philadelphia, USA). A 150 g powder batch of this formulation was produced by mixing the two powders in a 1 L stainless steel container for half a minute (70 rpm) in a lab-scale Turbula® mixer (WAB Maschinenfabrik, Muttenz, Switzerland). Immediately after mixing, the powder blend was tabletted on a single punch Diaf Tablet press. The tablets had a flat surface, diameter of 6 mm, weight of 100 mg and a thickness of approximately 2.6 mm.

A single tablet was selected from the tablet batch and fixed onto a microscope slide using cyanoacrylate glue. The tablet was first measured by NIR-CI directly on the flat tablet surface. The tablet was then sectioned using a microtomer (Leica EM Trim, Leica Mikrosysteme GmbH, Vienna, Austria) removing a top layer of approximately 200 µm. The newly exposed tablet surface was now analysed by NIR-CI and this process was repeated until 12 layers of 200 µm had been removed and a hyperspectral data cube obtained from each surface. For each NIR-CI surface analysis it was attempted to analyse the exact same surface location similar to the approach described by Clarke et al.[36]

A near-infrared line mapping system (Spectrum Spotlight 350 FT-NIR Microscope) was utilised to obtain the hyperspectral data cubes. For each measurement the same area of approximately 2 × 2 mm was analysed us-

ing a pixel size of 25 × 25 µm obtaining an image of 120 × 120 pixels (= 14400 spectra). Each spectrum was the average of 8 scans from wavelength region 7800-4000 cm^{-1} using a 16 cm^{-1} spectral resolution. The thirteen measurements (sample 1-13) were conducted in succession using the same background measurement for correction. The collected hyperspectral images were analysed by CLS (section 6.3) to obtain the chemical images.

The chemical images of API for all thirteen surfaces analysed of the single tablet is shown in Figure 42. The non-scaled chemical images are displayed because it makes each image easier to visually interpret but recall that this makes it more difficult to directly compare the thirteen chemical images.

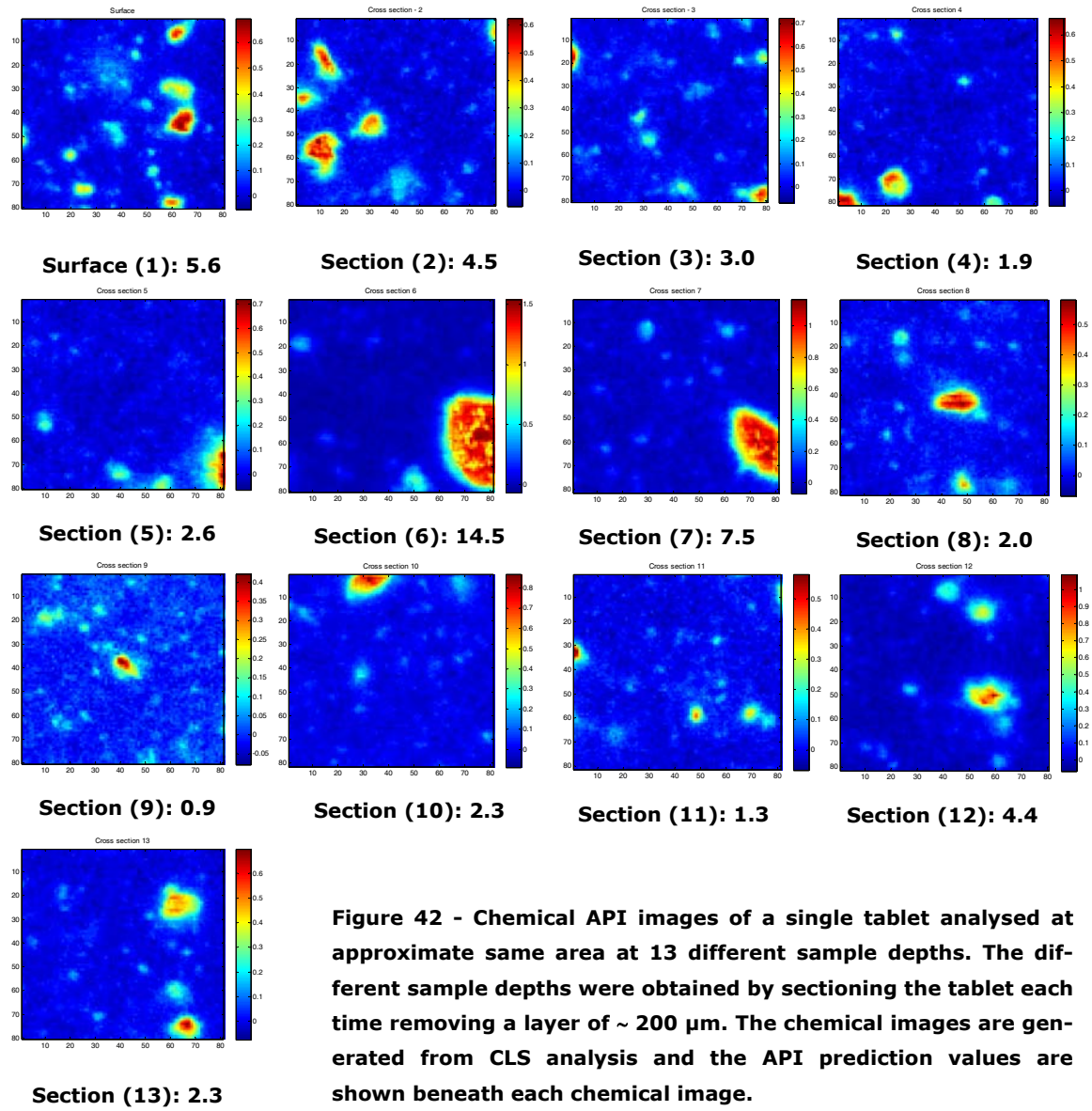


Figure 42 - Chemical API images of a single tablet analysed at approximate same area at 13 different sample depths. The different sample depths were obtained by sectioning the tablet each time removing a layer of ~ 200 µm. The chemical images are generated from CLS analysis and the API prediction values are shown beneath each chemical image.

The thirteen chemical images in Figure 42 clearly show the sampling issue of only analysing one surface of a single tablet, particularly in the case of heterogeneous tablets. The

chemical images differ from almost no API appearing in section (9) to the large clump of API in section (6). The estimated API content by CLS analysis varied from 0.92 to 14.4 with a mean value of 4.1 and a relative standard deviation on 90 %. The corresponding mean and standard deviation values for microcrystalline cellulose predictions were 75.0 and 8 %.

It is interesting to note how the smaller sized top and bottom of the large API clump from section (6) apparently appear in section (5) and (7), respectively. The API domain size statistics that would be calculated based on each of these chemical images would obviously be very different. This underlines the earlier point discussed on this matter; that determinations of particle/domain shapes and sizes can always only be an estimate when analysing one single chemical image.

The example shown here only analyses one single tablet, yet it addresses many of the sampling issues related to NIR-CI experiments. One of these sampling issues is representative sample size area, i.e. how large an area of a tablet should be analysed to be representative of the whole tablet. In this study a tablet surface area on 4 mm^2 out of a total tablet surface area of approximately 28 mm^2 was analysed. It is more than likely that measuring a larger sample area than 14 % of the total tablet surface area will improve the results and reduce the variation in content predictions; in particular for this apparently inhomogeneous tablet. However, even with a larger surface area analysed the very different results from the cross sections at different depths in Figure 42 indicate, that measuring one single surface is unlikely to be representative of the whole tablet. The solution may be to measure several cross sections of each tablet as performed in this experiment, to analyse several samples of each tablet batch (as in conventional pharmaceutical quality control) or even better to combine the two approaches. It will require a well-designed study and analysis of a lot more samples of different types to thoroughly investigate the sampling issues addressed here further. This study has provided inspiration for such further studies.

An obvious use of this cross sectioning approach is in formulation development of tablets. The time consuming process of sectioning and subsequent profound analysis of single tablets is not really an issue in formulation development. The cross-section analyses are therefore a powerful tool to scrutinize new formulations in the initial development phase to obtain detailed and valuable information about the distribution of components throughout the tablet.

With numerous cross sections of a single tablet it is possible to assemble them all in sequence and visualise the result as a three-dimensional volumetric chemical image. Prior to assembling, the binary images for the component of interest for each of the cross sections are first generated. The volumetric chemical image of API for the thirteen hyperspectral data cubes presented above are displayed in Figure 43. This shows the distribution of API throughout the tablet in a single image. Similar volumetric images can be made for each of the other components in a tablet. However, compared to the plot in Figure 43 the graphical computer plot of this volumetric chemical image is a much stronger visual tool, as it is possible to rotate the 3D plot and view the distribution from many different angles.

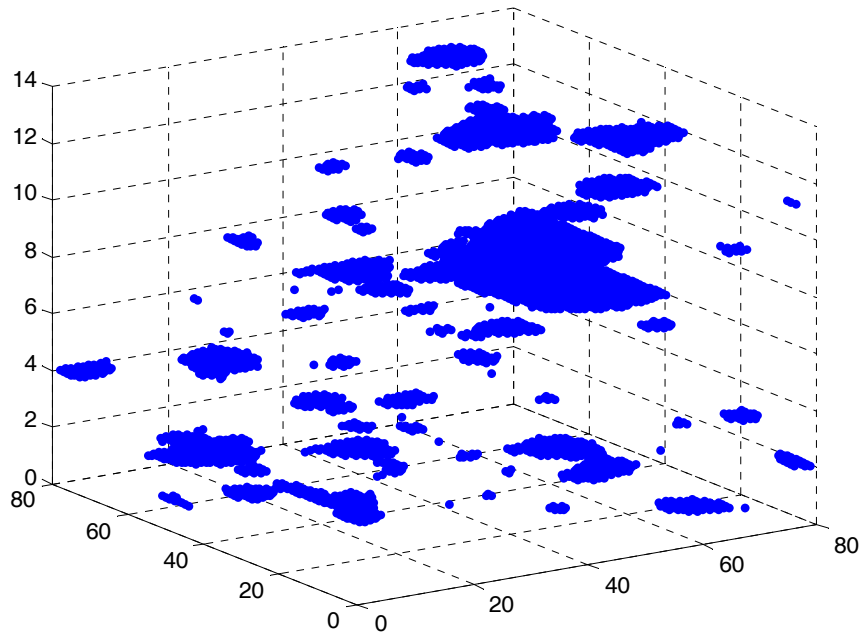


Figure 43 - Volumetric 3D chemical image of API in a tablet. The volumetric image is composed of thirteen NIR-CI analysed cross-sections of the tablet. The binary API image of each cross-section sample is created and assembled to produce the volumetric API image. It shows the distribution of API from the surface throughout to the bottom of the tablet.

CHAPTER 9

CONCLUSION AND FUTURE RESEARCH

9.1 Conclusion

The significant added value that near-infrared chemical imaging can provide in the development of new pharmaceutical tablet formulations was clearly demonstrated in this thesis. There are, however, a number of limitations and cautions that need to be considered when using this technology for pharmaceutical applications.

NIR-CI is capable of providing information about the composition and concentration, but what is unique for this technology is that it provides information on the size and spatial distribution of all components within a sample matrix. Having a thorough understanding of how to properly measure and analyse hyperspectral NIR data cubes is extremely important for obtaining this information. The NIR-CI data acquisition generates an immense amount of spectral data, and effective data processing methods and image processing tools are needed to extract accurate and meaningful information. All three steps in a complete NIR-CI experiment – i.e. data acquisition, data processing and image processing – are of equal importance for successful results. The emphasis in this thesis was on data and image processing.

A variety of different data processing methods have been described, and these methods constitute a 'data processing toolbox' that will be able to cover most NIR-CI applications in tablet formulation development. Furthermore, each of the steps typically involved in a full NIR-CI analysis is described in details, which provides a practical guidance for analysis of hyperspectral NIR data cubes. The choice of the appropriate data processing method will depend on the data set and the purpose of the analysis. Suggestions are given for which method to use and when.

Obviously, the aim of data processing is most often to generate the most accurate chemical images. However, the challenge for data processing methods, and generally for many aspects of NIR-CI, is that there exist no objective criteria for what the 'best' or most accurate results are. Such evaluation will for example require some kind of calibrated 'NIR-CI tablet'. In this thesis, the problem was solved by constructing and analysing a calibration data set. The data processing method providing the lowest prediction error from the subsequent regression analysis was considered to generate the most accurate chemical images. In such a study comparing three common data processing approaches, PLS1 was demonstrated to be the superior method, but with CLS as an excellent and relatively accurate alternative. Furthermore, the advantage in the general use of CLS is that it only requires the availability of the pure component spectra to work.

Moreover, the general applicability of using a calibration data set for accuracy and efficiency evaluations was further demonstrated. This was done by minor studies in which the most appropriate spectral pre-processing method and wavenumber range for data

processing analysis was determined. Other factors, such as instrumental settings and sample preparation, may also profitably be evaluated in this way.

In this thesis, it was emphasised and demonstrated that in order to fully exploit the potential of the NIR-CI technology it is necessary to move from the subjective interpretation of chemical images to a processing of the data into objective and quantitative information. Such objective measures are essential for discriminating similar samples and for deployment of the technology in QA/QC and for PAT applications.

Also demonstrated was the usefulness of the common image processing tools of histogram statistics and domain size statistics. However, their limitations were also exposed and there is still an unmet need to develop accurate methods that numerically describe the information contained in a chemical image, e.g. spatial arrangement, relative positioning of components and domain size statistics. This thesis provides a contribution to this important work. A new exciting tool was developed to evaluate the homogeneity of the generated chemical images. The principle builds upon the concept of histogram statistics but involves a much more extensive sampling of the chemical images ('macropixel analysis'). Different approaches to evaluating and extracting numerical values from the 'homogeneity curve' generated for each chemical image using this method were discussed. Homogeneity is always a major concern in the development of new tablet formulations, which is why this newly developed quantitative manner of describing homogeneity is seen as a promising and powerful image processing tool. Moreover, as discussed in the thesis it is plausible that the homogeneity concept can be linked to theoretical mixing theory of particulate systems, and it is possible that even more information can be extracted from chemical images using this approach (cf. section 9.2).

The structure of the tablet matrix is assessed as being the key factor in determining the overall quality of a tablet formulation. This makes the relevance and importance of the NIR-CI technology obvious, as the key information gained from NIR-CI is the capability of providing insight into the structure and function of the tablet matrix.

This was demonstrated in a tablet formulation development study which integrated NIR-CI and evaluated the effect on API distribution in tablets using different API particle sizes. Results from all the applied methods/tools were able to discriminate between batches with different ingoing API particle sizes. The study emphasises how the unique information obtained by NIR-CI and the processed data make it possible to evaluate and understand how the tablet matrix evolves as a function of the process or the component conditions used. The methodology approach presented in this application study can most likely be applied to other types of formulation development studies.

NIR-CI is still a maturing technology and there are some important factors, limitations and cautions to consider, as these may affect the overall quality of the final result, either individually or in combination. Only a few have been addressed and discussed in this work:

- Care must be taken in visual interpretation of chemical images. The goal is not high image contrast, but the right, i.e. most accurate, contrast, as concluded in Paper I.

- The chemical images can be quite sensitive to analytical settings such as the chosen pre-processing method, wavenumber range, and generation of pure component spectra.
- No appropriate objective method of threshold-value selection exists for generating binary images from chemical images. This will affect the domain size statistic results, which should be kept in mind when evaluating such results.
- Particle/domain size determinations from a single binary image must be evaluated with caution, as they can only ever be an estimate of the particle/domain sizes in the sample. Beware of over- and underestimations.
- Sampling issues should always be considered. How representative a single chemical image is of the whole tablet, and how representative a chemical imaged tablet is of the batch requires further investigations.
- Thoroughly and detailed understand aspects of data analysis methods, i.e. relate details of each method with the final result (chemical image).

For some of the existing factors/limitations there are no easy answers, if any. Nevertheless, it is important to be aware of them when considering the actual information that can be obtained from the chemical images. Some of the limitations are closely linked to some of the future investigations that should be performed to address these issues. Suggestions for such future investigations are presented in Section 9.2.

Despite the limitations and cautions pointed out, the unique information obtainable by NIR-CI is still invaluable for studying and gaining increased understanding of tablet formulations and other solid dosage forms. NIR-CI is probably most useful for comparing and discriminating formulations and for determining product changes in manufacturing or during storage. The current use is therefore primarily envisaged to be for development of new product formulations and for understanding manufacturing defects. To get adopted and deployed in QA/QC functions, or for at-line, in-line or on-line PAT applications, improvements are needed related to data acquisition speed, flexibility, instrument robustness, automated and robust data processing approaches, etc.

Finally, it should be emphasised that chemical imaging lead by NIR-CI is the only current analytical technique that can extract chemical information and information on component's size and spatial distribution from solid-phase samples, such as tablets or other solid dosage forms.

NIR-CI is a maturing technology, also in pharmaceutical applications, and this thesis adds a contribution to the knowledge foundation still being built. It is the intention and hope of the author that this thesis will provide both theoretical and practical knowledge useful for researchers working with NIR-CI, and in particular for applications related to new pharmaceutical formulations. The involvement of formulation scientists or process chemists is important, as detailed knowledge of the manufacturing process is often required to relate variations observed in a chemical image to the responsible unit operation.

9.2 Future research

The last few years increasing number of publications on the use of NIR-CI in pharmaceutical applications indicate that more and more pharmaceutical research groups are using and implementing the technology. The contributions have been trouble-shooting cases, product characterisation, merely data processing development, formulation development applications, studies of mechanistic understanding, reviews etc. (Table 1, Section 4.6). These studies and more of a similar kind are very welcomed as they all contribute to the knowledge basis for the NIR-CI technology.

An area in NIR-CI that needs more attention to further strengthen the value of the technology lies in understanding the correlation between NIR-CI data and product performance. It has been suggested that some of this information and knowledge are already available, e.g. at pharmaceutical organisations routinely employing NIR-CI, but that little incentive or inclination to publish this work exist for a variety of strategic reasons.[33]

The limitations and cautions described in the conclusion above (section 9.1) and also frequently described in literature[16,33,38,44] suggest the need to do more fundamental studies on NIR-CI. These studies should be designed and performed to clarify and obtain a better understanding of the fundamentals of the NIR-CI measuring technique. Hence, more NIR-CI studies on the fundamental understanding of the technology, trouble-shooting and formulation development experiments, mechanistic understanding of pharmaceutical processes etc. are all welcomed in the future. During this thesis work, ideas for different future studies and applications have emerged and suggestions for these are given below. Not all suggestions are equally straightforward to carry out but may promote important fundamental discussions or inspire to other relevant studies.

Suggestions for future studies:

1. Study the effect of sample preparation to determine the parameters providing the best sample surface. This will e.g. include testing different pharmaceutical materials (soft versus hard), comparing different sample preparation techniques, microtoming speed, type of cutting blade etc. The challenge is how to confirm the best sample preparation?
2. Study the impact of variation in angle presentation of sample surface to the measuring NIR microscope and how to detect when such variations occur.
3. Examine the effect of measuring the tablet sample surface compared to surfaces after sectioning. What are e.g. the effects of possible difference in density throughout a tablet?
4. Analyse tablets produced from same powder blend but manufactured with different compression force. The study could also include pure component compacts manufactured by different compression force (e.g. for CLS analysis) for further studying the aspect presented in this thesis (Section 6.3, CLS, Figure 18).
5. Evaluate effect on chemical images using different instrumental settings such as spectral resolution, number of scans and spatial resolution. The experiments

should be performed on the same sample and exact same analysed area. All these parameters have an impact on the total data acquisition time and it would therefore be valuable to obtain an idea of what instrumental settings will 1) provide the shortest acquisition time while still obtaining an adequate quality of the chemical images, and 2) provide an optimum as a compromise of chemical image quality and acquisition time.

6. Further investigate sampling issues discussed in Section 8.2, i.e. how representative is a single chemical image of a tablet and how representative is the tablet of a batch. The experiments could include comparison of results from different cross-sectioning layers of a tablet. For example, first analyse the entire tablet surface by NIR-CI and then divide the results up in different sub-area sizes and generate chemical images by data processing. Then cross-section the tablet by microtoming and repeat analysis. This can subsequently be repeated for additional cross-sectioning layers. Ideally, a calibrated NIR spectroscopy model is developed for the tablet formulation to non-destructively analyse the tablet by NIR spectroscopy prior to NIR-CI analysis and thus provide more accurate determinations of the bulk composition of the tablet.

7. Perform fundamental experiments to explain, and understand, what the NIR-CI instrument is actually measuring by the use of pharmaceutical model systems. The model systems could e.g. be 20:80 % w/w two-component formulations of a model API compound combined with excipients that 1) does not absorb NIR light (e.g. potassium chloride or calcium carbonate), 2) reflects light completely (e.g. titanium dioxide), and 3) absorbs light entirely (e.g. black iron oxide). For comparison and to mimic the usual situation, similar experiments with model system using typical excipients should be performed.[i] In fact, these studies have already been performed during this thesis work but were not evaluated by appropriate data and image processing methods.

8. It would in general be of great value if it was possible to compose model systems (of e.g. small 'bricks' or 'spheres' of well-characterised pure component material) where the exact composition, particle size and distribution was well-defined and could be precisely controlled. This would mimic the 'NIR-CI calibrated tablet' that has been in demand for many purposes throughout this thesis. These model systems could be used to study impact of particle sizes and forms, penetration depth, adjacencies of components etc., similar to the hypothetical situations presented in Figure 9, Section 4.5.

9. Construct and analyse a rather large calibration data set (e.g. 20 samples) of a tablet formulation composed of for example five components varying the concentration of each component. Ideally, a NIR spectroscopy method and HPLC analysis method should be available for reference analysis (for the API). Such an excellent data set will enable a number of experiments and factors to be studied, such as:

[i] This idea originates from Prof. Tony Moffat and Dr. R.D. Jee, The School of Pharmacy, University of London.

a. Compare different data processing methods for their ability to generate accurate chemical images, similar to (I) but this time for all components in the formulation (present in high and low concentrations). The data processing (e.g. PLS) calibration models could be generated with either HPLC reference data of the batch or even better from NIR spectroscopy reference data for each tablet (due to the non-destructive nature of NIR spectroscopy analysis).

b. Continue the detailed evaluation, scaling and interpretation of CLS prediction results from Section 6.3.

c. Compare the effects of using different spectral pre-processing or wavenumber ranges, similar to the minor study presented in Table 3, Section 6.3.

d. Sampling issues investigations, e.g. perform conventional content uniformity analysis on tablet batches according to the pharmacopoeias but analyse all tablets by NIR-CI prior to HPLC analysis.

e. With methods available to evaluate the final tablet formulation performance (such as dissolution, resistance to crushing etc.) it renders possible a number of experiments to study how formulation and process factors affect product performance.

10. Develop a general applicable methodology, where a combination of NIR spectroscopy and NIR-CI are integrated as the only two chemical analysis method used for a tablet formulation development study. The NIR spectroscopy should possibly be applied without developing a calibration model. The combination would rapidly and non-destructively provide a high level of information which would be useful in early formulation development, e.g. screening many different formulations, collecting many samples from several mixing studies etc., which normally requires a heavy wet-chemistry (typical HPLC) analytical work-load. The concept was introduced in (II) and similar approaches has earlier been described[18,49,110] but the method needs refinement for developing into a generic approach.

11. It is possible that NIR-CI can be used for different mechanistic powder compression investigations, studying e.g. particle arrangement, deformation, density distribution etc. However, it will require very accurate NIR-CI measurements and it is possible that Raman chemical imaging with the opportunity for higher spatial resolution is more appropriate for this purpose (or maybe a combination of both techniques in a chemical image fusion study[36]). A related type of experiment for mechanistic understanding of pharmaceutical processes was recently published measuring the distribution of density and tabletting force in pharmaceutical tablets by chemical imaging.[60]

12. More image processing tools to extract useful information from chemical images are desirable. As mentioned in section 7.4, work on simulated "chemical images" is on-going to fully understand the information that is contained within the homogeneity curves generated from macropixel analysis. Another approach could be to

determine the distances between API particles/domains (from binary images) as an estimate of the distribution uniformity (e.g. the Euclidean distance).

13. In continuation of latter suggestion (no. 12) specifically develop automated and objective method of threshold value selection. This step is essential for generating binary images and hence the subsequent extraction of information from these. This threshold-value method development could be facilitated using model samples composed of components having well-defined forms and sizes (see suggestion no. 8).
14. NIR-CI line mapping and global imaging systems are frequently theoretically compared in the literature and also in this thesis, but to the author's best knowledge no practical comparative study measuring the exact same sample(s) on both systems has been reported. Numerous experiments could be performed for a thorough comparison of the pros and cons for these two systems (and to further such studies Raman-CI could be included if available).
15. An interesting and obvious use of NIR-CI is for stability studies of solid dosage forms. NIR-CI is a non-destructive method and hence it is possible to measure the exact same area of the same tablet at different testing points and follow the changes over time in the tablet matrix, e.g. appearance of impurities, polymorph, solvate, or hydrate changes etc.

The suggestions given above for new NIR-CI research is mainly focused on fundamental experiments, methodology development or studies exploring what increased process and product understanding NIR-CI is capable of providing. This is in line with the primary objective of this project.

However, it is also relevant and one of the major goals for this technology to explore and eventually implement NIR-CI for (near) real time process monitoring and control. This means monitoring CQA and CPP during processing, i.e. develop NIR-CI into a complete PAT tool. There have been attempts going in that direction for the NIR-CI technology. El-Hagrasy et al.[18] designed a custom made NIR-CI set-up for powder blend homogeneity monitoring. A stationary NIR-CI system was set up close to a V-blender to collect images from two thirds of the top surface of the powder blend. During a 30 minutes blending process images were collected every second minute. The blending process had to be stopped for each image collection and the blender platform was rotated to obtain images from both sides of the blender. With this set-up it was possible to determine an optimal blending time between 14 and 16 minutes.

Bringing this concept even further, a prototype NIR-CI global imaging system has been mounted onto a V-blender for blend monitoring during processing (Figure 44).[85,111] In fact, this is a joint project of a pharmaceutical company and the U.S. Food and Drug Administration (FDA) which emphasises the high focus and beliefs in the NIR-CI technology and the urge to promote the PAT initiative to raise the quality standards in pharmaceutical manufacturing.[111] The prototype blend monitor uses an imaging fibre bundle to collect and transmit NIR spectral information from the blend back to the core NIR-CI instrumentation. The system is not yet fully developed and the blender still needs to be

halted to take measurements (non-destructively and non-invasively). However, the aim is to have the imaging fibre bundle mounted to take measurements during the blending process.

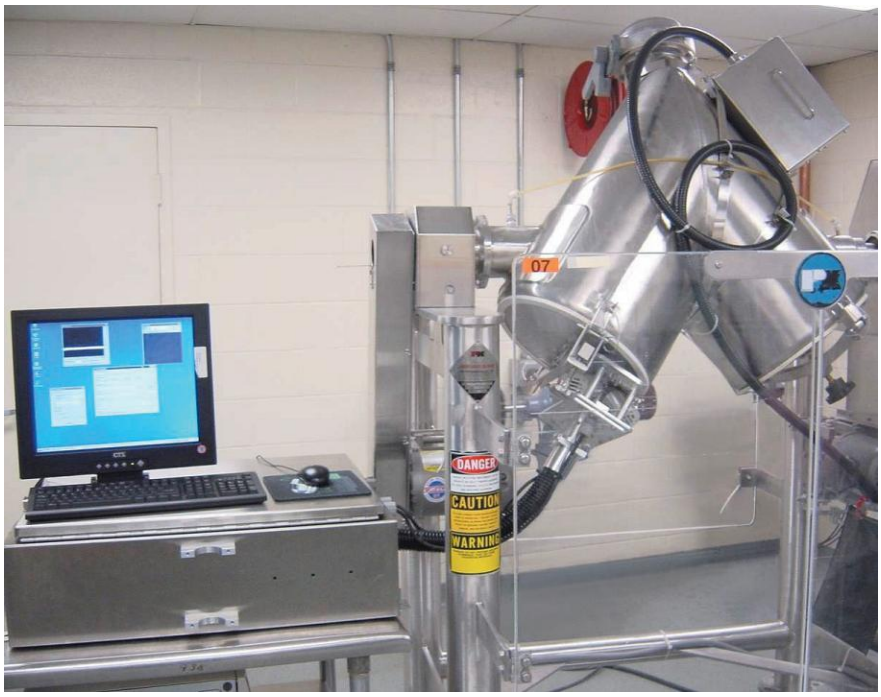


Figure 44 – A prototype NIR-CI global imaging blend monitor system mounted to a V-blender (Source: Lewis et al.[85]).

Currently, further development in instrument technology is needed to implement NIR-CI in manufacturing process environments. However, as described above work is on-going and it may not be long before the first commercial system is available. Once this occur a valuable tool for process monitoring and PAT applications will be available for pharmaceutical manufacturing of solid dosage forms. Until then, there is time for improvements and plenty of knowledge to be gained from the use of NIR-CI in laboratory environments.

REFERENCES

1. U.S. Food and Drug Administration (FDA), Guidance for Industry, PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, FDA (CDER), (2004), [document on the Internet; accessed 18 Dec 2008]. Available from: http://www.fda.gov/cder/guidance/6419fnl.pdf.

2. E.N. Lewis, J. Schoppelrei, E. Lee, Near-infrared chemical imaging and the PAT initiative, *Spectroscopy*, **19** (2004) 26-36.

3. R.C. Lyon, E.H. Jefferson, C.D. Ellison, L.F. Buhse, J.A. Spencer, M.M. Nasr, A.S. Hussain, Exploring pharmaceutical applications of near-infrared technology, *Am. Pharm. Rev.*, **6** (2003) 62-70.

4. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline Pharmaceutical Development Q8, *Step 5*, (2005).

5. E.N. Lewis, J.E. Carroll, F. Clarke, A near infrared view of pharmaceutical formulation analysis, *NIR News*, **12** (2001) 16-18.

6. M. Blanco, J. Coello, H. Iturriage, S. Maspoch, C. de la Pezuela, Near-infrared spectroscopy in the pharmaceutical industry, *Analyst*, **123** (1998) 135R-150R.

7. E.W. Ciurczak, J.K. Drennen III, Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy, Marcel Dekker Inc., New York, (2002).

8. G. Reich, Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications, *Adv. Drug Delivery Rev.*, **57** (2005) 1109-1143.

9. World Health Organization (WHO), Substandard and counterfeit medicines, Fact sheet N°275, November (2003), [document on the Internet; accessed 17 Dec 2008]. Available from: http://www.who.int/mediacentre/factsheets/2003/fs275/en/.

10. R. Barer, R.H. Cole, H.W. Thompson, Infra-red spectroscopy with the reflecting microscope in physics, chemistry and biology, *Nature*, **163** (1949) 198-201.

11. G.E. Peck, G.J. Baley, V.E. McCurdy, G.S. Banker, Tablet Formulation and Design, In: H.A. Lieberman, L. Lachman, J.B. Schwartz (Eds.), Pharmaceutical Dosage Forms: Tablets, Vol. 1, 2nd Ed., Marcel Dekker, Inc., New York, (1989) 75-130.

12. M.E. Aulton, Aulton's Pharmaceutics: The Design and Manufacture of Medicines, 3rd Edition, Churchill Livingstone, Elsevier, (2007).

13. European Pharmacopoeia 6th Edition (6.2), 07 Dosage Forms: Tablets, (2008).

14. C. Gendrin, Y. Roggo, C. Spiegel, C. Collet, Monitoring galenical process development by near infrared chemical imaging: One case study, *Eur. J. Pharm. Biopharm.*, **68** (2008) 828-837.

15. E.N. Lewis, J. Schoppelrei, E. Lee, L.H. Kidder, Near-infrared chemical imaging as a process analytical tool. In: K.A. Bakeev (Ed.), Process Analytical Technology, Blackwell Publishing Ltd, Oxford, (2005), 187-225.

16. D. Clark, M. Henson, F. LaPlant, S. Sasic, L. Zhang, Pharmaceutical Applications of Chemical Mapping and Imaging, In: D. E. Pivonka, J. M. Chalmers, P. R. Griffiths (Eds.), The Handbook of Vibrational Spectroscopy, Applications in Life, Pharmaceutical and Natural Sciences, Pharmaceutical Applications, John Wiley & Sons, Ltd., Vol. 5 (2007) 1-27.

17. S. Sasic, Chemical imaging of pharmaceutical granules by Raman global illumination and near-infrared mapping platforms, *Anal. Chim. Acta*, **611** (2008) 73-79.

18. A.S. El-Hagrasy, H.R. Morris, F. D'Amico, R.A. Lodder, J.K. Drennen III, Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity, *J. Pharm. Sci.*, **90** (2001) 1298-1307.

19. G. Reich, Potential of attenuated total reflection infrared and near-infrared spectroscopic imaging for quality assurance/quality control of solid pharmaceutical dosage forms, *Pharm. Ind.*, 64 (2002) 870-874.

20. A.S. Zidan, O.A. Sammour, M.A. Hammad, N.A. Megrab, M.J. Habib, M.A. Khan, Process analytical technology: Non-destructive assessment of anastrozole entrapment within PLGA microparticles by near infrared spectroscopy and chemical imaging, J. Microencapsulation, **25** (2008) 145-153.

21. M. P. Summers, Aulton M. E., Granulation, In: Aulton M. E. (Ed.), Aulton's Pharmaceutics: The Design and Manufacture of Medicines, 3rd Edition, Churchill Livingstone, Elsevier, (2007) 410-424.

22. H.G. Kristensen, T. Schaefer. Granulation – A review on pharmaceutical wet-granulation, *Drug Dev. Ind. Pharm.*, 13 (1987) 803-872.

23. European Pharmacopoeia 6th Edition (6.2), 2.9.8. Resistance to crushing of tablets, (2008).

24. European Pharmacopoeia 6th Edition (6.2), 2.9.7. Friability of uncoated tablets, (2008).

25. European Pharmacopoeia 6th Edition (6.2), 2.9.40. Uniformity of dosage units, (2008).

26. European Pharmacopoeia 6th Edition (6.2), 2.9.1. Disintegration of tablets and capsules, (2008).

27. European Pharmacopoeia 6th Edition (6.2), 2.9.3. Dissolution test for solid dosage forms, (2008).

28. United States Pharmacopeia, USP31-NF26, General chapters: <905> Uniformity of dosage units, (2008).

29. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Guidance for Industry: Powder Blends and Finished Dosage Units – Stratified In-Process Dosage Unit Sampling and Assessment. Draft Guidance, Pharmaceutical CGMPs, (2003).

30. H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise (Eds.), Near infrared spectroscopy – Principles, instruments, applications, Wiley-VCH, Weinheim (Germany), (2002).

31. E. Räsanen, N. Sandler, Near infrared spectroscopy in the development of solid dosage forms, *J. Pharm. Pharmacol.*, **59** (2007) 147-155.

32. J. Luypaert, D.L. Massart, Y. Vander Heyden, Near-infrared spectroscopy applications in pharmaceutical analysis, *Talanta*, **72** (2007) 865-883.

33. E.N. Lewis, J. Dubois, L.H. Kidder, K.S. Haber, Near infrared chemical imaging: Beyond the pictures. In: H. F. Grahn, P. Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons, Ltd., (2007), 335-361.

34. C. Gendrin, Y. Roggo, C. Collet, Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review, *J. Pharm. Biomed. Anal.*, 48 (2008) 533-553.

35. E. Lee, W. X. Huang, P. Chen, E. N. Lewis, V. Vivilecchia, High-throughput analysis of pharmaceutical tablet content uniformity by near-infrared chemical imaging, *Spectroscopy* **21** (2006) 24-32.

36. F.C. Clarke, M.J. Jamieson, D.A. Clark, S.V. Hammond, R.D. Jee, A.C. Moffat, Chemical image fusion. The synergy of FT-NIR and Raman mapping microscopy to enable a more complete visualization of pharmaceutical formulations, *Anal. Chem.*, **73** (2001), 2213-2220.

37. S.V. Hammond, F.C. Clarke, Near-infrared microspectroscopy, In: J.M. Chalmers, P.R. Griffiths (Eds.), The Handbook of Vibrational Spectroscopy, Sampling Techniques, Microscopy, John Wiley & Sons, Ltd., Vol. **2** (2002) 1405-1431.

38. F. LaPlant, Factors affecting NIR chemical images of solid dosage forms, *Am. Pharm. Rev.*, **7** (2004) 16-24.

39. J.A. Zeitler, P.F. Taday, D.A. Newham, M. Pepper, K.C. Gordon, T. Rades, Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting – a review. *J. Pharm. Pharmacol.*, **59** (2007) 209-223.

40. S.Y. Luk, N. Patel, M.C. Davies, Chemical imaging of pharmaceuticals by time-of-flight secondary ion mass spectrometry, *Spectrosc. Eur.*, **15** (2003) 14-18.

41. H.T.-Aubin, X.X. Zhu, NMR spectroscopy and imaging studies of pharmaceutical tablets made of starch, *Carbohydr. Polym.*, **75** (2009) 369-379.

42. J.A. Zeitler, L.F. Gladden, In-vitro tomography and non-destructive imaging at depth of pharmaceutical solid dosage forms, Eur. J. Pharm. Biopharm., **71** (2008) 2-22.

43. S.J. Hudak, K. Haber, G. Sando, L.H. Kidder, E.N. Lewis, Practical limits of spatial resolution in diffuse reflectance chemical imaging, *NIR news*, **18** (2007) 6-8.

44. D. Clark, S. Sasic, Chemical images: Technical approaches and issues, *Cytometry Part A*, **69A** (2006) 815-824.

45. F.C. Clarke, S.V. Hammond, R.D. Jee, A.C. Moffat, Determination of the information depth and sample size for the analysis of pharmaceutical materials using reflectance near-infrared microscopy, *Appl. Spectrosc.*, **56** (2002) 1475-1483.

46. M.J. Henson, L.Zhang, Drug characterization in low dosage pharmaceutical tablets using Raman microscopic mapping, *Appl. Spectrosc.*, **60** (2006) 1247-1255.

47. S.V. Hammond, NIR microspectroscopy and the control of quality in pharmaceutical production, *Eur. Pharm. Rev.*, **3** (1998) 47-51.

48. S.V. Hammond, F.C. Clarke, Near-infrared microspectroscopy, In: J.M. Chalmers, P.R. Griffiths (Eds.), The Handbook of Vibrational Spectroscopy, Sampling Techniques, Microscopy, John Wiley & Sons, Ltd., Vol. **2** (2002) 1405-1431.

49. R.C. Lyon, D.S. Lester, E.N. Lewis, E. Lee, L.X. Yu, E.H. Jefferson, A.S Hussain, Near-infrared spectral imaging for quality assurance of pharmaceutical products: Analysis of tablets to assess powder blend homogeneity, *AAPS PharmSciTech*, **3** (2002) 1-15.

50. F.W. Koehler, E. Lee, L.H. Kidder, E.N. Lewis, Near infrared spectroscopy: the practical chemical imaging solution, *Spectrosc. Eur.*, **14** (2002) 12-19.

51. F. Clarke, S. Hammond, NIR microscopy of pharmaceutical dosage forms, *Eur. Pharma Rev.*, **1** (2003) 41-50.

52. E.N. Lewis, E. Lee, L.H. Kidder, Combining imaging and spectroscopy: Solving problems with near infrared chemical imaging, *Microsc. Today* (2004), 8-12.

53. F. Clarke, Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy, *Vib. Spectrosc.*, **34** (2004) 25-35.

54. J. Burger, P. Geladi, Spectral pre-treatments of hyperspectral near infrared images: analysis of diffuse reflectance scattering, *J. Near Infrared Spectrosc.*, **15** (2007) 29-37.

55. H. Ma, C.A. Anderson, Optimisation of magnification levels for near infrared chemical imaging of blending of pharmaceutical powders. *J. Near Infrared Spectrosc.*, **15** (2007) 137-151.

56. L.R. Hilden, C.J. Pommier, S.I.F. Badawy, E.M. Friedman, NIR chemical imaging to guide/support BMS-561389 tablet formulation development. *Int. J. Pharm.*, **353** (2008) 283-290.

57. L.J. Makein, L.H. Kidder, E.N. Lewis, M. Valleri, Non-destructive evaluation of manufacturing process changes using near infrared chemical imaging, *NIR news*, **7** (2008) 11-15.

58. L. Weiyong, A. Woldu, R. Kelly, J. McCool, R. Bruce, H. Rasmussen, J. Cunningham, D. Winstead, Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging, *Int. J. Pharm.*, **350** (2008) 369-373.

59. H. Ma, C.A. Anderson, Characterization of pharmaceutical powder blends by NIR chemical imaging, *J. Pharm. Sci.*, **97** (2008) 369-373.

60. C.D. Ellison, B.J. Ennis, M.L. Hamad, R.C. Lyon, Measuring the distribution of density and tabletting force in pharmaceutical tablets by chemical imaging, *J. Pharm. Biomed. Anal.*, **48** (2008) 1-7.

61. A.A. Gowen, C.P. O'Donnell, P.J. Cullen, S.E.J. Bell, Recent applications of chemical imaging to pharmaceutical process monitoring and quality control, *Eur. J. Pharm. Biopharm*., **69** (2008) 10-22.

62. B.J. Westenberger, C.D. Ellison, A.S. Fussner, S. Jenney, R.E. Kolinski, T.G. Lipe, R.C. Lyon, T.W. Moore, L.K. Revelle, A.P. Smith, J.A. Spencer, K.D. Story, D.Y. Toler, A.M. Wokovich, L.F. Buhse, Quality assessment of internet pharmaceutical products using traditional and non-traditional analytical techniques, *Int. J. Pharm.*, **306** (2005) 56-70.

63. M.L. Hamad, C.D. Ellison, M.A. Khan, R.C. Lyon, Drug product characterization by macropixel analysis of chemical images, *J. Pharm. Sci.*, **96** (2007) 3390-3401.

64. N. Jovanovic, A. Gerich, A. Bouchard, W. Jiskoot, Near-infrared imaging for studying homogeneity of protein-sugar mixtures, *Pharm. Res.*, **23** (2006) 2002-2013.

65. C. Gendrin, Y. Roggo, C. Collet, Content uniformity of pharmaceutical solid dosage forms by hyperspectral imaging: A feasibility study, *Talanta* **73** (2007) 733-741.

66. T. Furukawa, H. Sato, H. Shinzawa, I. Noda, S. Ochiai, Evaluation of homogeneity of binary blends of poly(3-hydroxybutyrate) and poly(L-lactic acid) studied by near infrared chemical imaging (NIRCI), *Anal. Sci.*, **23** (2007) 871-876.

67. J. Dubois, J.-C. Wolff, J.K. Warrack, J. Schoppelrei, E.N. Lewis, NIR chemical imaging for counterfeit pharmaceutical products analysis, *Spectroscopy*, **22** (2007) 40-50.

68. M.A. Veronin, E. Lee, E.N. Lewis, "Insight" into drug quality: comparison of Simvastatin tablets from the US and Canada obtained via the internet, *Ann. Pharmacother.*, **41** (2007) 1111-1115.

69. E.N. Lewis, L.H. Kidder, E. Lee, NIR chemical imaging – near-infrared spectroscopy on steroids, *NIR News*, **16** (2005) 2-4.

70. L. Makein, Investigating coating heterogeneity with near infrared chemical imaging, Recorded web-seminar (15th Jan 2008), Malvern Instruments Ltd. Available from: http://www.malvern.com/LabEng/support/events/events.htm.

71. S. Sasic, An in-depth analysis of Raman and near-infrared chemical images of common pharmaceutical tablets, *Appl. Spectrosc.*, **61** (2007) 239-250.

72. J.M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco, Study of pharmaceutical samples by NIR chemical-image and multivariate analysis, *Trends Anal. Chem.*, **27** (2008) 696-713.

73. C. Cairos, J.M. Amigo, J. Coello, S. Maspoch, Enhanced autocorrelation maps in chemical images and application to pharmaceutical tablets, *Eur. J. Pharm. Sci.* (submitted).

74. C. Gendrin, Y. Roggo, C. Collet, Self-modelling curve resolution of near infrared imaging data, *J. Near Infrared Spectosc.*, **16** (2008) 151-157.

75. K. Awa, T. Okumura, H. Shinzawa, M. Otsuka, Y. Ozaki, Self-modelling curve resolution (SMCR) analysis of near-infrared (NIR) imaging data of pharmaceutical tablets, *Anal. Chim. Acta*, **619** (2008) 81-86.

76. A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis, *Trends Anal. Chem.*, **23** (2004) 70-79.

77. O. Kolomiets, U. Hoffmann, P. Geladi, H.W. Siesler, Quantitative determination of pharmaceutical drug formulations by near-infrared spectroscopic imaging, *Appl. Spectrosc.*, **62** (2008) 1200-1208.

78. K.L.A. Chan, S.V. Hammond, S.G. Kazarian, Applications of attenuated total reflection infrared spectroscopic imaging to pharmaceutical formulations, *Anal. Chem.*, **75** (2003) 2140-2146.

79. T.-H. Lee, S.-Y. Lin, Microspectroscopic FT-IR mapping system as a tool to assess blend homogeneity of drug-excipient mixtures, *Eur. J. Pharm. Sci.*, 23 (2004) 117-122.

80. K.L.A. Chan, N. Elkhider, S.G. Kazarian, Spectroscopic imaging of compacted pharmaceutical tablets, *Chem. Eng. Res. Des.*, **83** (2005) 1303-1310.

81. Y. Roggo, N. Jent, A. Edmond, P. Chalus, M. Ulmschneider, Characterizing process effects on pharmaceutical solid forms using near-infrared spectroscopy and infrared imaging, *Eur. J. Pharm. Biopharm.*, **61** (2005) 100-110.

82. Y. Roggo, A. Edmond, P. Chalus, M. Ulmschneider, Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms, *Anal. Chim. Acta*, **535** (2005) 79-87.

83. L. Zhang, M.J. Henson, S.S. Sekulic, Multivariate data analysis for Raman imaging of a model pharmaceutical tablet, Anal. Chim. Acta, **545** (2005) 262-278.

84. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, ICH Harmonised Tripartite Guideline Quality Risk Management Q9, *Step 5*, (2005).

85. E.N. Lewis, L.H. Kidder, E. Lee, NIR chemical imaging as a process analytical tool, *Innovations in pharmaceutical technology*, **17** (2005) 107-111.

86. J. Huang, H. Wium, K.B. Qvist, K.H. Esbensen, Multi-way methods in image analysis – relationships and applications, *Chemom. Intell. Lab. Syst.*, **66** (2003) 141-158.

87. ImageJ – Image processing and analysis in Java [homepage on the Internet, cited 16 Dec 2008]. Available from: http://rsbweb.nih.gov/ij/

88. JIMIA - Hyperspectral chemical imaging instrumentation and software tools for laboratory research and process automation [homepage on the Internet, cited 16 Dec 2008]. Available from: http://www.burgermetrics.com/Downloads/JIMIA.htm

89. Mathworks Inc., Matlab® version 7.5.0 R2007b (www.mathworks.com).

90. S. Wold, E. Johansson, M. Cocchi, PLS – Partial least-squares projections to latent structures. In: H. Kubinyi (Ed.), 3D-QSAR in Drug Design, Theory, Methods and Applications. Escom Science Publishers, Leiden, (1993) 523-550.

91. I.-G. Chong, C.-H. Jun, Performance of some variable selection methods when multicollinearity is present, *Chemom. Intell. Lab. Syst.*, **78** (2005) 103-112.

92. L. Nørgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, Interval partial least-squares regression (*i*PLS): A comparative chemometric study with an example from near-infrared spectroscopy, *Appl. Spectrosc.,* **54** (2000) 413-419.

93. M.W. Borer, X. Zhou, D.M. Hays, J.D. Hofer, K.C. White, Evaluation of key sources of variability in the measurement of pharmaceutical drug products by near infrared reflectance spectroscopy, *J. Pharm. Biomed. Anal.*, **17** (1998) 641-650.

94. A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.*, **36** (1964) 1627-1639.

95. R.J. Barnes, M.S. Dhanoa, S. Lister, Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra, *Appl. Spectrosc.*, **43** (1989) 772-777.

96. P. Geladi, D. MacDougall, H. Martens, Linearisation and scatter-correction for NIR reflectance spectra of meat, *Appl. Spectrosc.*, **39** (1985) 491-500.

97. P. Geladi, H. Isakson, L. Lindquist, S. Wold, K. Esbensen, Principal component analysis of multivariate images, *Chemom. Intell. Lab. Syst.*, **5** (1989) 209-220.

98. P. Geladi, H. Grahn, Multivariate image analysis, John Wiley & Sons, Chichester, (1996).

99. S. Baronti, A. Casini, F. Lotti, S. Porcinai, Principal component analysis of visible and near-infrared multispectral images of works of art, *Chemom. Intell. Lab. Syst.*, **39** (1997) 103-114.

100. H. Martens, T. Naes. Multivariate Calibration, John Wiley & Sons, Chichester, 1989.

101. N.B. Gallagher, Detection, classification, and quantification in hyperspectral images using classical least squares models, In: H.F. Grahn, P. Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons, Chichester, (2007) 181-202.

102. S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chem. Intell. Lab. Syst.*, **58** (2001) 109-130.

103. A. de Juan, R. Tauler, Multivariate curve resolution (MCR) from 2000: progress in concepts and applications, Crit. Rev. Anal. Chem., 36 (2006) 163-176.

104. N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.*, **9** (1979) 62-66.

105. Mathworks Inc., Image Processing Toolbox 6.0 (www.mathworks.com).

106. P.V. Danckwerts, Theory of mixtures and mixing, *Research (London)*, **6** (1953) 355-361.

107. P.M.C. Lacey, Developments in the theory of particle mixing, *J. Appl. Chem.*, **4** (1954) 257-268.

108. R. J. Lantz, J. B. Schwartz, Mixing, In: H. A. Lieberman, L. Lachman, J. B. Schwartz (Eds.), Pharmaceutical Dosage Forms: Tablets, Vol. 2, 2nd Ed., Marcel Dekker, Inc., New York, (1990) 1-53.

109. R. Weinekötter, H. Gericke, Mixing of solids, Particle technology series, Kluwer academic publishers, Dordrecht, (2000).

110. S.S. Sekulic, J. Wakeman, P. Doherty, P.A. Hailey, Automated system for the on-line moniroring of powder blending processes using near-infrared spectroscopy PartII. Qualitative approaches to blend evaluation, *J. Pharm. Biomed. Anal.*, **17** (1998) 1285-1309.

111. P. Taylor, Chemical imaging investigated for process monitoring [document on the Internet, cited 17 Dec 2008], in-pharmatechnologist.com. Available from: http://www.in-pharmatechnologist.com/Processing-QC/Chemical-imaging-investigated-for-process-monitoring.

ANVENDELSE AF NÆR-INFRARØD KEMISK BILLEDDANNELSE VED UDVIKLING AF FASTE DOSEREDE LÆGEMIDLER

Carsten Ravn[j]
ErhvervsPhD afhandling (January 2009)
Institut for Farmaci og Analytisk Kemi
Det Farmaceutiske Fakultet
Københavns Universitet, Danmark
&
Novo Nordisk A/S, Måløv, Danmark

Resumé (Abstract in Danish)

Emnet for denne afhandling er anvendelsen af nær-infrarød kemisk billeddannelse (NIR-CI) til analyse af tabletter i farmaceutisk formuleringsudvikling. NIR-CI er en teknologi i udvikling indenfor farmaceutisk analyse og til anvendelse ved udvikling af nye faste, doserede lægemiddelformuleringer. Teknologien kan give indsigt i og viden om tabletters struktur og funktionelle egenskaber, hvilket ikke umiddelbart kan erhverves ved brug af andre teknologier. Denne information er yderst værdifuld, idet mange kvalitetsegenskaber for tabletter er tæt knyttet til størrelsen og fordelingen af ikke kun den aktive komponent (API), men ligeledes de hjælpestoffer, der indgår i en tablet matrix.
Nær-infrarød kemisk billeddannelse er en teknologi baseret på nær-infrarød (NIR) spektroskopi, men i tillæg til identifikation og kvantificering af komponenterne bestemt ved NIR spektroskopi bibringer teknologien information om den rummelige fordeling af *alle* de spektroskopisk aktive kemiske komponenter i en prøve. I afhandlingen gives en kort beskrivelse af principperne for NIR-CI teknologien og det anvendte udstyr.

En NIR-CI måling producerer enorme mængder spektrale data og anvendelsen af multivariat data- samt billedanalyse-metoder er vigtige for at kunne fortolke og ekstrahere nøjagtig og relevant kemisk information fra de målte data. Målet er at kunne identificere, kvantificere og lokalisere de kemiske komponenter i en prøve og skabe et billede af størrelsen, formen og fordelingen af hver komponent i prøven (et kemisk billede).
Det primære formål var oprindeligt at udforske mulighederne for anvendelse af NIR-CI i farmaceutisk formuleringsarbejde og evaluere den øgede proces og produkt forståelse, der kunne opnås fra sådanne eksperimenter. Det blev dog tydeligt under den indledende projekt-fase, at der stadig var behov for grundlæggende teknologi-eksperimenter, herunder blandt andet udvikling af nøjagtige data- og billedanalyse-metoder. En stor del af denne afhandling beskriver derfor forskellige data- og billedanalyse-metoder til at analysere og ekstrahere relevant og nøjagtig information fra målte NIR-CI data. Der er redegjort for de enkelte trin i en typisk analyse af NIR-CI data og denne afhandling kan således anvendes som en vejledning til at udføre sådanne analyser.

---

[j] *Author contact details:* cacaravn@yahoo.dk; carstenravn@privat.dk

Udvikling af pålidelige og nøjagtige NIR-CI data analyse metoder vurderes ud fra deres evne til at producere nøjagtige kemiske billeder. Med dette formål blev tre almindeligt anvendt data analyse metoder sammenlignet. Kalibreringsprøver blev fremstillet og analyseret. Nøjagtigheden for tre metoder blev sammenlignet ud fra metodernes evne til at producere nøjagtige kemiske billeder, hvilket blev evalueret ved metodernes prædiktionsevne.

Resultater fra et farmaceutisk formuleringsudviklings-eksperiment med anvendelse af de udviklede NIR-CI data- og billedanalyse-metoder er ligeledes præsenteret. Kemiske billeder, histogram statistik, binære billeder, statistisk på partikel/domæne-størrelser og en metode til at evaluere homogeniteten i tabletter blev alle anvendt til at evaluere tabletter fra formuleringsudviklings-eksperimentet. Dette studie viser et eksempel på den unikke og uvurderlige information, som NIR-CI teknologien er i stand til at tilvejebringe for en formulerings-kemiker. Metoderne og metodikken i det aktuelle eksempel vil være generelt anvendelig for mange forskellige typer eksperimenter til udvikling af nye tablet og andre faste, doserede lægemiddelformuleringer.

NIR-CI teknologien er stadig i sin vorden og nogle af dens begrænsninger og forsigtigheder, der bør udvises, diskuteres i afhandlingen. Der diskuteres ligeledes nogle af de vigtige aspekter, der kræver fortsat udvikling og forbedring, for at denne NIR-CI teknologi skal finde udbredt anvendelse i farmaceutisk udvikling. Det fulde potentiale for anvendelse af NIR-CI teknologien i den farmaceutiske industri er endnu ikke fuldt udforsket, men denne afhandling bidrager med grundlæggende viden hertil. I afhandlingen demonstreres det, hvilken unik information, der kan opnås ved anvendelse af NIR-CI teknologien samt den værdifulde rolle teknologien kan forestilles at have ved udvikling af nye faste, doserede lægemiddelformuleringer.

PAPER I

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

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

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

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

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

ARTICLE INFO

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

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

ABSTRACT

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

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

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

1. Introduction

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

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

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

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

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

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

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

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

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

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

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

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

2. Materials and methods

2.1. Materials

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

2.2. Samples

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

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

2.3. Data acquisition

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

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

2.3.1. Spectral correction

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

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

2.3.2. Conversion to absorbance

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

2.3.3. Unfold 3D hyperspectral data cube

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

2.4. Data processing

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

2.4.1. Variable selection

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

2.4.2. Spectral preprocessing

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

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

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

2.4.3. Calibration methods

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3. Results and discussion

3.1. Pure compound spectra and mean spectra of calibration samples

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

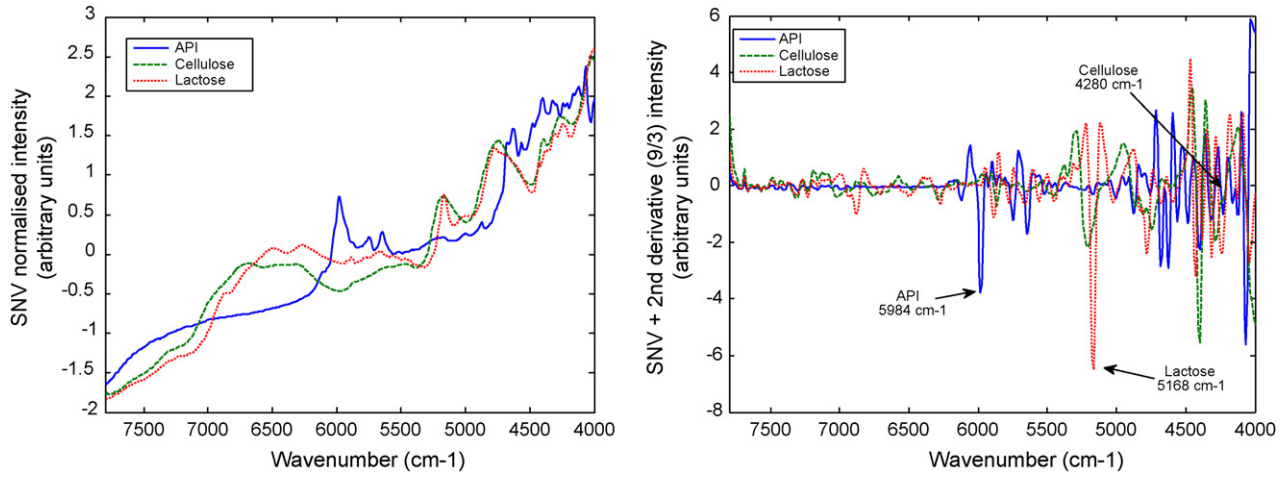


Fig. 1. Pure compound reference spectra of the three major compounds constituting the calibration batches. Spectra are preprocessed by SNV (left) and SNV followed by second derivative (9/3) (right). The arrows point out the wavenumbers used for the single wavenumber analysis.

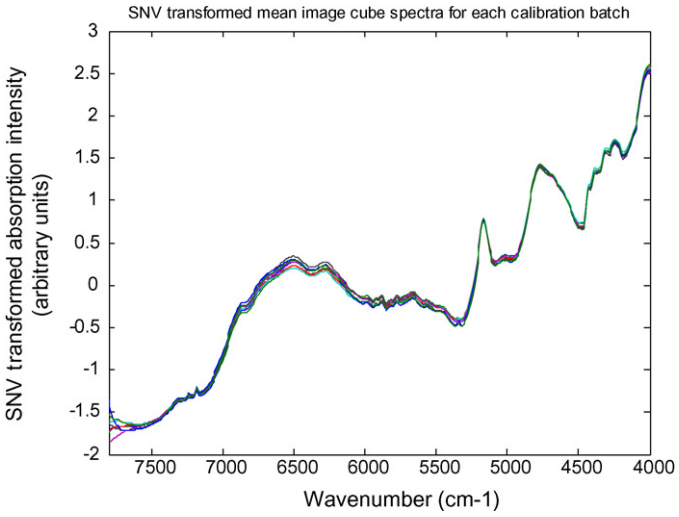


Fig. 2. Image cube mean spectra of the 9 calibration batches normalised by standard normal variate (SNV).

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

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

3.2. Prediction of concentrations

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

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

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

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

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

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

3.3. Chemical images

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

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

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

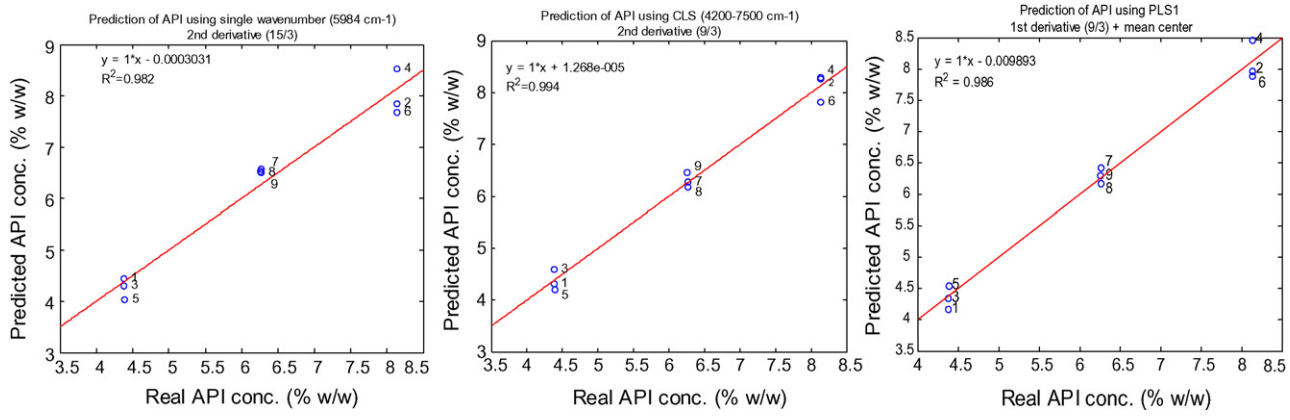


Fig. 3. Linear regression results of API concentration for single wavenumber method (left), CLS (middle) and PLS1 (right). Results are shown for the preprocessing methods giving the best regression and concentration predictions.

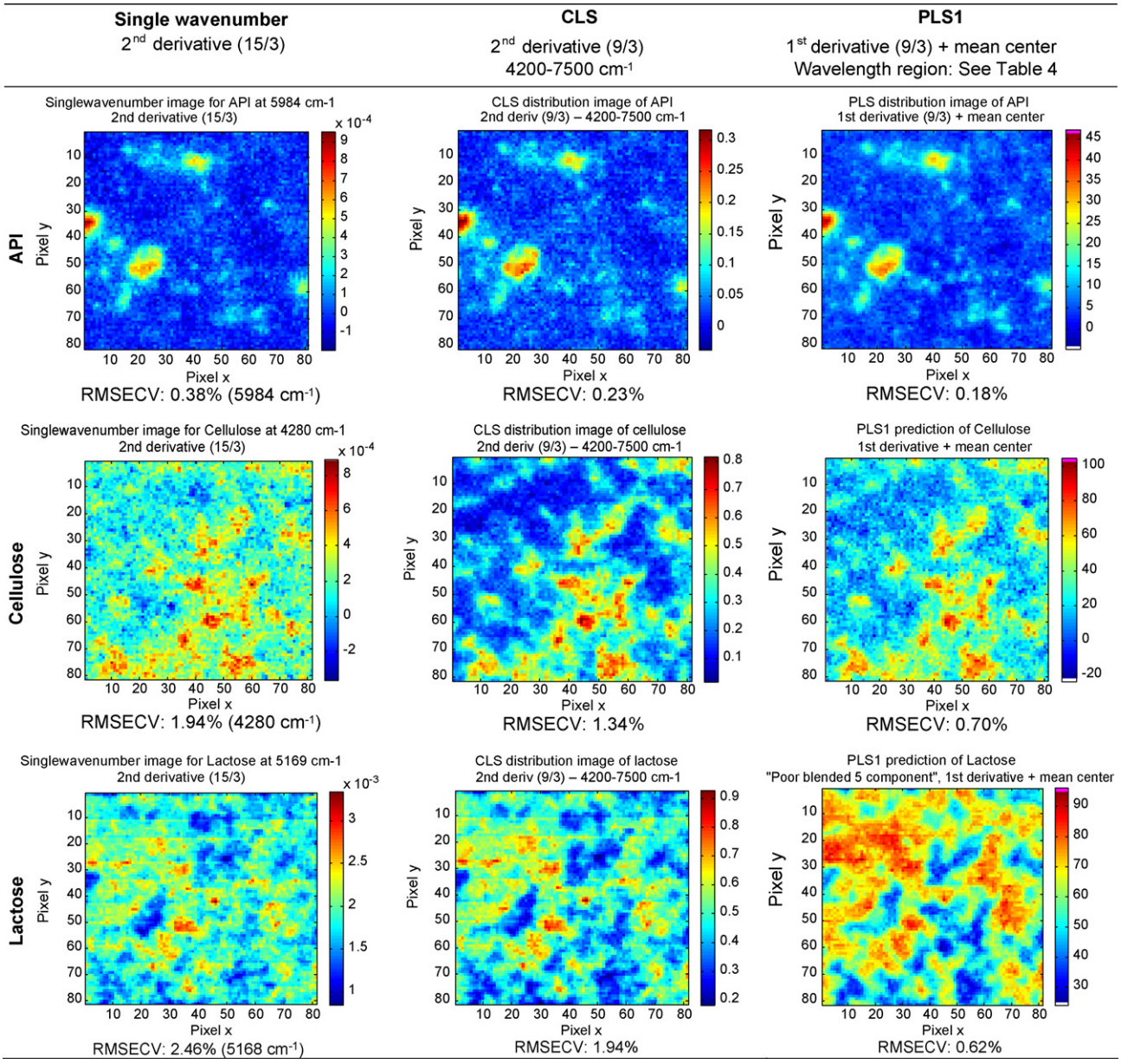


Fig. 4. Chemical images of a tablet of batch 9 composition (cf. Table 2) generated from data processing using single wavenumber, CLS or PLS1. The chemical images show the distribution of the three major compounds API, cellulose and lactose. The prediction errors (RMSECV) from Table 4 are shown beneath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

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

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

4. Conclusion

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

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

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

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

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

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

References

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

PAPER II

Application of near-infrared chemical imaging in formulation development of solid dosage forms

Manuscript status: Ready for submission.

Carsten Ravn[a,b,*], José Manuel Amigo[c], Carlos Cairós[d], Jørn Møller-Sonnergaard[a], Erik Skibsted[b], Claus Cornett[a], Rasmus Bro[d]

[a]Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark

[b]CMC Formulation and Analysis, Novo Nordisk A/S, Novo Nordisk Park, 2760 Maaloev, Denmark

[c]Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark

[d]Unidad de Química Analítica. Departament de Química. Facultat de Ciències. Universitat Autònoma de Barcelona. 08193 Bellaterra, Barcelona. España.

[*] Corresponding author. *Email address*: carstenravn@privat.dk; cacaravn@yahoo.dk

Abstract

This study demonstrates how near-infrared chemical imaging (NIR-CI) can be used as a valuable analytical tool in formulation development of tablets or other solid dosage forms using conventional NIR spectroscopy for comparison.

NIR-CI was used to study the API domain size and its distribution uniformity within a tablet formulation. A two-component powder blend consisting of the active pharmaceutical ingredient (API) and microcrystalline cellulose was used to manufacture tablets by a direct compression process. Three different tablet batches were produced which only varied by the ingoing API particle size. Images showing the spatial distribution of API (chemical images) for four tablets per batch were obtained by Classical Least Squares (CLS) and evaluated visually and by their histogram statistics. The CLS method was clearly capable of generating chemical images revealing the spatial distribution information of API and estimating the concentration within each chemical image. Two image processing tools were utilised to obtain objective and numerical information from the chemical API images about the number and size of API domains and homogeneity of the API distribution. The latter was obtained from a novel promising image processing method to evaluate homogeneity directly from the chemical images. The results from applied methods reflected a higher number of smaller API domains more uniformly distributed in tablets prepared with small in-going API particles compared to a lower number of larger API domains unevenly distributed in tablets prepared with larger in-going API particle size. For comparison, tablets were also analysed by conventional NIR spectroscopy and the homogeneity of API in the tablets evaluated by this approach confirmed the findings from NIR-CI. This suggests NIR-CI and NIR spectroscopy as a beneficial and powerful combination in formulation development of tablets.

Keywords: *formulation development; tablets; solid dosage forms; near infrared chemical imaging (NIR-CI); hyperspectral imaging; image analysis; homogeneity; content uniformity; classical least squares (CLS); uniformity of distribution; macropixel analysis*; *homogeneity curves.*

Introduction

Formulation development of tablets comprise a series of experimental studies that together form the basis for selecting the optimally formulated drug product regarding quality, therapeutic performance and production costs. Much time and effort is spend on developing new tablet formulations and its quality is indeed determined by the structure of the evolving matrix. The quality of powder samples and/or tablets produced from these studies is traditionally evaluated by conventional analytical techniques (HPLC, mass spectrometry, dissolution etc.) to determine identity, strength, purity, and quality. Unfortunately, these tests destroy the formulated tablet matrix, are time consuming and only provide a single value for one tablet or a group of tablets. No insight into the internal structure and spatial distribution of components in the final product matrix is obtained. Information, which is valuable and can elucidate intra-tablet variations helping the formulator to see how the process actually performed. Spectroscopic or chemical imaging techniques such as near infrared chemical imaging (NIR-CI) can provide this information (Hammond, 1998; Lewis et al., 2001; Lewis et al., 2004).

Investigating the homogeneity in pharmaceutical powder blends or tablets is always a main concern in formulation development and is related to a uniform distribution of the active pharmaceutical ingredient (API) throughout the blend/tablets. There is no definitive definition of homogeneity of a powder blend or tablet batch. Homogeneity of tablets is indirectly assessed by calculating an acceptance value based on the mean and standard deviation of individual active substance content from 10 or 30 dosage units and comparing the results with predetermined acceptance criteria (USP31-NF26 S1 <905>; Ph.Eur 6^{th} Ed. <2.9.40>). As the content uniformity of tablets primarily depends on the uniformity of the powder blend, it is also important to determine blend uniformity. This is the basis of the FDA guidance describing how sampling and assessment of uniformity in powder blends and finished dosage units can be conducted to comply with good manufacturing practice requirements (U.S. FDA Guidance for Industry, Stratified sampling, 2003).

The blend uniformity obtained during a mixing process may be affected by the subsequent process steps such as transporting the mixing container to the tablet press, emptying the blend into the hopper, the movement in the hopper, filling of powder into the die during the tabletting process etc. The blend uniformity or homogeneity is therefore best determined as late in the manufacturing process as possible, i.e. determining content uniformity of the final dosage form. The quality of the formulation in this study is therefore assessed by analysing the final tablets.

The employment and usefulness of particularly near infrared (NIR) spectroscopy (Blanco et al., 1998; Ciurczak and Drennen, 2002; Lyon et al., 2003; Reich, 2005; Luypaert et al., 2007) but also increasingly near-infrared chemical imaging (NIR-CI) (Hammond, 1998; Lewis et al., 2001; Koehler et al., 2002; Clarke and Hammond, 2003; Clarke, 2004; Lewis et al., 2004; Lewis et al., 2005a; Reich, 2005; Gowen et al., 2008) in pharmaceutical development and manufacturing of solid dosage forms is well documented. NIR spectroscopy measures a bulk average NIR spectrum and reflects an average composition of the sample. NIR chemical imaging builds on the NIR spectroscopy technology and adds spatial information to the spectral information by combining it with digital imaging. In NIR-CI, a NIR spectrum is recorded in each pixel of the sample image, resulting in a hyperspectral data cube. By utilising appropriate data processing methods this immense amount of spectral information can be translated into a set of chemical images, providing information not only on the composition and concentrations but most importantly on the distribution of all spectroscopically active components (API and excipients) within each sample.

Several univariate and multivariate approaches for data processing hyperspectral data cubes obtained from NIR-CI analysis into chemical images have been developed and compared for pharmaceutical samples (de Juan et al., 2004; Jovanovic et al., 2006; Sasic, 2007; Gendrin et al., 2007; Clark et al., 2007; Amigo et al., 2008; Ravn et al., 2008). The information immediately apparent from chemical images can be very convincing but care must be taken in the subjective and qualitative interpretation of these resulting chemical images. Developing methods to derive objective and numerical informa-

tion contained in the images is therefore essential for the advancement of the NIR-CI technology; however there has been a lack of development of these image processing tools probably because it is not so straightforward. The reported image processing tools of chemical images is so far confined to the two most obvious tools being particle/domain size statistics and histogram statistics. The particle/domain size statistics relates to calculating the number, size and shapes of domains from threshold generated binary images (Clarke and Hammond, 2003; Clarke, 2004; Lewis et al., 2005b; Henson and Zhang, 2006; Clark et al., 2007; Sasic, 2007; Lewis et al., 2007; Ma and Anderson, 2008; Li et al., 2008; Hilden et al., 2008). The histogram statistics concerns representing all pixel intensity values in a single chemical image as a histogram, where the mean value is taken as an estimate of the bulk content of the component and the standard deviation as a quantitative measure of the overall sample homogeneity (Lyon et al., 2002, Lewis et al., 2007, Gendrin et al., 2008; Jovanovic, 2006; Ma and Anderson, 2008, El-Hagrasy et al., 2001, Furukawa et al., 2007). This homogeneity concept from histogram statistics introduced for NIR-CI analysis is different from the traditional homogeneity concept described earlier. The thousands of individual measurements obtained from microscopic pixel locations of each sample provide data for intra-sample statistical analysis. The homogeneity is thus pertaining to each sample, i.e. intra-tablet homogeneity, more than the overall batch as in conventional homogeneity testing. However, it can be shown that the current NIR-CI homogeneity concept from histogram statistics is not always adequate, so a new method has been developed. This new method and a pharmaceutical application hereof is described and presented in this paper.

Three tablet batches were produced to illustrate a general formulation development case. The difference between the batches was the in-going API particle size The API domain size and its distribution uniformity within the tablets were evaluated from NIR chemical imaging and conventional NIR spectroscopy analysis.

Experimental Section

Samples

A two-component powder blend consisting of 5 % (w/w) active pharmaceutical ingredient (API) (a proprietary compound) and 95 % (w/w) silicified microcrystalline cellulose (ProSolv SMCC® HD90, JRS Pharma, Germany) was used to produce three different batches. The three batches were produced with the following different API particle size fractions: (1) < 125 µm; (2) 125 – 355 µm; and (3) > 355 µm, in the following referred to as batch 1, batch 2 and batch 3, respectively. The API particle size fractions were obtained from a mechanical agitation dry sieving process using sieve sizes of 500, 355, 180, 125, 90 and 63 µm. The residue on the collection pan was also used and the fractions within each of the three selected intervals was collected and gathered. Each two-component powder blend of 150 g was loaded in a 1 L stainless-steel container (~ to 50 % of total container volume) and mixed on a lab-scale Turbula® mixer (WAB Maschinenfabrik, Muttenz, Switzerland) for 5 minutes at 70 rpm. Immediately after mixing, the two-component powder blend was tabletted on a single punch Diaf tablet press. The tablets manufactured had a flat surface, diameter of 6 mm and aiming at target weight 100 mg and target crushing strength 75 N.

Pure compound reference samples of the API and microcrystalline cellulose (MCC) were also produced for the classical least squares (CLS) image analysis. Approximately 250 mg of each pure compound material was compressed into 8 mm diameter compacts on a hydraulic tablet press using 10 kN pressure for 10 seconds. The compacts were analysed similar to the pharmaceutical tablets and used to generate pure compound reference spectra for the CLS analysis.

NIR-CI analysis

A near infrared line mapping system (Spectrum Spotlight 350 FT-NIR Microscope, PerkinElmer, UK) was utilised to obtain the hyperspectral raw data. Four tablets from each of the three batches were analysed by NIR-CI. Each tablet was moved across the

field of view of a 16-element MCT linear detector array and the 2D image was built up line by line in steps determined by the pixel size. The third dimension was added as NIR diffuse reflectance spectra were collected for each of the 16 detectors to obtain the 3D hyperspectral data cube. Tablets were fixed in a suitable rubber-ring holder and analyzed on the tablet surface. An area of 3 x 3 mm was analysed for each tablet using a pixel size of 25 x 25 µm thus obtaining images of 120 by 120 pixels with a total of 14400 spectra (= pixels). Each spectrum was the average of 8 scans from wavelength region 7800-4000 cm^{-1} using a 16 cm^{-1} spectral resolution. The NIR diffuse reflectance spectra from the data acquisition were first background corrected ($R=R_{sample}/R_{background}$) using the high reflectance standard Spectralon™ (Labsphere, Inc., North Sutton, NH) as background reference and then converted into absorbance (A) by the relation $A = \log_{10}(1/R)$.

Chemical images using CLS

The processing of the hyperspectral absorbance data cube into chemical images was conducted using Classical Least Squares (CLS) (Martens and Naes, 1989; Gallagher, 2007). CLS is often used in simpler spectroscopic applications but its usefulness has also been demonstrated for analysing hyperspectral NIR data of pharmaceutical samples (Gendrin et al., 2007; Gendrin et al., 2008, Amigo et al., 2008, Ravn et al., 2008). CLS estimates concentrations from pure spectra and generates chemical or concentration images rather than abstract score images (Gallagher, 2007). A prerequisite for using CLS is that *all* the spectra of the pure compounds constituting the sample are available. This is almost always the case in developing a new tablet formulation. An important implication of using pure component spectra in CLS is that the concentrations of the chemical images vary from approximately zero to one because the pure component spectra are arbitrarily set to concentration one. The CLS prediction values for a component are thus fractions or percentages of the pure component sample, which mostly though approximates the mean percentage of the component in the actual sample.

All CLS analyses were performed in Matlab® 7.5.0 (R2007b) software (Mathworks, www.mathworks.com). Prior to the CLS analysis the 3D hyperspectral absorbance data

cubes were unfolded to obtain a two-dimensional matrix. Each pure compound spectrum was calculated as the mean spectrum of all pure compact sample spectra. The hyper-spectral data cube and the pure spectra were both pre-processed using a first derivative Savitzky-Golay transform with a nine-point filter width and polynomial order three (denoted '1st derivative (9/3)'). The noisy wavenumber ends were also removed reducing the spectral range to 7500-4200 cm^{-1} (Ravn et al., 2008).

NIR spectroscopy

Transmission NIR spectra were measured through both sides of thirty tablets from each batch and an average spectrum was calculated for each tablet. The NIR transmission spectra were collected using an FT-NIR Bruker Multi Purpose Analyzer instrument (Bruker Optics, Ettlingen, Germany). A spectral resolution of 16 cm^{-1} and 64 scans per spectrum were used. The wavelength region from 12500 cm^{-1} to 5800 cm^{-1} was measured. Air was used to record a background spectrum.

The resulting NIR spectra were analysed by two approaches used as reference analysis methods to determine a measure of the tablet uniformity for each batch, i.e. an overall measure of the batch homogeneity. The first method was principal component analysis (PCA) using PLS_Toolbox (version 4.1.1, www.eigenvector.com). The transmission spectra were pre-processed by multiplicative scatter correction (MSC) followed by mean-centering prior to PCA analysis. The standard deviation of the PCA score values for the thirty tablets from each batch was calculated assuming correlation between one of the principal component (PC) score values and the API concentrations. This was ensured by using the principal component that had a loading vector most similar to the API spectrum. In the second method, a single standard deviation value of the thirty spectra from each batch was calculated similar to earlier described approaches (Sekulic et al., 1998; El-Hagrasy et al., 2001). The spectra were pre-processed by a second derivative (9/3) transform and a wavenumber region around the major contribution from the API was selected (defined from the pure API spectrum). The standard deviation of the absorbance values for each wavenumber was calculated over the thirty pre-processed spectra, result-

ing in a standard deviation spectrum. The pooled standard deviation over all wavenumbers of the standard deviation spectrum for each batch was finally determined and reported as a single value. The results obtained using these two approaches on the NIR spectroscopy data were compared with the NIR-CI results.

Binary images and domain statistics

To calculate the particle/domain statistics of a NIR-CI analysed sample the chemical image was translated into a binary image reflecting either presence or absence of a component. This involves setting a threshold on the intensity or 'concentration' scale of the chemical image. Setting the threshold is not an easy task and to the author's knowledge no objective method has been reported to select such a threshold in hyperspectral imaging of pharmaceutical products. A general, objective method of automatic image threshold selection was introduced by Otsu (1979) but it was not found useful for the samples in this study. Until appropriate NIR-CI spatial and spectral standards are available it is not possible to confirm the appropriateness of an objective threshold selecting method. Thus, currently the best way seems to determine the threshold as a function of the sample and problem investigated. In this study, however, an objective and well-defined method for determining the threshold value to generate binary images of the API is suggested. The threshold value was determined from analysing the hyperspectral data cube of the pure API compact by CLS using only the mean API spectrum from this sample. The histogram plot of the resulting pure chemical API image shows a symmetrical distribution around the mean of 1.0 (Fig. 1). The spread of the data around the mean was used to represent an estimate of the methods precision assuming a uniform pure API compact. The threshold value was determined as equal to three times the standard deviation (Std. = 0.051, Fig. 1) from the CLS analysis of the pure API image; resulting in a threshold value of 0.15. The interpretation of this threshold selecting approach is that a concentration value of 0.15 in a chemical image could within statistical probability (three standard deviations) just as well be regarded as zero (or vice versa) as found in the 'precision'

analysis of the pure API image. So in order to statistically confirm API in a chemical image the concentration/intensity values must be above 0.15.

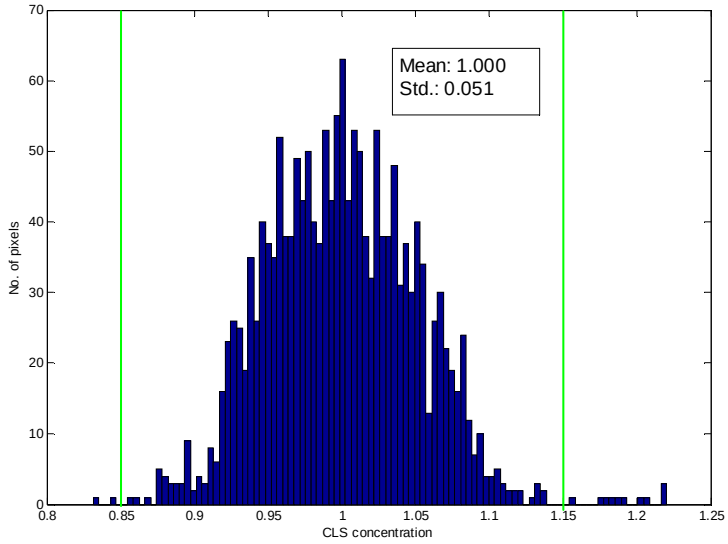


Fig. 1. Histogram representation of the chemical API image from analysing the pure API compact by CLS using only the mean API spectrum. The threshold limit is set as three time the standard deviation from this analysis (= 0.15).

The domain statistics of the binary images were calculated from in-house routines in Matlab® 7.5.0 (R2007b) software using functions from the Image Processing Toolbox 6.0 (Mathworks, www.mathworks.com). The domain statistics calculated were the number of domains, mean equivalent diameter and the area covered in percentages. The mean equivalent diameter represents the diameter of a circle with the same area as the region of the domain of interest. The area covered is the number of pixels of the calculated component in percent of the total number of image pixels. The percent area covered thus gives an estimate of the component concentration based on the binary image approach.

NIR-CI homogeneity measure

In addition to the domain statistics, the uniformity of API distribution of each chemical image was evaluated by yet another image processing method to extract numerical information from the chemical images. This was done by sampling the chemical images by several moving squares (called 'macropixels') varying in size from the inherent pixel resolution up to the size of the entire chemical image. For example, an image consisting

of 30 by 30 pixels could be sampled by a macropixel of 10 by 10 pixels (Fig. 2). The chemical image is sampled by all the macropixels possible to position in the chemical image, i.e. 441 squares in Fig. 2. This methodology of extensively sampling a chemical image by all possible macropixels ('macropixel analysis') was first introduced by Hamad et al. (2007) and the method was specifically referred to as 'Continuous-Level Moving Block' (CLMB). However, the approach used in this study to evaluate the macropixels is different. The measure of homogeneity is calculated by the standard deviation of the mean CLS prediction for each macropixel, i.e. the standard deviation of 441 mean CLS predicted API concentrations from the 441 possible macropixels in the example shown (Fig. 2). This new NIR-CI homogeneity approach ensures a more extensive and detailed sampling of each image, which may provide a better and more comprehensive basis of evaluating the homogeneity of samples.

In this study, the standard deviation was calculated for the quadratic macropixels of 1, 2, 3, 4, 6, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110 and 120 pixel length and height. With the standard deviation calculated for each of these macropixels a *homogeneity curve* was constructed for each image with the standard deviation on the y-axis and the squared macropixel size (pixel length × pixel height) on the x-axis. All data analysis for this methodology was performed in Matlab® 7.5.0 (R2007b) software (Mathworks, www.mathworks.com) using in-house written routines.

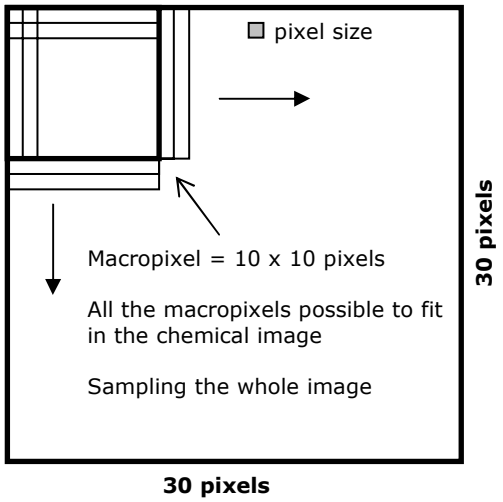


Fig. 2. Illustration of the sampling principle for the novel NIR-CI homogeneity measure method. The entire chemical image is sampled by all the macropixels it is possible to position in the chemical image. The standard deviation of the mean CLS prediction values for each macropixel size is calculated and plotted as a function of the macropixel sizes. This generates a *homogeneity curve*.

Results and discussion

Chemical images and prediction of API content in the tablets

The CLS generated chemical API images of the four tablets from each of the three batches having different ingoing API particle sizes are shown in Fig. 3. The colour-scale images display the CLS predictions in each pixel and visualises the abundance, distribution and domain size of API within the analysed tablets. As the formulation is made up of only two components the high intensity values (red) indicate areas of high API content and the low intensity values indicate low API content and therefore most likely high cellulose concentration (blue). The mean value and standard deviation of all CLS pixel predictions of API for each tablet are listed in Table 1 (histogram statistics for each chemical API image). In Table 1, also the mean of the mean prediction values for the four tablets from each batch and its standard deviation is shown. This standard deviation (denoted Std-b) reflects the variation in API content prediction 'between' all four tablets analysed for each batch. Finally, the mean of the four standard deviation values (Std.) for each batch is calculated and presented in Table 1. This standard deviation (denoted Std-w)

reflects a weighted mean-value of the variation in pixel content predictions 'within' each of the four tablets from each batch.

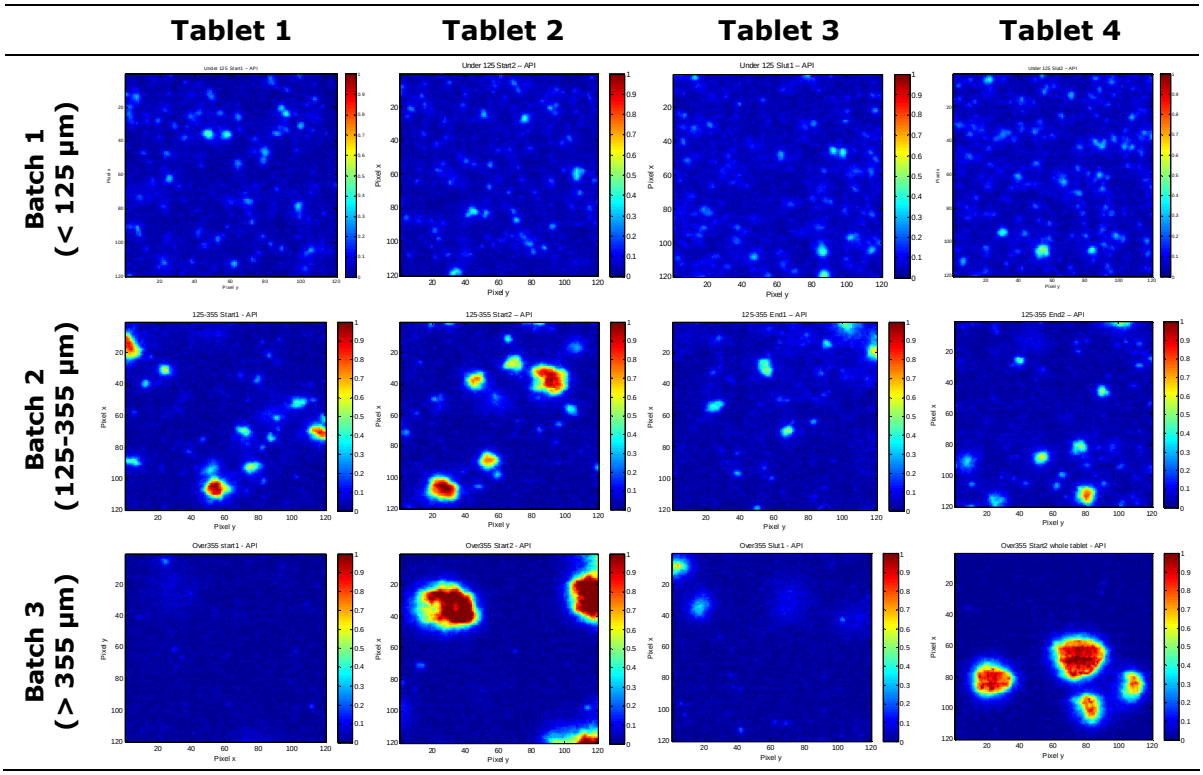


Fig. 3. Chemical images of API generated by CLS. The chemical images (3x3 mm) are of the four tablets from each of the 3 batches with different in-going API particle sizes: < 125 µm (batch 1, upper row); 125-355 µm (batch 2, middle row); and > 355 µm (batch 3, lower row). All images are scaled from 0 to 1.

	Batch 1 (<125 µm)		Batch 2 (125-355 µm)		Batch 3 (>355 µm)	
	Mean	Std.	Mean	Std.	Mean	Std.
Tablet 1	6.0	4.3	7.2	11.0	2.5	2.0
Tablet 2	6.2	3.9	8.8	14.5	12.5	25.3
Tablet 3	7.1	4.2	5.9	5.8	3.4	4.3
Tablet 4	7.6	4.7	6.4	7.1	10.2	21.2
Mean	6.7		7.1		7.2	
Std-b (between)	0.76		1.3		4.9	
Std-w (within)		4.3		9.6		13.2

Table 1. Histogram statistics of the CLS generated chemical API images from Fig. 3 showing the mean value and standard deviation (Std.) of the API content predictions for each of four tablets within the three batches (CLS predictions × 100). The standard deviation Std-b is the standard deviation of the four mean values for each batch. The standard deviation Std-w is the mean of four standard deviations (Std.) for each batch.

Just by visual inspection of the chemical API images valuable information about the distribution and relative content of API and thus the overall blend quality are obtained (Fig. 3). The distribution mapping of API clearly illustrates what could be expected from the knowledge of the ingoing API particle sizes. A high number of small domain sizes well distributed over the sample surface are observed for batch 1 with the finer API material (< 125 µm). As the particle size increases for the ingoing API material, the API domain sizes in the chemical images also increases and the number of domains decreases (batch 2 and 3). However, for the four images of batch 3 with in-going API particle size above 355 µm only images of tablet 2 and 4 display chemical images with large API domains. The reason is that there simply are no (large) API domains in the 3 x 3 mm sampled area of tablet 1 and 3. This is confirmed by the low CLS concentration predictions of 2.5 %[1] and 3.4 % for tablet 1 and 3, respectively with hardly any API compared to 12.5 % and 10.2 % for tablet 2 and 4, respectively having large API domains (Table 1, Batch 3).

These results are in accordance with what intuitively would be expected from the knowledge of the ingoing API particle sizes. However, other less conceivable results might have been obtained and it is therefore important to have the results verified in the final dosage form as demonstrated here for the chemical images. For example, it is well known that small particles become more cohesive with greater risk of agglomeration, which could have been observed for the smallest particle size fraction (Clark et al., 2007). Also, large agglomerates of API can be broken down during the mixing process and could therefore have shown smaller particles than the ingoing size for batch 3. This clearly illustrates the unique and invaluable information that can be provided by NIR-CI and which helps the formulator visualise how the process actually performed.

The qualitative conclusions that can be drawn from the visual inspection of the chemical images are fast and very informative but also subjective and non-operational. For less distinct and illustrative chemical images than presented here it could be a lot harder to visually distinguish blend quality between samples. It emphasises the need to translate

[1] Note, that the unit '%' for CLS predictions (× 100) means percentage of the pure component compact used for CLS analysis and should not be mistaken by the common perception of '%' as relative content in the actual sample. Mostly, though, the CLS prediction of a component is approximate to the mean percentage of the component in the actual sample.

the information from the chemical images into objective and numerical values. The most obvious method to obtain quantitative measures is from the statistical analysis of a chemical image represented as a histogram, where the mean value provides an estimate of the component abundance and the standard deviation is an indication of the component homogeneity at the utilised spatial resolution. The qualitative results described above are all well reflected by the histogram statistics for all API chemical images using CLS (Table 1). The mean of the mean values for all four tablets for each batch are quite similar but the standard deviation (Std-b) increases from 0.76 % to 1.3 % and to 4.9 % for batch 1, batch 2 and batch 3, respectively. This reflects the increasing variation in content prediction between the four tablets from each batch and suggests an increasing in-homogeneity of tablets from batch 1 through to batch 3. The standard deviation (Std-w) that for each batch reflects the variation in pixel content predictions within the tablets also increases from batch 1 through to batch 3 (4.3 %, 9.6 % and 13.2 %[1], respectively). These results also indicate a decreasing uniform distribution of API in the tablets from batch 1 to batch 3.

NIR spectroscopy results

The visual inspection and histogram statistics indicated increasing in-homogeneity with increasing ingoing API particle size fraction. This is supported by the PCA analysis of the transmission NIR spectroscopy data. The score values for principal component 3 (PC3) are shown for the thirty tablets from each of the three batches (Fig. 4). Principal component 3 was utilised because the loading of PC3 resembled the pure API spectrum (Fig. 5). This indicates that the score values of principal component 3 will reflect the variation in API concentration. It must be noted though, that this method is quite ad hoc. There is no exact scientific justification for scores being equal to concentrations when the loading and the spectrum of a constituent are almost identical. The actual principal component may not only reflect the constituent content but also other things in the tablet which are not desirable. However, the method can be used as an approximate qualitative approach to evaluate the uniformity of a constituent in the tablets. In this context, the PCA analysis of

the NIR spectroscopy data illustrates the increasing in-homogeneity of the API in the tablets for the 3 batches. The calculated standard deviation between the score values shown in Fig. 4 for each batch confirm this with increasing values of 0.02, 0.06 and 0.10 for batch 1, batch 2 and batch 3, respectively.

The second approach used to evaluate the NIR spectroscopy data confirmed the results from the PCA analysis. As described for the method in the experimental section, batch homogeneity was determined by comparing the second derivative spectra of thirty tablets from each batch and by calculating the pooled mean standard deviation. The results showed a progressively increase of the mean standard deviation on 0.0061, 0.0085 and 0.0094 with the increasing ingoing API particles size for batch 1, batch 2 and batch 3, respectively.

NIR spectroscopy is a fast, non-destructive method for obtaining reference data to compare with the imaging data. The results demonstrate the beneficial use of NIR-CI and NIR spectroscopy in conjunction to provide a rapid and very informative approach to evaluate pharmaceutical tablets. The promising results suggest that NIR spectroscopy and NIR-CI can be used together alone for certain experimental studies in formulation development work without the need for slower conventional analytical techniques. However, once again is must be underlined that the two methods are ad hoc and cannot be assumed to generally apply to all formulations. The methods are based upon the assumptions that either the principal component score values or selected spectral regions of API mainly reflect the variation in API concentration and not other characteristics of the tablets. These assumptions must be used with caution and evaluated case-by-case for each formulation.

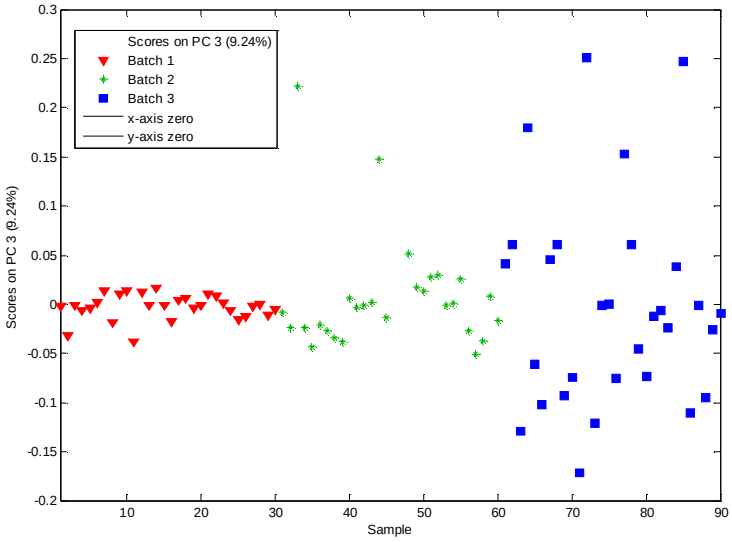


Fig. 4. Results from PCA analysis of the NIR spectroscopy data of 30 tablets for each of the three batches (28 tablets for batch 2 – two outliers removed). The plot shows the score values of principal component 3 (PC3). The PC3 was selected due to the high correlation with the pure API spectrum (cf. Fig. 5).

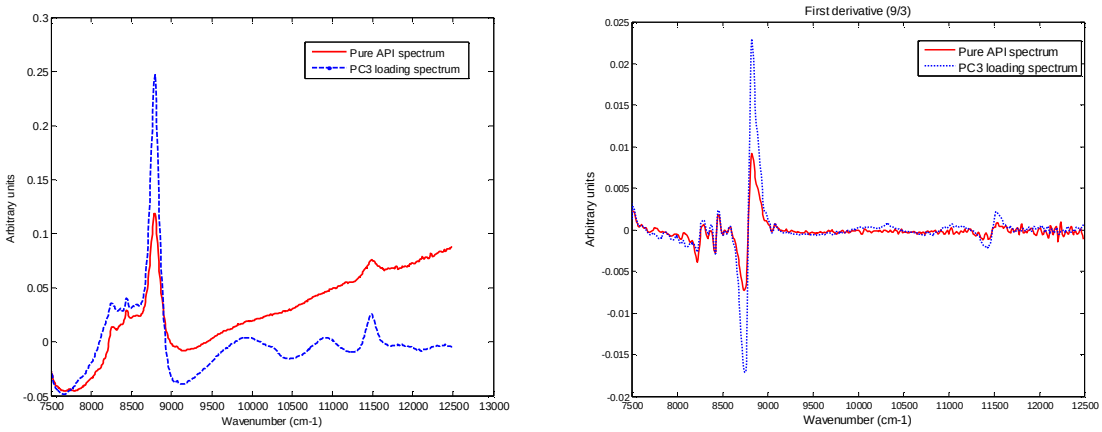


Fig. 5. Comparison of the pure API NIR absorbance spectrum with the loading spectrum for principal component 3 (left) and the corresponding first derivative (9/3) spectra (right).

Binary images and domain statistics

With the evident visual differences in API domain sizes between the three batches in this study, domain statistics is an obvious image processing tool to compare and evaluate the chemical images. To perform this quantitative analysis it is necessary initially to create a binary image. From binary images domain statistics such as number of particles and their sizes can be determined.

Discriminating between one large particle, agglomerated particles or closely spaced but separate particles in binary images is often not possible. It is therefore more the domain size distribution rather than the particle size distribution that is determined and hence the term 'domain statistics' is used. Even so, the domain statistics are useful to determine if any changes in domain size or form happen during the manufacturing process and to compare different product/processes.

It should also be noted that the spatial resolution decides the smallest domain size possible to determine. It therefore makes most sense to measure domain/particle sizes for components where the ingoing particle size is larger than the pixel resolution. Otherwise, what is determined as one particle/domain may actually be made up of many small particles or there may be many particles smaller than the pixel size that will not be included in the calculation. While this domain statistics after all only is an approach of the real particle sizes in the solid-phase mixtures it provides a fine estimate and a valuable tool for comparing different formulations. Further, to the author's knowledge chemical imaging is the only non-destructive method that can measure this unique domain size information within solid mixtures.

The binary images of 'tablet 2' analysed from each of the three batches using the calculated threshold value of 0.15 are presented in Fig. 6. From these binary images the effect of different ingoing API particle sizes on its size and distribution in the final tablet is even more evident compared to the corresponding chemical images in Fig. 3. Table 2 summarizes the calculated API domain size statistics above the spatial resolution (25 x 25 μm) for all twelve binary images using the threshold value 0.15. The results clearly confirm the earlier visual observations with the number of API domains decreasing but the size of domains increasing from batch 1 to batch 3. The results for batch 3 are differing even more from the other batches if the two tablets (no. 1 and 3) with hardly any API are excluded (cf. results in brackets, Table 2). The calculated domain statistics demonstrate one possible approach on how to extract numerical information from the generated chemical images to distinguish between samples.

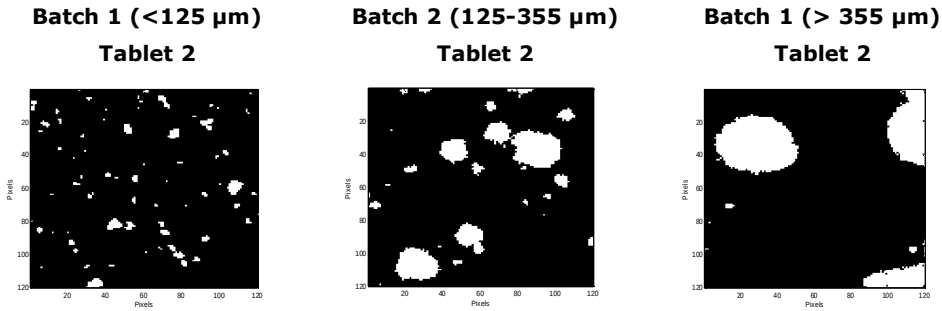


Fig. 6. Binary images of 'tablet 2' from each of the three batches with different ingoing API particle sizes. Similar binary images were made for all twelve tablets. The same threshold value of 0.15 was used to create all binary images.

	No. of domains			**Mean equivalent diameter (µm)**			**Area covered (%)**		
	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3
Tablet 1	69	30	2	71	127	69	4.1	8.2	0.1
Tablet 2	68	25	9	66	158	299	3.6	11.0	16.4
Tablet 3	71	31	7	73	95	124	4.5	4.7	1.9
Tablet 4	104	42	4	73	96	659	6.6	6.0	15.9
Mean	78	32	6 (7)	71	119	288 (479)	4.7	7.5	8.6 (16.2)
RSD (%)	22	22	57	5	25	92	28	37	102

Table 2. Domain statistics of API based on binary images generated from CLS chemical images of API and a threshold value of 0.15. The mean results only calculated for tablet 2 and 4 for batch 3 are shown in brackets, i.e. excluding the two tablets with hardly any API.

NIR-CI homogeneity measure results

The mean standard deviation (std-w) of the API predictions for the four images increased from 4.3 % to 9.6 % to 13.2 % for the three batches with increasing in-going API particle size fraction (Table 1). These results were based on the inherent pixel size of 25 x 25 µm, but this measure of homogeneity only tells something about the homogeneity at this certain spatial resolution or macropixel size. In the presented method of a homogeneity measure for chemical images the function of spatial resolution is taken into account in the calculation. This ensures a more extensive and detailed sampling of each image, which may provide a more comprehensive estimate of the component homogeneity within each sample.

The resulting homogeneity curves from this novel method obtained for the twelve tablet samples in this study are shown in Fig. 7.

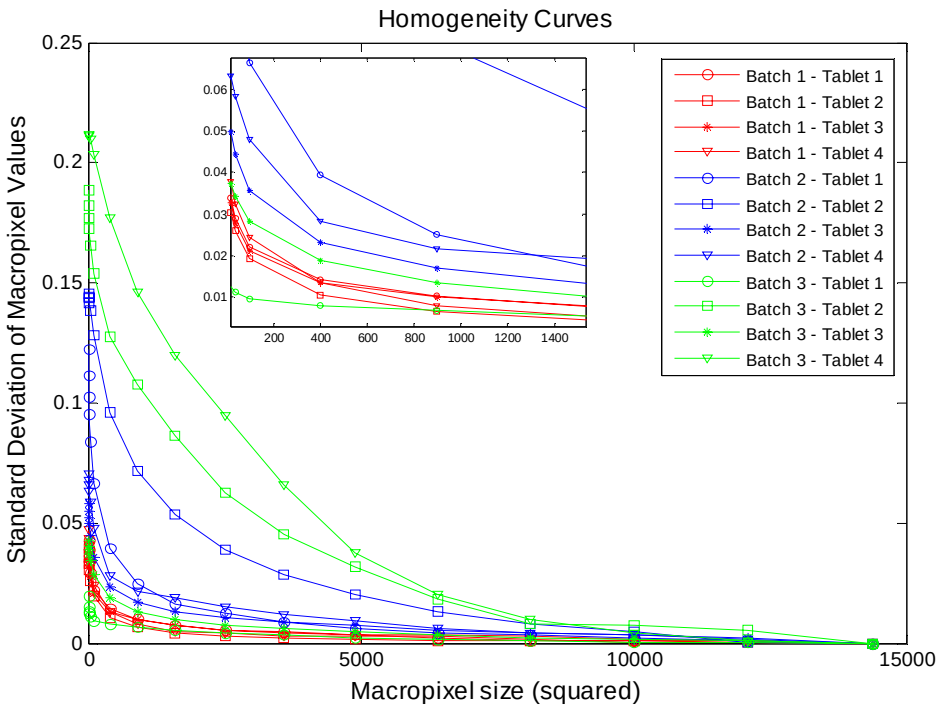


Fig. 7. Homogeneity curves showing standard deviations of CLS predictions versus squared macropixel size. The curves are shown for each of four tablets from each of the three batches. The insert enlarges the area around origin.

Fig. 7 shows that the homogeneity curves were generally lying lower (smaller Std. values) for tablet samples with smaller API domains more evenly spread out on the tablet surface compared to tablet samples from batch 3, where larger API domains were distributed less evenly throughout the tablet surface. The two curves for tablet 1 and 3 for batch 3 break this pattern but they actually just confirm the functionality of this methodology. As mentioned before, there were hardly any API in these samples and they therefore erroneously appear well blended. These homogeneity curves are therefore found among the homogeneity curves for the more homogenous batch 1 (see insert, Fig. 7). Such deviating samples would always easily be captured by comparing the homogeneity curves with the low CLS predictions of API (Table 1) or assay values from other conventional analytical methods.

The visual evaluation of the shape and position of the homogeneity curves would as demonstrated in Fig. 7 in many cases be sufficient for the purpose of comparing the homogeneity of different products. But not all products compared by their homogeneity

curves of the chemical images may be so easy to visually differentiate as for the very illustrative and distinct samples in this study. It would therefore also for this method be useful to move beyond this subjective and qualitative analysis and determine a numerical value as a quantitative measure for this homogeneity parameter. An option to numerical assess homogeneity from these curves would simply be to define a macropixel size and set a specification limit (standard deviation) for which the homogeneity curve values need to be below. This univariate approach is actually similar to the histogram statistics using the inherent spatial resolution but the advantage for the macropixel method is that the intra-tablet homogeneity can be evaluated at any suitable level of scrutiny (i.e. macropixel size). A tablet determined inhomogeneous when sampled with the original pixel size could easily be determined homogenous using other macropixels. And note that a sample is considered homogenous once it complies with acceptance criteria defined by its purpose of use. With the homogeneity curves it is possible to select the acceptance criteria most suitable for the particular product of interest. In practise this would for example mean that for a specific product a standard deviation below 0.005 for squared macropixel size 3600 on the homogeneity curve is assessed satisfactory, which could be based on experience from several development experiments. This evaluation would of course be based on correlation to different product performance results evaluated by other analytical tests. This approach is in accordance with the current homogeneity concept, stating that a blend or tablet batch is considered homogenous if the content uniformity results comply with the acceptance criteria determined by its purpose of use. Selecting the appropriate macropixel for calculation will thus depend on the degree of homogeneity (quality) needed which depends on the purpose of use for each product. It is still a challenge how to select the appropriate macropixel size and standard deviation value as specification limit for each product. However, selecting these parameters is probably most useful for routine quality control purposes, whereas in early formulation development it would be sufficient to compare the standard deviations between products at a certain sampling square without necessarily setting a specification limit.

Another option for numerical evaluation would be to use all the information contained within the homogeneity curves. As the goal is to obtain as low standard deviations between macropixels as possible an obvious evaluation approach is simply to determine the area under each homogeneity curve (AUHC). This method has the advantage that it includes all the information contained in the homogeneity curves. The results from such a calculation by trapezoidal integration are presented in Table 3. These AUHC results also confirm the earlier findings as the area under the curve increases, i.e. higher standard deviations, with increasing ingoing API particle size.

AUHC	**Batch 1**	**Batch 2**	**Batch 3**
Tablet 1	49	114	(30)
Tablet 2	30	305	457
Tablet 3	53	99	(67)
Tablet 4	42	122	594

Table 3. The area under the homogeneity curves (AUHC) from Fig. 7 calculated by trapezoidal integration. The results from the deviating tablet 1 and 3 from batch 3 are shown in brackets.

More knowledge and experience from many different types of products are needed to learn more about how these homogeneity curves derived from chemical images can and should be evaluated and interpreted. However, from the homogeneity curve results obtained in this study, the novel homogeneity approach of chemical images seems promising. Another important image processing tool for evaluating and extracting useful information from chemical images is hereby available.

Conclusion

It is demonstrated in this study, how NIR-CI can be used as a valuable analytical tool to study tablets from formulation development experiments. The results emphasise how the unique information obtainable by NIR-CI on particular the spatial distribution of components makes it possible to determine and understand how the final tablet matrix evolves as a function of the process or ingredient conditions used. For the present formu-

lation experiment it was specifically revealed, how the difference in in-going API particle size affected its size and distribution within the final tablets. However, the experimental approach can most likely be used for many different types of formulation development experiments. NIR-CI was clearly capable of visually revealing the spatial distribution of API and estimating the API concentration in the chemical images. Two image processing methods were also employed. The domain size statistics such as number and size of API domains were calculated from binary images and a novel method to evaluate the homogeneity degree of API within each analysed sample was developed. To advance the NIR-CI technology there is a need to develop such image processing tools to extract useful information from the chemical images preferable by objective and numerically results. Both image processing methods used contributed to this by providing quantitative measures that proved capable of distinguishing the quality of the different tablets and batches.

Traditional NIR spectroscopy and NIR chemical imaging have both proven valuable in pharmaceutical analysis. The benefit of using these two techniques in conjunction was shown in this study where the NIR spectroscopy results confirmed the NIR-CI findings. It suggests NIR spectroscopy and NIR-CI to be used in combination and possibly as the only analytical techniques for some formulation development studies. This may be very useful in early formulation development, for example in screening different formulations, and opens up for a reduced analytical package and thus reduced time and cost for formulation development studies and can also be useful for troubleshooting at later stages. The applied evaluation methods of the NIR spectroscopy data are however ad hoc and not generally valid methods and should therefore be used and evaluated carefully.

Much can still be learnt about the NIR-CI technology and its application to pharmaceutical samples. Despite the limitations and room for improvement in all steps of a full NIR-CI pharmaceutical analysis, the technique provides unique information from solid samples that are invaluable for many pharmaceutical applications. This study adds to the increasing number of NIR-CI applications in pharmaceutical analysis and contributes to the important knowledge foundation being built for the NIR-CI technology.

References

Amigo, J.M., Cruz J., Bautista, M., Maspoch, S., Coello, J., Blanco, M., 2008. Study of pharmaceutical samples by NIR chemical-image and multivariate analysis. Trends Anal. Chem. 27, 696-713.

Blanco, M., Coello, J., Iturriage, H., Maspoch, S., de la Pezuela, C., 1998. Near-infrared spectroscopy in the pharmaceutical industry. Analyst 123, 135R-150R.

Ciurczak, E.W., Drennen, III J.K. (Eds), 2002. Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy, Marcel Dekker Inc., New York.

Clark, D., Henson, M., LaPlant, F., Sasic, S., Zhang, L., 2007. Pharmaceutical Applications of Chemical Mapping and Imaging, In: Pivonka, D.E., Chalmers J.M., Griffiths P.R. (Eds.), The Handbook of Vibrational Spectroscopy, Applications in Life, Pharmaceutical and Natural Sciences, Pharmaceutical Applications, John Wiley & Sons, Ltd. Vol. 5, 1-27.

Clarke, F.C., Hammond, S.V., Jee, R.D., Moffat, A.C., 2002. Determination of the information depth and sample size for the analysis of pharmaceutical materials using reflectance near-infrared microscopy. Appl. Spectrosc. 56, 1475-1483.

Clarke, F., Hammond, S., 2003. NIR microscopy of pharmaceutical dosage forms. Eur. Pharm. Rev. 8, 41-50.

Clarke, F., 2004. Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. Vib. Spectrosc. 34, 25-35.

Juan, A. de, Tauler, R., Dyson, R., Marcolli, C., Rault, M., Maeder, M., 2004. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. Trends Anal. Chem. 23, 70-79.

El-Hagrasy, A.S., Morris, H.R., D'Amico, F., Lodder, R.A., Drennen, III J.K., 2001. Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity. J. Pharm. Sci. 90, 1298-1307.

European Pharmacopoeia 6th Edition (6.2), 2008. <2.9.40> Uniformity of dosage units.

Furukawa, T., Sato, H., Shinzawa, H., Noda, I., Ochiai, S., 2007. Evaluation of homogeneity of binary blends of poly(3-hydroxybutyrate) and poly(L-lactic acid) studied by near infrared chemical imaging (NIRCI). Anal. Sci. 23, 871-876.

Gallagher, N.B., 2007. Detection, Classification, and Quantification in Hyperspectral Images Using Classical Least Squares Models. In: Grahn, H.F., Geladi, P. (Eds.), Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons Ltd., 181-202.

Gendrin, C., Roggo, Y., Collet, C., 2007. Content uniformity of pharmaceutical solid dosage forms by hyperspectral imaging: a feasibility study. Talanta 73, 733-741.

Gendrin, C., Roggo, Y., Spiegel, C., Collet, C., 2008. Monitoring galenical process development by near infrared chemical imaging: one case study. Eur. J. Pharm. Biopharm. 68, 828-837.

Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Bell, S.E.J., 2008. Recent applications of chemical imaging to pharmaceutical process monitoring and quality control. Eur. J. Pharm. Biopharm. 69, 10-22.

Henson, M.J., Zhang L., 2006. Drug Characterization in Low Dosage Pharmaceutical Tablets Using Raman Microscopic Mapping. Appl. Spectrosc. 60, 1247-1255.

Hilden, L.R., Pommier, C.J., Badawy, S.I.F., Friedman, E.M., 2008. NIR chemical imaging to guide/support BMS-561389 tablet formulation development. Int. J. Pharm. 353, 283-290.

Jovanovic, N., Gerich, A., Bouchard, A., Jiskoot, W., 2006. Near-infrared imaging for studying homogeneity of protein-sugar mixtures. Pharm. Res. 23, 2002-2013.

Hammond, S.V., 1998. NIR microspectroscopy and the control of quality in pharmaceutical production. Eur. Pharm. Rev. 3, 47-51.

Koehler, F.W., Lee, E., Kidder, L.H., Lewis, E.N., 2002. Near infrared spectroscopy: the practical chemical imaging solution. Spectrosc. Eur. 14, 12-19.

LaPlant, F., 2004. Factors affecting NIR chemical images of solid dosage forms. Am. Pharm. Rev. 7, 16-24.

Lewis, E.N., Carroll, J.E., Clarke, F., 2001. A near infrared view of pharmaceutical formulation analysis. NIR News 12, 16-18.

Lewis, E.N., Schoppelrei, J., Lee, E., 2004. Near-infrared chemical imaging and the PAT initiative. Spectroscopy 19, 26-36.

Lewis, E.N., Schoppelrei, J., Lee, E., Kidder, L.H., 2005a. Near-infrared chemical imaging as a process analytical tool. In: Bakeev, K.A. (Ed.), Process Analytical Technology, Blackwell Publishing Ltd, Oxford, 187-225.

Lewis, E.N., Kidder, L.H., Lee, E., 2005b. NIR chemical imaging – near-infrared spectroscopy on steroids. NIR News 16, 2-4.

Lewis, N.E., Dubois, J., Kidder, L.H., Haber, K.S., 2007. Near Infrared Chemical Imaging: Beyond the Pictures. In: Techniques and Applications of Hyperspectral Image Analysis, Grahn, H.F., Geladi P. (Eds.), John Wiley & Sons Ltd., 335-361.

Li, W., Woldu, A., Kelly, R., McCool, J., Bruce, R., Rasmussen, H., Cunningham, J., Winstead, D., 2008. Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging. Int. J. Pharm. 350, 369-373.

Luypaert, J., Massart, D.L., Vander Heyden, Y., 2007. Near-infrared spectroscopy applications in pharmaceutical analysis. Talanta 72, 865-883.

Lyon, R.C., Lester, D.S., Lewis, E.N., Lee, E., Yu, L.X., Jefferson, E.H., Hussain A.S., 2002. Near-infrared spectral imaging for quality assurance of pharmaceutical products: analysis of tablets to assess powder blend homogeneity. AAPSPharmSciTech 3, 1-15.

Lyon, R.C., Jefferson, E.H., Ellison, C.D., Buhse, L.F., Spencer, J.A., Nasr, M.M., Hussain, A.S., 2003. Exploring pharmaceutical applications of near-infrared technology. Am. Pharm. Rev. 6, 62-70.

Ma, H., Anderson, C.A., 2008. Characterization of pharmaceutical powder blends by NIR chemical imaging. J. Pharm. Sci. 97, 369-373.

Martens, H., Naes, T., 1989. Multivariate Calibration, John Wiley & Sons, Chichester.

Mathworks Inc., Matlab® version 7.2.0 R2006a (www.mathworks.com).

Otsu N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62-66.

PLS_Toolbox version 4.1.1, Eigenvector Research Inc., (www.eigenvector.com).

Ravn, C., Skibsted, E., Bro, R., 2008. Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms – comparing common calibration approaches, J. Pharm. Biomed. Anal. 48, 554-561.

Reich, G., 2005. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. Adv. Drug Delivery Rev. 57, 1109-1143.

Sasic, S., 2007. An in-depth analysis of Raman and near-infrared chemical images of common pharmaceutical tablets. Appl. Spectrosc. 61, 239-250.

Sekulic, S.S., Wakeman, J., Doherty P., Hailey P.A., 1998. Automated system for the on-line moniroring of powder blending processes using near-infrared spectroscopy PartII. Qualitative approaches to blend evaluation. J. Pharm. Biomed. Anal. 17, 1285-1309.

United States Pharmacopeia, 2008. USP31-NF26, General chapters: <905> Uniformity of dosage units.

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), 2003. Guidance for Industry: Powder Blends and Finished Dosage Units – Stratified In-Process Dosage Unit Sampling and Assessment. Draft Guidance, Pharmaceutical CGMPs.

PAPER III

European Journal of Pharmaceutical Sciences 37 (2009) 76–82

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

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

Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging

José Manuel Amigo [a,*], Carsten Ravn [b,c]

[a] Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
[b] Faculty of Pharmaceutical Sciences, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark
[c] Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark

ARTICLE INFO

Article history:
Received 4 September 2008
Received in revised form
19 November 2008
Accepted 7 January 2009
Available online 17 January 2009

Keywords:
Pharmaceutical tablets
Direct quantification
Spatial distribution in surfaces
CLS
MCR-ALS
Augmented MCR-ALS

ABSTRACT

Near Infrared Chemical Imaging (NIR-CI) is an attractive technique in pharmaceutical development and manufacturing, where new and more robust methods for assessment of the quality of the final dosage products are continuously demanded. The pharmaceutical manufacturing process of tablets is usually composed by several unit operations such as blending, granulation, compression, etc. Having reliable, robust and timely information about the development of the process is mandatory to assure the quality of the final product.

One of the main advantages of NIR-CI is the capability of recording a great amount of spectral information in short time. To extract the relevant information from NIR-CI images, several Chemometric methods, like Partial Least Squares Regression, have been demonstrated to be powerful tools. Nevertheless, these methods require a calibration phase. Developing new methods that do not need any prior calibration would be a welcome development.

In this context, we study the potential usefulness of Classical Least Squares and Multivariate Curve Resolution models to provide quantitative and spatial information of all the ingredients in a complex tablet matrix composed of five components without the development of any previous calibration model. The distribution of the analytes in the surfaces, as well as the quantitative determination of the five components is studied and tested.

1. Introduction

The analytical techniques of hyperspectral imaging, like Near Infrared Chemical Imaging (NIR-CI) or Raman Chemical Imaging (Raman-CI), comprise a group of attractive techniques in the development and manufacturing of pharmaceutical tablets (Chan et al., 2003; Shah et al., 2007; Zhang et al., 2005) where more robust methods are continuously demanded for assessment of the quality of the final products. This attractiveness is due to the capability of the mentioned techniques to obtain a great amount of spectral and spatial information in defined surfaces of pharmaceutical products (Furukawa et al., 2007). The possibility of recording one spectral profile in a wide wavelength range for each defined pixel area of the tablet poses new challenges for the development of robust and reliable methodologies of data analysis to extract all the desirable information.

Basically, the expectations on Chemical Imaging are focused on obtaining quantitative information about the content of each component in the tablet (Chevallier et al., 2006; Zhang et al., 2005) providing at the same time reliable information about the distribution of the component in the surface of the tablet (Furukawa et al., 2007; Lyon et al., 2002) to assure the quality of the final product.

As mentioned above, the great interest of CI techniques relies in the capability to obtain thousand of spectra of a defined surface of the tablet. To extract the quantitative information, powerful mathematical tools are demanded. In this context, several Chemometric methods, like Principal Components Regression (PCR) (Hamilton et al., 2002), Partial Least Squares Regression (PLS) (Burger and Geladi, 2006; Furukawa et al., 2007; Jovanovic et al., 2006; Lied et al., 2000; Roggo et al., 2005; Svensson et al., 2006) or its variant PLS2 (Gendrin et al., 2007; Li et al., 2008), have already been demonstrated to be powerful tools. Nevertheless, these methods require a previous calibration phase, being tedious and time-consuming. Consequently, the development of new methods that do not need any prior calibration stage would be a welcome development.

To overcome the problem of the previous calibration stage, several alternatives can be proposed (Amigo et al., 2008), being

* *Corresponding author*. Tel.: +45 35 33 25 70; fax: +45 35 28 32 45l.
E-mail address: jmar@life.ku.dk (J.M. Amigo).

0928-0987/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejps.2009.01.001

Classical Least Squares (CLS) (Chan et al., 2005; Gendrin et al., 2007), Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) (de Juan et al., 2004; Zhang et al., 2005) and MCR-ALS in augmented fashion (Amigo et al., 2008; de Juan et al., 2004) the most promising techniques.

In this context, we study and test the potential usefulness of Classical Least Squares and Multivariate Curve Resolution models to provide quantitative and spatial information about the distribution of each analyte in the surface of a complex pharmaceutical mixture composed of five components without the development of any previous calibration model and by using Near Infrared Chemical Imaging.

2. A brief theoretical background

An exhaustive theoretical background of the proposed models will not be given here but readers are encouraged to look for more information in the supplied references.

When a NIR-Chemical Image is measured for a single tablet, the structure of the obtained dataset is a three-dimensional array known as hyperspectral cube, **D**, whose dimensions are $(X \times Y \times \lambda)$ (Zhang et al., 2005). X and Y refer to the spatial dimensions of the surface; whereas λ refers to the spectral pattern measured in a whole wavelength range for each xyth pixel of the tablet.

The basis of CLS and MCR-ALS could be argued to be the same. Both methods assume that the absorbance follows a linear behaviour with the concentration (Beer–Lambert Law) and that the sum of the individual absorbances for each component equals the total absorbance for each pixel (Eq. (1)).

$$a_{xy\lambda} = c_{1xy}\varepsilon_{1\lambda} + c_{2xy}\varepsilon_{2\lambda} + c_{3xy}\varepsilon_{3\lambda} + \ldots + c_{fxy}\varepsilon_{f\lambda} \quad (1)$$

where $a_{xy\lambda}$ is the absorbance at each xyth pixel dimension and at each wavelength (λ) and, c_{fxy} and $\varepsilon_{f\lambda}$ the concentration and molar absorptivity, respectively, for each f component of the surface.

Mathematically speaking, both are bilinear models. So a previous stage of unfolding is mandatory in order to adapt the three-dimensional structure to bilinear models (Fig. 1). Consequently, the dimensions of the unfolded image **D** are $(XY \times \lambda)$.

The main difference between CLS and MCR-ALS relies in the way of obtaining the quantitative information. CLS is aimed to obtain the concentration of each component by direct regression of **D** by using the pure spectra (Chan et al., 2005; Gendrin et al., 2007). Despite the easiness of calculating this concentration matrix, CLS has one main drawback: it works perfectly if the spectra of the components are known and, what is more important, if there is no other variability source in the sample (for example, interaction between components, distribution of moisture and hydration/adsorption water, etc.) that can promote other features in the sample not related to the pure components. CLS is constrained to model all the variability of the sample by using just the pure spectra. For this question, a more feasible method would be preferred.

Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) (de Juan and Tauler, 2006; de Juan et al., 2004; Tauler, 1995) decomposes the matrix **D** $(XY \times \lambda)$ into the product of two matrices, **C** $(XY \times F)$, containing the concentration profiles and $\mathbf{S^T}$ $(F \times \lambda)$, containing the spectral profiles for each F component (Eq. (2)):

$$\mathbf{D} = \mathbf{C}\mathbf{S^T} + \mathbf{E} \quad (2)$$

where **E** $(XY \times \lambda)$ contains the experimental error. The main difference from CLS is that MCR-ALS works by iteratively optimizing the matrices **C** and $\mathbf{S^T}$. Thus, the result is not constrained to the pure spectra, allowing MCR-ALS to cope with minor sources of variability that may be in the tablet.

To start the iterations, MCR-ALS needs initial estimates, as well as several constraints based on chemical knowledge of the sample studied. Using the pure spectra of the components (if they are available) as initial estimations becomes one of the most attractive ways of including information about the tablet. Several constraints based on chemical knowledge or on mathematical features of the data can be imposed to the iterations, being the most useful in the quantification of mixtures by means of Chemical Image analysis are: (1) non-negativity in concentration and/or spectral profiles, which imposes that concentration profiles of the components are supposed to be always positives and (2) closure, where each pixel can be supposed to accomplish a constant mass balance of 1, representing 100% of global concentration.

The main disadvantage of MCR-ALS in Chemical Imaging is the lack of selectivity in the surface (Amigo et al., 2008); i.e., to obtain good results, each component must have a selective area in the tablet, accounting for the variability of the concentration in the sample. This lack of selectivity can be associated to a rank-deficiency problem. MCR-ALS only works properly if the rank of

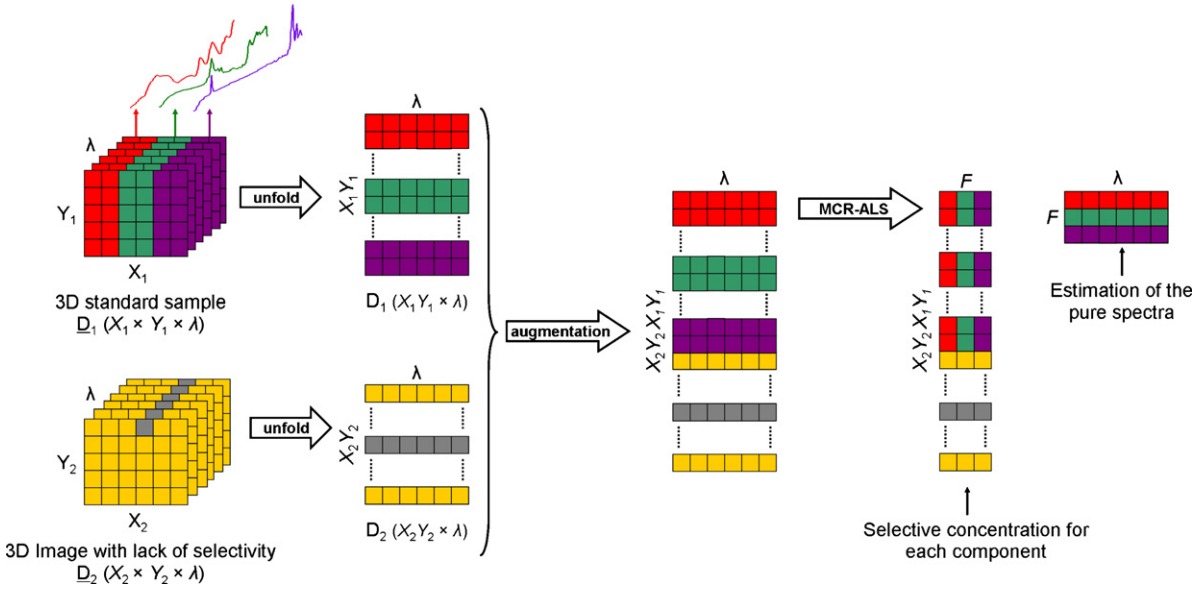


Fig. 1. Augmentation of images and MCR-ALS analysis. The example is illustrated by assuming three components in the samples.

the surface, e.g. the number of mathematical components, is equal to the number of chemical components present in the tablet. Several alternatives has been already proposed to study the chemical rank of images, like Principal Component Analysis (PCA), or more specifically, the Fixed Size Image Window-Evolving Factor Analysis (FSIW-EFA) (de Juan et al., 2004). But, still, this is a complicated issue to solve when the quantification of minor components is pursued.

To overcome the problem of lack of selectivity in MCR-ALS, an alternative has already been formulated: the augmentation of the original sample with more images. This method, proposed by de Juan et al. (2004), treats to analyse several samples simultaneously, profiting the qualities of one to each other (de Juan and Tauler, 2006). The main feature of the matrix augmentation is that the unfolded target sample ($\mathbf{D_2}$) is augmented with an image without lack of selectivity problems (full-rank, or standard image, $\mathbf{D_1}$) in a column-wise fashion (Eq. (3)):

$$\begin{bmatrix}\mathbf{D_1}\\ \mathbf{D_2}\end{bmatrix} = \begin{bmatrix}\mathbf{C_1}\\ \mathbf{C_2}\end{bmatrix}[\mathbf{S}]^{\mathbf{T}} + \begin{bmatrix}\mathbf{E_1}\\ \mathbf{E_2}\end{bmatrix} \quad (3)$$

where $\mathbf{D_1}$ ($X_1Y_1 \times \lambda$) represents the full-rank image and $\mathbf{D_2}$ ($X_2Y_2 \times \lambda$) is the rank-deficient image (the target image). X_1, Y_1 and X_2, Y_2 represent the dimensions of the first and the second unfolded image, respectively. $\mathbf{C_1}$ ($X_1Y_1 \times F$) and $\mathbf{C_2}$ ($X_2Y_2 \times F$) are the concentration surfaces simultaneously obtained for each image and $\mathbf{S^T}$ ($F \times \lambda$) is the transpose of the global spectral profiles. $\mathbf{E_1}$ ($X_1Y_1 \times \lambda$) and $\mathbf{E_2}$ ($X_2Y_2 \times \lambda$) correspond to the residual matrices.

This "standard" image is composed by specific sections of each pure component of the tablet (Fig. 1), and can be easily provided by creating it with images of the pure components. Working with this standard sample, much more valuable and reliable information is modelled by MCR-ALS at the same time, since experimental variability is also reflected in the augmented image, allowing MCR-ALS to get hold of selective information in the iterations and, thus, minimizing the possible lack of selectivity in the target sample.

3. Experimental

3.1. Reagents and instruments

Six replicates of a five-compound conventional pharmaceutical tablet formulation were used to produce the dataset. The nominal composition was Active Pharmaceutical Ingredient (API: 6.3%, w/w), MicroCrystalline Cellulose (MCC: 20.0%, w/w), Lactose (Lact: 71.5%, w/w), and the common lubricants Magnesium Stearate (0.75%, w/w) and Talc (1.5%, w/w).

The mixing of dry-blend formulation was performed in a drum-mixer and compressed into tablets of 175 mg on a six-punch station rotary tablet press. A flat punch-set was used to obtain a flat sample surface of 8 mm of diameter and 2.6 mm of thickness. Pure compound reference tablets of the five components were also produced by compressing 250 mg of each into 8-mm diameter wafers on a hydraulic tablet press using 10 kN pressure during 10 s. The standard sample used for augmentation in MCR-ALS was performed by sectioning a representative area of 10 square pixels of each pure component image and rearranging the five areas in a new image with 500 spectra (100 for each component).

3.2. Data acquisition and preprocessing

Each tablet was fixed onto a microscope slide using cyanoacrylate adhesive and measured directly on the flat tablet surface. Samples were analysed on a NIR line mapping system (Spectrum Spotlight 350 FT-NIR Microscope, PerkinElmer, UK) from which 16 spectra were collected in each acquisition from a linear MCT detector array. An area of 2 mm × 2 mm were analysed using pixel size 25 μm × 25 μm thus obtaining a total of 6400 spectra (=pixels) for each image. Each spectrum was the average of 8 scans from wavelength region 7800–4000 cm^{-1} using 16 cm^{-1} spectral resolution. More information about the data collection can be found in reference, Ravn et al. (2008).

One previous step of unfolding the 3D hyperspectral data cube to a 2D matrix is mandatory to adapt the three-dimensional data cube (hyperspectral image) for further preprocessing and modelling steps with bidimensional methods (Amigo et al., 2008; Ravn et al., 2008) (First step in Fig. 1).

Standard Normal Variate (SNV) and Savitzky–Golay smoothing of the spectral pattern with a window size of 15 and a second polynomial order were applied to the unfolded six samples as well as to the standard sample.

3.3. Models application and validation of the results

CLS, Savitzky–Golay smoothing and SNV algorithms were implemented in a routine named BACRA working under MatLab v. 7.5 (MathWorks, Massachusetts). The software and detailed easy-to-follow instructions are freely available from the authors or via Internet (http://www.models.kvl.dk/users/jose_manuel_amigo/index.htm). For MCR-ALS and augmented-MCR-ALS application, the software provided by Tauler et al., was used (http://www.ub.es/gesq/mcr.htm; Jaumot et al., 2005).

For MCR-ALS analysis, non-negativity of concentration profiles and closure constraints were imposed. Pure spectra of the five components were used as initial estimations. The percentage of lack of fit (% LOF) between the obtained results and the original data gives a measure of the fit quality and has been calculated as follows:

$$\%LOF = 100 \times \sqrt{\frac{\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{\lambda=1}^{\lambda} e_{xy\lambda}^2}{\sum\sum\sum d_{xy\lambda}^2}} \quad (4)$$

where $e_{xy\lambda}$ and $d_{xy\lambda}$ represent the $xy\lambda$th terms of the residuals and original dataset, respectively.

The precision and the predictive capabilities of the different models were evaluated by calculating the Root Mean Square Error of Prediction (RMSEP) and the Standard Error of Prediction (SEP) between the mean value of concentration obtained for each component in each sample and its nominal concentration accordingly with Eqs. (5) and (6), respectively.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}{n}} \quad (5)$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i - bias)^2}{n-1}} \quad (6)$$

where $\hat{c}_i$ denotes the mean value of concentration obtained for each component in the surface of each n sample and c_i is the nominal concentration. The *bias* was calculated as follows:

$$bias = \frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)}{n} \quad (7)$$

To evaluate the distribution of the analytes in the tablet, histograms of the concentration surfaces were studied and the Standard Deviation Between Pixels (SDBP) (Furukawa et al., 2007; Jovanovic et al., 2006) was calculated (Eq. (8)) accounting for the dispersion of the analyte in the surface of the sample:

$$SDBP = \sqrt{\frac{\sum_{x=1}^{X}\sum_{y=1}^{Y}(\hat{c}_{xy} - \bar{c})^2}{XY-1}} \quad (8)$$

Table 1
CLS, MCR-ALS and augmented MCR-ALS mean concentrations (%) obtained for the six replicates.

	API[a] 6.2[d]	MgSt[b] 0.75[d]	MCC 20[d]	Lact 71.49[d]	Talc[c] 1.5[d]
CLS					
Mean	8.0 (20.2)[e]	41.2 (20.3)	18.7 (3.3)	80.2 (3.1)	50.2 (0.2)
RMSEP	18.7	44.9	3.5	9.5	48.7
SEP	20.3	20.4	3.3	3.2	0.2
MCR-ALS					
Mean	1.0 (1.4)	0.1 (0.1)	22.2 (3.4)	76.0 (3.2)	–
RMSEP	4.9	0.6	3.7	6.1	–
SEP	25.1	0.2	10.2	43.5	–
Augmented MCR-ALS					
Mean	4.6 (0.4)	3.8 (0.2)	19.2 (2.2)	71.1 (3.2)	1.7 (0.2)
RMSEP	1.6	3.1	2.5	2.8	0.3
SEP	0.4	0.2	2.6	2.9	0.2

[a] MCR-ALS: values obtained without sample 2.
[b] MCR-ALS: values obtained without samples 2, 3, 4 and 5.
[c] MCR-ALS: not possible to obtain an estimation of the concentration for Talc.
[d] Nominal.
[e] Standard deviation in parenthesis.

where $\hat{c}_{xy}$ is the calculated concentration in each xyth pixel, $\bar{c}$ represents the mean value of the concentration obtained for each analyte and XY denotes the total number of pixels.

4. Results and discussion

4.1. Quantitative results

CLS results for the six samples are depicted in Table 1. At first sight, it can be observed that good values of mean concentration and standard deviation have been obtained for the two major components (MCC and Lact). RMSEP and SEP values obtained for MCC are slightly high, but still under a good predictive capability. For Lact, RMSEP value is higher than SEP, indicating certain overfitting of the model.

Despite these reasonably good results for MCC and Lact, results obtained for the minor components (API, MgSt and Talc) did not cover all the expectations. Results are far away from being representative of the nominal concentration of these components in the tablet. The reason may be found taking into consideration the mutual correlation between components (Fig. 2). To illustrate this, the correlation coefficients between pure spectra have been calculated and are shown in Table 2.

Table 2
Correlation coefficients between pure spectra.

	API	MgSt	MCC	Lact	Talc
API	1.0000	0.9377	0.9072	0.9086	0.9539
MgSt		1.0000	0.9243	0.9233	0.9503
MCC			1.0000	0.9829	0.9291
Lact				1.0000	0.9187
Talc					1.0000

In general all the correlation coefficients present a high value (more than 0.90). Due to their similar chemical structure, the highest value of correlation is found between MCC and Lact whereas the lowest values are between the minor components (API, MgSt and Talc) and the major components (MCC and Lact). One point to consider is that the concentration of the two major components represents more than 90% of the sample. Both facts make CLS unable to obtain reliable estimations of the minor components, with CLS being highly affected for the high values of correlation between the components of the sample.

For MCR-ALS models the percentage of lack of fit (%LOF) and explained variance were of 3.80% and 99.87%, respectively, indicating that the models are of good quality. Nevertheless, looking at the quantitative results in Table 1, it can be observed that the

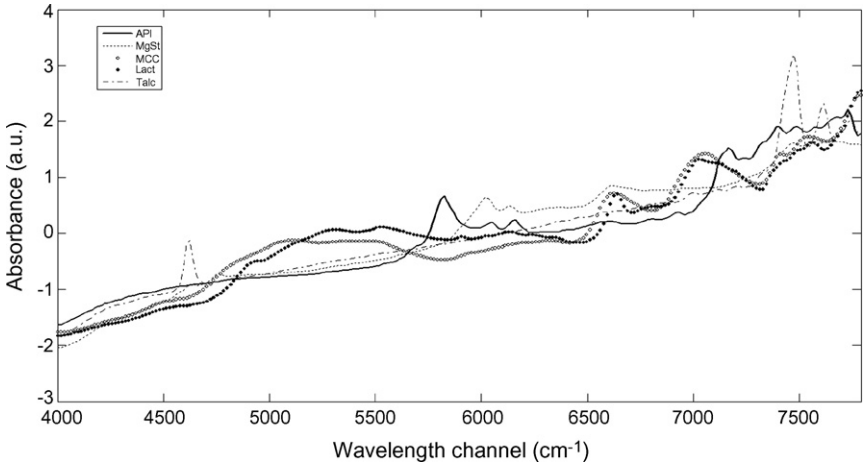


Fig. 2. Pure spectra of the five components composing the samples (after SNV and smoothing pretreatments).

concentrations obtained with MCR-ALS did not correspond to the nominal composition of the tablets. For major components, the calculated concentration covers the expected range. Nevertheless, their RMSEP and SEP values indicate that the predictive capability of the model is not of good enough quality. Attending to the minor components it is remarkable that it was unable to quantify Talc in any of the six samples; whereas it was only possible to quantify API and MgSt in one and five of the samples, respectively. These bad results for minor components reflect the above-mentioned problem of lack of selectivity (see Section 2).

The results obtained using the augmented version of MCR-ALS are highly worth mentioning (Table 1). The percentage of Lack Of Fit (%LOF) and explained variance were of 3.90% and 99.80%, respectively. The concentrations obtained for major components (Lact and MCC) are in the same quality as those obtained with CLS. Nevertheless, the low values obtained for RMSEP and SEP indicate the high predictive capability of augmented MCR-ALS. The good results obtained for Talc are especially remarkable. Its calculated concentration and standard deviation contained the nominal concentration in the tablet. Also good result of RMSEP and SEP were obtained. For API, lower prediction was obtained; whereas a higher concentration was obtained for MgSt. Nevertheless, their standard deviations as well as their RMSEP and SEP values were of high quality, indicating a good predictive capability of the model. These results may conclude that augmented MCR-ALS is a good methodology to offer reliable and semi-quantitative information about the concentration of minor components in tablets, being very accurate in the quantification of major components.

It should be noted that none of the methods actually provide satisfactory quantification results in relation to the European standard practice that the active ingredient in drug products should range between 95% and 105% of the declared content. Nevertheless, measuring an area of 2 mm × 2 mm and the surface of the tablet would also be inadequate for that purpose. The aim is to provide accurate/reliable quantitative results to obtain the most accurate information of the spatial distribution of components. To evaluate the quantitative results from the different models in this study we compare with the theoretical concentrations which is obviously and approximation.

Table 3
Standard Deviation Between Pixels (SDBP) for the concentrations obtained with augmented MCR-ALS.

	API	MgSt	MCC	Lact	Talc
Sample 1	2.9	2.2	6.2	7.4	1.4
Sample 2	2.6	2.1	9.1	10.1	1.4
Sample 3	4.1	2.4	9.3	10.9	1.5
Sample 4	2.8	2.2	10.2	11.0	1.4
Sample 5	2.5	2.3	11.0	11.7	1.2
Sample 6	4.7	2.2	8.8	9.8	1.3

4.2. Spatial distribution of the analytes in the surface

As an illustrative example, the concentration surfaces for sample 1 obtained with the three methodologies proposed are depicted in Fig. 3. The concentration surfaces obtained with CLS (upper row) denoted the lack of predictive capability of this method for MgSt and Talc. Looking at the concentration maps obtained with MCR-ALS (middle) similar conclusions can be extrapolated from the CLS results.

As we expected from the quantitative results, the augmented version of MCR-ALS offered a more reliable semi-quantitative overview of the distribution of each component in the sample (Fig. 3, bottom). This distribution was studied by calculating the SDBP (Table 3) and depicting the histograms of the distribution for each sample (Fig. 4). Several conclusions can be extracted from the histograms: in general, very close distribution histograms were obtained in the six samples, giving an idea of the similar distribution of the analytes between the samples. Moreover, most of the distributions presented a Gaussian shape. The distribution of the minor components is characterized by having a high number of pixels of 0 concentration, being their distribution very close to this concentration. Another remarkable information is that histograms of API and MgSt are highly correlated in the six samples. The

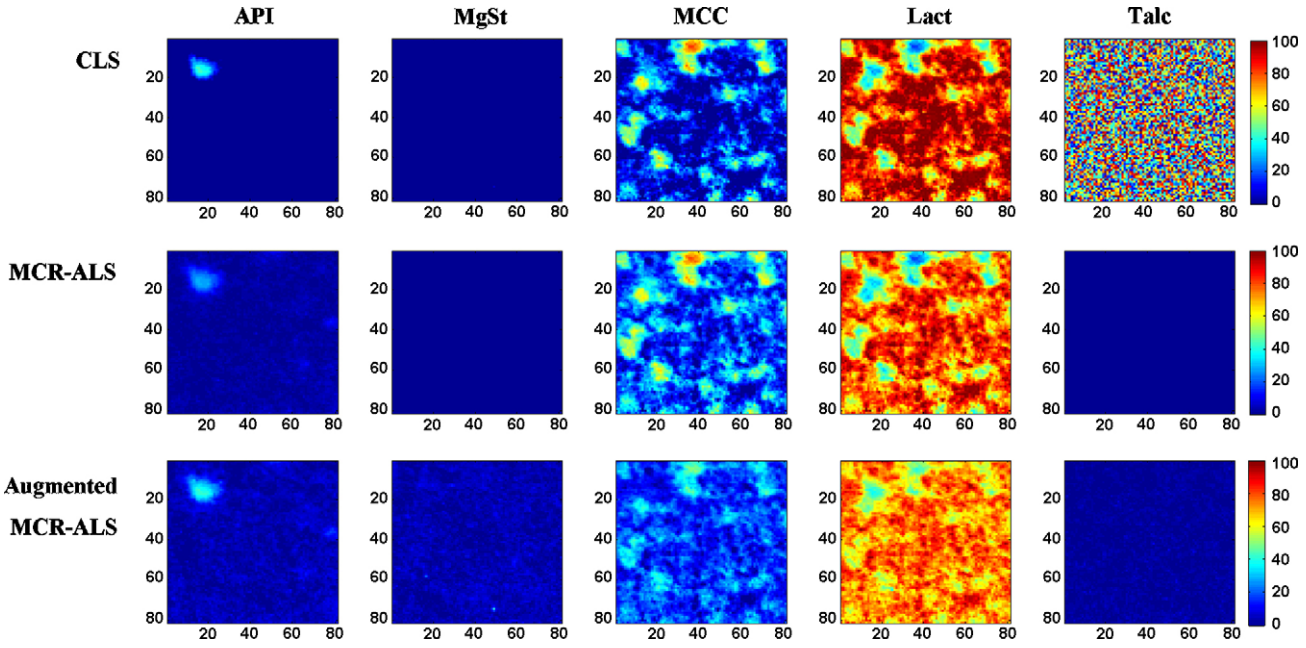


Fig. 3. Concentration surfaces obtained for one of the tablets by applying CLS (upper), MCR-ALS (middle) and augmented MCR-ALS (below). All images are scaled to same color intensity values.

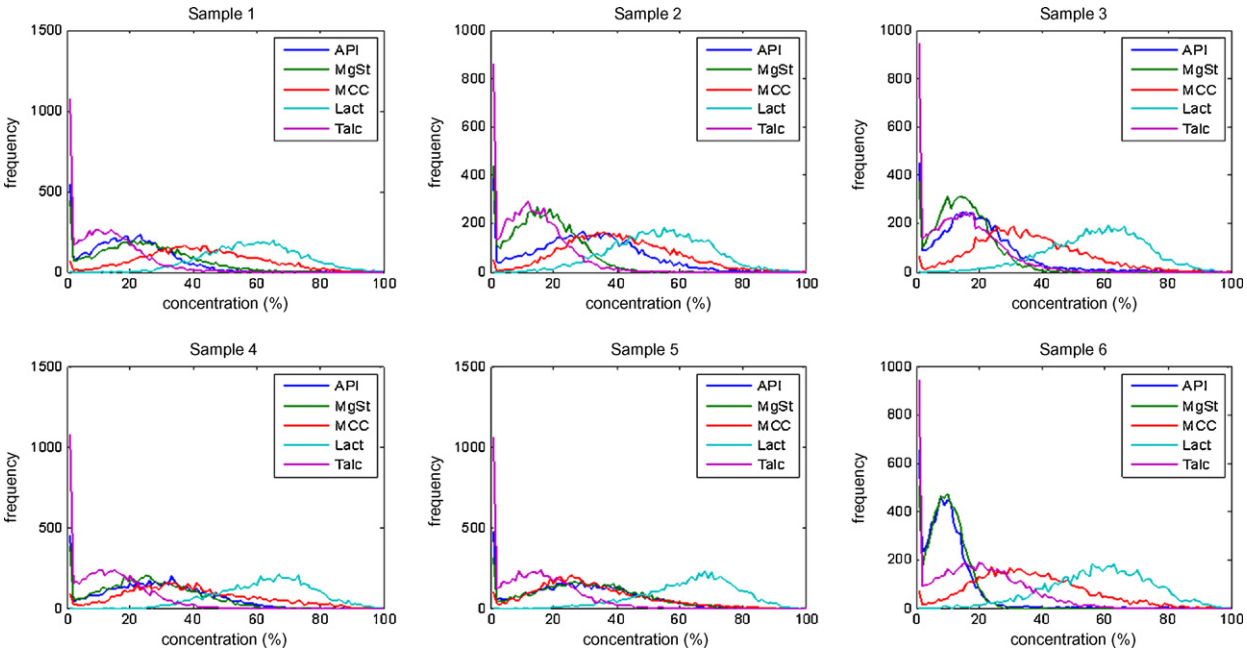


Fig. 4. Histograms of concentrations obtained for the six samples by means of augmented MCR-ALS.

distribution of the major components was as expected, covering most of the concentration range.

One special feature of the obtained histograms is that none of the Gaussian distribution showed a thin shape. This fact is corroborated with the high values of SDPB obtained (Table 3) in all the samples, indicating a high level of dispersion in all the samples for major components, specially.

5. Conclusions and further perspectives

In this work, several methods have been tested to overcome the possibility of direct quantification of major and minor components in pharmaceutical samples by using NIR-Chemical Imaging without a previous stage of calibration.

The quantitative results obtained denoted that the most robust and reliable results have been obtained by means of the augmented version of Multivariate Curve Resolution. This working methodology has been demonstrated to offer reliable and quantitative information about the concentration of major components, as well as a semi-quantitative estimation of minor components in tablets. These features allow the augmented version of MCR-ALS to become a very useful methodology in the field of, for example, blending processes monitoring.

It is also remarkable the usefulness of studying the distribution of each analyte in the surface of the tablet. This target has been overcome by using the histograms and SDBP of the results obtained with augmented MCR-ALS. These two parameters have offered an estimation of the real distribution of each analyte in the sample, being very useful when the total homogeneous distribution in the tablet in a blending process is pursued.

Acknowledgements

José Manuel Amigo wants to thank the Danish Research Council for Technology and Production Sciences for his post-doctoral fellowship.

References

Amigo, J.M., Bautista, M., Cruz, J., Coello, J., Maspoch, S., Blanco, M., 2008. Study of pharmaceutical samples by NIR chemical image and multivariate analysis. Trends in Analytical Chemistry 27 (8), 696–713.

Burger, J., Geladi, P., 2006. Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. Analyst 131 (10), 1152–1160.

Chan, K.L.A., Elkhider, N., Kazarian, S.G., 2005. Spectroscopic imaging of compacted pharmaceutical tablets. Chemical Engineering Research & Design 83 (A11), 1303–1310.

Chan, K.L.A., Hammond, S.V., Kazarian, S.G., 2003. Applications of attenuated total reflection infrared spectroscopic imaging to pharmaceutical formulations. Analytical Chemistry 75 (9), 2140–2146.

Chevallier, S., Bertrand, D., Kohler, A., Courcoux, P., 2006. Application of PLS-DA in multivariate image analysis. Journal of Chemometrics 20 (5), 221–229.

de Juan, A., Tauler, R., 2006. Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. Critical Reviews in Analytical Chemistry 36 (3–4), 163–176.

de Juan, A., Tauler, R., Dyson, R., Marcolli, C., Rault, M., Maeder, M., 2004. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. Trac-Trends in Analytical Chemistry 23 (1), 70–79.

Furukawa, T., Sato, H., Shinzawa, H., Noda, I., Ochiai, S., 2007. Evaluation of homogeneity of binary blends of poly(3-hydroxybutyrate) and poly(L-lactic acid) studied by near infrared chemical imaging (NIRCI). Analytical Sciences 23 (7), 871–876.

Gendrin, C., Roggo, Y., Collet, C., 2007. Content uniformity of pharmaceutical solid dosage forms by near infrared hyperspectral imaging: a feasibility study. Talanta 73 (4), 733–741.

Hamilton, S.J., Lowell, A.E., Lodder, RA, 2002. Hyperspectral techniques in analysis of oral dosage forms. Journal of Biomedical Optics 7 (4), 561–570, http://www.models.kvl.dk/users/jose_manuel_amigo/index.htm. http://www.ub.es/gesq/mcr.htm.

Jaumot, J., Gargallo, R., de Juan, A., Tauler, R., 2005. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. Chemometrics and Intelligent Laboratory Systems 76 (1), 101–110.

Jovanovic, N., Gerich, A., Bouchard, A., Jiskoot, W., 2006. Near-infrared imaging for studying homogeneity of protein–sugar mixtures. Pharmaceutical Research 23 (9), 2002–2013.

Li, W.Y., Woldu, A., Kelly, R., McCool, J., Bruce, R., Rasmussen, H., Cunningham, J., Winstead, D., 2008. Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging. International Journal of Pharmaceutics 350 (1–2), 369–373.

Lied, T.T., Geladi, P., Esbensen, K.H., 2000. Multivariate image regression (MIR): implementation of image PLSR-first forays. Journal of Chemometrics 14 (5–6), 585–598.

Lyon, R.C., Lester, D.S., Lewis, E.N., Lee, E., Yu, L.X., Jefferson, E.H., Hussain, A.S., 2002. Near-infrared spectral imaging for quality assurance of pharmaceutical

products: analysis of tablets to assess powder blend homogeneity. AAPS PharmSciTech 3 (3), 1–17.
Ravn, C., Bro, R., Skibsted, E., 2008. Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—comparing common calibration approaches. Journal of Pharmaceutical and Biomedical Analysis 48, 554–561.
Roggo, Y., Edmond, A., Chalus, P., Ulmschneider, M., 2005. Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms. Analytica Chimica Acta 535 (1–2), 79–87.
Shah, R.B., Tawakkul, M.A., Khan, M.A., 2007. Process analytical technology: chemometric analysis of Raman and near infra-red spectroscopic data for predicting physical properties of extended release matrix tablets. Journal of Pharmaceutical Sciences 96 (5), 1356–1365.
Svensson, O., Abrahamsson, K., Engelbrektsson, J., Nicholas, M., Wikstrom, H., Josefson, M., 2006. An evaluation of 2D-wavelet filters for estimation of differences in textures of pharmaceutical tablets. Chemometrics and Intelligent Laboratory Systems 84 (1–2), 3–8.
Tauler, R., 1995. Multivariate curve resolution applied to second order data. Chemometrics and Intelligent Laboratory Systems 30 (1), 133–146.
Zhang, L., Henson, M.J., Sekulic, S.S., 2005. Multivariate data analysis for Raman imaging of a model pharmaceutical tablet. Analytica Chimica Acta 545 (2), 262–278.

PAPER IV

International Journal of Pharmaceutics xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

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

Note

A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging

José Manuel Amigo [a], Carsten Ravn [b,c], Neal B. Gallagher [d], Rasmus Bro [a,*]

[a] Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
[b] Faculty of Pharmaceutical Sciences, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark
[c] Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark
[d] Eigenvector Research Inc., Manson, WA 98831, USA

ARTICLE INFO

Article history:
Received 23 October 2008
Received in revised form 14 February 2009
Accepted 17 February 2009
Available online xxx

Keywords:
Hyperspectral imaging
CLS
PLS-DA
PLS-Classification
PLS-Class

ABSTRACT

In hyperspectral analysis, PLS-discriminant analysis (PLS-DA) is being increasingly used in conjunction with pure spectra where it is often referred to as PLS-Classification (PLS-Class). PLS-Class has been presented as a novel approach making it possible to obtain qualitative information about the distribution of the compounds in each pixel using little a priori knowledge about the image (only the pure spectrum of each compound is needed). In this short note it is shown that the PLS-Class model is the same as a straightforward classical least squares (CLS) model and it is highlighted that it is more appropriate to view this approach as CLS rather than PLS-DA. A real example illustrates the results of applying both PLS-Class and CLS.

1. Introduction

The powerful combination of hyperspectral images and multivariate analysis has not been overlooked by the pharmaceutical industry and the number of pharmaceutical applications has increased significantly the past few years (Gowen et al., 2007). One of the major challenges is selecting the appropriate data analysis methodology for extracting desired information from hyperspectral images. In hyperspectral images, a high-resolution spectrum is obtained for each pixel and for example, principal component analysis (PCA) (Amigo et al., 2008; Kohler et al., 2007) has been used for unsupervised pixel classification, multivariate curve resolution—alternating least squares (MCR-ALS) (Amigo and Ravn, 2009; de Juan et al., 2004) has been used for chemical mapping and interpretation of images, and partial least squares (PLS) regression has been used for quantification (Ravn et al., 2008).

PLS-discriminant analysis (PLS-DA) (Chevallier et al., 2006) is being increasingly used in conjunction with pure spectra where it is often referred to as PLS-Classification (PLS-Class) or 'PLS2 using a library of pure component spectra' (see, for example references Clarke, 2004; Furukawa et al., 2007; Henson and Zhang, 2006; Lyon et al., 2002; Ma and Anderson, 2008; Weiyong et al., 2008; Westenberger et al., 2005). This method is based on the development of a PLS2 calibration model with a calibration matrix with one pure spectrum for each chemical component and a dummy matrix as target matrix. It has been advocated as attractive because it merges the easiness of obtaining a calibration set (just the pure spectra of the components) and the exploratory and visualizing properties of the PLS2 model. In this paper, it is demonstrated that the PLS-Class approach is equivalent to performing classical least squares (CLS) (Amigo et al., 2008; Gallagher, 2007; Martens and Naes, 1984; Naes and Martens, 1984). This is shown both mathematically and by means a practical example.

2. PLS-Classification in hyperspectral imaging and its similarity with CLS

PLS-DA is a PLS2-based model, and therefore, the general model form can be written as Eq. (1):

$$\boldsymbol{X}\hat{\boldsymbol{B}} = \boldsymbol{Y} \tag{1}$$

where $\mathbf{X}$ is an $M \times J$ matrix of calibration spectra with a corresponding $M \times I$ class membership matrix $\mathbf{Y}$ and where residuals are avoided for simplicity. The matrix $\hat{\mathbf{B}}$ holds the estimated regression coefficients. The number of spectra used for calibration is M and the number of spectral channels is J. The I columns of $\mathbf{Y}$ (a dummy matrix) correspond to class memberships in classes $i = 1, \ldots, I$. Rows with a 1 in column i indicates membership in class i. The model

* Corresponding author.
E-mail address: rb@life.ku.dk (R. Bro).

0378-5173/$ – see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.ijpharm.2009.02.014

is applied to new data $\mathbf{X_{new}}$ (e.g., all the spectra in a hyperspectral image) by using the estimated $\hat{\mathbf{B}}$ (Martens and Naes, 1989) (Eq. (2)):

$$\hat{\mathbf{Y}}_{\mathbf{new}} = \mathbf{X_{new}}\hat{\mathbf{B}} \quad (2)$$

where $\hat{\mathbf{Y}}_{\mathbf{new}}$ is the estimated class membership matrix for the new data.

Because it can be difficult in hyperspectral imaging to identify classes a priori, a modification of the PLS-DA approach that uses measured pure component spectra for $\mathbf{X}$ can be employed. To avoid confusion with PLS-DA, this approach is called PLS-Class. In PLS-Class, the calibration matrix now contains the pure component spectra $\mathbf{S}$ ($I \times J$) and $\mathbf{Y}$ is replaced with a corresponding dummy matrix $\mathbf{I}$ ($I \times I$) which is an identity matrix. The PLS-DA model then becomes

$$\mathbf{S}\hat{\mathbf{B}}_{PLS\text{-}Class} = \mathbf{I} \quad (3)$$

The regression matrix $\hat{\mathbf{B}}_{PLS\text{-}Class}$ can then be estimated using a PLS algorithm. As there are typically only I samples in the data, it is not possible to calculate a PLS-Class model with more components than the number of classes. Even when there are more samples than the number of classes, the chemical rank will usually be I because additional samples are just replicates of the pure spectra. Analogously, it is not meaningful to calculate the PLS-Class model with fewer components than the number of classes as it would imply that the I analytes do not have distinct spectral features. Hence, the PLS-Class model is really not a low-rank regression model typical of PLS. Note that the use of this PLS-Class model is only one specific application of PLS-DA. Mostly, PLS-DA is used in applications where low-rank models are expected to outperform full rank models.

Eq. (3) is in the form of an inverse least squares (ILS) model and $\hat{\mathbf{B}}_{PLS\text{-}Class}$ can be estimated using any ILS regression algorithm that handles rank deficient systems of equations. This is because the least squares estimate of $\hat{\mathbf{B}}_{PLS\text{-}Class}$ is

$$\hat{\mathbf{B}}_{PLS\text{-}Class} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{I} \quad (4)$$

However, for spectroscopic applications, it is typical that $J > I$ (and often $J \gg I$). Therefore the $J \times J$ matrix $\mathbf{S}^T\mathbf{S}$ is at most rank I and is not invertible. PLS is a common method for estimating $\hat{\mathbf{B}}_{PLS\text{-}Class}$, but principal component regression (PCR) could easily be employed as well. To see this, define the singular value decomposition of the pure component spectra as

$$\mathbf{S} = \mathbf{U}\Lambda\mathbf{V}^T \quad (5)$$

where $\mathbf{U}(I \times I)$ and $\mathbf{V}(J \times J)$ are orthogonal matrices of left- and right-singular vectors respectively, and $\mathbf{\Lambda}$ ($I \times J$) is a diagonal matrix. Because $J > I$, only the first I singular values are non-zero. Keeping only the factors that correspond to non-zero singular values (and considering that $\mathbf{U}^T\mathbf{U}$ is $\mathbf{I}$) gives

$$\mathbf{S}^T\mathbf{S} = \mathbf{V}\Lambda\mathbf{U}^T\mathbf{U}\Lambda\mathbf{V}^T = \mathbf{V}\Lambda^2\mathbf{V}^T \quad (6)$$

Therefore, the pseudo-inverse is given by Eq. (7):

$$(\mathbf{S}^T\mathbf{S})^{\dagger} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T \quad (7)$$

and the least squares estimate is

$$\hat{\mathbf{B}}_{PLS\text{-}Class} = (\mathbf{S}^T\mathbf{S})^{\dagger}\mathbf{S}^T\mathbf{I} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T\mathbf{S}^T\mathbf{I} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T\mathbf{S}^T \quad (8)$$

Eq. (8) is a PCR model, but because the weights identified for an I-component PLS model span the same space as $\mathbf{V}$, $\hat{\mathbf{B}}_{PLS\text{-}Class}$ identified by PCR and PLS will be identical for this application.

Substituting Eq. (5) into the last expression of Eq. (8) gives

$$\hat{\mathbf{B}}_{PLS\text{-}Class} = \mathbf{V}\Lambda^{-2}\mathbf{V}^T\mathbf{V}\Lambda\mathbf{U}^T = \mathbf{V}\Lambda^{-1}\mathbf{U}^T = \mathbf{V}\Lambda^{-1}\mathbf{U}^T\mathbf{U}\Lambda^{-1}\mathbf{V}^T\mathbf{V}\Lambda^{-1}\mathbf{U}^T$$
$$= \mathbf{V}\Lambda^{-1}\mathbf{U}^T(\mathbf{U}\Lambda\mathbf{V}^T\mathbf{V}\Lambda\mathbf{U}^T)^{-1} = \mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (9)$$

Classical least squares regression is a well-known method for calibration. It is very useful in hyperspectral analysis because of the simple chemical interpretation it allows (Gallagher, 2007; Ravn et al., 2008). The only requirement of CLS is that the pure spectra of all the analytes must be available and that any mixture spectrum can be described as a linear combination of these spectra. Once the pure spectra are known ($\mathbf{S}$) the concentrations in a hyperspectral unfolded sample, $\mathbf{x_{new}}$, can be easily calculated by direct regression as is indicated in Eq. (10):

$$\hat{\mathbf{y}}_{\boldsymbol{new}} = \mathbf{x}_{\boldsymbol{new}}\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (10)$$

Comparing the last two equations (Eqs. (9) and (10)) it can be noticed that the last expression of Eq. (9) [$\hat{\mathbf{B}}_{PLS\text{-}Class} = \mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1}$] is the CLS estimate of $\hat{\mathbf{B}}_{PLS\text{-}Class}$. Therefore, the CLS and PLS-Class models are identical and will provide *exactly* the same results.

3. Example

3.1. Experimental

The data set used in this demonstration is described in detail in Amigo and Ravn (2009) and Ravn et al. (2008). A five-compound conventional pharmaceutical tablet formulation was used to produce the data set analysed (active pharmaceutical ingredient (API),

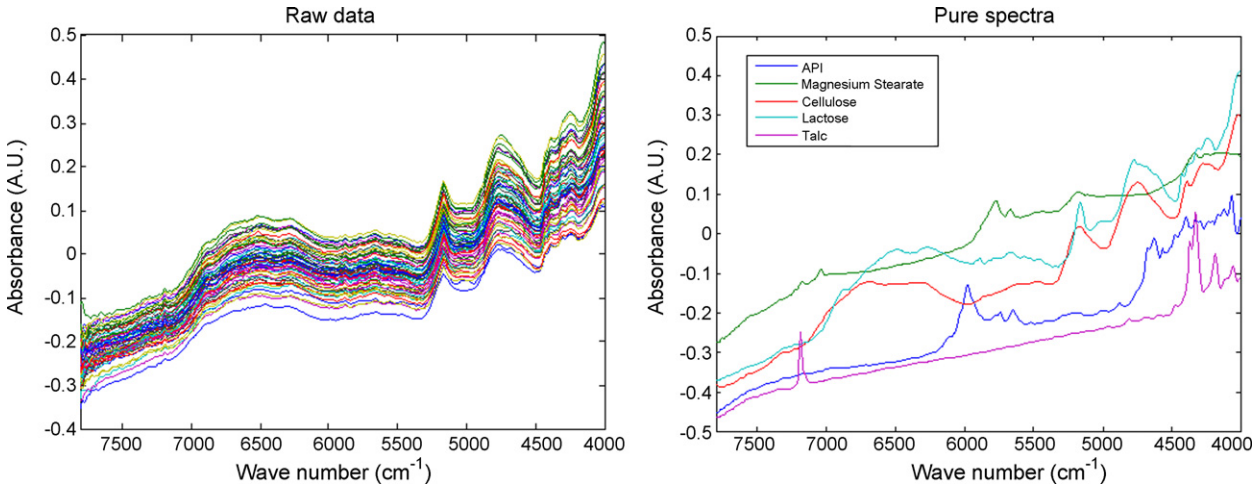


Fig. 1. Unfolded sample and pure spectra.

Please cite this article in press as: Amigo, J.M., et al., A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging. Int J Pharm (2009), doi:10.1016/j.ijpharm.2009.02.014

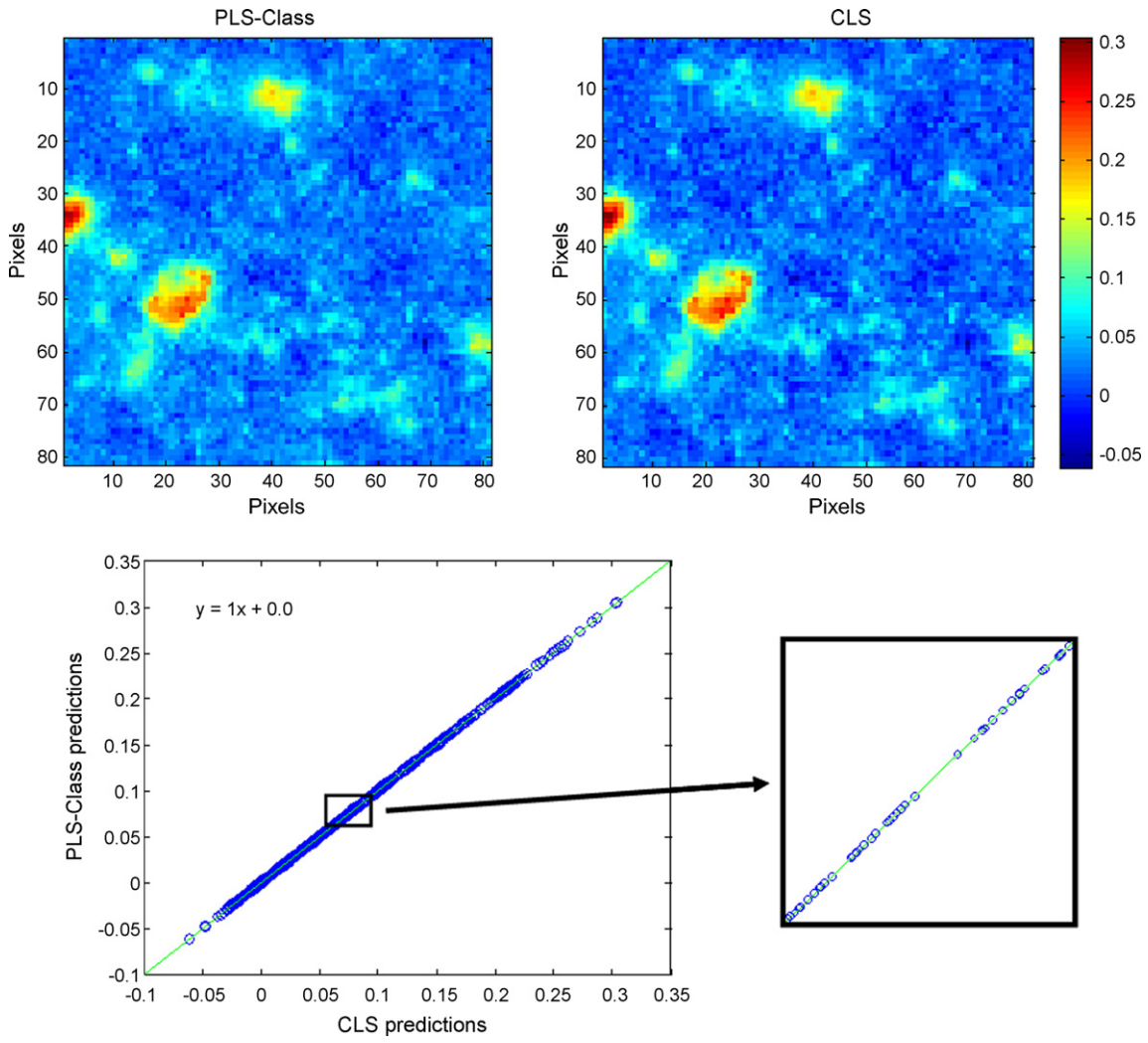


Fig. 2. Upper: concentration surface for API predicted by PLS-Class and CLS methods. Bottom: CLS predictions against PLS-Class predictions for API.

6.3%; microcrystalline cellulose, 20.0% (w/w); lactose, 71.5%; magnesium stearate, 0.75%; and talc, 1.5% (w/w)). Pure compound reference tablets of the five components were also produced. The sample selected was analysed on a NIR line mapping system (Spectrum Spotlight 350 FT-NIR Microscope, PerkinElmer, UK). An area of 2 mm × 2 mm was analysed using pixel size 25 μm × 25 μm, covering 6561 pixels in total. The spectrum of each pixel was collected from wavelength region 7800–4000 cm^{-1} using a 16 cm^{-1} spectral resolution, providing a total of 476 wavelength channels (Fig. 1).

The PLS-Class calibration model was performed by using the pure five spectra as **X**, and constructing a square diagonal dummy matrix **I** with five classes, denoting as 1 the belonging of each spectrum to each class (as in Eq. (4)). The prediction of the five components of the tablet was performed with this PLS-Class model. On the other hand, the CLS prediction was developed by direct regression of $\mathbf{X}_{\mathbf{new}}$ by using the pure spectra matrix **S**.

PLS-Class algorithm was performed by using the PLS2 algorithm implemented in PLS-Toolbox (Eigenvector Research Inc.). CLS for hyperspectral analysis algorithm from references (Amigo et al., 2008) was used. This algorithm is freely available at http://www.models.kvl.dk/users/jose_manuel_amigo/index.htm (February 2009). Both algorithms work under MatLab v. 7.5 (MatLab(R)).

3.2. Results and discussion

Only results for the prediction of the active pharmaceutical ingredient are depicted in Fig. 2. The color denotes the intensity for the concentration of API obtained for each pixel (moving from deep blue, low concentration to light red, highest concentration). As can be observed, the predictions are exactly the same for the two methods and the same relationship was obtained for the other four components. This result verifies that both methods offer the same response.

In a practical application, the PLS-Class model involves two different steps (calculation of the calibration model, $\hat{\mathbf{B}}_{PLS\text{-}Class}$, and prediction of the concentrations of the new sample). CLS, on the other hand, is typically performed by just a direct regression of the new sample onto the pure spectra.

4. Conclusion

It has been demonstrated, mathematically and with an example, that the PLS classification methodology (PLS-Class) provides *exactly* the same results as CLS. There are many advantages in avoiding the use of the PLS engine for building models as above. First of all, the PLS-Class model is not a low-rank model and therefore does not provide the usual benefits of PLS. More importantly, CLS is

Please cite this article in press as: Amigo, J.M., et al., A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging. Int J Pharm (2009), doi:10.1016/j.ijpharm.2009.02.014

a very well-defined method and it offers theoretical and practical tools, e.g., for calculating well-defined uncertainties of estimated parameters (see Appendix A and Draper and Smith, 1981; Gallagher, 2007).

While the comparison in this paper has been using hyperspectral samples, the results are generic and valid for any type of data. The use of the term PLS-Class model, though, is almost exclusively seen in hyperspectral imaging.

Acknowledgement

José Manuel Amigo wants to thank the Danish Research Council for Technology and Production Sciences for his post-doctoral fellowship.

Appendix A. Theoretical background of the calculation of uncertainty boundaries in CLS model

Having an unfolded hyperspectral sample, $\mathbf{X}(M \times J)$, and the pure spectra matrix, $\mathbf{S}$ $(J \times I)$, the CLS model to calculate the individual concentration of each component in each pixel can be expressed as follows:

$$\hat{\mathbf{y}}^T = \mathbf{x}^T\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (A.1)$$

The errors in the individual measurements of $\mathbf{x}$ are assumed to dominate the estimation error. The differential of Eq. (A.1) is

$$(d\hat{\mathbf{y}}^T) = (d\mathbf{x}^T)\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (A.2)$$

Therefore, the error covariance for the CLS model estimate of the concentrations is given by (Draper and Smith, 1981):

$$E(d\hat{\mathbf{y}}\ d\hat{\mathbf{y}}^T) = (\mathbf{S}\mathbf{S}^T)^{-1}\mathbf{S}^T\,E(d\mathbf{x}\ d\mathbf{x}^T)\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (A.3)$$

where $E()$ is the expectation operator. For CLS, it is assumed that the noise on each channel is of similar magnitude and uncorrelated, therefore:

$$E(d\mathbf{x}\ d\mathbf{x}^T) = \sigma^2\mathbf{I} \quad (A.4)$$

Substituting Eq. (A.4) into the variance of the estimation gives

$$E(d\hat{\mathbf{y}}\,d\hat{\mathbf{y}}^T) = \sigma^2(\mathbf{S}\mathbf{S}^T)^{-1}\mathbf{S}\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1} \quad (A.5)$$

$$E(d\hat{\mathbf{y}}\ d\hat{\mathbf{y}}^T) = \sigma^2(\mathbf{S}\mathbf{S}^T)^{-1} \quad (A.6)$$

Eq. (A.6) accounts for the final estimation of the uncertainty in CLS models. Further information can be found in the supplied references (Draper and Smith, 1981; Gallagher, 2007).

References

Amigo, J.M., Bautista, M., Cruz, J., Coello, J., Maspoch, S., Blanco, M., 2008. Study of pharmaceutical samples by NIR chemical image and multivariate analysis. Trends Anal. Chem. 27, 696–713.
Amigo, J.M., Ravn, C., Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging. Eur. J. Pharm. Sci. (2009), doi:10.1016/j.ejps.2009.01.001.
Chevallier, S., Bertrand, D., Kohler, A., Courcoux, P., 2006. Application of PLS-DA in multivariate image analysis. J. Chemometr. 20, 221–229.
Clarke, F., 2004. Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. Vib. Spectrosc. 34, 25–35.
de Juan, A., Tauler, R., Dyson, R., Marcolli, C., Rault, M., Maeder, M., 2004. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. Trac-Trends Anal. Chem. 23, 70–79.
Draper, N., Smith, H., 1981. Applied Regression Analysis, 2nd edition. John Wiley & Sons, Ltd., New York.
Furukawa, T., Sato, H., Shinzawa, H., Noda, I., Ochiai, S., 2007. Evaluation of homogeneity of binary blends of poly(3-hydroxybutyrate) and poly(L-lactic acid) studied by near infrared chemical imaging (NIRCI). Anal. Sci. 23, 871–876.
Gallagher, N.B., 2007. Detection, classification and quantification in hyperspectral images using classical least squares models. In: Grahn, H., Geladi, P. (Eds.), Techniques and Applications of Hyperspectral Image Analysis. John Wiley & Sons, West Sussex, England, pp. 181–201.
Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M., 2007. Hyperspectral imaging—an emerging process analytical tool for food quality and safety control. Trends Food Sci. Technol. 18, 590–598.
Henson, M.J., Zhang, L., 2006. Drug characterization in low dosage pharmaceutical tablets using Raman microscopic mapping. Appl. Spectrosc. 60, 1247–1255, http://www.models.kvl.dk/users/jose_manuel_amigo/index.htm.
Kohler, A., Bertrand, D., Martens, H., Hannesson, K., Kirschner, C., Ofstad, R., 2007. Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information. Anal. Bioanal. Chem. 389, 1143–1153.
Lyon, R.C., Lester, D.S., Lewis, E.N., Lee, E., Yu, L.X., Jefferson, E.H., Hussain, A.S., 2002. Near-infrared spectral imaging for quality assurance of pharmaceutical products: analysis of tablets to assess powder blend homogeneity. AAPS Pharm. Sci. Tech. 3, 1–17.
Ma, H., Anderson, C.A., 2008. Characterization of pharmaceutical powder blends by NIR chemical imaging. J. Pharm. Sci. 97, 369–373.
Martens, H., Naes, T., 1984. Multivariate calibration. 1. Concepts and distinctions. Trac-Trends Anal. Chem. 3, 204–210.
Martens, H., Naes, T., 1989. Multivariate Calibration. John Wiley & Sons, Chichester.
MatLab(R),2006. Version 7.0. The Matworks, Inc., Massachusetts.
Naes, T., Martens, H., 1984. Multivariate calibration. 2. Chemometric methods. Trac-Trends Anal. Chem. 3, 266–271.
PLS-Toolbox 2004. Version 3.5. Eigenverctor Research, WA, USA.
Ravn, C., Bro, R., Skibsted, E., 2008. Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—comparing common calibration approaches. J. Pharm. Biomed. Anal. 48, 554–561.
Weiyong, L., Woldu, A., Kelly, R., 2008. Measurement of drug agglomerates in powder blending simulation samples by near infrared chemical imaging. Int. J. Pharm. 350, 369–373.
Westenberger, B.J., Ellison, C.D., Fussner, A.S., Jenney, S., Kolinski, T.G., Lipe, T.G., Lyon, R.C., Moore, T.W., Revelle, L.K., Smith, A.P., Spencer, J.A., Story, K.D., Toler, D.Y., Wokovich, A.M., Buhse, L.F., 2005. Quality assessment of internet pharmaceutical products using traditional and non-traditional analytical techniques. Int. J. Pharm. 306, 56–70.

Please cite this article in press as: Amigo, J.M., et al., A comparison of a common approach to partial least squares-discriminant analysis and classical least squares in hyperspectral imaging. Int J Pharm (2009), doi:10.1016/j.ijpharm.2009.02.014