



Emma Klotz, BSc

# **Quantification of 2D Printed Medicine with Near Infrared Spectroscopy**

## **MASTER'S THESIS**

to achieve the university degree of

Diplom-Ingenieurin

Master's degree programme: Biomedical Engineering

submitted to

**Graz University of Technology**

Supervisor

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Hermann Scharfetter

Institute of Medical Engineering

Dr.-techn. Otto Scheibelhofer, Research Center Pharmaceutical Engineering GmbH  
DI Miriam Wimmer-Teubenbacher, Research Center Pharmaceutical Engineering GmbH

Graz, March 2018

## **AFFIDAVIT**

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

---

Date

---

Signature

# Acknowledgement

I would like to thank the RCPE GmbH with for giving me the opportunity to work on this project and to conduct my thesis work at their facilities. I want to especially thank Miriam and Otto, who have supervised and guided me along the way, even after leaving the company. They have always taken the time and made me feel welcome to ask questions and help me along in the process. Furthermore, I would like to thank Kai and Sandra for their input regarding the conclusion of the thesis. I would also like to thank Assoc. Prof. Hermann Scharfetter for I want thank the new friends I've made at the RCPE, who have included and welcomed me from the beginning and made work not feel like soley work, and my friends who have accompanied me during the years of university. They have made Graz feel like home for the past years, it has truly been a pleasure.

However, most of all I have to thank my family, my mother Marta, my father Erwin and my sister Sophie. They have supported me my whole life, providing me with everything I've ever needed, and more.

Emma Klotz, Graz 18/01/2018

# Abstract

## Quantification of 2D Printed Medicine with Near Infrared Spectroscopy

In this work three “inks” consisting of sodium picosulfate solved in water, carvedilol solved in ethanol and metformin hydrochloride solved in water were formulated and printed on a number of substrates namely: EVA, HPMC, HGC, icing sheet, potato starch wafer and HGC with 2% titanium dioxide using the pico system of Scienion’s sciFLEXARRAY printer.

After a drying period of one day the samples were measured using EVK hyperspectral imaging system, which measures wavelengths ranging from 732 - 1700 nm.

The samples were then displayed in MATLAB<sup>®</sup>. After comparing all samples the combination of metformin hydrochloride printed on gelatin with 2% titanium dioxide was chosen to investigate the possibility of implementing NIR spectroscopy as PAT technology for printed medicine.

A calibration model was created with a number of spots with increasing drops per spot using projection to latent structures as implemented in MATLAB<sup>®</sup> using the SIMPLS algorithm.

In order to create an appropriate response matrix Raman measurements were conducted, by revealing the distribution of API within the sample drops.

To determine the viability of the model HPLC measurements were conducted on both the calibration and the test samples.

Key Words: Printed Medicine, NIR Spectroscopy, PLSR



# Abstract

## Quantifizierung von 2D gedruckter Medizin mittels Nah-Infrarot Spektroskopie

In dieser Arbeit wurden drei "Tinten" mit dem pico-system des sciFLEXXARRAY Druckers von Scienion auf verschiedene Substrate gedruckt. Die Tinten bestanden aus den aktiven pharmazeutischen Ingredienzien Natriumpicosulfat und Metformin Hydrochlorid gelöst in Wasser, und Carvedilol gelöst in Ethanol. Als Substrate wurden: EVA, HPMC, HGC, Glasurblätter, Waffeln aus Kartoffelstärke und HGC mit eingebettetem Titandioxid verwendet.

Nach einer eintägigen Trocknungszeit werden die Proben mit einem hyperspektralen Bildanalysesystem von EVK im Wellenlängenbereich von 931 - 1700 nm vermessen. Nach einem Vergleich aller Kombinationen von Tinten und Substraten wird Metformin Hydrochlorid auf HGC mit eingebettetem Titandioxid ausgewählt, um zu untersuchen, ob es möglich ist NIR Spektroskopie als PAT Technologie zu verwenden.

Ein Kalibrationsmodell wird durch Projektion auf latente Strukturen in Matlab erstellt.

Um die Qualität des Modells zu bewerten wird die vorhergesagte Menge an Metformin Hydrochlorid mit dem Ergebnis einer HPLC Messung verglichen.

Schlüsselwörter: gedruckte Medizin, NIR Spectroskopie, PLSR

# Contents

<b>Acknowledgement</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Abbreviations</b>	<b>viii</b>
<b>1. Introduction</b>	<b>1</b>
1.1. Printed Medicine . . . . .	1
1.1.1. Quality Control . . . . .	3
1.2. NIR Spectroscopy . . . . .	4
1.2.1. Hyperspectral Imaging . . . . .	7
1.3. Multivariate Data Analysis . . . . .	7
1.3.1. NIPALS . . . . .	8
1.3.2. SIMPLS . . . . .	10
1.3.3. Challenges . . . . .	11
1.4. Aim . . . . .	11
<b>2. Methods</b>	<b>12</b>
2.1. Printing . . . . .	12
2.1.1. Printer . . . . .	12
2.1.2. Ink Formulations . . . . .	13
2.1.3. Substrates . . . . .	14
2.2. Near Infrared Spectrometer Setup . . . . .	15
2.3. Measurements . . . . .	16
2.4. Resolution . . . . .	18
2.4.1. Spatial Resolution (Pixelvalue y) . . . . .	18
2.4.2. Temporal Resolution (Pixelvalue x) . . . . .	18
2.5. Data Preprocessing . . . . .	20
2.5.1. Defect Pixel Correction . . . . .	20
2.5.2. Standard Normal Variate . . . . .	21
2.5.3. Savitzky-Golay Filtering . . . . .	21
2.6. Quantification . . . . .	22
2.6.1. NIPALS Algorithm . . . . .	23
2.6.2. SIMPLS Algorithm . . . . .	24
2.7. Reference Methods . . . . .	26
2.7.1. Raman Spectroscopy . . . . .	26
2.7.2. High-Performance Liquid Chromatography . . . . .	27

## Contents

<b>3. Experiments</b>	<b>28</b>
3.1. Basic Modelling . . . . .	30
3.2. Refinement . . . . .	31
3.3. Testing . . . . .	33
3.4. Final Process . . . . .	34
3.4.1. NIR measurement . . . . .	35
3.4.2. Data processing . . . . .	35
3.4.3. Backprojection and Prediction . . . . .	37
<b>4. Results</b>	<b>39</b>
<b>5. Discussion</b>	<b>51</b>
5.1. Conclusion . . . . .	54
<b>A. Fits for HPLC Results</b>	<b>56</b>
<b>B. Tables for created models</b>	<b>57</b>
<b>C. Conversion Wavelength Variable and nm</b>	<b>65</b>
<b>List of Figures</b>	<b>67</b>
<b>List of Tables</b>	<b>70</b>
<b>Bibliography</b>	<b>72</b>

## Abbreviations

<b>API</b>	active pharmaceutical ingredient
<b>ODFs</b>	orodispersible films
<b>DoD</b>	drop on demand
<b>PAT</b>	process analytical technology
<b>HPLC</b>	high-performance liquid chromatography
<b>NIRS</b>	near-infrared spectroscopy
<b>PCA</b>	principal component analysis
<b>PLS</b>	partial least squares
<b>NIPALS</b>	nonlinear iterative partial least squares
<b>FOV</b>	field of view
<b>listerine</b>	Listerine Pocketpaks <sup>®</sup> , Oral Care Strips
<b>EVA</b>	Ethylene vinyl acetate
<b>HGC</b>	Gelatin film strips
<b>HGC+TiO<sub>2</sub></b>	Gelatin film strips with 2% titanium dioxide
<b>HPMC</b>	Hydroxypropylmethylcellulose
<b>NIR</b>	near infrared
<b>HSI</b>	hyperspectral imaging
<b>USAF</b>	US Air Force
<b>FWHM</b>	full width half maximum
<b>SNV</b>	standard normal variate
<b>PLS</b>	projections to latent structures
<b>ROI</b>	region of interest

# 1. Introduction

## 1.1. Printed Medicine

Under the term printed medicine the technique of dispensing an active pharmaceutical ingredient (API) on orodispersible films (ODFs) is described.

Traditional SDFs, such as tablets, can be problematic for patients with swallowing disorders and dysphagia, just like neonatal and pediatric populations. To improve drug delivery for these patient groups orodispersible film have gained popularity in the pharmaceutical industry. While the soluble films forgo the problem of swallowing, it further paves the way for personalized medicine. Through printing the dosage can be tailored to each individual patient's need based on age, size and specific characteristics, this would eliminate the need for cutting tablets into pieces [1–4].

Moreover, since these ODFs are thin and dissolve quickly the need for water is eliminated, which is an advantage for i.e. travellers or patients in regions where clean drinking water is not readily available.

While ODFs were introduced quite early on in the 1960s, the first to gain popularity were the Listerine® PocketPaks, thin breath-freshening films, introduced by Pfizer in 2001. The EU first approved prescription drug delivered by ODF in 2010 [5,6].

Additionally, to help patients with swallowing difficulties the rapid absorption is useful for problems with the need for fast relief like motion sickness or allergic reactions, since pre-gastric absorption can enhance bioavailability and unwanted effects. [6,7]

Ideally ODFs should be thin, flexible, but still show enough tensile strength to withstand stress from mouth activity. Furthermore, swelling should not be excessive to avoid discomfort. Moreover, the substrate has to keep the API stable. [8,9]

Various polymers can be used in the film's preparation, these polymers provide the film's strength. Popular polymers and substrates are: edible icing sheets, hydroxypropyl cellulose (HPC), hydroxyl propyl methyl cellulose (HPMC), starches and modified starches, pullulan, polyethylene oxide, pectin, gelatin, carboxy methyl cellulose, polyvinylpyrrolidone (PVP) and crosslinked PVP, alginates, poly vinyl alcohol, maltodextrin and polyox. [2,10]

In this thesis HPMC, edible icing sheets, gelatin, listerine, ethylene vinyl acetate and starch are used.

While the formulation of orodispersible films is well established within the pharmaceutical industry, the printing of medicine is still facing technical difficulties. As mentioned before, this technique shows great potential, due to its variability in dosage. Tablets are mass produced with predefined dosages, and it is not possible to

## 1. Introduction

cut tablets and keep the variation in dosage at an acceptable level. Two popular methods of printing drop on demand (DoD), as is needed for the required accuracy of printing pharmaceuticals, are thermal ink-jet printing and piezoelectric printing. [11] Thermal ink-jet printing uses heat pulses to dispense fluid. A heat element consisting of a resistor heats up through electrical pulses, with this heat a vapor bubble forms, as this bubble expands the volume within the reservoir creating a droplet, and forcing it out of the nozzle. When this air bubble collapses a vacuum is created, causing more ink to flow from the ink reservoir into the nozzle.

Piezoelectric printing is based on the piezoelectric effect, which describes the change polarization due to mechanical stress. The piezoelectric crystal is positioned at the ink reservoir's back, through an applied voltage, the element deforms creating pressure waves, which causes ink to be released through the nozzle. After the element has returned to its starting state, the nozzle refills with ink. Schematics for both these processes are shown in figure 1.1. [11–13]

In 2012 Takala et al. wanted to compare thermal and piezo-electric ink jet printing

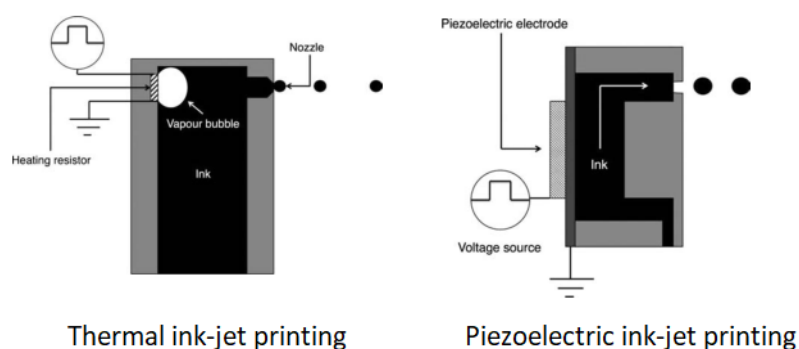


Fig. 1.1.: Schematic for a thermal and a piezoelectric printhead, as shown by Gerard Cummins. [13]

regarding their suitability to print pharmaceuticals, using commercial printers. In their study the piezoelectric printhead was unable to print any substance. [12] Other studies such as those of Lee et al. or Scoutaris et al. used piezoelectrically driven printers both showing promising results. For this work a piezoelectric ink-jet system is used. [14, 15]

Challenges that have yet to be fully overcome are posed by the required viscosity and surface tension from the ink solution. While viscosity needs to be low enough for the fluid to be jetted, it has to be viscous enough not to run through the nozzle at unwanted time points (leading to errors in the drop on demand system, and subsequently in dosage). The ink also requires enough surface tension to form spherical drops and resist leakage through the print nozzle. One solution to decrease problems of high viscosity fluids is to use heated piezoelectric ink-jet systems, this in turn can lead to degradation of APIs due to their temperature sensitivity. Furthermore, high-dose dosage forms might be limited due to the substrate's capacity and certain API's limited solubility. In addition the drop volume is not constant, and can not be determined for each drop individually for time consumption reasons, this leads to a variation in dosage and requires a tool for checking the dispensed dosage. One instance, where printed medicines on ODFs are not an option are drugs which

## 1. Introduction

irritate the oral mucosa. [1, 11, 16, 17]

The advantages and possibilities of printed medicine, however, outweigh the disadvantages and call for action to improve the technique. These advantages, as mentioned above include (as described by [9, 11, 12, 16–19]):

- easy form factor for low dosage medicines (no need for padding material to increase the size of a tablet)
- possible complex drug release
- multiple pharmaceutical agents possible in one form factor
- high reproducibility given an appropriate quality control system or improved certainty regarding the DoD systems
- waste minimization
- ease of administration for pediatric, geriatric and patients suffering from dysphagia
- no need for water
- reduced unwanted effects through pre-gastric absorption
- higher bioavailability due to pre-gastric absorption
- rapid relief for problems, in need of sudden relief like allergic attack or coughing
- cost effectiveness
- improved production efficiency by allowing real-time release testing (with appropriate feedback loops)
- reduce mass production
- ability to produce personalized medicine, attuned for each individual's needs

### 1.1.1. Quality Control

Although printing of medicine is a promising technology there is a need for an appropriate process analytical technology (PAT). The FDA defines PAT as a system that analyses and controls manufacturing processes in a timely fashion for critical quality parameters, to ensure acceptable product quality. [20]

This means that a PAT system overcomes the traditional separation between analytical chemistry and the manufacturing process. It focuses on building quality into the product (QbD) through at-line, in-line or on-line measurements of process quality using chemometric approaches, i.e. multivariate statistical analysis. Figure 1.2 shows a schematic displaying the differences between at-, in- and on-line processes. This means technologies need to be quick and non-destructive to the sample, resulting in the fact, that traditional techniques such as high-performance liquid chromatography (HPLC) and mass spectrometry are not suitable. Moreover, the technologies should be automated giving the opportunity for uninterrupted operation. Technologies in use for PAT are NIR spectroscopy for moisture determination, X-ray spectroscopy, microwaves for moisture determination and photoacoustic spectroscopy. [20]

To extract knowledge about the process and quality from the chemical information delivered by these technologies mathematical models are used. The field of projecting

## 1. Introduction

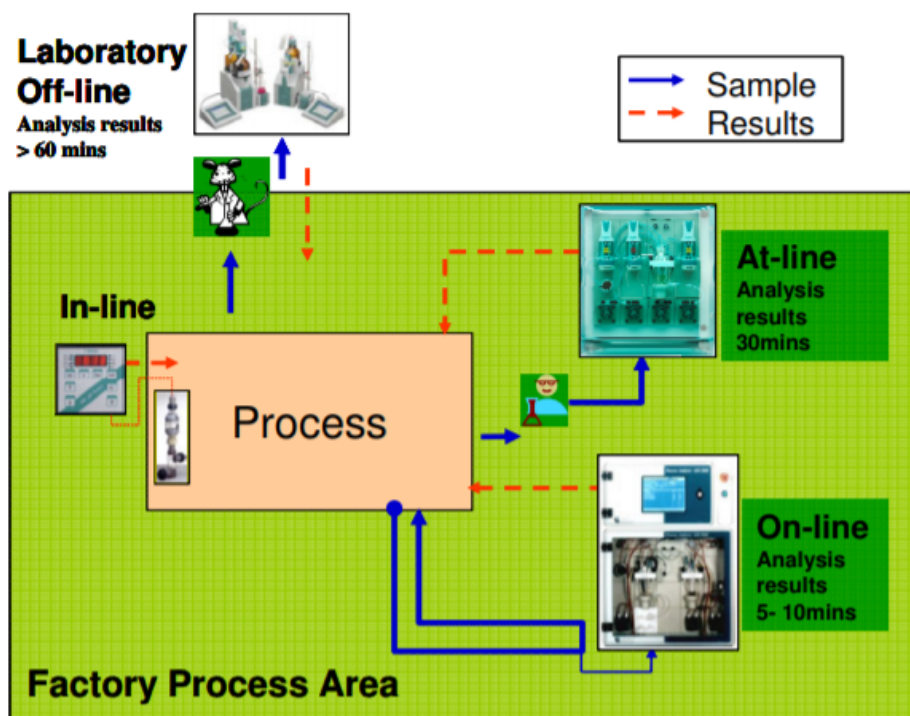


Fig. 1.2.: Schematic displaying differences between in- on- and at-line processes, as described by [21].

chemical information to a few dimensions using multivariate analysis is called *chemometrics*. [20,22]

## 1.2. NIR Spectroscopy

NIR chemical imaging (NIR-CI) (applying near-infrared spectroscopy (NIRS) to chemical information) has been used for rapid characterization of raw materials, with the application of models it should be possible to quantify the information. NIRS is the most common PAT analyzer in manufacturing of solid dosage forms. Hyperspectral image (HSI) acquisition is used more and more to be able to get a two dimensional spatial positions in addition to the spectral information. [23–25]

The NIR spectrum is a small region within the electromagnetic spectrum in the region between 780 - 2500 nm. NIR spectra exist due to molecular vibration induced by infrared radiation. When light with the energy equaling one of a molecule's vibrational energy level differences interacts with it, energy is absorbed and the molecule enters a higher energy state. Light's energy is related to it's wavelength with equation (1.1), where  $h$  is Planck's constant,  $c$  is the speed of light,  $f$  is frequency, and  $\lambda$  the wavelength.

$$E = hf = \frac{hc}{\lambda} \quad (1.1)$$



## 1. Introduction

As Blanco et al. describes, vibrational frequency for a diatomic molecule can be simulated using a harmonic oscillator model, then the variation of the potential energy with bond distance shows a parabolic curve centred around the equilibrium state with evenly spaced vibrational energy levels (as shown in figure 1.3). At higher

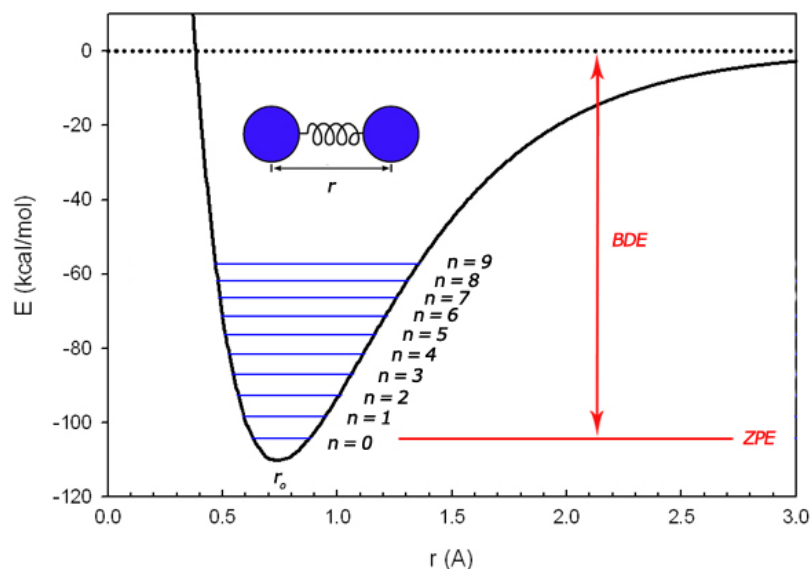


Fig. 1.3.: Morse curve: showing the relationship between energy and the internuclear distance  $r$  in a diatomic molecule as shown in [26]. ZPE is the zero point energy, it is the lowest possible energy level a molecule can have. BDE is the bond dissociation energy.

energy levels (shown by the blue lines) the distance variation ( $r$ ) between the atoms increases, when enough energy is transferred to the molecule the atoms disassociate, as soon as the bond dissociation energy is reached. The light's energy required to elevate a molecule from it's ground state ( $n = 0$ ) needs to match the difference in the energy state exactly. The energy  $E_n$  between the discrete energy levels is described by equation (1.2). Here  $n$  is the vibrational quantum number,  $k$  the bonding force constant (a measure of rigidity of a chemical bond in its normal equilibrium position) and  $m$  the effective mass ( $m = \frac{m_1 m_2}{m_1 + m_2}$ ). [27]

$$E_n = \left( n + \frac{1}{2} \right) \hbar \sqrt{\frac{k}{m}} \quad (1.2)$$

In reality vibrations are not harmonic and energy levels not completely evenly spaced, this is why in figure 1.3 the parabola is slanted to the right.

To improve the model higher order terms and a non harmonicity constant need to be involved. Through this inclusion it is shown, that frequencies that are multiples of the fundamental frequency exist, these are called overtones. Absorption bands within the NIR region are related to overtones and combinations of fundamental vibrations of function groups of -CH-, -NH-, -OH-. [27, 28] Figure 1.4 shows the absorption bands as a function of the wavelength. In this region stretching and bending are the two vibrational modes, meaning the distances and bond angles changes

## 1. Introduction

continuously. [29] Due to the fact, that the vibrations within this region are over-

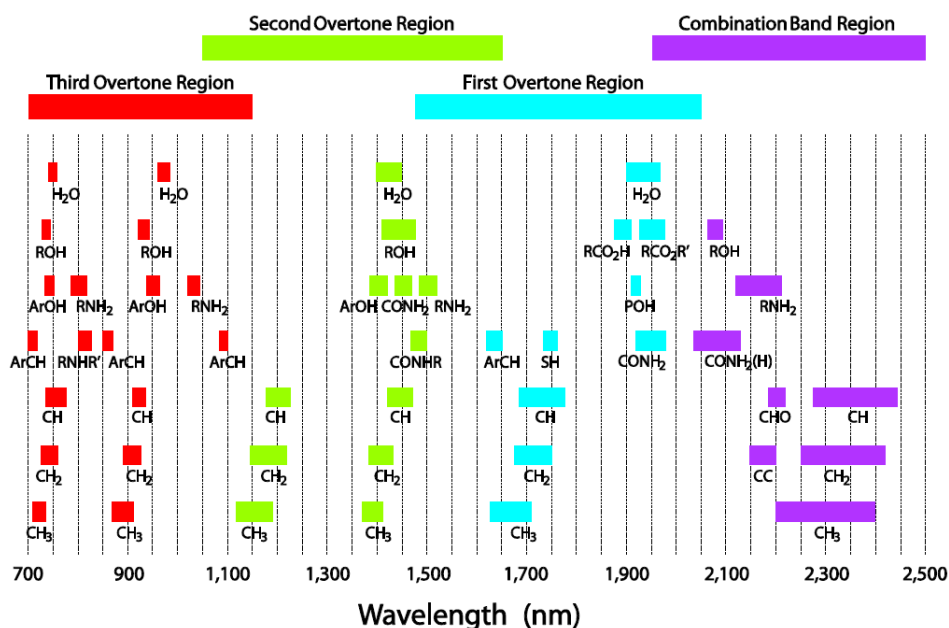


Fig. 1.4.: NIR functional groups absorption bands as a function of the wavelength.

tones, absorption coefficients are rather low, since transition probability decreases significantly with increased vibrational quantum number. This leads to restricted sensitivity but enables reflectance measurements by increasing penetration depth. NIRS can be implemented as a reflectance or transmission mode. In the reflectance mode reflected light is detected after it has interacted with the sample, in transmission mode transmitted light is detected, therefore the detected light is directly proportional to incident - absorbed light.

To extract absorbance and concentration information from reflected light Beer's law is used [27]

$$A = \log \frac{1}{R} = a'c \quad (1.3)$$

Summarized the advantages of NIRS are listed below [30]:

- minimal sample preparation
- measurement within seconds
- cost effective
- non destructive
- measurement of solid samples in reflectance mode
- easy implementation of HSI

A major drawback of this technique is the fact that low energy levels result in spectra with broad peaks, increasing the chance of overlapping absorption bands and therefore more complex attainability of information. For this reason it is standard practice to create a calibration model based on measurements with known information. To link the two chemometrics is used, to extract information of new samples the model, based on training data, is applied to the sample. [30]

### 1.2.1. Hyperspectral Imaging

Hyperspectral imaging, also known as spectroscopic chemical imaging, combines conventional imaging (defined as creation of a representative reproduction of an object) and spectroscopy to attain spectral and spatial information from an object. [23, 25, 28, 31]

A conventional image consists of a two-dimensional data matrix that can be described as a function  $f(x, y)$  where  $x$  and  $y$  are orthogonal coordinates and  $I$  is the intensity at each datapoint called pixel. Spectroscopic results depend on the wavelength. HSI is the combination of these two techniques: a hyperspectral image is described by two orthogonal spatial directions but  $I$  is described not by one intensity value but an entire (NIR-)spectrum. [32]

Therefore the hyperspectral image is not a two-dimensional data matrix but is described by three dimension, this 3D data matrix is called a *hypercube*. [25]

Gowen et al. describes multiple methods of measuring a hypercube. One, called the "staring imager" where the wavelength is modified using a tuneable filter in dependence of two spatial dimensions. The other method is called "push broom scanning", here the multiple wavelengths are measured simultaneously in dependence of one spatial dimension. [25] Push broom setup usually involve an additional element, such as a moving linear stage, to measure the second spatial dimension. The system measures one line of sample after another while the sample is moved at a constant rate. The line can then be subdivided into pixels. [31]

HSI allows the determination of chemical information for a specific spatial plane, not an average composition as for non-imaging NIR spectroscopy. [28] Based on the principle of HSI, especially push broom technologies, it becomes apparent, that it would be an appropriate technology to implement as a PAT system.

Hyperspectral data is multidimensional (two spatial and one spectral dimension with multiple variables in each dimension), for analysis dimensionality needs to be reduced while retaining relevant information. This is where chemometrics and multivariate analysis comes into play.

Before applying multivariate analysis data is preprocessed to remove non chemical biases such as scattering effects or baseline drifts.

## 1.3. Multivariate Data Analysis

As mentioned above multivariate data analysis is needed to reduce data's dimensionality to relevant information. An important tool for this reduction are projection methods such as principal component analysis (PCA) and partial least squares (PLS). These methods take cases of a  $K$ -dimensional space and projects them onto a lower-dimensional plane, called hyper-plane. This is practicable for summarizing and visualizing a data set, classification and finding quantitative relationships

## 1. Introduction

among variables. Multivariate data analysis and projection of data onto a hyper-plane are established methods for PAT and quality control in the food processing industry. [22, 33, 34]

PCA is useful to get an overview over data, it helps to recognize groups within a larger data set and can further be useful to classify new observations, since they exhibit similar patterns.

For a quantification regression modeling is used, two data blocks, usually denoted  $X$  and  $Y$ , are compared aiming for the ability to predict  $Y$  from a new observation data set  $X_2$ . This model is achieved through the PLS method. [33] For this work  $X$  is the measured hypercube and  $Y$  the concentration, or mass, of API as a function of the two spatial dimensions.

The PLS approach was formulated in the 1970s by Herman Wold, and shortly after slightly modified by Svante Wold and Harald Martens who also started to interpret PLS as Projection to Latent Structures (as opposed to Partial Least Squares) for a more descriptive meaning. [33, 35]

Assuming  $X$  to consist of  $K$  variables and  $N$  observations, the predictors, and as it is in this present case the response matrix  $Y$  to be one dimensional  $X$  and  $Y$  have the dimensions  $(N * K)$  and  $(N * 1)$ , respectively.  $N$  is determined by the number of pixels measured by the HSI system,  $K$  is determined by the spectrum. The number of variables is the number of measured spectral wavelength (I of subsection 1.2.1).  $Y$  must always have the same number of observation points but can have another number of responses, when for example more than one chemical information has to be determined, therefore  $Y$  could have the dimension of  $(N * M)$ .

There are multiple algorithms to calculate these hyper-planes efficiently, one is the nonlinear iterative partial least squares (NIPALS) another the SIMPLS algorithm.

### 1.3.1. NIPALS

The NIPALS algorithm is an established algorithm to estimate the parameters iteratively. The steps are described subsection 2.6.1 and in several publications. [35–38] The data exists in two separate coordinate system, in  $X$  and  $Y$ , respectively, for both the number of coordinates equals the number of variables (columns). The algorithm's task is to find a coordinate system that describes the relationship between the observations in the two existing systems. This is done by finding a number of components. Figure 1.5 taken from [33] describes the relationship for an exemplary data set where  $X$  has the dimension  $(21 * 3)$  and  $Y$   $(21 * 1)$ . The first component is a line in the  $X$ -space approximates the point swarm and correlates well with the  $Y$ -vector (of matrix if  $Y$  is multidimensional). When projecting the sample onto the line, the distance between the projected point and the line is termed as the score  $t_1$ . The scores  $T$  can be thought of as new variable that describes the information in the original  $X$ -variables and can be used to estimate the residuals  $F$  by applying the score and a weight matrix  $C$  (dimension  $(N * 1)$ ).  $F$  describes the variation that is left unexplained by the components calculated to this point. An ideal model would explain the entire variation in  $Y$ , so  $F$  would be zero.

## 1. Introduction

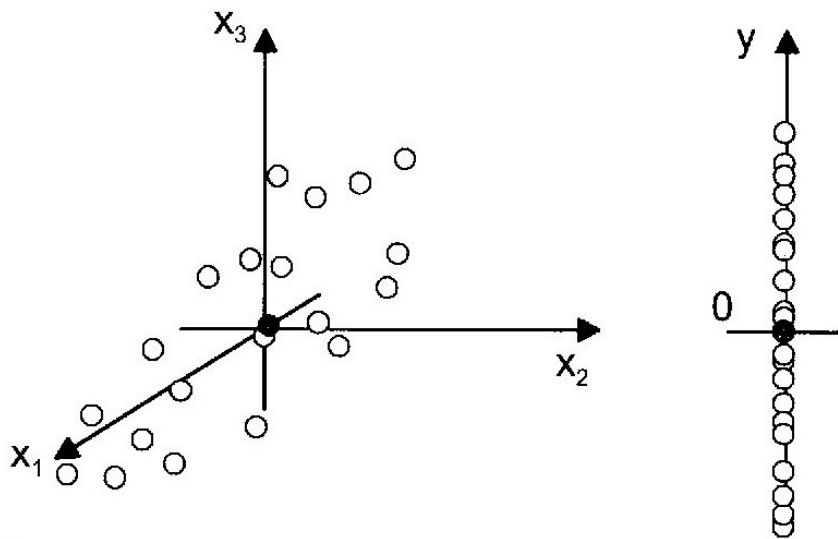


Fig. 1.5.: For the regression observations can be understood as two data swarms in two separate spaces, through the regression these are related to each other. In this example the data has been mean-centered, meaning the coordinate systems pass through the average point (dark gray) of each point swarm.

To add the second component, a second line, orthogonal to the first is added to describe the  $X$ -data better and still provides a good correlation with the residuals  $F$  of the first component. The projection to calculation of scores is repeated to create  $t_2$ , which with a second column of the weight matrix  $C$  (dimension  $(N * 2)$ ), can be applied to the  $Y$ -data to create the new residual  $F$  matrix.

This process is repeated until most of the variation is explained, therefore  $F$  converges.

Overall it can be summarized that scores matrices called  $T$  and  $U$  are calculated, these store information related to the observations. The weights matrix  $C$  and the  $X$ -loading  $P$  store information related to variables for the  $Y$ - and  $X$  data, respectively. An indicator for how well these variables model  $X$  and  $Y$  residuals, denoted  $E$  and  $F$  can be calculated. Equations (1.4) and (1.7) describe the relationships between these matrices. [33]

$$Y = TC^T + F \quad (1.4)$$

$$\text{with} \quad (1.5)$$

$$T = XW^* \quad (1.6)$$

$$X = TP^T + E \quad (1.7)$$

Figure 1.6 shows the relationship and dimensions of variables calculated during the PLS algorithm as shown in [33] on page 80.

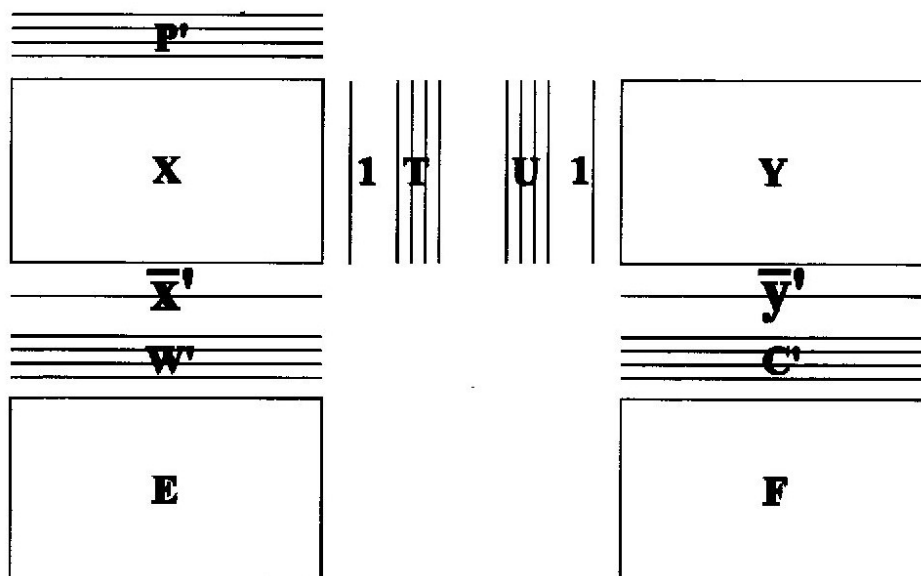


Fig. 1.6.: Matrix relationships among the PLS computation.  $1 * \bar{x}$  and  $1 * \bar{y}$  are the averages along each column, that stem from the mean centering step.  $T$  and  $U$  are the  $X$ - and  $Y$ -scores (containing information about the observations and their similarities with respect to the given problem). The weights  $W^*$  and  $C$  contain information about how the variables combine to relay the relation between  $X$  and  $Y$ .  $E$  and  $F$  contain the variation in the data not contained in the model, called residuals.

### 1.3.2. SIMPLS

De Jong et al. proposed an alternative PLS algorithm called the SIMPLS algorithm, that is implemented in the MATLAB<sup>®</sup> function. Their objective being finding a predictive linear model  $\hat{Y} = \mathbf{X}\mathbf{B}$  (meaning  $F$  in (1.4) equals zero). The algorithm is changed and shortened to avoid calculating weight  $W^*$ . The algorithm is described in more detail in subsection 2.6.2.

De Jong points out several advantages of the SIMPLS algorithm mainly the direct calculation of weights (instead of using deflated  $X$  matrices). He shows that SIMPLS and NIPALS are equivalent for  $M = 1$  and only slightly different for multivariate  $Y$ . [39]

Due to the fact that in this work only a one dimensional  $Y$  matrix is used it is fitting, that MATLAB<sup>®</sup> uses the SIMPLS and does not need to be adapted.

The calculated regression coefficients give a weight to each variable (i.e. wavelength), these coefficients can be applied to new observations. Through this application (multiplication) the spectrum for each observation is related to the response value (e.g. the concentration).

### 1.3.3. Challenges

A first challenge to create a viable calibration model is the determination of the response matrix  $Y$ , to minimize the residuals it gives. Naturally, the accuracy of  $Y$  is dependent on a suitable reference method. It has been proposed that an inverse calibration might be useful. This means the concentration  $Y$  used for the calibration is a regressed value. [34]

Another challenge is the determination of the right number of components used for the calibration model. The risk here is to overfit or underfit. With overfitting the model finds correlations where none are present. For example noise could be correlated by chance, this will lead to a suboptimal result when the model is applied to an independent data set. Underfitting the model is another risk, where not all aspects of the data are explained within it.

Validation is a tool to estimate the prediction error, therefore cross-validation can be used. Cross-validation separates the data into segments, then the main data is used for model creation and the smaller data segments are used for validation. The number of values within a segment should be roughly the same for all iterations, and every data value should be used for validation once. [40]

### 1.4. Aim

The aim of this thesis is to research the feasibility of using NIR spectroscopy with a push broom HSI system as a PAT tool for printed medicine. During the printing process knowledge about the drop volume of the printing system is limited, therefore, the actual mass of dispensed API cannot be known accurately. It is necessary to find a way of deducing the actual concentration of API based on spectroscopic data, by applying regression coefficients calculated on a calibration data set to new measurements and predicting the mass correctly, comparing it to HPLC results.





## 2. Methods

Table 2.1.: Parameters used for target setup.

Parameter	Description
Start point	displacement in x and y direction respectively from upper left corner
Field gap	space between fields in x and y direction respectively
No. of Fields	number of fields in x and y direction
Spots per Field	number of spots per field in x and y direction
Spot pitch	distance between spots in field in x and y direction respectively

trickling to another than the defined position.

In the target setup page of Scienion's software, the user may define a number of parameters to create the desired printing pattern described in table 2.1.

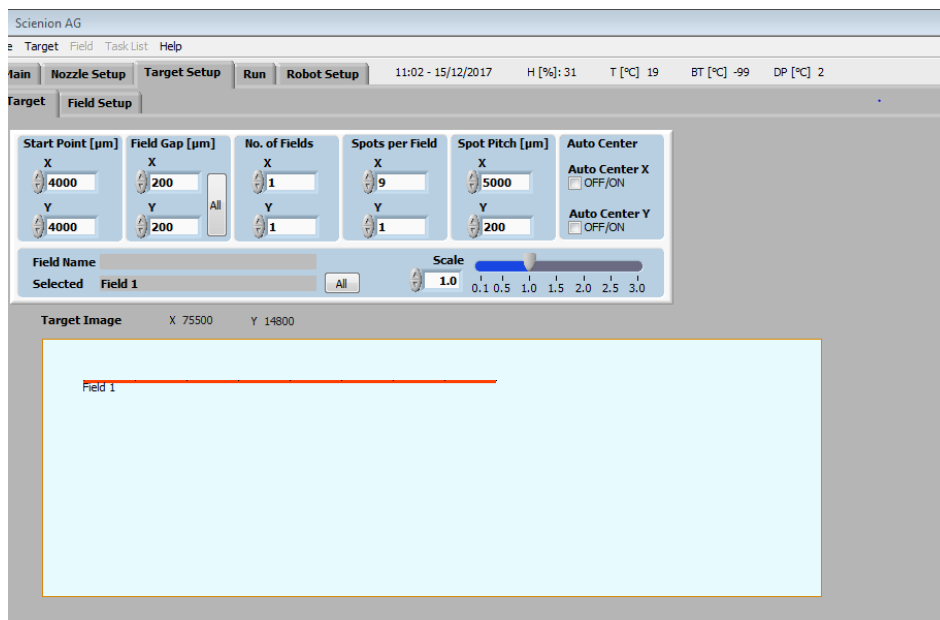


Fig. 2.2.: Target setup menu with in Scienion's software.

### 2.1.2. Ink Formulations

Three inks were formulated and tested before choosing one for further analysis and quantification. The different substances are listed in table 2.2, the corresponding material data safety sheet can be found in the appendix.

## 2. Methods

Table 2.2.: Formulated inks.

API	Solvent	Concentration (mg/ml)
Sodium picosulfate	water	200
Metformin hydrochloride	water	250
Carvedilol	ethanol	10

The differences in concentration stem from the API's solubility, and these concentration have been tested to create a stable drop when printing. Metformin hydrochloride (Sigma Aldrich, St. Louis, USA) and sodium picosulfate (Shenzhen Nexconn Pharmatechs Ltd., Shanghai, China) are soluble in water up to  $1.06 \cdot 10^3 \text{ mg/ml}$  at  $25^\circ\text{C}$  (est) and over  $1 \text{ mg/l}$ , respectively. Carvedilol (Hersteller) is practically insoluble in water [41] therefore ethanol is used which is also a good solvent. It's disadvantage is that the preparation effort for the printer is higher, since the system liquid needs to be changed as well.

Sodium picosulfate is a laxative used to treat constipation or preparing patients before colonoscopies or surgeries to clean the large bowel.

Metformin is a medication for overweight type 2 diabetes patients.

Carvedilol is a beta blocker use on patients suffering from congestive heart failure or left ventricular dysfunction or high blood pressure.

### 2.1.3. Substrates

As described in 2.1 several substrates are used to choose one for further analysis. The substrates are

- Wafer Paper Sheets with the ingredients potato starch and water (Print4you Cake Toppers, UK) (starch)
- Listerine Pocketpaks<sup>®</sup>, Oral Care Strips (listerine)
- Ethylene vinyl acetate (EVA),
- Gelatin film strips (HGC) (Capsugel, Bornem, Belgium)
- Gelatin film strips with 2% titanium dioxide (HGC+TiO<sub>2</sub>) (Capsugel, Bornem, Belgium)
- Hydroxypropylmethylcellulose (HPMC) (Capsugel, Bornem, Belgium): is an ingredient of a water-soluble cellulose ether derivative used for controlled-release preparations.
- Edible icing sheet containing water, cornstarch (bulking agent, corn syrup, corn syrup solids, cellulose (firming agent), sorbitol, glycerin, vegetable oil, gum arabic, vanilla flavor enhancer, titanium dioxide and citric acid (preservative) (DECO Enterprises Ltd, Sutton Valence, UK) (*icing*)

All of these substrates, are materials used in pharmaceutical sciences.

Quantification is done for one ink in combination with one substrate. In order to decide which combination of ink and substrate to use several calibration curves are

## 2. Methods

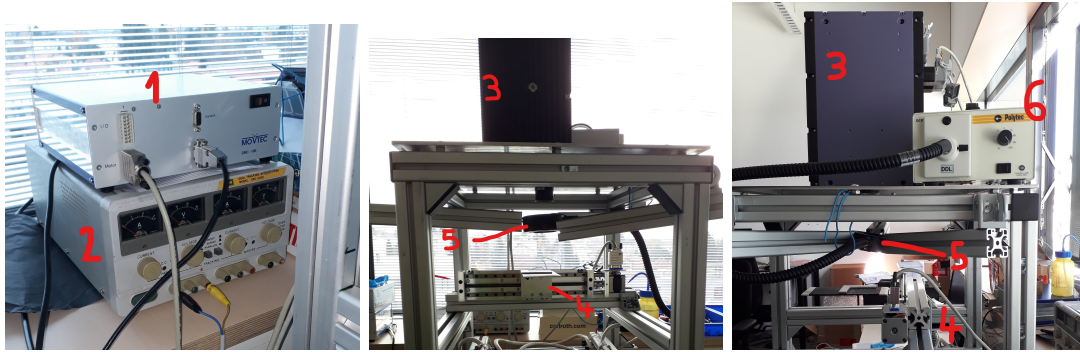


Fig. 2.3.: Setup for NIR measurements. (1) Linear stage control, (2) power supply, (3) helios sensor, (4) linear stage, (5) fiber optig ring light guide, (6) Light source.

printed on different combinations of inks and substrates. The ink formulations and substrates to choose from are described in sections 2.1.2 and 2.1.3, respectively. The printing pattern consists of 7 spots with a varying number of drops dispensed at each spot.

### 2.2. Near Infrared Spectrometer Setup

For the near infrared (NIR) measurements the HELIOS NIR G2-320 Class by EVK (EVK DI Kerschhaggl GmbH, Raaba, Austria) is used. The sensor is powered by a 24 V DC external power supply.

This hyperspectral imaging (HSI) system measures a spectral range with 256 pixels spanning 900 -1700 nm, a sample is measured one line at a time that is subdivided into 320 pixels, the maximum measurement frequency is 333 Hz. To improve resolution a KOWA extension ring of 5 mm is added to the sensor's lens.

Underneath the lens a linear stage is mounted, which is used to move the samples at a constant rate to be able to measure the entire sample. The MovTec SMC-100 stage is hooked up to a computer where the NanoPro V17 Software is used to steer it. A platform covered in blackened aluminum, is attached to the stage, where the samples can be positioned.

For illumination a tungsten- halogen lamp is used (Polytec GmbH, Waldbronn, Germany) to which a fiber optic ring light guide (Edmund Optics GmbH, York, UK) is attached. The ring light guide is positioned just below the HSI lens. This setup allows for shadow less illumination of the samples.

The setup can be seen in figure 2.3. Although the individual parts existed this setup was designed and assembeled before any measurements could be undertaken.

## 2.3. Measurements

As can be seen in section 2.2 all measurements are performed in reflectance mode. This means the light source is positioned in the same direction as the spectrometer in respect to the sample. The light hits the sample and has a penetration depth of about  $500 \mu\text{m}$ . Depending on the chemical configuration of the molecules within the sample, light is reflected due to vibrations in the molecule. The reflected light then hits the spectrometer. To calculate the absorbance units from the measurements equation (2.1) is used

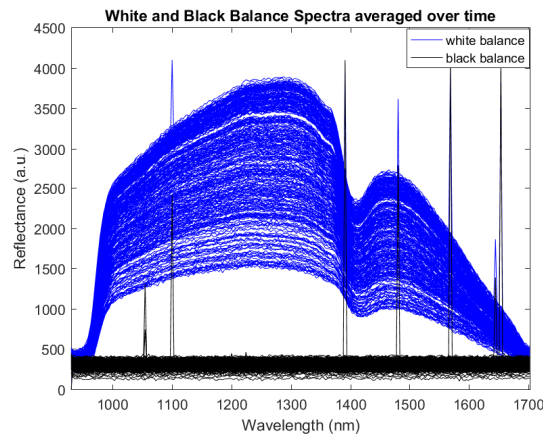
$$A = -\log_{10} \left( \frac{x - B}{W - B} \right) \quad (2.1)$$

where  $x$  is the reflectance spectrum ( $\text{max } 320 \times 256$ ),  $B$  the black reference spectrum and  $W$  the white reference spectrum.

The white and black reference spectra are measured before the samples with the same parameters by measuring a reflective teflon sheet (which does not absorb NIR light [27] and covering the lens, respectively (see figure 2.4(a)). For these measure-



(a) Teflon sheet and lens cover used for measurement of white and black reference spectrum.



(b) All spectra for white and black reference used for the calculation of absorption units.

Fig. 2.4.: Tools used to create white and black reference spectra, and the spectra averaged along the time dimension.

ments the linear stage is not used, the stage remains static until at least 500 frames have been measured. The measurements are averaged to create the black and white reference spectra used for the calculation in equation (2.1). An example for such spectra can be seen in figure 2.4(b). The figure shows 320 lines for each measurement representing the 320 spatial resolution pixels mentioned above. Outliers exist due to defect pixels in the CCD chip matrix, these values are corrected in the preprocessing stage of the data (see section 2.5).

The hyperspectral data is saved in a data cube with two spatial dimensions  $x$  and  $y$ ,

## 2. Methods

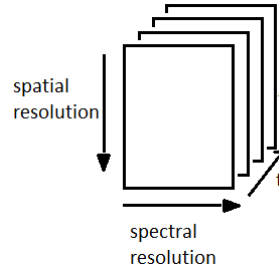


Fig. 2.5.: Schematic representation of the measurement's data set.

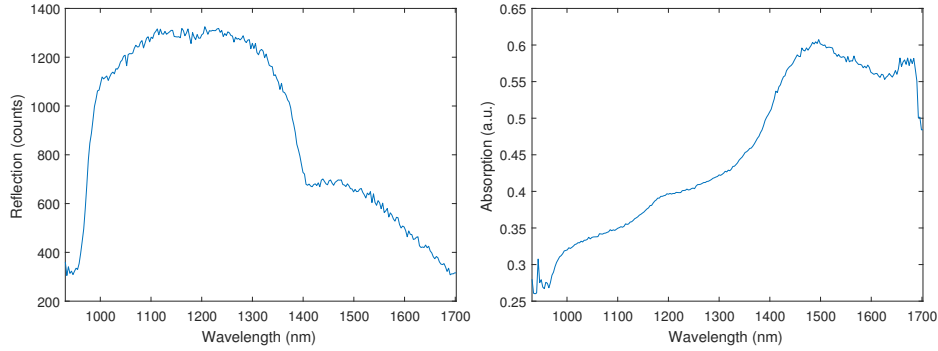


Fig. 2.6.: Measurement spectrum (reflection)(left panel) and calculated absorption (right panel) of a random pixel.

and a spectral dimension. The temporal direction  $t$ , will be denoted as pixelvalue  $x$  in this work, the spatial dimension as pixelvalue  $y$  (in the figure noted as spatial resolution). Figure 2.5 shows a schematic representation of this datacube. This means the intensity of a certain pixel is a function of time  $t$ , space  $x$ , and wavelength  $\lambda$ .

For the measurements exposure time is set high at  $9000 \mu\text{s}$  to ensure enough light reaches the sensor, while the frequency is kept rather low at 10 Hz.

During the measurement of samples, the specimen is secured on blackened aluminum, which has a very high absorbance, keeping them as flat as possible as to avoid distance variation during measurement. The sample is positioned where illumination is the highest. The speed of the linear stage is kept at 2 mm/s. How data changes when going from reflectance to absorption using the aforementioned equation can be seen in figure 2.6. Here a random pixel of a measurement of sodium picosulfate on HGC+TiO<sub>2</sub> is displayed. The conversion from reflectance to absorption units is done, to later get a unit proportional to the concentration based on Beer's law:

$$A = -\log\left(\frac{I}{I_0}\right) = \epsilon cd \quad (2.2)$$

where  $d$  is the light's path length,  $c$  is the concentration and  $\epsilon$  is an extinction coefficient. This is not always ideal since path length in reflection mode is not known, due to Beer Lambert usually being defined for a single chemical compound with known extinction.

## 2.4. Resolution

### 2.4.1. Spatial Resolution (Pixelvalue y)

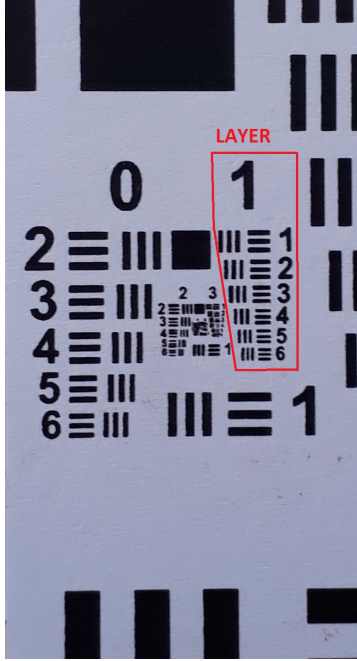


Fig. 2.7.: USAF test chart used for the determination of spatial resolution

To determine the resolution in the spatial direction a 1951 US Air Force (USAF) resolution test chart as seen in figure 2.7 is used. The target is measured without using the linear stage. Multiple time frames are measured with an exposure time of  $9000 \mu\text{s}$  and a frequency of 10 Hz. The measurements are averaged to reduce noise effects and a mean black reference spectrum is subtracted. Furthermore, the different wavelengths are averaged to be able to plot the reflection's cross-section, as can be seen in figure 2.8(b). No further processing or smoothing are performed. Figure 2.8(a) shows the surface plot for the measurement of layer one in figure 2.7 and the additional element 1 below, the color represents the mean intensity of all wavelengths.

The lines at about 170 pixels (fig. 2.8(a)), represent the one in figure 2.7, while the other absorption peaks represent the elements 1-6 of layer 1, where an element is the triple lines. Elements 1-3 are clearly separable, while the lines at element four are no longer. Therefore, it can be concluded that layer 1 element three represents the resolution limit.

Resolution according to the test chart is given by equation (2.3), where  $G$  is the group (the number above each column of elements, and  $e$  is the element number (given next to the triple lines), the resolution is given in  $\text{lp}/\text{mm}$ . A line pair (lp) consists of a black and a white line [42].

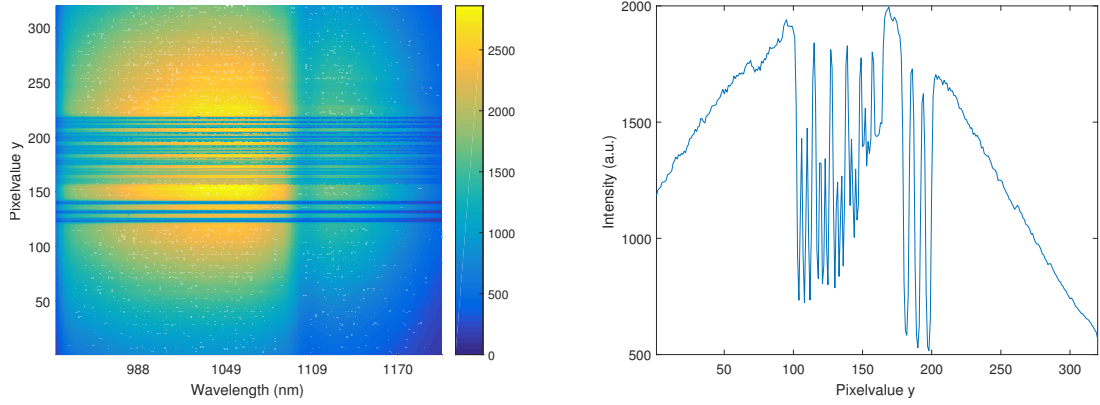
$$\text{Resolution}(\text{lp}/\text{mm}) = 2^{G+(e-1)/6} = 2.5\text{lp}/\text{mm} \quad (2.3)$$

A line width in this element is  $228 \mu\text{m}$  and is represented in two measurement points, as can be seen in figure 2.8(b), resulting in a resolution of  $R_y = \frac{0.22887\text{mm}}{2\text{pixels}} = 0.114 \frac{\text{mm}}{\text{px}}$ . The same can be confirmed counting the number of measurements for 2.5 line pairs ( $\frac{1\text{mm}}{9\text{pixels}} = 0.111 \frac{\text{mm}}{\text{px}}$ ).

### 2.4.2. Temporal Resolution (Pixelvalue x)

The second dimension in figures such as the center panel in 2.9 is created by plotting the subsequently measured lines (third dimension in 2.5) next to each other. The resolution in this direction is a function of the measurement and linear stage velocity as well as the basic resolution.

## 2. Methods



(a) Surface plot of measurement of layer one, color represents the intensity of the reflection. (b) Cross-section for the measurement wavelengths averaged.

Fig. 2.8.: Measurement

To find this value a single drawn line is moved through the sensor's measuring line to approximate the resolution in the temporal dimension, this direction will be called pixelvalue  $x$ . The line can be seen in the left panel of figure 2.9. The paper is moved at a velocity of 0.2 mm/s and measured with a frequency of  $f=10$  Hz.  $b_{line}$  is the line's width measured using the Senterra optical microscope (Bruker, Billerica, USA) and it's measuring tool, the measured  $b_{line}$  equals 285  $\mu\text{m}$ .

The middle panel of figure 2.9 shows the surface plot for the absorption when the wavelengths are averaged, the right panel shows the absorption along the pixelvalue  $x$  axis.

To calculate an approximation for this resolution the assumption is made that the real line ( $f_{line}$ ), the systems point spread function (PSF) ( $f_{PSF}$ ) and the measured line ( $f_{meas}$ ) can be displayed as a Gauss function. This allows the linewidth to be expressed as the standard deviation  $\sigma$  of Gauss functions, based on the rules of convolution (as described in (2.5)).

$$f_{meas} = f_{line} * f_{PSF} \quad (2.4)$$

$$\sigma_{meas}^2 = \sigma_{line}^2 + \sigma_{PSF}^2 \quad (2.5)$$

The width of the drawn and measured line can be measured (left and right panel in 2.9). This allows the deduction of the PSF.

$$\sigma_{PSF} = \sqrt{\sigma_{meas}^2 - \sigma_{line}^2} \quad (2.6)$$

To find  $\sigma_{meas}$  the number of pixels at full width half maximum (FWHM)  $n_{FWHM}$  can be counted (right panel figure 2.9). Gaussian curves have the relationship:  $FWHM = 2\sqrt{2\ln(2)}\sigma$ .

The line's width ( $b_{line}$ ) is known. The blackness drops drastically towards the line's border, therefore the assumption is made that the measured width contains 99% of



## 2. Methods

the line's Gauss function, this corresponds to  $6\sigma$ .

To convert the width in pixels to mm a factor is needed, this factor is determined by the frequency of measurement  $f$  and the speed of the linear stage  $v$ :  $k = v/f(mm)$ . This results in:

$$\sigma_{PSF} = \sqrt{\left(\frac{n_{FWHM}}{2\sqrt{2\ln(2)}} \cdot k\right)^2 - \left(\frac{b_{line}}{6}\right)^2} \quad (2.7)$$

$n_{FWHM}$  is 19,  $k$  equals 0.02 mm, and as mentioned before  $b_{line}$  was measured at 0.285 mm. Using equation (2.7) this results in a resolution in the x direction of  $R_x = 0.154 \text{ mm}/px$ .

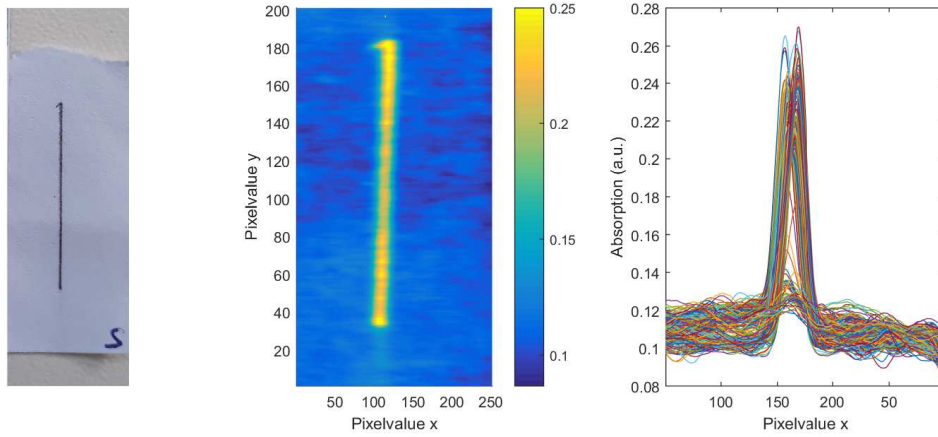


Fig. 2.9.: Photo of pencil line and its NIR measurement shown at the mean wavelength.

## 2.5. Data Preprocessing

### 2.5.1. Defect Pixel Correction

As can be seen in figure 2.4(b) there are several outliers. The mean black reference spectrum as shown in the figure is chosen to perform the detection.

Herein the mean value of each wavelength and subsequently the median absolute deviation is calculated. The threshold for an outlier detection is set to 7 times the median deviation, the value was chosen through trial and error (eq. (2.8)). These outliers are then interpolated linearly, setting the value to the mean of the two neighboring points. If the outlier happens to be at a border the value is set to its neighboring value.

$$OL = (x - \bar{x}) > (7 \cdot median(|x - \bar{x}|)) \quad (2.8)$$

After the indices of defect pixels are known, they are corrected in all measurements as described above. To ensure the correct calculation of the absorbance units (equation (2.1)) it is important to check the denominator for any values that become zero



## 2. Methods

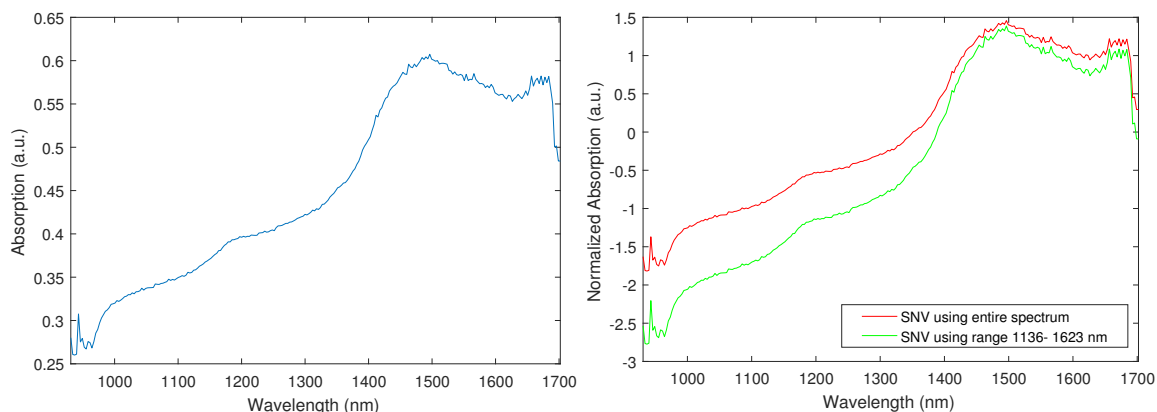


Fig. 2.10.: Calculated absorption spectrum (left panel) and normalized absorption after standard normal variate (SNV) (right panel).

after subtraction, if this is the case the value is interpolated linearly. Furthermore any values below zero after the division must be detected, these are changed to the smallest possible value in MATLAB<sup>®</sup> eps before applying the logarithm.

### 2.5.2. Standard Normal Variate

The first step in the processing pipeline after absorbance calculation is applying the standard normal variate. This is a popular method for scatter correction, since it does not require a reference signal [43].

Equation (2.9) is implemented in MATLAB<sup>®</sup>, where  $x_{orig}$  is the original two dimensional spectrum (as in space and wavelength),  $\bar{x}_{orig}$  is the mean value of each spectrum (along the spectral dimension), and  $\sigma_{x_{orig}}$  is the standard deviation ( $320 \times 1$  vector repeated 256 times).

$$x_{corr} = \frac{x_{orig} - \bar{x}_{orig}}{\sigma_{x_{orig}}} \quad (2.9)$$

Since certain spectral values outside the region of our interest might influence the mean and standard deviation too much, it is possible to only use a range of wavelength values for the correction. If a range is not given the entire spectral range is used. The effect of using only a certain range can be seen in figure 2.10. In this thesis a spectral range of 1158 - 1557 nm was used.

### 2.5.3. Savitzky-Golay Filtering

According to Rinnan et al. derivatives are capable of removing additive effects in spectra and have been used in analytical spectroscopy for a long time.

For derivation and smoothing the very popular method of using a Savitzky-Golay filter (SG-filter) is applied. This filter smoothes and derives the spectra by fitting a low order polynomial to a subset of data and estimating the derivation for this fit [43] [44]. MATLAB<sup>®</sup> provides the `sgolayfilt()` function. Here it is important to

## 2. Methods

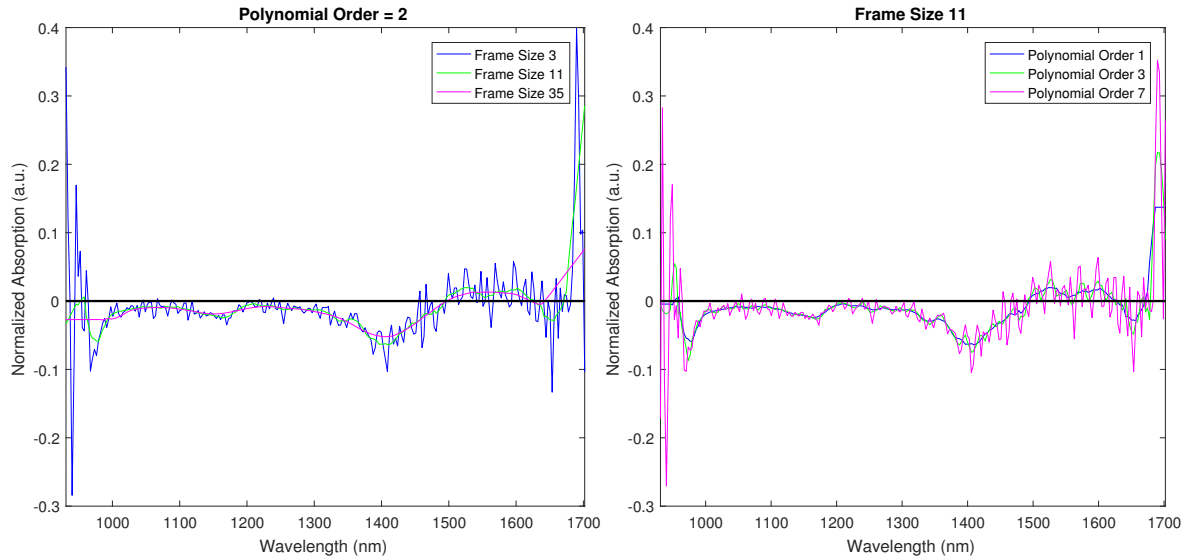


Fig. 2.11.: Differences in filter result by varying frame size (left) and polynomial order (right), keeping the polynomial order constant at 2 and frame size constant at 11 points, respectively.

find the right parameters for the frame size and polynomial order, to reduce noise through smoothing and still retaining all relevant information. The effect of changing frame size and polynomial order can be observed in figure 2.11. Since the filter uses the subset of frame size to estimate the central point of the derivation the result is smoother the bigger the frame size is. Contrary to this for the polynomial order the higher the order is, the noisier the signal becomes due to the more complex fitted curve.

To find the right parameters is not an easy task and requires a lot of trial and error. In this thesis multiple parameters were tried, for the final results a polynom of order 2 was fitted to a 35 step size window.

### 2.6. Quantification

For the quantification a PLS regression or projections to latent structures (PLS) is used. This method combines features of PCA and multiple linear regression. It can be helpful to predict a set of dependent variables from a large set of independent variables [37]. The algorithm projects the data to a lower-dimensional hyperplane, this allows the finding of quantitative relationships among variables by decomposing a response  $\mathbf{Y}$  and an observation matrix  $\mathbf{X}$  by maximizing the covariance between the two [33] [37].

The observation matrix  $\mathbf{X}$  consists of  $N$  observations and  $M$  dependent variables i.e. the NIR measurements, where  $N$  observations are the spatial and temporal information (the data set of figure 2.5 is rearranged as shown in figure 2.12) and the dependent variables are the wavelengths. The response matrix  $\mathbf{Y}$  has the dimension  $N \times K$ , to each observation the predicted value is matched i.e. the concentration.

For the quantification of the printed medicine a calibration curve (see chapter 3) is

## 2. Methods

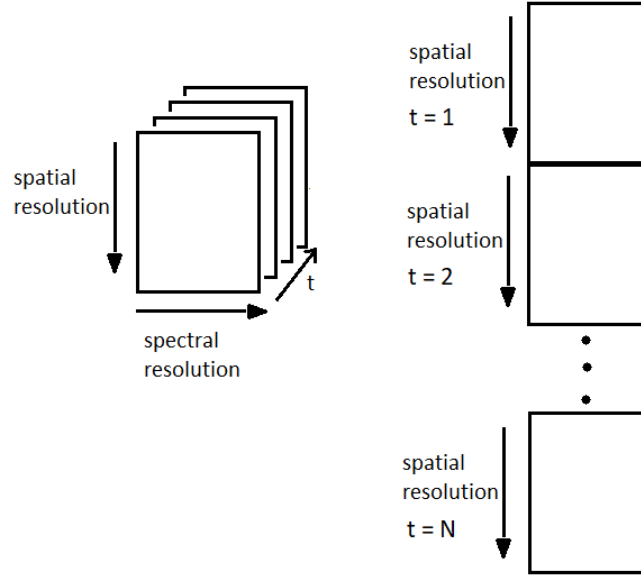


Fig. 2.12.: Schematic of how data set is rearranged to reduce it from three to two dimensions.

printed, measured and preprocessed as described in section 2.5 and then rearranged in two dimensions and declared as  $\mathbf{X}$ . Since only one parameter, i.e. the concentration, needs to be quantified the response matrix is of size  $N \times 1$ , each observation is matched to one value.

$\mathbf{Y}$  is created by mapping the measurement's spots onto a new matrix and setting each value to the estimated mass of API in that observation. The mass can be calculated with equation (2.10), where  $v_{drop}$  is the drop volume recorded by the printer and  $c_{ink}$  is the ink's concentration.

$$m_{API} = v_{drop}(\text{ml}) \cdot c_{ink}(\text{mg/ml}) \quad (2.10)$$

It is important to predict the mass as accurately as possible, which automatically improves the explained variance and by maximizing the correlation between the data sets and therefore the model. To gather the distribution of API within a dispensed spot, a Raman measurement is performed (see section 2.7.1). The advantage of Raman spectrometers over NIR spectrometers is the proportionality of intensity and concentration, as well as its insensitivity to the surface composition in comparison to NIR. The disadvantage, however, is its expenditure of time, while the NIR measurements measures  $20 \mu\text{m}$  in  $9 \text{ ms}$ , the mapping of a  $(1 \times 1) \text{ mm}$  area took several hours. With  $\mathbf{X}$  and  $\mathbf{Y}$  set the regression, to estimate the parameters can be started, as mentioned in the introduction, either the NIPALS or the SIMPLS algorithm could be applied. The following two subsections describe these two algorithms in more detail.

### 2.6.1. NIPALS Algorithm

The algorithm consists of the following steps after mean centering.

## 2. Methods

1. Start:  $\mathbf{u}$  to the first column of  $\mathbf{Y}$
2.  $\mathbf{W}^* = \frac{\mathbf{X}^T \mathbf{u}}{\mathbf{u}^T \mathbf{u}}$
3. scale  $\mathbf{W}$  to length one ( $\mathbf{W}^* = \frac{\mathbf{W}^*}{|\mathbf{W}^*|}$ )
4.  $\mathbf{T} = \mathbf{XW}^*$
5.  $\mathbf{C} = \frac{\mathbf{Y}^T \mathbf{T}}{\mathbf{T}^T \mathbf{T}}$
6. scale  $\mathbf{C}$  to length one
7.  $\mathbf{U} = \frac{\mathbf{Y}^T \mathbf{C}}{\mathbf{C}^T \mathbf{C}}$
8. repeat 2-8 until convergence (for  $M > 1$ )

With  $\mathbf{U}, \mathbf{C}$  and  $\mathbf{W}^*$  the X- and Y loadings ( $\mathbf{P}$  and  $\mathbf{Q}$ ) can be calculated due to the relationship shown in (2.11). This means that the X-scores  $\mathbf{T}$  with the estimated weights  $\mathbf{C}$  describe  $\mathbf{Y}$  with small residuals  $\mathbf{F}$ . The loadings are used to calculate  $\mathbf{E}$  and  $\mathbf{F}$ , the residuals used for the next iteration.

$$\mathbf{Y} = \mathbf{TC}^T + \mathbf{F} \quad (2.11)$$

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (2.12)$$

$$\mathbf{P} = \frac{\mathbf{X}^T \mathbf{T}}{\mathbf{T}^T \mathbf{T}} \quad (2.13)$$

$$\mathbf{Q} = \frac{\mathbf{Y}^T \mathbf{U}}{\mathbf{U}^T \mathbf{U}} \quad (2.14)$$

$$\mathbf{E} = \mathbf{X} - \mathbf{TP}' \quad (2.15)$$

$$\mathbf{F} = \mathbf{Y} - \mathbf{XW}^* \mathbf{C}^T \quad (2.16)$$

In the first iteration all calculated variables are vectors, the first X-score  $\mathbf{t}_1$  can be thought of as a new, latent variable, which reflects the information in the original X-variables, this is the first component, to add the second, the algorithm is repeated with  $\mathbf{E}$  and  $\mathbf{F}$  as input. With each iteration, and added component  $\mathbf{F}$  will decrease until a convergence. Therefore,  $\mathbf{F}$  indicates is an indicator for the model's quality, since it describes what is not explained with the regression terms. It should be as small as possible.

$\mathbf{W}^* \mathbf{C}^T$  is also called the PLS-regression coefficient  $\mathbf{B}$ . According to Wold et al. the changing (called deflation) of  $\mathbf{Y}$  is optional since the results are equivalent. [35, 38] For an ideal model  $\mathbf{Y}$  can be described as:  $\mathbf{Y} = \mathbf{XB}$ .

### 2.6.2. SIMPLS Algorithm

To shorten the NIPALS algorithm De Jong et al. propose to find X-scores  $\mathbf{t}_a$  that maximize the covariance with the corresponding Y-loadings  $\mathbf{u}_a$ . In their system the weights  $\mathbf{W}^*$  are not calculated but a generalized inverse of  $\mathbf{P}' \mathbf{R}$  is used (see (2.17)).

## 2. Methods

[39]

$$\mathbf{T} = \mathbf{X}_0 \mathbf{R} \quad (2.17)$$

Here  $\mathbf{X}_0$  is the original centered data instead of the depleted  $\mathbf{X}$  matrix since these correspond to linear combinations of the centered  $X$  variables.

With this adaptation neither the weight  $\mathbf{W}^*$  nor the depleted matrices  $\mathbf{E}$  and  $\mathbf{F}$  need be calculated. [29]

The pseudocode of the SIMPLS algorithm as described by de Jong [39] is shown in figure 2.13

```

INPUT  $n \times p$  matrix  $\mathbf{X}$ ,
         $n \times m$  matrix  $\mathbf{Y}$ ,
        number of factors  $A$ .

 $\mathbf{Y}_0 = \mathbf{Y} - \text{MEAN}(\mathbf{Y})$            center  $\mathbf{Y}$ 
 $\mathbf{S} = \mathbf{X}' * \mathbf{Y}_0$                    cross-product
For  $a = 1, \dots, A$                 per dimension
     $q =$  dominant eigenvector of  $\mathbf{S}' * \mathbf{S}$    $Y$  block factor weights
     $r = \mathbf{S} * q$                     $X$  block factor weights
     $t = \mathbf{X} * r$                     $X$  block factor scores
     $t = t - \text{MEAN}(t)$               center scores
     $normt = \text{SQRT}(t' * t)$         compute norm
     $t = t / normt$                   normalize scores
     $r = r / normt$                   adapt weights accordingly
     $p = \mathbf{X}' * t$                     $X$  block factor loadings
     $q = \mathbf{Y}_0' * t$                    $Y$  block factor loadings
     $u = \mathbf{Y}_0' * q$                    $Y$  block factor scores
     $v = p$                            initialize orthogonal loadings
    if  $a > 1$  then
         $v = v - \mathbf{V} * (\mathbf{V}' * p)$     make  $v \perp$  previous loadings
         $u = u - \mathbf{T} * (\mathbf{T}' * u)$     make  $u \perp$  previous  $t'$  values
    end
     $v = v / \text{SQRT}(v' * v)$           normalize orthogonal loadings
     $\mathbf{S} = \mathbf{S} - v * (v' * \mathbf{S})$       deflate  $\mathbf{S}$  with respect to current loadings
    Store  $r, t, p, q, u,$  and  $v$  into
    into  $\mathbf{R}, \mathbf{T}, \mathbf{P}, \mathbf{Q}, \mathbf{U},$  and  $\mathbf{V}$ , respectively.
End

 $\mathbf{B} = \mathbf{R} * \mathbf{Q}'$                    regression coefficients
 $h = \text{DIAG}(\mathbf{T} * \mathbf{T}') + 1/n$        leverages of objects
 $\text{varX} = \text{DIAG}(\mathbf{P}' * \mathbf{P}) / (n - 1)$  variance explained for  $X$  variables
 $\text{varY} = \text{DIAG}(\mathbf{Q}' * \mathbf{Q}) / (n - 1)$  variance explained for  $Y$  variables

```

Fig. 2.13.: Pseudocode for the SIMPLS algorithm as published by De Jong [39].

The MATLAB<sup>®</sup> function `plsregress` uses the SIMPLS algorithm. As input the measured NIR data, the created concentration matrix and the number of components to do the regression with and a cross-validation partitioning are used. The vector of regression coefficients, denoted as `beta` is extended by one variable (column) that contains the means, this could be interpreted as the  $0^{\text{th}}$  component. Therefore, when applying `beta`, the data matrix (NIR- measurements) has to be extended, with a column of ones.

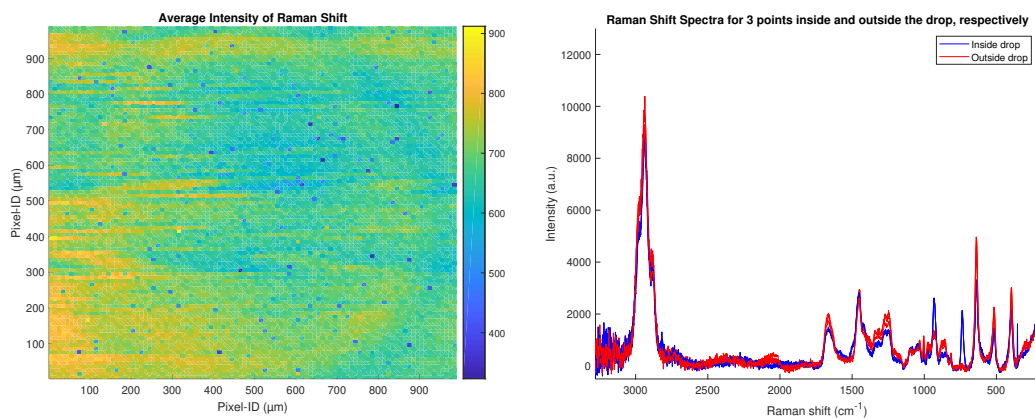
## 2.7. Reference Methods

### 2.7.1. Raman Spectroscopy

As mentioned above, a Raman shift measurement is performed on two drops. For the measurement the RamanStation 400F (PerkinElmer, Waltham, US) is used. This station uses a 350 mW near infrared 785 nm laser, and measures a spectral range of 95- 3500  $\text{cm}^{-1}$ .

The drop is positioned on the stage, and then moved below the laser by using the Spectrum<sup>TM</sup> software and the preinstalled USB CCD video camera. During the measurement an area of 1x1 mm with a spacing of 0.01 mm is mapped. Each point is scanned 3 times with an exposure time of 6 seconds, resulting in a total time of 50 hours.

Based on the spectra (comparing inside and outside the drop) it can be determined that metformin hydrochloride has a peak at 735  $\text{cm}^{-1}$ . With the fsmloader by Ben Perston it is possible to load the PerkinElmer files into MATLAB<sup>®</sup> [45]. Next the data corresponding to the wavelengths around this peak are averaged (730 - 740), and the cross-section at this wavelength is plotted to examine the intensity and therefore the concentration. Figure 2.14(a) shows the mean intensity over all wavelengths for a drop of metformin hydrochloride ink on HGC+TiO<sub>2</sub>. Here the drop's border is vaguely visible, the determination of the peak follows by comparing spectra inside and outside the drop, as shown in figure 2.14(b). At 735  $\text{cm}^{-1}$  the curves inside the drop show a peak, while the outside does not. Since the Raman measurement was conducted days after printing, it is assumed that any water has evaporated. The resulting cross-section and the recreation of the response matrix used for the



(a) Mean Intensity

(b) Spectra at randomly chosen points inside and outside the drop

Fig. 2.14.: Raman shift measurement for a drop of the metformin hydrochloride ink (see section 2.1.2) on HGC+TiO<sub>2</sub>.

regression can be seen in chapter 3.

## 2. Methods

### 2.7.2. High-Performance Liquid Chromatography

High-performance liquid chromatography (HPLC) is used to compare the predicted and actual API content for the printed samples. HPLC reveals very accurate results and is a common method for the determination of API content, but it is a destructive method, and therefore not ideal for a quality control.

In order to extract the API, the printed samples are dissolved in 5 ml of deionized water (Hersteller) and shaken at room temperature and 300 rpm for 2 hours. SP content of the printed samples is determined by using HPLC with UV-detection. Flow rate is set to 0.4 ml/min, injection temperature at 20°C, column temperature at 30°C, and detection wavelength at 233 nm. A 90:10 mixture of 1 mM ammoniumacetat and ACN is used as mobile phase for 10 minutes. A Waters Alliance 2695 (MassLynx) (Waters Cooperation, Millford MA, USA) equipped with a Agilent Zorbax SB-CN (narrow-bore 2.1x150 mm, 5 µm Part. Nr. 883700-905) was used.

The results of the HPLC measurements are used to asses the model's prediction values.

### 3. Experiments

As mentioned in section 2.1.2 and 2.1.3 there were multiple APIs and substrates but only one is chosen to perform the quantification. Figure 3.1 show the API's and substrate's NIR- spectra, respectively. Here, the raw spectra without data preprocessing (except defect pixel correction) are plotted. For the decision of which combination of

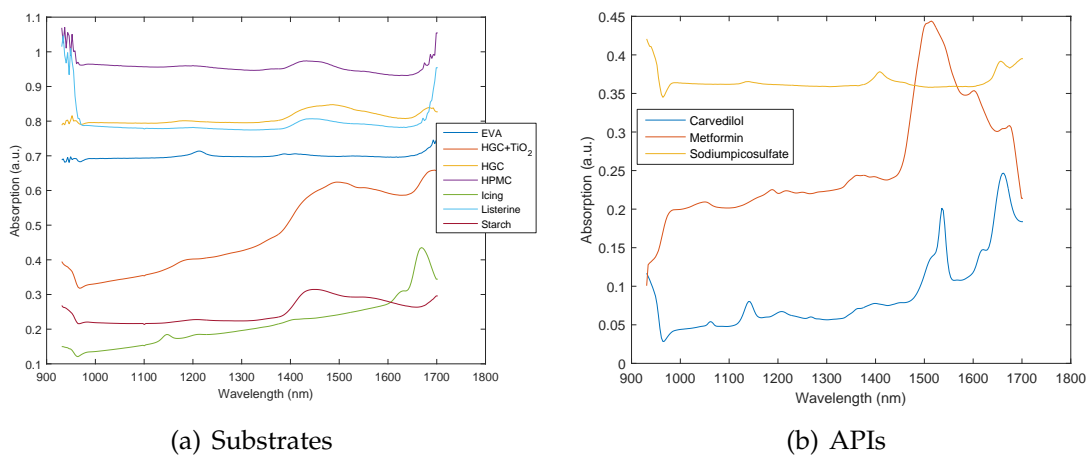


Fig. 3.1.: NIR- spectra for all substrates and APIs measured separately.

API and substrate to use, all combinations are printed with the same pattern. The pattern chosen is a calibration curve with 7 spots each with a different number of drops per spot. Table 3.1 shows the parameters used to print these calibrations in the target setup tab as shown in figure 2.2. In the field setup sub-tab the number of drops are chosen, this results in a printing pattern as shown in figure 3.2. Important to note is, that in the NIR measurement surface plot the order of spots is reversed due to the direction of the moving linear stage.

Table 3.1.: Target page in target setup tab in Scienion AG parameters.

Parameter	Value X	Value Y
No. of Fields	1	1
Spots per Field	7	1
Field Gap [ $\mu\text{m}$ ]	200	200
Spot Pitch [ $\mu\text{m}$ ]	5000	200
Startpoint	<i>variable</i>	<i>variable</i>



### 3. Experiments

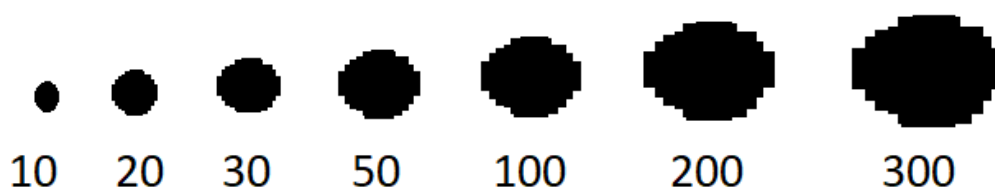


Fig. 3.2.: Printing pattern schematic: 7 spots with increasing number of drops per spot ranging from 10 to 300 drops.

Figure 3.3 shows all substrates with the printed carvedilol ink. The linear stages velocity here was set to 0.1 mm/s resulting in not being able to get all calibration spots in one data set. Colorcoding represents the absorption at carvedilol's peak at 1539 nm (compare figure 3.1). For the creation of these images preprocessing has not been applied. Figure 3.4 shows pattern of sodium picosulfate ink printed

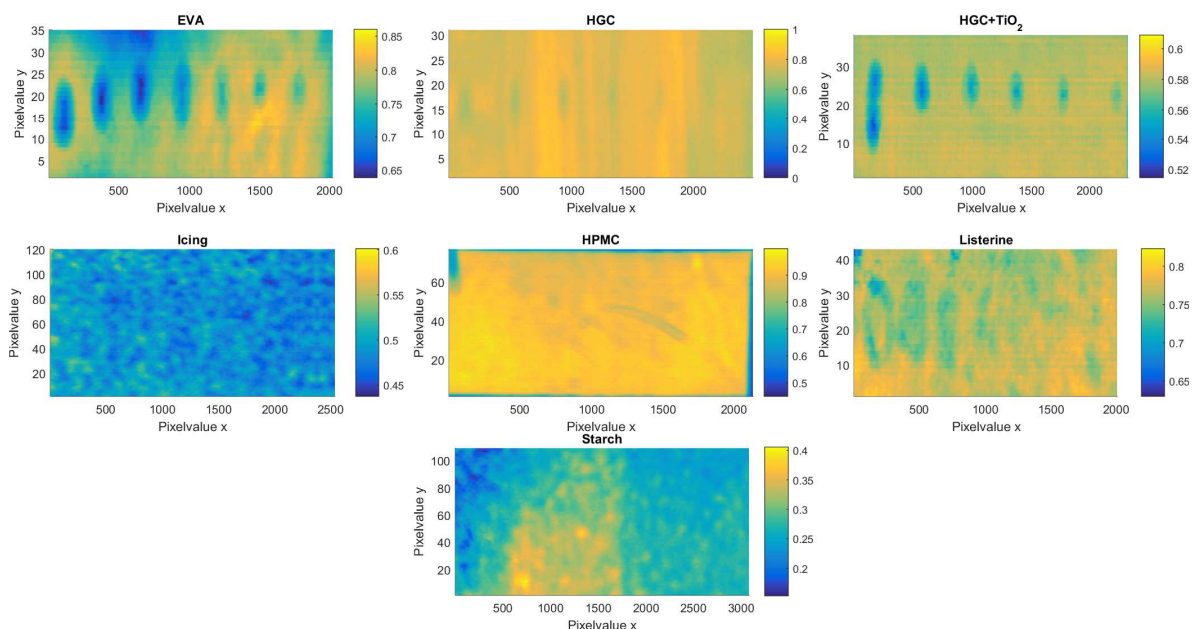


Fig. 3.3.: 9.9 mg Carvedilol in 10 ml ethanol printed on all substrates, ethanol is also used as system liquid for the printer.

on all substrates. Here ink had a concentration of 200.5 mg/ml, the samples were measured with a frequency of 10 Hz and an exposure time of 9 ms. The linear stage was operated at 1 mm/s. All plots are displayed at the 1409 nm peak.

Figure 3.5 shows all substrates with metformin printed on it at a concentration of 250.04 mg/ml. During the NIR measurement the stage was moved at 0.2 mm/s, the measurement frequency was set to 10 Hz and exposure time was 9 ms.

Based on figures 3.3-3.5 starch, icing sheet and listerine are disregarded for further analysis, since listerine and icing break very easily and starch absorbs the ink to such an extend, that it is no longer possible to see the drops after the drying process. For sodium picosulfate HPMC, HGC and HGC+TiO<sub>2</sub> could work for quantification since

### 3. Experiments

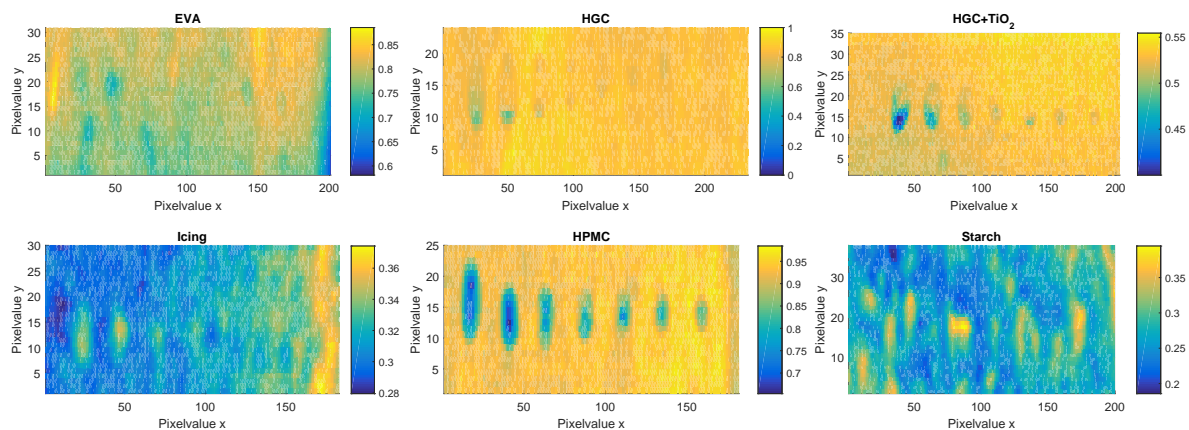


Fig. 3.4.: sodium picosulfate ink on all substrates except listerine since the water created holes.

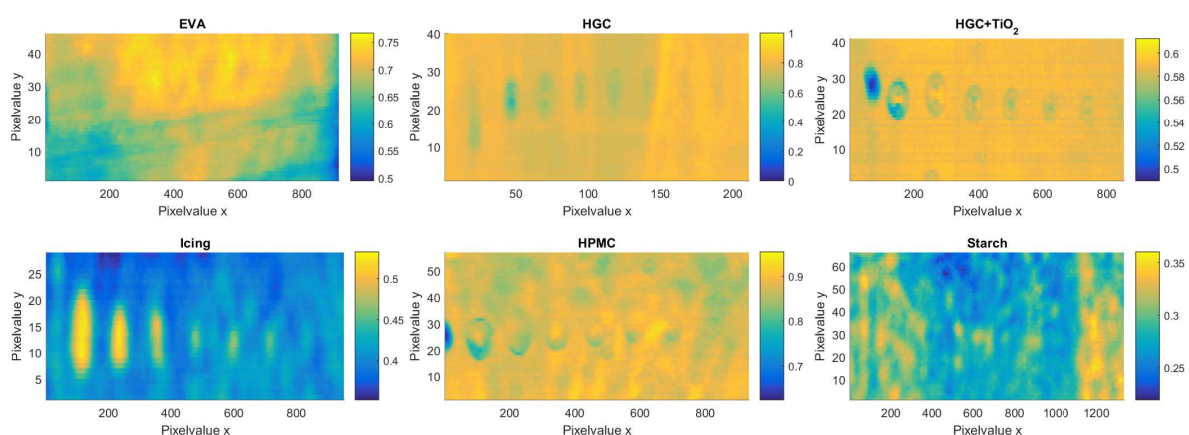


Fig. 3.5.: Metformin ink on all substrates except listerine displayed at 1515 nm.

borders are clearly visible (which they need to be to create a better response matrix). For metformin hydrochloride HGC+TiO<sub>2</sub> is clearly the best substrate showing very clear borders. Therefore this combination was chosen for continuing work.

### 3.1. Basic Modelling

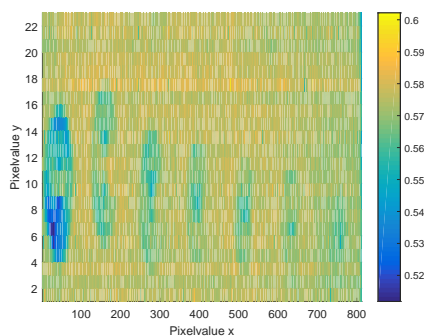


Fig. 3.6.: Calibration curve.

The first attempt of creating a calibration model was done with figure 3.6. This is the absorption plot displayed at 1515 nm (same as upper right panel in figure 3.5). The first 15 and last 11 discrete measured wavelengths are deleted due to excessive noise (leaving a range of 976 - 1669 nm) (not shown in figure 3.6, since this shows the two spatial dimensions at 1515 nm). This figure is also the basis for the Y matrix creation. This is done by creating a mask based on figure 3.6,

### 3. Experiments

where the width of the spot is estimated for each pixelvalue  $y$ . Then all pixels within the border, therefore within the estimated spot, are set to the concentration (calculated with equation (3.1)), while every other pixel is set to zero.  $\bar{v}_{drop}$  is the mean drop volume in ml,  $c$  is API concentration in the ink in mg/ml,  $n_{drops}$  is the number of drops on the spot and  $n_{pixel}$  is the number of pixels within the spot.

$$m_{API,spot} = \frac{\bar{v}_{drop} \cdot c \cdot n_{drops}}{n_{pixel}} \quad (3.1)$$

The 10 drop spot (as shown in figure 3.2) was ignored, since it is so small and barely visible, that it is very hard to define edges.

The data is preprocessed as described in section 2.5. For the standard normal variate (SNV) correction values between 1182 and 1665 nm are used. During the Savitzky-Golay filtering step a second order polynomial is fitted using a 21 point window (chosen after examination of the resulting spectra and advice by more experienced colleagues).

After applying the `plsregress` function its result is applied to the NIR- data to create a prediction matrix where each value represents the mass of API on the respective pixel. Predicting the mass of API, on the data that has been used to create the model is called backprojection. The backprojection should give realistic result as to where API is present. If the input, therefore the designed response matrix, is incorrect, the regression is not able to create a realistic model. Subsequently, the backprojection predicts unrealistic values.

With the first designed response matrix, results are not realistic and point to necessary refinement of the model.

## 3.2. Refinement

Instead of estimating the number of pixelvalues  $x$  contained within the spot (spot's width along  $x$ ) for each pixelvalue  $y$  (as described in 3.1), and therefore to make it more adaptable to other measurements, ellipses are fitted by defining only the center and length of axis setting the value of calculated, estimated mass if condition (3.2) is true ( $(x,y)$  is the test point;  $(h,k)$  center point,  $r_x$  and  $r_y$  half minor and major axis).

$$\frac{(x-h)^2}{r_x^2} + \frac{(y-k)^2}{r_y^2} \leq 1 \quad (3.2)$$

After calculating the mass of API per spot (table 3.2), this value is divided by the number of pixels within the designed ellipses, to give the mass of API per pixel. This resulted in a  $Y$  matrix as shown in figure 3.7.

Figure 3.8 shows the image of the 300 drop spot taken by the Senterra optical microscope. Here it could be assumed, that the coffeeing effect is present, this is

### 3. Experiments

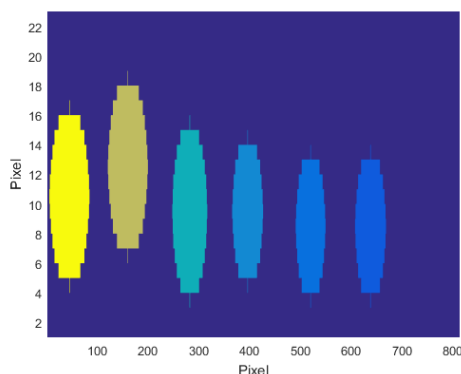


Fig. 3.7.: Modelled  $Y$  matrix with constant concentration within spots from left to right 300,200,100,50,30,20 drops per spot each with corresponding concentration shown in 3.2 calculated with (3.1).

Table 3.2.: Masses of API on each spot calculated with (3.1).

$$\begin{aligned} m_{300} &= 3.53 \cdot 10^{-5} \text{ mg} & m_{50} &= 9.82 \cdot 10^{-6} \text{ mg} \\ m_{200} &= 2.51 \cdot 10^{-5} \text{ mg} & m_{30} &= 6.07 \cdot 10^{-6} \text{ mg} \\ m_{100} &= 1.53 \cdot 10^{-5} \text{ mg} & m_{20} &= 4.30 \cdot 10^{-6} \text{ mg} \end{aligned}$$

why the Raman measurement was taken to confirm the suspicion. Furthermore,

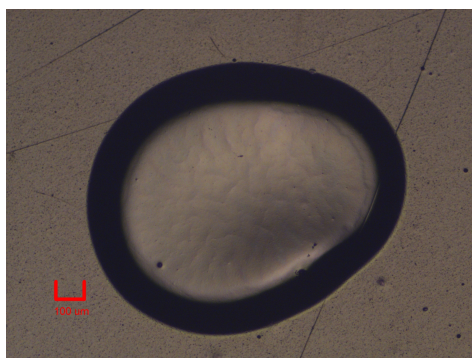


Fig. 3.8.: 300 drops spot taken taken using the Senterra microscope

as described in chapter 2 the Raman measurement was taken to determine the API distribution within the drop. This is achieved by taking the intensity image at  $735 \text{ cm}^{-1}$  and averaging values between  $450$  and  $550 \mu\text{m}$  of one axis and plotting the intensity value along the other direction as shown in figure 3.9. This shows a relatively constant API concentration across the drop with a rapid decrease towards the border. Trying to emulate this the concentration matrix is changed accordingly. This change is done by creating one smaller ellipse (shorter axis and a higher API value) with rings around it with decreasing values. Additionally, it is important, to keep the sum of dispensed and modelled API the same for each spot.

### 3. Experiments

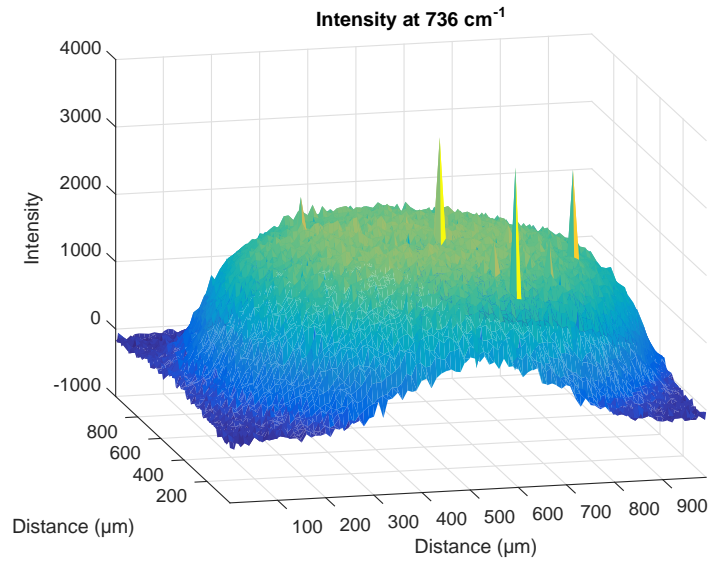


Fig. 3.9.: Raman shift intensity at API peak (metformin).

### 3.3. Testing

To test the created model new samples are printed. Here 17 fields are printed each with 17 spots spaced 300 μm apart. This pattern is repeated with different layers and with a different number of drops per spot. A sketch of this printing pattern can be seen in figure 3.10. The first test pattern consisted of 4 squares, each with one drop

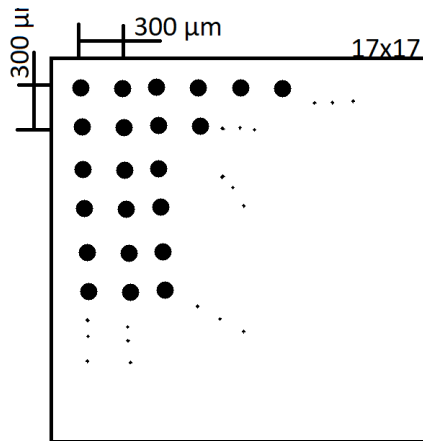


Fig. 3.10.: printing pattern for test model.

per spot at one, 10, 20 and 30 layers (see figure 3.11).

### 3. Experiments

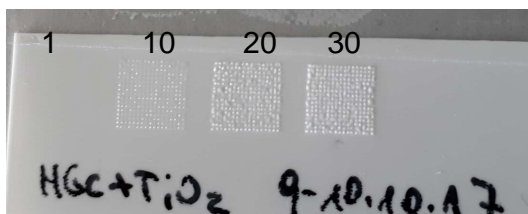
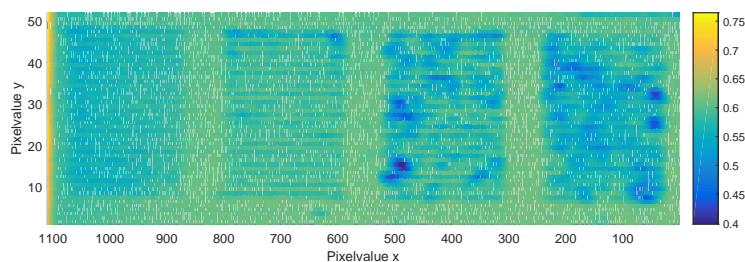


Fig. 3.11.: Absorption displays at 1515 nm (no preprocessing) and photo of first test sample, 4 squares at 1,10,20 and 30 layers, each layer at 1 drop per spot.

## 3.4. Final Process

In order to apply all the lessons learned from the first attempts new samples (calibration and testing samples) are printed.

### Ink Formulation

The new metformin hydrochloride ink is formulated weighing in 2.5003 g and dissolving it in 10 ml of purified water, resulting in a concentration of  $c_{\text{ink}}=250.03 \text{ mg/ml}$ . The ink is filtered once to reduce the clogging-risk during printing.

### Printing

After filling 100 ml of air and subsequently 250 ml of ink into the Scienion printer its camera is used to check if a stable, round drop is dispensed. This was the case when setting the printer parameters to 90 V, with pulse duration of 55 ms. 10000 drops are dispensed while every 500 drops the drop volume is measured. This run reveals a mean drop volume of 413.7 pl with a deviation of 2% (see figure 3.12). The calibration is printed using one field in Y direction, 9 spots in X direction with a spot pitch of 3000  $\mu\text{m}$  (target setup menu shown in figure 2.2), the number of drops is increased with each spot (10, 20, 30, 50, 100, 200, 300, 400, 500, respectively)(setting done in the subtab "Field Setup" (tab Target Setup) within the Scienion software). Additionally three test samples are printed all using the printing pattern shown in figure 3.10. One was printed at one drop per spot at 10, 15, 20, 25 and 30 layers on a square, respectively, one at one drop per spot at 11, 15, 20, 25 and 30 layers and one at 5 drops per spot for 10, 15, 20, 25 and 30 layers, respectively.



### 3. Experiments

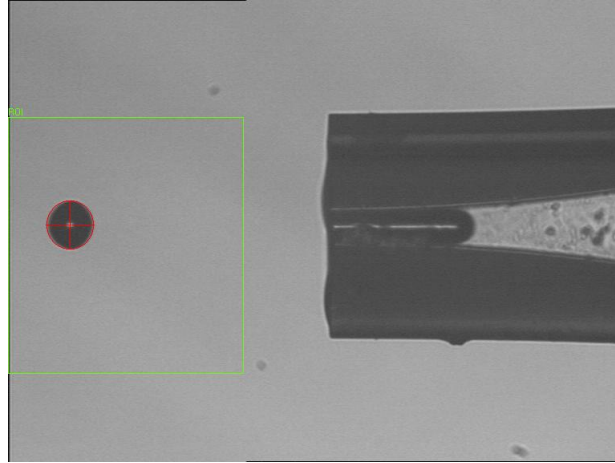


Fig. 3.12.: Measurement of drop volume by the scienion software.

#### 3.4.1. NIR measurement

All samples are fixed on blackened aluminum using tape, positioned on the linear stage and moved at a velocity of  $v=0.2$  mm/s while illuminated with the tungsten light source for the measurement. Spectra at 256 wavelengths along 320 pixels are measured 10 times a second ( $f=10$  Hz) with an integration time of 9 ms. An example is shown in figure 3.13.



Fig. 3.13.: Sample (Testpattern 1 drop per spot, [11,15,20,25,30] layers) on the blackened aluminum, as it is then measured.

#### 3.4.2. Data processing

All measurements are saved, loaded into MATLAB<sup>®</sup> and absorption units are calculated using mean white and black reference spectra.

By plotting absorption as a function of the spatial and the temporal dimension (third dimension of 2.5 as pixelvalue  $x$ , and spatial dimension as pixelvalue  $y$ ) the region of interest (ROI) can be kept while the majority of the data can be deleted to reduce needed computing power. This process is shown in figure 3.14. Before applying the

### 3. Experiments

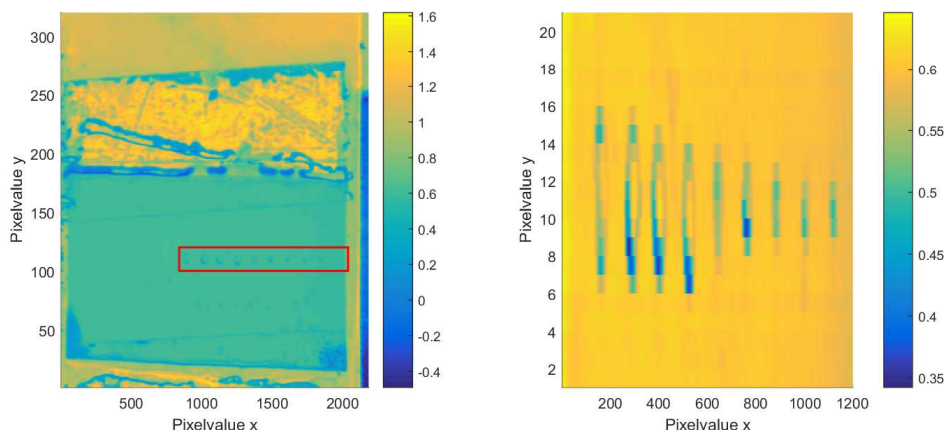


Fig. 3.14.: *Left panel:* Extraction of ROI from entire measurement. *Right panel:* the extracted ROI containing the calibration spots. The colorbars represent absorption at 1515 nm.

SNV and the SG Filter the first 15 and last 16 measured wavelengths are discarded due to excessive noise, reducing the wavelength range to 973 - 1653 nm. The SNV is applied by using the 1140 - 1533 nm range. In the Savitzky Golay filtering step a second order polynomial is fitted to a window of size 35, after attempting several other parameter combinations.

Figure 3.15 shows the normalized absorption at 1515 nm as well as the effect of preprocessing on the spectra shown exemplarily for the spectra at 270 - 280 and 10 - 11 for pixelvalue x and y, respectively.

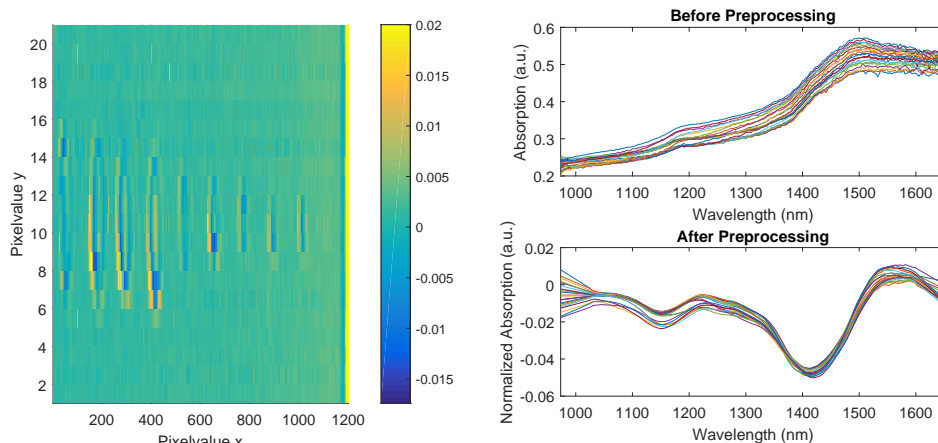


Fig. 3.15.: Effects of data preprocessing on spectra and normalized and derived absorption at 1515 nm.

#### Predictor Matrix

For creation of the predictor matrix, the spot's centers and borders are estimated and ellipses are created. HPLC is only conducted on the spots down to 50 drops (not on 30, 20, 10 drops), therefore only these spots are created in the concentration matrix since only there the mean mass of API per spot is known. The mean mass of API per pixel is calculated by dividing the total mass by the number of pixels in the spot. Then ellipses of different sizes are creating, assigning different values to each one (representing the mass of API at that pixel) are created for each spot to create



### 3. Experiments

the reduced mass towards the border to emulate the distribution revealed by the Raman measurement. The predictor matrix created using HPLC and Raman results is shown in figure 3.16.

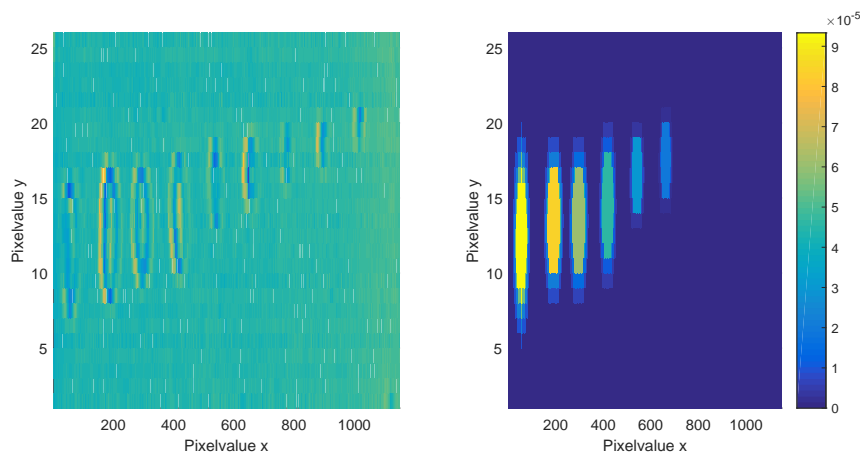


Fig. 3.16.: Preprocessed data displayed at 1515 nm (left, colorbar representing normalized absorption in the left panel of figure 3.15) and the created concentration matrix (response matrix) with color representing the mass of API in mg. A smaller ellipse than the actual drop is estimated in the center and rings are put around it with different amplitudes to emulate the drop of API towards the border.

#### PLS Regression

The `plsregress` function expects at least three inputs, namely the two two-dimensional data sets to do minimization with and the number of components to be used. The number of components can be estimated by the number of components in the mixture (i.e. water, API, substrate, noise) and by examining the explained variance of the components. Figure 4.2(b) shows how the explained variance in  $Y$  (therefore the concentration matrix) increases with the number of components up to 9 and then remains constant, even before that the explained Variance does not increase much. Then 5 components were chosen since after 5 the explained variance increases only slightly. A lower number of components is beneficial to decreasing the probability of finding correlations by chance, and added bonus is that less components also mean that less computing power is needed.

#### 3.4.3. Backprojection and Prediction

After a first model assessment using the percentage of explained variance, examination of weights per component and comparing fitted (the predicted) and observed (response- reference) data, it is of use to backproject the starting data, to see where which values are predicted.

The prediction is created by applying the model beta to the preprocessed data set using multiplication. It is important to note, as described in chapter 2, that beta has a different dimension which is why the preprocessed measurement data needs

### 3. Experiments

to be extended with one column at the beginning (constisting of all ones). The preprocessing before applying the model has to be the same as was used for the model creation.

The result of model application is vector containig the predicted value (mass of API for each pixel), the prediction can be displayed as a function of pixelvalue  $x$  and  $y$ , where the amplitude represents the API.

For quantification ROIs are extracted and the single masses are summed up, to calculate the total dispensed API on the spot. For the standard deviation the two dimensional standard deviation of noise is calculated, therefore, a ROI where nothing was printed is extracted. The value is corrected with the size of the noise ROI.

Furthermore the total dispensed API per sample is calculated by adding all predicted values and all HPLC values, with this the deviation is calculated using equation (3.3).

$$Deviation = \left( 1 - \frac{mass_{prediction}}{mass_{HPLC}} \right) \cdot 100 \quad (3.3)$$

## 4. Results

The HPLC results are shown in tables 4.1 and 4.2 and figure 4.1. The calibration's values are used for the creation of the response matrix. Sample 1 consists of the test pattern printed with one drop per spot 11, 15, 20, 25 and 30 times. Sample 2 consists of the test pattern with five drops per spot 10, 15, 20, 25 and 30 times and sample 3 printed at one drop per spot for 10, 15, 20, 25 and 30 layers.

Table 4.1.: HPLC results for the calibration and the three printed test samples. Plots for these values can be seen in figure 4.1. For the calibration

sample name	drops on one spot	API ( $\mu\text{g}$ )
calibration	50	3.85
	100	6.18
	200	16.46
	300	24.06
	400	33.05
	500	38.26

Figure 4.2 shows the first model creation of the calibration, applying the first few adaptations, described in section 3.4 and using the measurement and the created response matrix shown in figure 3.16. Here the entire measured sample is used without creating the spots not measured in the HPLC (so spots with 30, 20 and 10 drops) (response matrix figure 3.16(b)). Panel (a) shows the calculated weights for each component in dependence of the wavelength variable (Variables are the discrete measured wavelengths). (b) shows how the variance decreases with higher number of components. (c) compares the values estimated in the response matrix (based on HPLC and Raman measurements) (observed) and the predicted values (fitted). (d) shows the regression coefficient in dependence of the wavelength variables, when the variable is applied to measurements (e) to predict Y from X as shown in (f).

Figure 4.3 shows the predicted mass for sample one as a function of pixelvalues x and y, additionally it shows one rectangle used to sum up the single rectangles with the different number of layers and subsequently the number of dispensed drops.

This summation of the predicted value on the entire rectangle displayed as a function of layers is shown in figure 4.4 (a),(b) and (c). Panel (d) shows the escalation curve as a function of the number of drops per field since sample 1 was printed with 1 drop/spot and sample 2 at 5 drops/spot. The printer data added to the first three panels is calculated by taking the mean drop volume (measured before and after the

## 4. Results

Table 4.2.: HPLC results for the three printed test samples. Plots for these values can be seen in figure 4.1. Each layer consists of  $17 \times 17$  spots as shown in figure 3.10, for samples 1 and 3 one drop was dispensed on each spot, for sample 2 five drops were dispensed on each spot, respectively .

sample name	no. of layers	API ( $\mu\text{g}$ )
sample 1	11	372.02
	15	386.16
	20	486.49
	25	544.58
	30	683.84
sample 2	10	1249.10
	15	1771.22
	20	2335.47
	25	2967.00
	30	3729.48
sample 3	10	240.92
	15	363.75
	20	492.49
	25	650.84
	30	738.17

run) and multiplying the number of drops (289 or 1445 per field), then calculating the mass using (2.10).

For each panel the deviation between the prediction and HPLC analysis is calculated.

In the attempt of improving the prediction the measurement data (figure 4.2(e)) is decreased in size as to not include spots without HPLC result i.e. spots of 30, 20, and 10 layers. Keeping everything else the same the explained variance is improved and raised to 69.74%. The variation in the comparison of fitted and estimated mass still shows a lot of variance but somewhat better.

Figure 4.6 shows the comparison of the HPLC data and the prediction for sample 1 and 2, and the comparison of sample 3 prediction with its HPLC and Printer result. Quantitative results for this model can be seen in the appendix table B.2.

Another refinement consisted of doing an inverse calibration as described by Kessler on page 103. [34] Herein the prediction of an earlier model is used as the response matrix to create a new model. This means for a new model panel (f) of figure 4.2 is used as response matrix for a new model. Based on the appearance of this backprojection it can be assumed that the prediction is accurate. Furthermore, the total mass as predicted differs only  $2.3 \mu\text{g}$  from the HPLC result. When approximating the first spot to calculate its predicted mass the difference between the HPLC value and the predicted is only  $3.5 \mu\text{g}$  and only  $2 \mu\text{g}$  using the values within the ROI of the first modelled spot (figure 3.16). Based on the appearance and these preliminary

## 4. Results

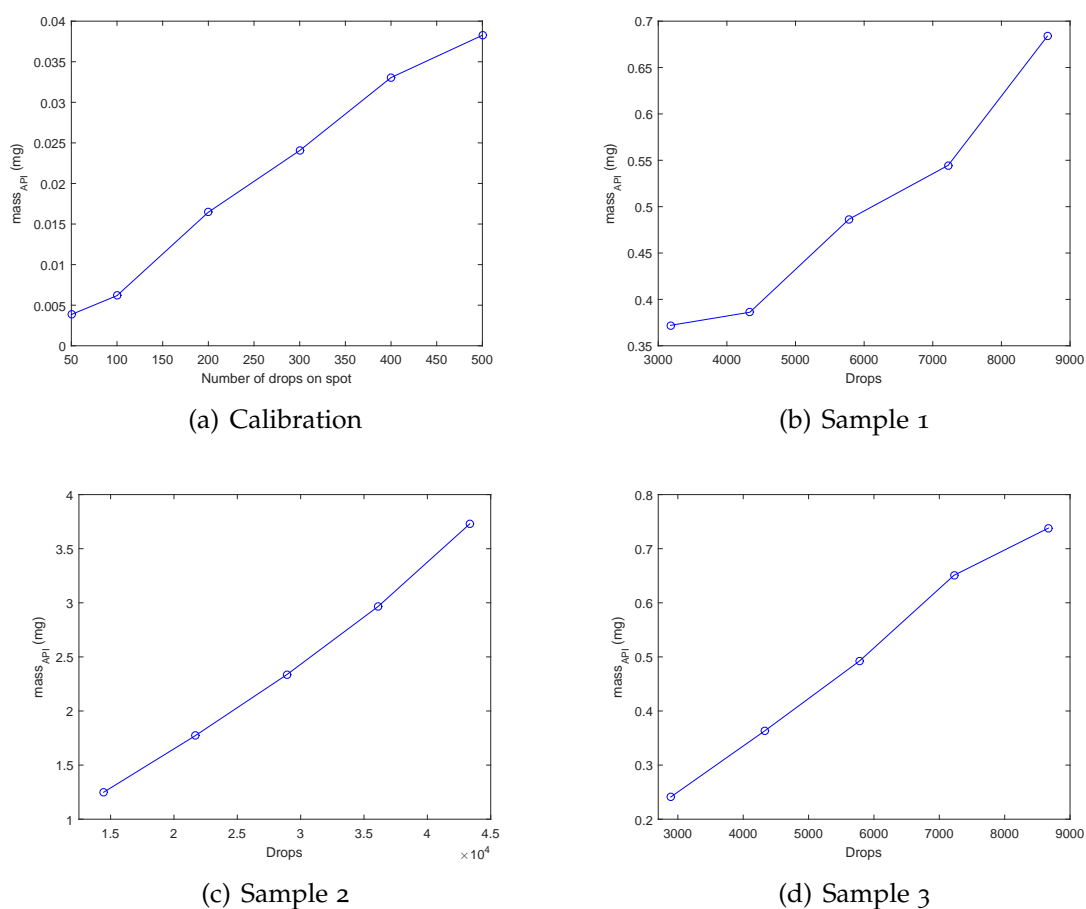
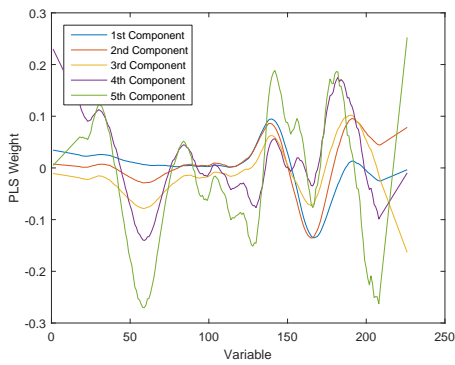


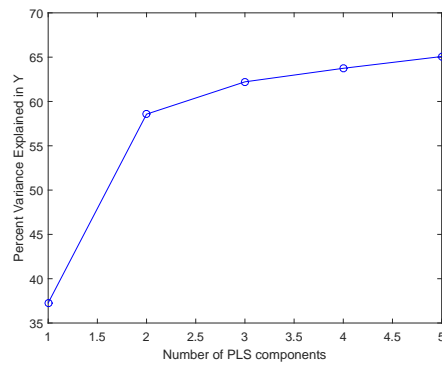
Fig. 4.1.: HPLC results plotted as a function of the number of drops on the spot/field.

values, the prediction was assumed to be better than the response matrix modelled beforehand, and therefore used to create a new model. This, unsurprisingly brought a significant improvement in the explained variance, and the comparison between fitted and the reference mass. At the same time though it also showed a big improvement when applying this model to the test samples, that were not used for the model creation. The plots comparing the printer, HPLC and prediction data are shown in figure 4.7.

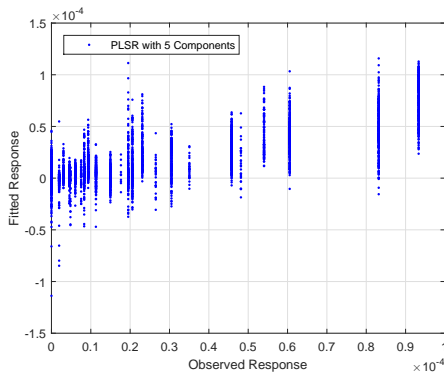
## 4. Results



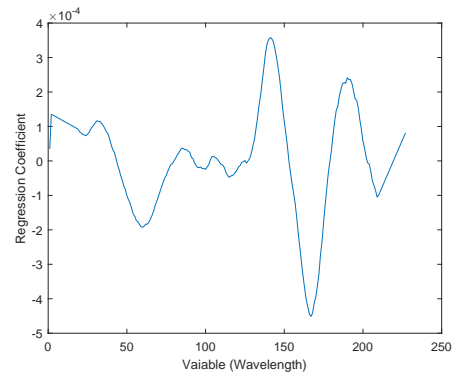
(a) PLS Weights for the 5 PLS components. Note the similarity to the absorbance spectra by comparing with Metformin in figure 3.1



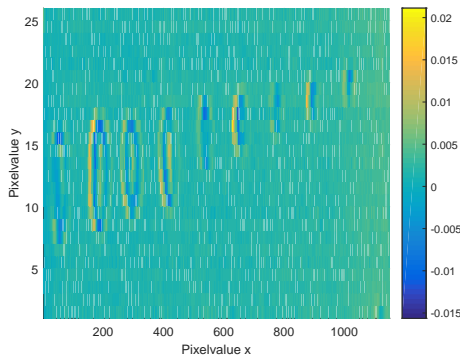
(b) Explained variance in response matrix.



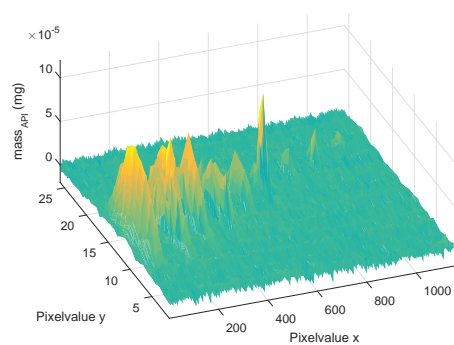
(c) Fitted vs. observed response. In the ideal case this would be a straight line with slope 1. The elongated point clouds result from assigning the same observed value to several measured pixels (in the response matrix), that however, lead to different predicted values.



(d) Regression coefficients (beta). Each wavelength variable has a certain weight (amplitude) that is applied to measurements to predict the concentration from them.



(e) Normalized absorbance at 1515 nm



(f) Backprojection: the spots are clearly visible, with higher concentration in their center.

Fig. 4.2.: First model results modelling spots of 500, 400, 300, 200, 100 and 50 spots and interpreting the results.

## 4. Results

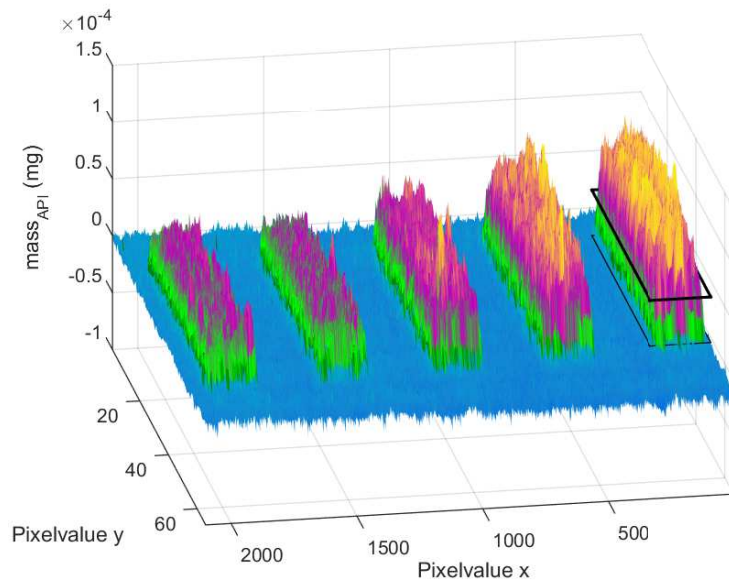
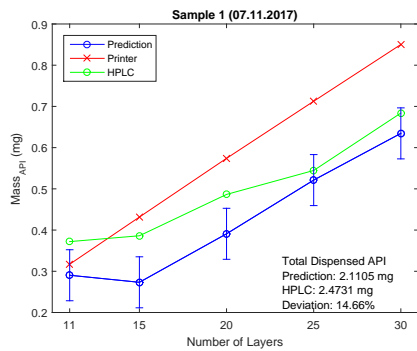
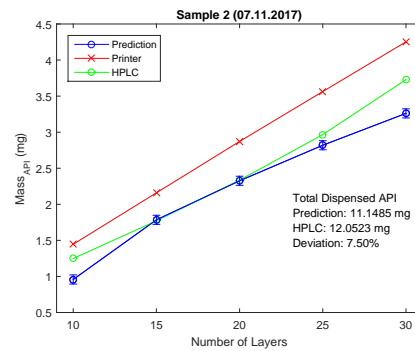


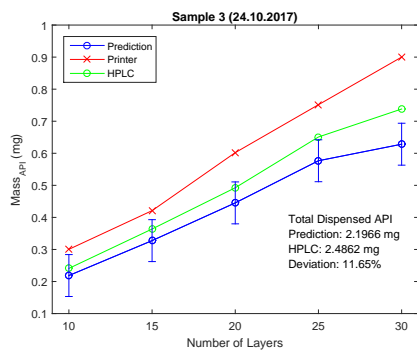
Fig. 4.3.: Prediction of sample 1, rectangle used to sum up total mass on testing surface added for sample with 30 layers.



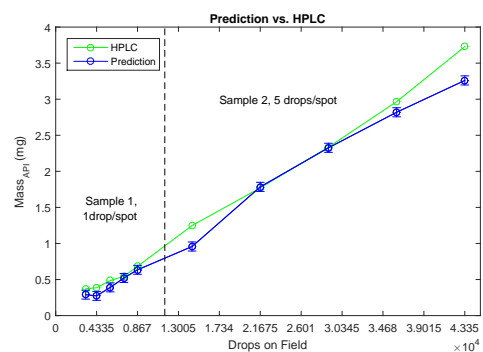
(a) Sample 1 mass as a function of the number of layers



(b) Sample 2 mass as a function of the number of layers



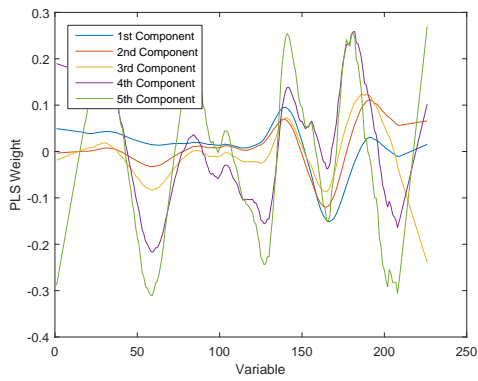
(c) Sample 3 mass as a function of number of layers



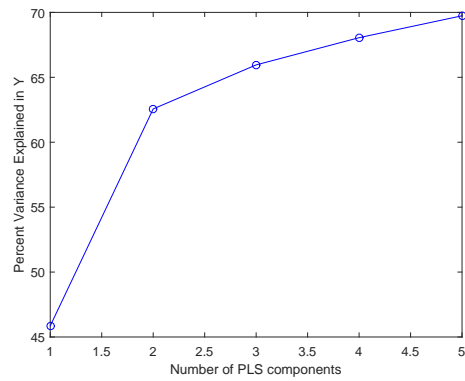
(d) Sample 1 and 2 mass as a function of the number of drops

Fig. 4.4.: Test one comparison with theoretical printer data, HPLC and predicted data.

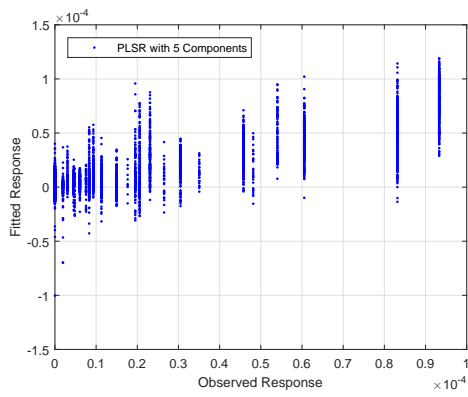
## 4. Results



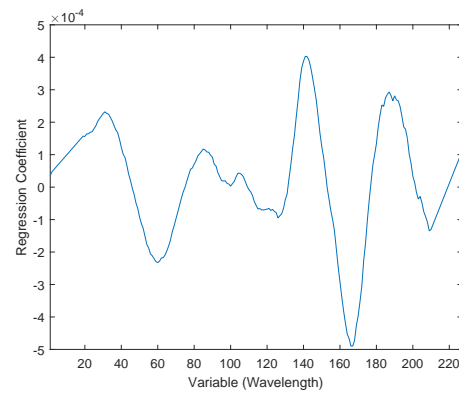
(a) PLS Weights components



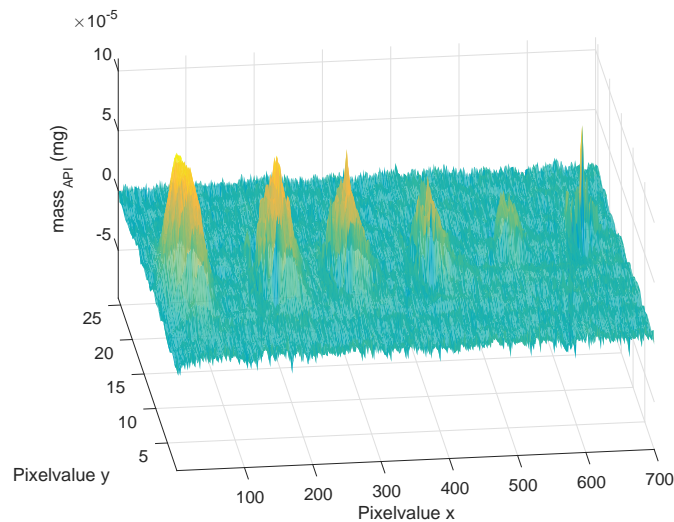
(b) Explained Variance in response matrix



(c) Fitted vs observed masses, still too many pixels are assigned the same value, where the prediction, the fit, reveals different concentrations.



(d) Regression Coefficients

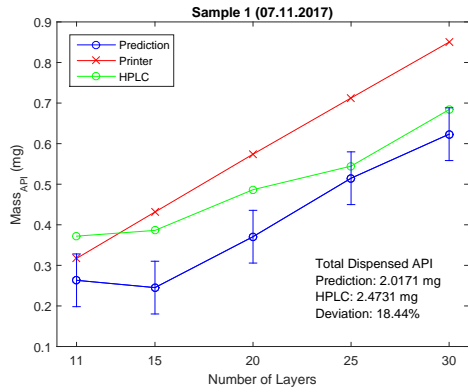


(e) Backprojection

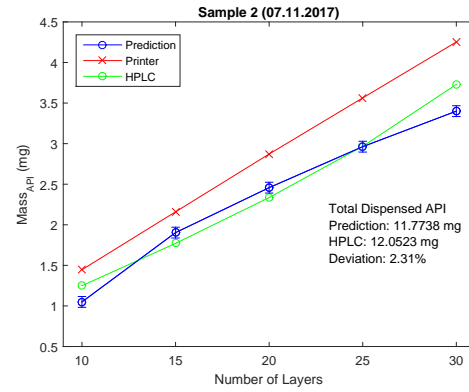
Fig. 4.5.: Modified model from the first model (figure 4.2) excluding the spots without HPLC data.



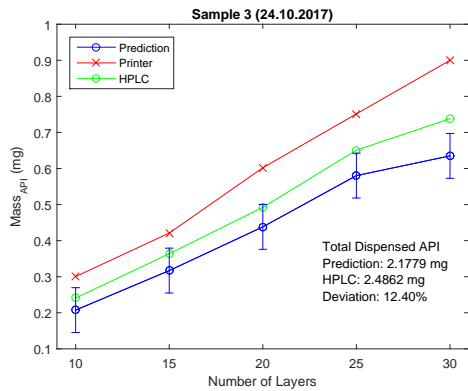
## 4. Results



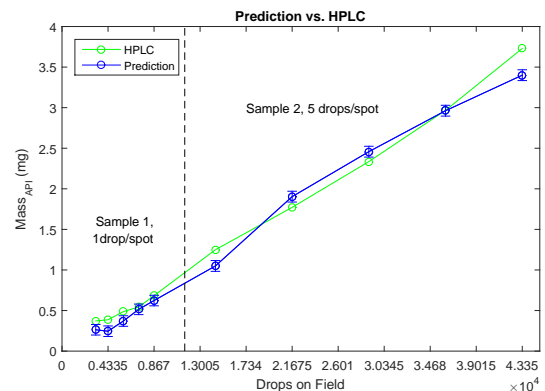
(a) Sample 1 mass as a function of the number of layers



(b) Sample 2 mass as a function of the number of layers



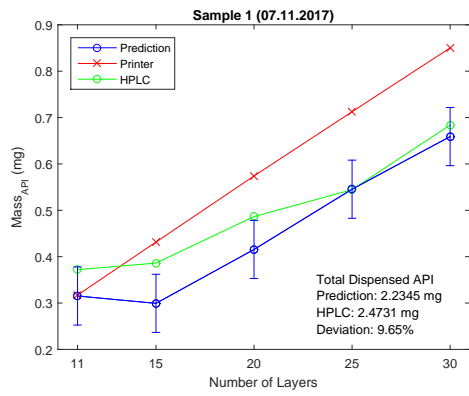
(c) Sample 3 as a function of the number of layers



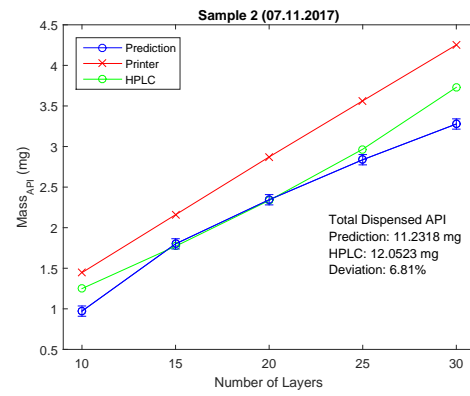
(d) Sample 1 and 2 as a function of the number of drops dispensed

Fig. 4.6.: Results for test without spots 30,20, 10 drops.

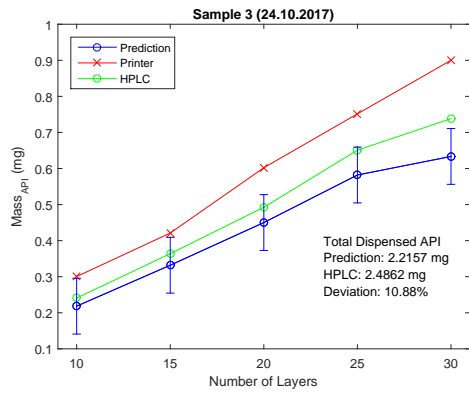
## 4. Results



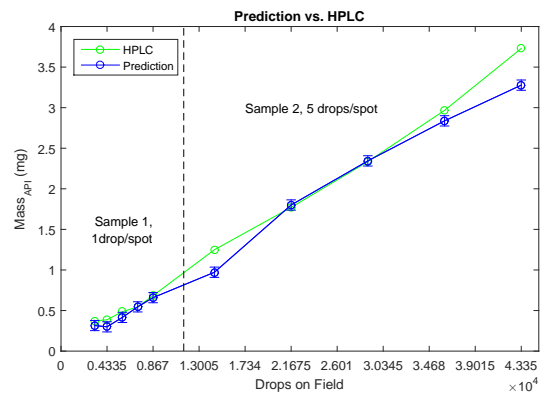
(a) Sample 1



(b) Sample 2



(c) Sample 3



(d) Sample 1 and 2 combined

Fig. 4.7.: Results when using figure 4.4(f) as response matrix.

#### 4. Results

Through further adaptations the final result is shown in figure 4.8. For this model the SNV is applied using a wavelength range of 1200-1600 nm, and a window size for the Savitzky-Golay filter of 35. The response matrix checking each modelled spot for the right mass (comparing it to HPLC), the difference of the modelled and expected mass on the calibration sample is .259 ng. Then the backprojection of the first step is used to calculate weights and the regression coefficients used for the prediction by giving the backprojected data as response matrix after setting negative values to zero.

Figure 4.9 shows 26 vectors (pixelvalues y of figure 4.8) values along the x values axis, as a function of the predicted mass. Figure 4.10 shows the results when applying the regression coefficients depicted in figure 4.8(d).

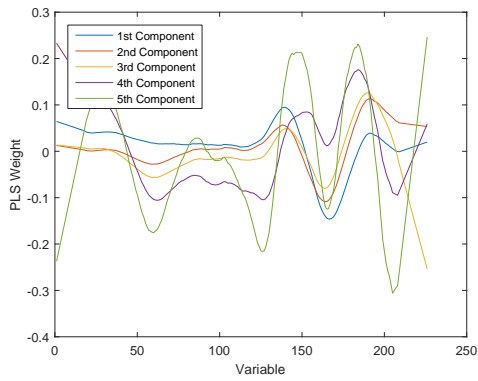
Figure 4.11 shows the rectangles used to calculate the values shown in figure 4.10. This figure shows the surface plot for the prediction, where the amplitude (color) represents the mass of API on the pixel, exemplary for sample 1.

Table 4.3 shows the quantitative values for figures 4.10.

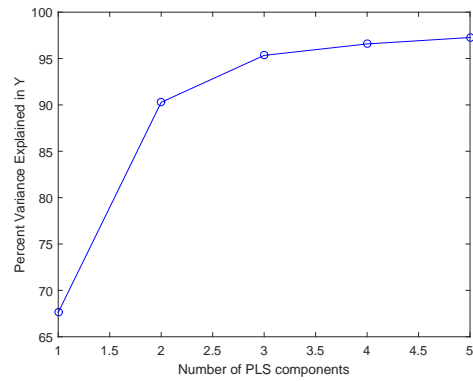
Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	320.61	13.82
1.15	431.53	386.16	308.95	19.99
1.20	574.02	486.49	432.57	11.08
1.25	712.54	544.58	577.26	-6.00
1.30	850.20	683.84	684.57	-0.11
2.10	1446.08	1249.10	1104.34	11.59
2.15	2159.63	1771.22	1959.43	-10.63
2.20	2871.92	2335.47	2515.15	-7.69
2.25	3563.44	2967.00	3019.55	-1.77
2.30	4252.42	3729.48	3450.72	7.47
3.10	300.42	240.92	205.71	14.61
3.15	421.05	363.75	325.44	10.53
3.20	602.10	492.49	449.85	8.66
3.25	751.49	650.84	594.99	8.58
3.30	900.09	738.17	652.47	11.61

Table 4.3.: Quantitative results shown for figure 4.10

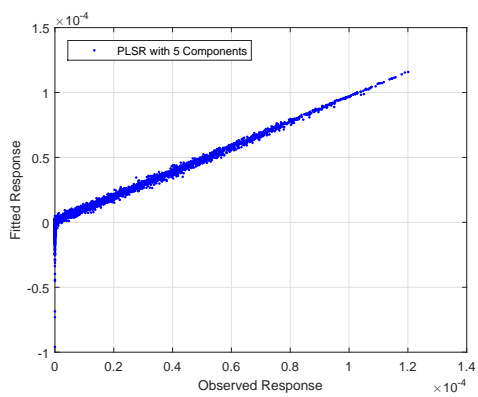
## 4. Results



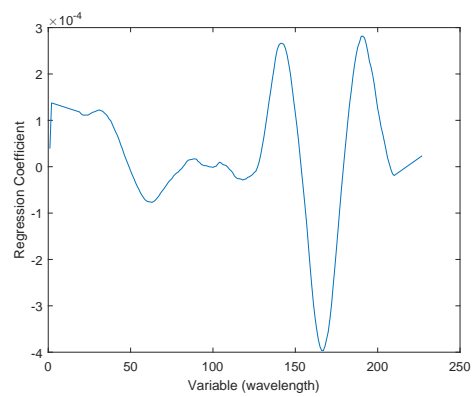
(a) PLS Weights for 5 components



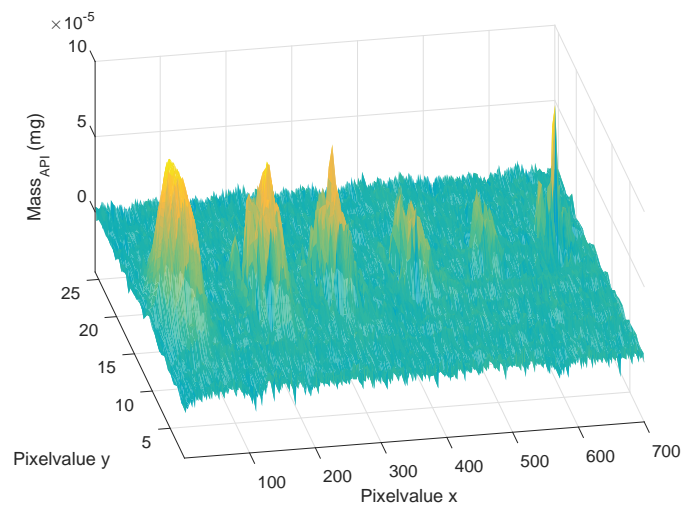
(b) Explained variance within Y



(c) Fitted vs. observed response: Due to a more refined response matrix (based on the backprojection) fit and assigned value overlap in most cases with only a small variation. Negative values within the backprojection, the observed response are set to zero, since it is not realistic to have a negative mass of API.



(d) Regression Coefficient



(e) Backprojection

Fig. 4.8.: Model after further adaptation, using a first prediction as the new response matrix but setting negative values to zero.

## 4. Results

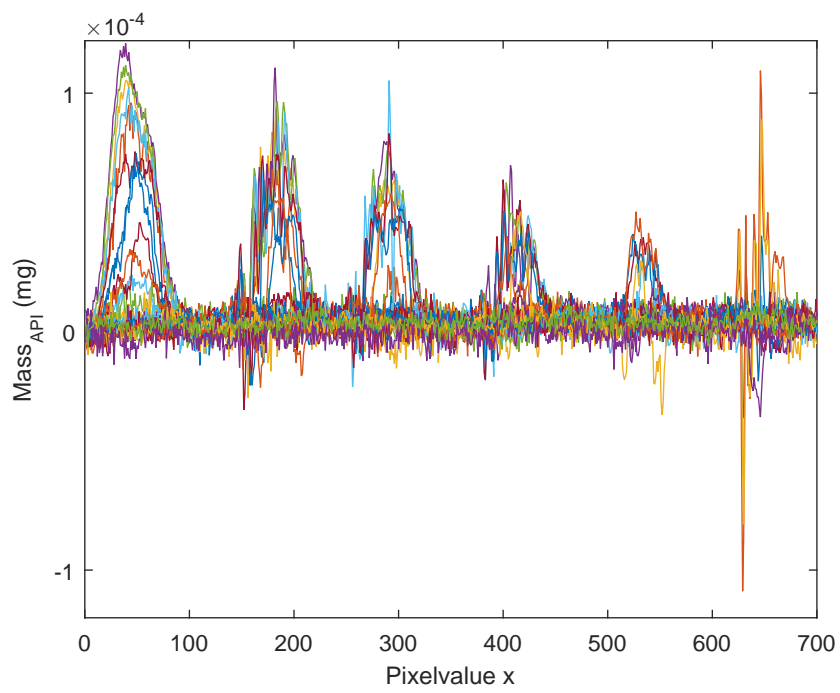


Fig. 4.9.: The 26 pixelvalues  $y$  plotted along the  $x$ -axis to show the prevalence of negative values.

## 4. Results

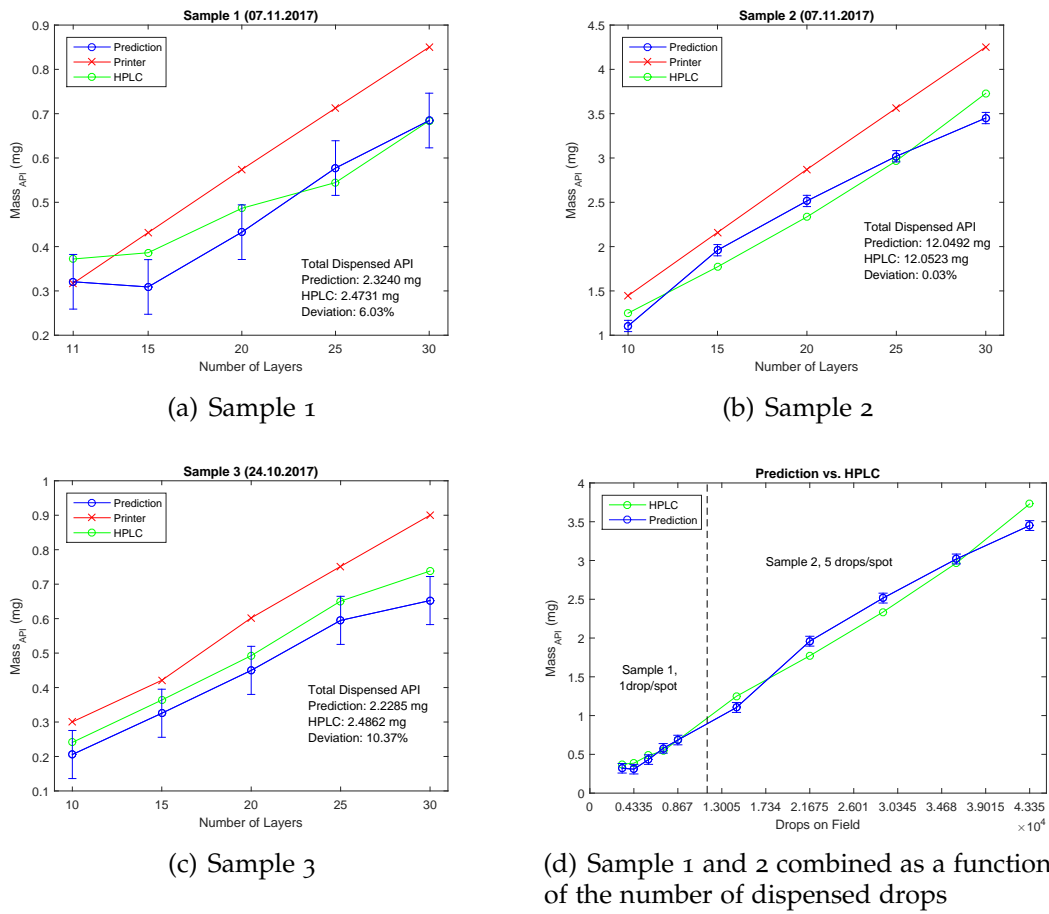


Fig. 4.10.: Final result comparing the predicted and HPLC data applying the regression coefficients in 4.8.

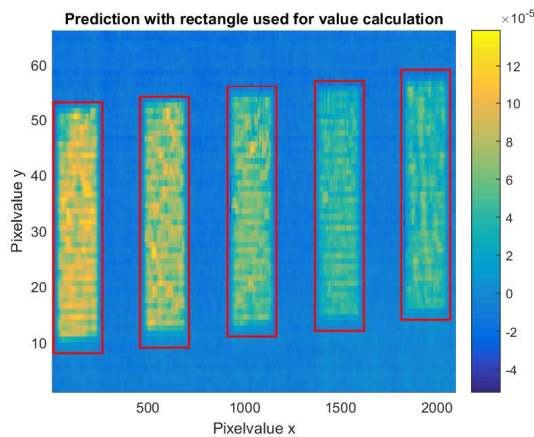


Fig. 4.11.: Sample 1 surfaceplot of prediction, with added ROIs of summation used for quantification and comparison with HPLC. The colorbar represents mass of API in mg.

## 5. Discussion

Considering figures 3.3, 3.4 and 3.5 in addition to metformin hydrochloride the calibration and following quantification of test samples could have been done for several combinations of substrates and ink:

- HGC + sodium picosulfate
- HGC+TiO<sub>2</sub> + sodium picosulfate
- HPMC + sodium picosulfate
- HGC + metformin hydrochloride

But only one combination was required for the first feasibility test, due to time limitations.

For sodium picosulfate and metformin hydrochloride listerine was not measured since the film disintegrated during the drying period. Presumably the water dissolved the substrate in the first printing tests, listerine was therefore deemed inappropriate for further testing and disregarded. The ethanol in the carvedilol ink did not have the same effect, but the drops dissolved too much, making it almost impossible to distinguish drops.

Surprisingly the carvedilol spots could not be identified on the HPMC film, one reason could be the reduced concentration another one swelling of the substrate.

Icing and starch wafers are problematic for all three inks, this might be due to their tendency of absorbing liquids easily. Then the ink is absorbed and API is distributed so much, that drops are no longer visible. Also both these substrates show a very low absorbance, therefore the light might be reflected from the substrate's molecules before detecting API.

Based on the absorption spectrum of EVA it was expected to be a suitable substrate, having a very constant absorption along the entire measured spectrum. The printed calibration curves showed the contrary though. In the 2D image (3.4 and 3.5) none of the drops are visible, and the substrate does not show a constant signal at 1539 nm and 1515 nm, respectively. One possible explanation is that the samples were not completely flat on the blackened aluminum, thus changing the light's preferred reflection direction significantly.

As mentioned in section 3.4 all measurements were taken using the entirety of the chip, the 320 pixels in the y direction, then choosing the ROI by examining the data set and cutting it accordingly. The process is shown for the calibration in figure 3.16. This step has been done for all measurements, since the printed samples are smaller than the 36 mm range. When the setup is used as a PAT it would be beneficial to reduce the size of the measured line in the push broom spectrometer. This can be done in HeliosViewer software. Thus, reducing the amount of data and computing power and shortening the process before beginning the actual working process.

## 5. Discussion

Moreover, because it can be hard to find the actual printed samples within the 2D displayed data due to its small size (as can be seen in figure 3.14).

When examining the values for the calibration in figure 4.1 and table 4.2 it is seen that the presence of API increases in an almost linear pattern with some variation. While with 100 drops the content is lower than would be expected the 200, 300, and 400 spots have a higher API content than a linear fit would predict. Figure 4.1(a) with an added a fit and the corresponding residuals can be seen figure A.1(a).

More surprising is the course within the samples, especially in sample 1 and 3. Here, the doubling in drops (15 layers to 30 layers) does not yield a doubled API content. Using a linear fit, the residual is smallest for the third data point (i.e. 20 layers) the neighboring data-points seem to have a significantly lower API content than would be expected (fits shown in figure A.1(b) and A.1(c))

Due to the mostly linear course in all four panels of figure 4.1 it can be assumed that sample 1 was degraded in some way, for example by scratching off some API, in retrospect it is not possible to determine at what point in the processes this might have happened.

For a first judgement of the model a backprojection is done, if here the result does not overlap with the expectation of seeing the same calibration curve, it can be assumed, that something went wrong in the regression process.

When examining the first backprojection (applying the first regression coefficients (figure 4.2(d)) to the measurement (e)) the result is plausible (figure 4.8(f)). The spots with a higher presence of metformin hydrochloride are higher and bigger than the ones with lower number of drops. Even though the smaller spots with 30, 20, and 10 drops are not represented in the response matrix (see figure 3.16) they are detected in the backprojection since metformin hydrochloride is detected at a low rate.

The fact, that these spots are not noted in the response matrix is one factor why the explained variance is rather low at only about 65% and why in figure (c) the fitted values show a high scattering.

One error, however, can be assumed at pixelvalue  $x$  about 600, where the predicted mass has its maximum, it is not likely, that in reality metformin has accumulated this much at this one spot.

When applying this model to the test prints higher amplitudes, therefore higher presence of API is detected correctly as visible in figure 4.3. For the quantitative result all values within a rectangle are added and then compared to the HPLC and estimated printer data result (figure 4.4), an example for such a ROI is shown in 4.3. When comparing printer data and HPLC it is easily visible, that the estimation using the drop volume and ink concentration is not a viable method of predicting printed API dosage. This might be due to the system not printing every drop as planned clogging for a short time and then unclogging through pressure build up, or clogging part of the nozzle. Another reason could be that through the filtering of the ink to reduce the risk of clogging during printing the API concentration is lowered to some extend, leading to the overprediction of API content.

The assumption, that sample 1 was degraded lowering API content especially on the fields of 15 and 25 layers is further supported. Before the NIR measurement



## 5. Discussion

some lint was stuck on the 15 layers field of sample 1, it was removed. During the careful removing it can be assumed that some API stuck to the lint reducing the perceived printed API content. For the data point of 25 layers on sample 1 it can be assumed, that some degradation happened during the storing period between the two measurements or in the sample preparation process for the HPLC.

The prediction for sample 3 follows the line of HPLC well, but divergence increases with an increasing number of drops.

When creating an escalation curve, by combining sample 1 and 2 (printed in one go) and plotting it as a function of drops. In the higher section (sample 2) prediction accuracy seems to be lower even though overall the deviation between printed and predicted mass is lower, this stems from the fact, that the underestimation tends to go into an overestimation at 15 and 20 layers. It is not clear if the first data point of sample 2 is the outlier or data points two and three, respectively.

The quantitative values (appendix) show the deviation between HPLC and prediction are in a wide range and many in a range above 15%.

This first result is promising but needs improvement.

Comparing panel (c) of both figure 4.2 and 4.5 the changes seem to have not made a major difference, comparing panels (d) though, it can be seen that the coefficients changed. The maximum increases, while also the last few variables increased- getting a more significant role in the prediction. Examining the backprojection the prevalence of negative values is surprising. Moreover, the 50 drop spot shows most certainly an error, with one pixel showing an extreme outlier.

When applying this new model prediction accuracy is not improved by much, still the curve of the prediction follows actual content but underestimates the mass.

After various changes with varying degrees of improvements of prediction a final model was created.

For the final model the regression is done twice, once with the same technique as described above with the created response matrix, and a second time with the predicted backprojected data, as described in chapter 4.

The model could probably be further improved by reducing the error in the 50 drop-spot.

Every model shows a surprisingly high number of negative values both in the back-projection as well as the test prediction, figure 4.9 shows these for the backprojection. The greatest problems exist at the spot's borders and for the 50 drop spot.

The results of applying the model shown in figure 4.8 are shown in figure 4.10 and table 4.3. Here all values deviate from the HPLC result only within 15% except for the second data point of sample 1. This data point shows the most deviation in all variations of the model. This could be due to some lint still stuck to the sample, even though not visible by eye or due to the disturbance created through removing the piece.

Overall it can be said, that the model works better in smaller concentrations, therefore

## 5. Discussion

for sample 1 and 3. This could be due to the fact, that the calibration was done with a lower API content. While the maximum predicted mass per pixel in the calibration is just over  $0.1 \mu\text{g}$ , all pixels with dispensed API in sample 3 have a higher dosage, with the maximum at over  $0.45 \mu\text{g}$ . This is not the case for sample 1, where its maximal predicted value is at  $0.14 \mu\text{g}$ , the same goes for sample 3. Additionally, a lower API content could allow for better light access to lower layers and thus better information on overall absorption.

To improve the prediction at higher dosages it may be necessary to move the calibration range to higher dosages as well. One risk of increasing the dispensed drops in one go is the deterioration of the substrate during the printing process. The substrate could swell up or the spot geometry changes. Swelling can appear due to the absorption of water in a short amount of time. The geometry could change due to breaking surface tension of the drop thus increases the spot's in area enough to combine with other spots (avoidable by increasing spot pitch) or that the per pixel API mass is decreased significantly once more.

To avoid this, one idea could be to avoid using single spots as calibration, but square or rectangle shape with either a different number of layers (similarly to our test samples) or a different number of drops.

Another improvement could be the reduction of negative values within the "empty region" of the prediction, and a reduction of noise.

As can be seen in figure 4.11 it can be hard to find the right dimensions to set the ROI to. While the width and height are easily determinable using measuring frequency and the linear stage's speed and the resolution in y direction, respectively, it can be hard to find the right origin for this rectangle. Relief for this problem might be brought with a measurement setup specific for this use, which once setup is not moved or changed, with this the measured area is more adapted and measurement timing more perfected.

One factor that might influence the results, but was not considered in this work is the drying period between printing the layers. This might influence how the API is deposited within the fields/ spots. The drying time between each printing cycle decreased with a decreasing number of fields to print.

### 5.1. Conclusion

In conclusion it can be said that NIR spectroscopy is a promising method for quality control of printed medicine, giving results within 15% deviation to the destructive

## 5. Discussion

reference method of HPLC, although improvements must be made for the application within the pharma - industry. A first improvement could be brought on by a permanent NIR measurement setup used for this purpose exclusively, allowing for the same light situation and geometry in every measurement. Furthermore, the drying effect during the printing process needs to be examined. Also the minimum time needed between printing and the NIR measurement for all solvent effects need to be investigated.

It would be beneficial to find a compound, that can be added to every API-solvent mixture without interfering with it's affects on the body. This compound would serve as a tracer to be able to calculate API content based on the tracer content. This means the tracer needs to be completely soluble in the solvent, and mixed in the ink homogeneously.

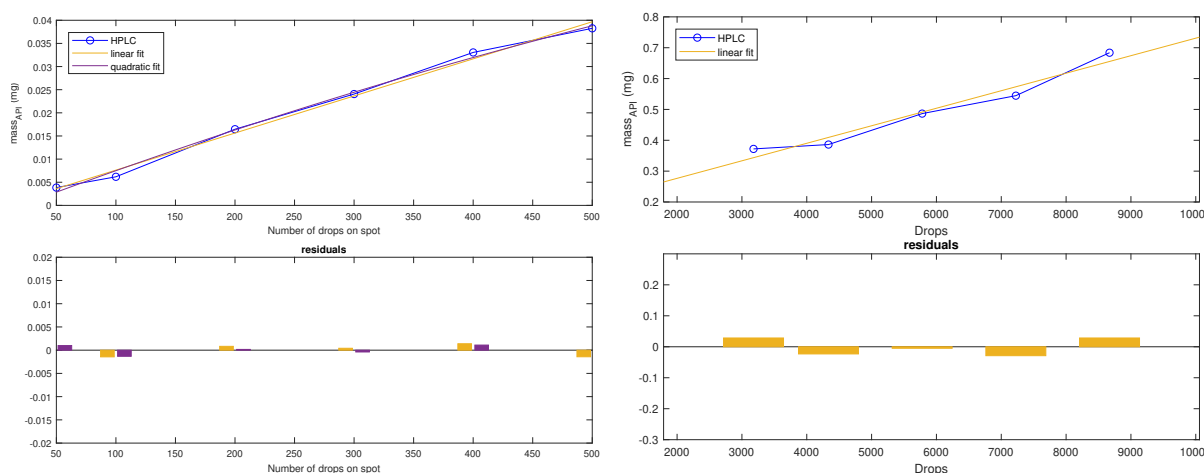
Herein a possible effect based on different substrates needs to be considered and reviewed.

Ideally NIR spectroscopy could be used as a in-line process analytical tool, measuring the dispensed API content after each printing step and adjusting and the, i.e. number of layers and drops accordingly. A processing line could be developed where the ink is printed, dried and measured, if not enough API has been detected the film is brought back to the printer and another printing step is added, repeating the process until the desired API dosage has been positioned on the film.

# Appendix A.

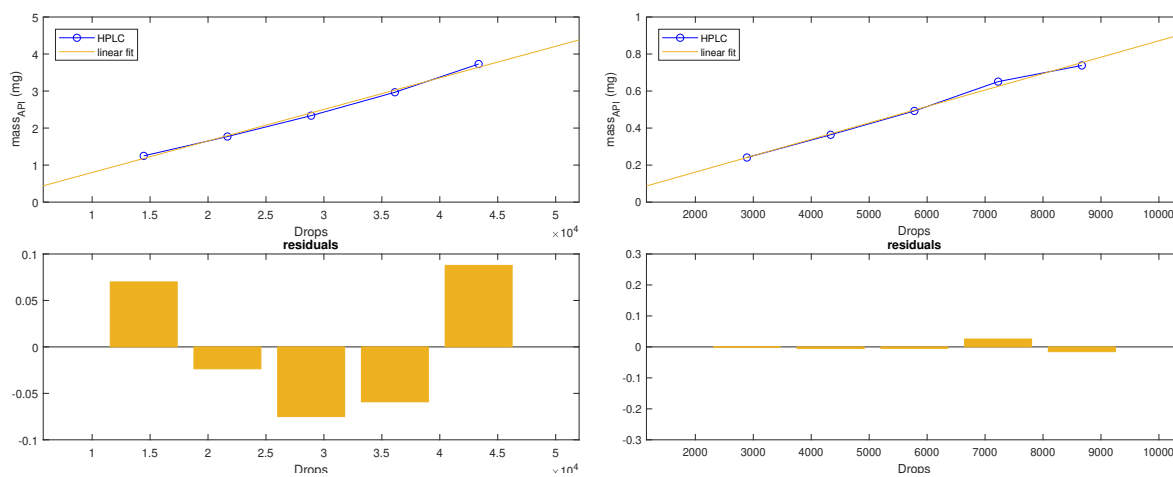
## Fits for HPLC Results

Figures A.1(a)- A.1(d) show the HPLC results for all samples, including the calibration sample with a linear fit (quadratic fit added for the calibration).



(a) Calibration Sample

(b) Sample 1



(c) Sample 2

(d) Sample 3

Fig. A.1.: HPLC results for all samples and their linear fit and residuals created using the MATLAB<sup>®</sup> basic fitting tool

# Appendix B.

## Tables for created models

Tables B.1 through B.14 show the quantitative results for some of the created models. For the first test (Table B.1) SNV range was chosen to be variable 70 - 200 and a window size of 35 for the SG filter.

Name	Printer $\mu\text{g}$	HPLC $\mu\text{g}$	Prediction $\mu\text{g}$	Deviation (HPLC & Prediction) %
1.11	317.14	372.02	290.41	21.94
1.15	431.53	386.16	273.28	29.23
1.20	574.02	486.49	390.94	19.64
1.25	712.54	544.58	521.25	4.28
1.30	850.20	683.84	634.59	7.20
2.10	1446.08	1249.10	957.93	23.31
2.15	2159.63	1771.22	1784.68	-0.76
2.20	2871.92	2335.47	2326.28	0.39
2.25	3563.44	2967.00	2819.15	4.98
2.30	4252.42	3729.48	3260.51	12.57
3.10	300.42	240.92	218.66	9.24
3.15	421.05	363.75	327.57	9.95
3.20	602.10	492.49	445.30	9.58
3.25	751.49	650.84	576.74	11.39
3.30	900.09	738.17	628.34	14.88

Table B.1.: API mass according to the estimated drop volume, HPLC and the prediction, and the deviation between the HPLC results and the prediction for test 1. SNV range: 70-200; SG filter: polynomial order 2, window size 35

Table B.3 shows the quantitative results for figure 4.7.

model 16:

For table B.14 it was tried to delete the 50 drop spot after the first backprojection (where an error was).

Appendix B. Tables for created models

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	263.18	29.26
1.15	431.53	386.16	245.18	36.51
1.20	574.02	486.49	370.66	23.81
1.25	712.54	544.58	514.76	5.48
1.30	850.20	683.84	623.27	8.86
2.10	1446.08	1249.10	1049.52	15.98
2.15	2159.63	1771.22	1903.51	-7.47
2.20	2871.92	2335.47	2457.64	-5.23
2.25	3563.44	2967.00	2962.71	0.14
2.30	4252.42	3729.48	3400.44	8.82
3.10	300.42	240.92	207.41	13.91
3.15	421.05	363.75	316.97	12.86
3.20	602.10	492.49	438.18	11.03
3.25	751.49	650.84	580.35	10.83
3.30	900.09	738.17	635.02	13.97

Table B.2.: Results of test 2, here SNV and SG parameters and ellipses form and amplitude are the same as for test 1 but not the entire measurement was used for the calibration since spots with fewer than 50 drops where not modelled, they were excluded from the calibration.

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	315.19	15.28
1.15	431.53	386.16	299.30	22.49
1.20	574.02	486.49	415.52	14.59
1.25	712.54	544.58	545.47	-0.16
1.30	850.20	683.84	658.98	3.64
2.10	1446.08	1249.10	971.91	22.19
2.15	2159.63	1771.22	1800.49	-1.65
2.20	2871.92	2335.47	2344.26	-0.38
2.25	3563.44	2967.00	2837.50	4.36
2.30	4252.42	3729.48	3277.65	12.11
3.10	300.42	240.92	218.34	9.37
3.15	421.05	363.75	331.84	8.77
3.20	602.10	492.49	450.14	8.60
3.25	751.49	650.84	581.98	10.58
3.30	900.09	738.17	633.43	14.19

Table B.3.: Results of test 3, here the backprojection of test 1 was used as the response matrix (inverse calibration).

Appendix B. Tables for created models

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	263.18	29.26
1.15	431.53	386.16	245.18	36.51
1.20	574.02	486.49	370.66	23.81
1.25	712.54	544.58	514.76	5.48
1.30	850.20	683.84	623.27	8.86
2.10	1446.08	1249.10	1049.52	15.98
2.15	2159.63	1771.22	1903.51	-7.47
2.20	2871.92	2335.47	2457.64	-5.23
2.25	3563.44	2967.00	2962.71	0.14
2.30	4252.42	3729.48	3400.44	8.82
3.10	300.42	240.92	207.41	13.91
3.15	421.05	363.75	316.97	12.86
3.20	602.10	492.49	438.18	11.03
3.25	751.49	650.84	580.35	10.83
3.30	900.09	738.17	635.02	13.97

Table B.4.: Results of test 4, here the spots with fewer than 50 drops were excluded but only after the preprocessing step.

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	285.34	23.30
1.15	431.53	386.16	272.54	29.42
1.20	574.02	486.49	397.72	18.25
1.25	712.54	544.58	541.42	0.58
1.30	850.20	683.84	652.59	4.57
2.10	1446.08	1249.10	1058.28	15.28
2.15	2159.63	1771.22	1926.33	-8.76
2.20	2871.92	2335.47	2489.04	-6.58
2.25	3563.44	2967.00	2999.28	-1.09
2.30	4252.42	3729.48	3442.11	7.71
3.10	300.42	240.92	224.55	6.79
3.15	421.05	363.75	337.41	7.24
3.20	602.10	492.49	459.93	6.61
3.25	751.49	650.84	602.60	7.41
3.30	900.09	738.17	654.17	11.38

Table B.5.: Results of test 5: uses the backprojection of test 4 as response matrix.

Appendix B. Tables for created models

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	308.31	17.12
1.15	431.53	386.16	293.03	24.12
1.20	574.02	486.49	403.63	17.03
1.25	712.54	544.58	527.35	3.16
1.30	850.20	683.84	631.00	7.73
2.10	1446.08	1249.10	932.27	25.36
2.15	2159.63	1771.22	1714.89	3.18
2.20	2871.92	2335.47	2226.31	4.67
2.25	3563.44	2967.00	2690.24	9.33
2.30	4252.42	3729.48	3102.35	16.82
3.10	300.42	240.92	218.30	9.39
3.15	421.05	363.75	322.24	11.41
3.20	602.10	492.49	433.61	11.95
3.25	751.49	650.84	559.10	14.10
3.30	900.09	738.17	607.81	17.66

Table B.6.: Results of test 6: uses results the backprojection of the first model with negative values set to zero

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	285.17	23.35
1.15	431.53	386.16	273.06	29.29
1.20	574.02	486.49	401.70	17.43
1.25	712.54	544.58	548.56	-0.73
1.30	850.20	683.84	662.46	3.13
2.10	1446.08	1249.10	1079.74	13.56
2.15	2159.63	1771.22	1960.85	-10.71
2.20	2871.92	2335.47	2531.03	-8.37
2.25	3563.44	2967.00	3049.61	-2.78
2.30	4252.42	3729.48	3500.12	6.15
3.10	300.42	240.92	217.03	9.92
3.15	421.05	363.75	332.91	8.48
3.20	602.10	492.49	457.95	7.01
3.25	751.49	650.84	602.72	7.39
3.30	900.09	738.17	655.35	11.22

Table B.7.: Results of test 7: uses backprojection based on test 2 as response matrix with negative values of the fit set to zero.



Appendix B. Tables for created models

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	256.20	31.13
1.15	431.53	386.16	238.81	38.16
1.20	574.02	486.49	361.18	25.76
1.25	712.54	544.58	501.03	8.00
1.30	850.20	683.84	604.76	11.56
2.10	1446.08	1249.10	1015.85	18.67
2.15	2159.63	1771.22	1843.36	-4.07
2.20	2871.92	2335.47	2378.62	-1.85
2.25	3563.44	2967.00	2866.77	3.38
2.30	4252.42	3729.48	3289.82	11.79
3.10	300.42	240.92	202.45	15.97
3.15	421.05	363.75	306.74	15.67
3.20	602.10	492.49	424.11	13.89
3.25	751.49	650.84	561.94	13.66
3.30	900.09	738.17	614.76	16.72

Table B.8.: Results of test 8: uses spots down to 50 drops but changes the values within the ellipses.

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	277.05	25.53
1.15	431.53	386.16	265.07	31.36
1.20	574.02	486.49	390.11	19.81
1.25	712.54	544.58	532.29	2.26
1.30	850.20	683.84	641.12	6.25
2.10	1446.08	1249.10	1043.58	16.45
2.15	2159.63	1771.22	1895.35	-7.01
2.20	2871.92	2335.47	2445.13	-4.70
2.25	3563.44	2967.00	2945.40	0.73
2.30	4252.42	3729.48	3379.86	9.37
3.10	300.42	240.92	210.97	12.43
3.15	421.05	363.75	321.32	11.66
3.20	602.10	492.49	442.12	10.23
3.25	751.49	650.84	582.12	10.56
3.30	900.09	738.17	632.84	14.27

Table B.9.: Results of test 9: uses the backprojection of test 8 as response matrix.

Appendix B. Tables for created models

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	282.32	24.11
1.15	431.53	386.16	265.58	31.23
1.20	574.02	486.49	379.71	21.95
1.25	712.54	544.58	505.53	7.17
1.30	850.20	683.84	614.19	10.18
2.10	1446.08	1249.10	923.50	26.07
2.15	2159.63	1771.22	1722.14	2.77
2.20	2871.92	2335.47	2244.29	3.90
2.25	3563.44	2967.00	2719.63	8.34
2.30	4252.42	3729.48	3145.61	15.66
3.10	300.42	240.92	213.04	11.57
3.15	421.05	363.75	316.85	12.89
3.20	602.10	492.49	430.50	12.59
3.25	751.49	650.84	557.35	14.36
3.30	900.09	738.17	607.02	17.77

Table B.10.: Results of test 10: combines test 1 (entire sample) with the adapted response matrix new values used for test 8.

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	305.94	17.76
1.15	431.53	386.16	290.13	24.87
1.20	574.02	486.49	402.35	17.30
1.25	712.54	544.58	527.40	3.15
1.30	850.20	683.84	636.07	6.99
2.10	1446.08	1249.10	935.20	25.13
2.15	2159.63	1771.22	1733.13	2.15
2.20	2871.92	2335.47	2256.12	3.40
2.25	3563.44	2967.00	2730.61	7.97
2.30	4252.42	3729.48	3154.18	15.43
3.10	300.42	240.92	211.89	12.05
3.15	421.05	363.75	320.15	11.99
3.20	602.10	492.49	433.99	11.88
3.25	751.49	650.84	560.83	13.83
3.30	900.09	738.17	610.23	17.33

Table B.11.: Results of test 11: used backprojection of test 12 as response matrix.

Appendix B. Tables for created models

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	257.00	30.92
1.15	431.53	386.16	239.54	37.97
1.20	574.02	486.49	362.71	25.44
1.25	712.54	544.58	503.46	7.55
1.30	850.20	683.84	607.01	11.24
2.10	1446.08	1249.10	1021.17	18.25
2.15	2159.63	1771.22	1853.26	-4.63
2.20	2871.92	2335.47	2391.38	-2.39
2.25	3563.44	2967.00	2882.35	2.85
2.30	4252.42	3729.48	3307.48	11.32
3.10	300.42	240.92	202.17	16.08
3.15	421.05	363.75	306.47	15.75
3.20	602.10	492.49	424.38	13.83
3.25	751.49	650.84	563.15	13.47
3.30	900.09	738.17	616.29	16.51

Table B.12.: Results of test 12: expands SNV range to 70-210, further changes to the values within the modelled ellipses in the response matrix.

Name	Printer	HPLC	Prediction	Deviation (HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	306.00	17.75
1.15	431.53	386.16	295.40	23.50
1.20	574.02	486.49	420.40	13.59
1.25	712.54	544.58	566.55	-4.03
1.30	850.20	683.84	674.20	1.41
2.10	1446.08	1249.10	1104.47	11.58
2.15	2159.63	1771.22	1963.52	-10.86
2.20	2871.92	2335.47	2521.02	-7.95
2.25	3563.44	2967.00	3027.36	-2.03
2.30	4252.42	3729.48	3458.42	7.27
3.10	300.42	240.92	190.46	20.95
3.15	421.05	363.75	310.93	14.52
3.20	602.10	492.49	436.37	11.40
3.25	751.49	650.84	582.71	10.47
3.30	900.09	738.17	640.26	13.26

Table B.13.: Results of test 13: same as test 1 but uses more components for the regression

Appendix B. Tables for created models

Name	Printer	HPLC	Prediction	Deviation ( HPLC & Prediction)
	$\mu\text{g}$	$\mu\text{g}$	$\mu\text{g}$	%
1.11	317.14	372.02	306.79	17.54
1.15	431.53	386.16	296.24	23.29
1.20	574.02	486.49	421.19	13.42
1.25	712.54	544.58	567.45	-4.20
1.30	850.20	683.84	675.62	1.20
2.10	1446.08	1249.10	1105.78	11.47
2.15	2159.63	1771.22	1966.28	-11.01
2.20	2871.92	2335.47	2525.10	-8.12
2.25	3563.44	2967.00	3032.51	-2.21
2.30	4252.42	3729.48	3464.61	7.10
3.10	300.42	240.92	189.90	21.18
3.15	421.05	363.75	311.13	14.47
3.20	602.10	492.49	436.86	11.29
3.25	751.49	650.84	583.44	10.36
3.30	900.09	738.17	641.23	13.13

Table B.14.: Results of test 14: does not include spot with 50 drops, since errors in the backprojection are often in this spot.

Further models with small variations in the ellipses , preprocessing steps etc. were created, that are not listed here.

## Appendix C.

# Conversion Wavelength Variable and nm

According to EVKs documentation the 256 wavelengths measured by the system, shows a linear relationship with the wavelengths from 931 - 1702 nm, table C.1 shows the conversion between the variable (v) and the wavelength in nm ( $\lambda$ ).

V	$\lambda$	V	$\lambda$	V	$\lambda$	V	$\lambda$
1	931	2	934	3	937	4	940
5	943	6	946	7	949	8	952
9	955	10	958	11	961	12	964
13	967	14	970	15	973	16	976
17	979	18	982	19	985	20	988
21	991	22	995	23	998	24	1001
25	1004	26	1007	27	1010	28	1013
29	1016	30	1019	31	1022	32	1025
33	1028	34	1031	35	1034	36	1037
37	1040	38	1043	39	1046	40	1049
41	1052	42	1055	43	1058	44	1061
45	1064	46	1067	47	1070	48	1073
49	1076	50	1079	51	1082	52	1085
53	1088	54	1091	55	1094	56	1097
57	1100	58	1103	59	1106	60	1109
61	1112	62	1115	63	1118	64	1122
65	1125	66	1128	67	1131	68	1134
69	1137	70	1140	71	1143	72	1146
73	1149	74	1152	75	1155	76	1158
77	1161	78	1164	79	1167	80	1170
81	1173	82	1176	83	1179	84	1182
85	1185	86	1188	87	1191	88	1194
89	1197	90	1200	91	1203	92	1206
93	1209	94	1212	95	1215	96	1218
97	1221	98	1224	99	1227	100	1230
101	1233	102	1236	103	1239	104	1242

Appendix C. Conversion Wavelength Variable and nm

105	1245	106	1248	107	1252	108	1255
109	1258	110	1261	111	1264	112	1267
113	1270	114	1273	115	1276	116	1279
117	1282	118	1285	119	1288	120	1291
121	1294	122	1297	123	1300	124	1303
125	1306	126	1309	127	1312	128	1315
129	1318	130	1321	131	1324	132	1327
133	1330	134	1333	135	1336	136	1339
137	1342	138	1345	139	1348	140	1351
141	1354	142	1357	143	1360	144	1363
145	1366	146	1369	147	1372	148	1375
149	1378	150	1382	151	1385	152	1388
153	1391	154	1394	155	1397	156	1400
157	1403	158	1406	159	1409	160	1412
161	1415	162	1418	163	1421	164	1424
165	1427	166	1430	167	1433	168	1436
169	1439	170	1442	171	1445	172	1448
173	1451	174	1454	175	1457	176	1460
177	1463	178	1466	179	1469	180	1472
181	1475	182	1478	183	1481	184	1484
185	1487	186	1490	187	1493	188	1496
189	1499	190	1502	191	1505	192	1509
193	1512	194	1515	195	1518	196	1521
197	1524	198	1527	199	1530	200	1533
201	1536	202	1539	203	1542	204	1545
205	1548	206	1551	207	1554	208	1557
209	1560	210	1563	211	1566	212	1569
213	1572	214	1575	215	1578	216	1581
217	1584	218	1587	219	1590	220	1593
221	1596	222	1599	223	1602	224	1605
225	1608	226	1611	227	1614	228	1617
229	1620	230	1623	231	1626	232	1629
233	1632	234	1635	235	1639	236	1642
237	1645	238	1648	239	1651	240	1654
241	1657	242	1660	243	1663	244	1666
245	1669	246	1672	247	1675	248	1678
249	1681	250	1684	251	1687	252	1690
253	1693	254	1696	255	1699	256	1702

Table C.1.: Conversion of variable number (V) to nm ( $\lambda$ ) for wavelength

# List of Figures

1.1.	Schematic for a thermal and a piezoelectric printhead, as shown by Gerard Cummins. [13]	2
1.2.	Schematic displaying differences between in- on- and at-line processes, as described by [21].	4
1.3.	Morse curve: showing the relationship between energy and the internuclear distance $r$ in a diatomic molecule as shown in [26]. ZPE is the zero point energy, it is the lowest possible energy level a molecule can have. BDE is the bond dissociation energy.	5
1.4.	NIR functional groups absorption bands as a function of the wavelength.	6
1.5.	For the regression observations can be understood as two data swarms in two separate spaces, through the regression these are related to each other. In this example the data has been mean-centered, meaning the coordinate systems pass through the average point (dark gray) of each point swarm.	9
1.6.	Matrix relationships among the PLS computation. $1 * \bar{x}$ and $1 * \bar{y}$ are the averages along each column, that stem from the mean centering step. $\mathbf{T}$ and $\mathbf{U}$ are the X-and Y-scores (containing information about the observations and their similarities with respect to the given problem. The weights $\mathbf{W}^*$ and $\mathbf{C}$ contain information about how the variables combine to relay the relation between $\mathbf{X}$ and $\mathbf{Y}$ . $\mathbf{E}$ and $\mathbf{F}$ contain the variation in the data not contained in the model, called residuals.	10
2.1.	Sciencion software with adjusters for voltage, pulse length, LED Delay and frequency.	12
2.2.	Target setup menu with in Sciencion's software.	13
2.3.	Setup for NIR measurements. (1) Linear stage control, (2) power supply, (3) helios sensor, (4) linear stage, (5) fiber optic ring light guide, (6) Light source.	15
2.4.	Tools used to create white and black reference spectra, and the spectra averaged along the time dimension.	16
2.5.	Schematic representation of the measurement's data set.	17
2.6.	Measurement spectrum (reflection)(left panel) and calculated absorption (right panel) of a random pixel.	17
2.7.	USAF test chart used for the determination of spatial resolution	18
2.8.	Measurement	19
2.9.	Photo of pencil line and its NIR measurement shown at the mean wavelength.	20

## List of Figures

2.10.	Calculated absorption spectrum (left panel) and normalized absorption after SNV (right panel). . . . .	21
2.11.	Differences in filter result by varying frame size (left) and polynomial order (right), keeping the polynomial order constant at 2 and frame size constant at 11 points, respectively. . . . .	22
2.12.	Schematic of how data set is rearranged to reduce it from three to two dimensions. . . . .	23
2.13.	Pseudocode for the SIMPLS algorithm as published by De Jong [39]. . . . .	25
2.14.	Raman shift measurement for a drop of the metformin hydrochloride ink (see section 2.1.2) on HGC+TiO <sub>2</sub> . . . . .	26
3.1.	NIR- spectra for all substrates and APIs measured seperatly. . . . .	28
3.2.	Printing pattern schematic: 7 spots with increasing number of drops per spot ranging from 10 to 300 drops. . . . .	29
3.3.	9.9 mg Carvedilol in 10 ml ethanol printed on all substrates, ethanol is also used as system liquid for the printer. . . . .	29
3.4.	sodium picosulfate ink on all substrates except listerine since the water created holes. . . . .	30
3.5.	Metformin ink on all substrates except listerine displayed at 1515 nm. . . . .	30
3.6.	Calibration curve. . . . .	30
3.7.	Modelled Y matrix with constant concentration within spots from left to right 300,200,100,50,30,20 drops per spot each with corresponding concentration shown in 3.2 calculated with (3.1). . . . .	32
3.8.	300 drops spot taken taken using the Senterra microscope . . . . .	32
3.9.	Raman shift intensity at API peak (metformin). . . . .	33
3.10.	printing pattern for test model. . . . .	33
3.11.	Absorption displays at 1515 nm (no preprocessing) and photo of first test sample, 4 squares at 1,10,20 and 30 layers, each layer at 1 drop per spot. . . . .	34
3.12.	Measurement of drop volume by the scienion software. . . . .	35
3.13.	Sample (Testpattern 1 drop per spot, [11,15,20,25,30] layers) on the blackened aluminum, as it is then measured. . . . .	35
3.14.	<i>Left panel:</i> Extraction of ROI from entire measurement. <i>Right panel:</i> the extracted ROI containing the calibration spots. The colorbars represent absorption at 1515 nm. . . . .	36
3.15.	Effects of data preprocessing on spectra and normalized and derived absorption at 1515 nm. . . . .	36
3.16.	Preprocessed data displayed at 1515 nm (left, colorbar representing normalized absorption in the left panel of figure 3.15) and the created concentration matrix (response matrix) with color representing the mass of API in mg. A smaller ellipse than the actual drop is estimated in the center and rings are put around it with different amplitudes to emulate the drop of API towards the border. . . . .	37
4.1.	HPLC results plotted as a function of the number of drops on the spot/field. . . . .	41



## List of Figures

4.2.	First model results modelling spots of 500, 400, 300, 200, 100 and 50 spots and interpreting the results. . . . .	42
4.3.	Prediction of sample 1, rectangle used to sum up total mass on testing surface added for sample with 30 layers. . . . .	43
4.4.	Test one comparison with theoretical printer data, HPLC and predicted data. . . . .	43
4.5.	Modified model from the first model (figure 4.2) excluding the spots without HPLC data. . . . .	44
4.6.	Results for test without spots 30,20, 10 drops. . . . .	45
4.7.	Results when using figure 4.4(f) as response matrix. . . . .	46
4.8.	Model after further adaptation, using a first prediction as the new response matrix but setting negative values to zero. . . . .	48
4.9.	The 26 pixelvalues y plotted along the x-axis to show the prevalence of negative values. . . . .	49
4.10.	Final result comparing the predicted and HPLC data applying the regression coefficients in 4.8. . . . .	50
4.11.	Sample 1 surfaceplot of prediction, with added ROIs of summation used for quantification and comparison with HPLC. The colorbar represents mass of API in mg. . . . .	50
A.1.	HPLC results for all samples and their linear fit and residuals created using the MATLAB <sup>®</sup> basic fitting tool . . . . .	56

# List of Tables

2.1.	Parameters used for target setup. . . . .	13
2.2.	Formulated inks. . . . .	14
3.1.	Target page in target setup tab in Scienion AG parameters. . . . .	28
3.2.	Masses of API on each spot calculated with (3.1). . . . .	32
4.1.	HPLC results for the calibration and the three printed test samples. Plots for these values can be seen in figure 4.1. For the calibration . . .	39
4.2.	HPLC results for the three printed test samples. Plots for these values can be seen in figure 4.1. Each layer consists of 17x17 spots as shown in figure 3.10, for samples 1 and 3 one drop was dispensed on each spot, for sample 2 five drops were dispensed on each spot, respectively .	40
4.3.	Quantitative results shown for figure 4.10 . . . . .	47
B.1.	API mass according to the estimated drop volume, HPLC and the prediction, and the deviation between the HPLC results and the prediction for test 1. SNV range: 70-200; SG filter: polynomial order 2, window size 35 . . . . .	57
B.2.	Results of test 2, here SNV and SG parameters and ellipses form and amplitude are the same as for test 1 but not the entire measurement was used for the calibration since spots with fewer than 50 drops where not modelled, they were excluded from the calibration. . . . .	58
B.3.	Results of test 3, here the backprojection of test 1 was used as the response matrix (inverse calibration). . . . .	58
B.4.	Results of test 4, here the spots with fewer than 50 drops were excluded but only after the preprocessing step. . . . .	59
B.5.	Results of test 5: uses the backprojection of test 4 as response matrix. .	59
B.6.	Results of test 6: uses results the backprojection of the first model with negative values set to zero . . . . .	60
B.7.	Results of test 7: uses backprojection based on test 2 as response matrix with negative values of the fit set to zero. . . . .	60
B.8.	Results of test 8: uses spots down to 50 drops but changes the values within the ellipses. . . . .	61
B.9.	Results of test 9: uses the backprojection of test 8 as response matrix. .	61
B.10.	Results of test 10: combines test 1 (entire sample) with the adapted response matrix new values used for test 8. . . . .	62
B.11.	Results of test 11: used backprojection of test 12 as response matrix. .	62
B.12.	Results of test 12: expands SNV range to 70-210, further changes to the values within the modelled ellipses in the response matrix. . . . .	63

List of Tables

B.13. Results of test 13: same as test 1 but uses more components for the regression . . . . . 63

B.14. Results of test 14: does not include spot with 50 drops, since errors in the backprojection are often in this spot. . . . . 64

C.1. Conversion of variable number (V) to nm ( $\lambda$ ) for wavelength . . . . . 66

# Bibliography

- [1] C Planchette, H Pichler, M Wimmer-Teubenbacher, M Gruber, H Gruber-Woelfler, S Mohr, C Tetyczka, W-K Hsiao, A Paudel, E Roblegg, et al. Printing medicines as orodispersible dosage forms: effect of substrate on the printed micro-structure. *International journal of pharmaceutics*, 509(1):518–527, 2016.
- [2] H Vakili, J O Nyman, N Genina, M Preis, and N Sandler. Application of a colorimetric technique in quality control for printed pediatric orodispersible drug delivery systems containing propranolol hydrochloride. *International journal of pharmaceutics*, 511(1):606–618, 2016.
- [3] M Slavkova and J Breitzkreutz. Orodispersible drug formulations for children and elderly. *European Journal of Pharmaceutical Sciences*, 75:2–9, 2015.
- [4] S Stegemann, M Gosch, and J Breitzkreutz. Swallowing dysfunction and dysphagia is an unrecognized challenge for oral drug therapy. *International journal of pharmaceutics*, 430(1):197–206, 2012.
- [5] E M Hoffmann, A Breitenbach, and J Breitzkreutz. Advances in orodispersible films for drug delivery. *Expert opinion on drug delivery*, 8(3):299–316, 2011.
- [6] A F Borges, C Silva, J FJ Coelho, and S Simões. Oral films: current status and future perspectives: I—galenical development and quality attributes. *Journal of Controlled Release*, 206:1–19, 2015.
- [7] M Radicioni, C Castiglioni, A Giori, I Cupone, V Frangione, and S Rovati. Bioequivalence study of a new sildenafil 100 mg orodispersible film compared to the conventional film-coated 100 mg tablet administered to healthy male volunteers. *Drug design, development and therapy*, 11:1183, 2017.
- [8] K K Peh and C F Wong. Polymeric films as vehicle for buccal delivery: swelling, mechanical, and bioadhesive properties. *J Pharm Pharm Sci*, 2(2):53–61, 1999.
- [9] M Preis and J M Rosenholm. Printable nanomedicines: the future of customized drug delivery?, 2017.
- [10] M A Mahaparale, S S T Shivnikar, K V Pawar, and N Prashant. Fast dissolving oral films: an innovative drug delivery system. *IJRRPAS*, 2(3):482–496, 2012.
- [11] M Alomari, F H Mohamed, A W Basit, and S Gaisford. Personalised dosing: printing a dose of one’s own medicine. *International journal of pharmaceutics*, 494(2):568–577, 2015.

## Bibliography

- [12] M Takala, H Helkiö, J Sundholm, N Genina, P Kiviluoma, T Widmaier, N Sandler, and P Kuosmanen. Ink-jet printing of pharmaceuticals. In *Proceedings of the 8th International DAAAM Baltic Conference*, 2012.
- [13] G Cummins and M PY Desmulliez. Inkjet printing of conductive materials: a review. *Circuit World*, 38(4):193–213, 2012.
- [14] B K Lee, Y H Yun, J S Choi, Y C Choi, J D Kim, and Y W Cho. Fabrication of drug-loaded polymer microparticles with arbitrary geometries using a piezoelectric inkjet printing system. *International journal of pharmaceutics*, 427(2):305–310, 2012.
- [15] N Scoutaris, M R Alexander, P R Gellert, and C J Roberts. Inkjet printing as a novel medicine formulation technique. *Journal of controlled release*, 156(2):179–185, 2011.
- [16] J G Khinast, A Zimmer, G Brenn, W Bauer, B Eitzinger, D Strohmeier, N Schroedl, M M Gruber, and C Voura. Printable medicines: a microdosing device for producing personalized medicines. *Pharmaceutical Technology Europe*, 23(1), 2011.
- [17] P B Patil and S K Shrivastava. Fast dissolving oral films: an innovative drug delivery system. *Structure*, 20(70):50–500, 2012.
- [18] R D Boehm, P R Miller, J Daniels, S Stafslie, and R J Narayan. Inkjet printing for pharmaceutical applications. *Materials Today*, 17(5):247–252, 2014.
- [19] K M Raval. Overview on oral strip. *Journal of drug discovery and therapeutics*, 1(03), 2013.
- [20] M L Balboni. Process analytical technology. *Pharm Tech*, 10:54–66, 2003.
- [21] J Huiskamp. What are the differences between in-line, on-line, at-line and off-line analysers?, September 2009. accessed 16.01.2018.
- [22] D K Pedersen and S B Engelsen. Monitoring industrial food processes using spectroscopy & chemometrics. *New Food*, 2:9–13, 2001.
- [23] K A Bakeev. *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. John Wiley & Sons, 2010.
- [24] P R Wahl, G Fruhmann, S Sacher, S Straka, Gand Sowinski, and J G Khinast. Pat for tableting: inline monitoring of api and excipients via nir spectroscopy. *European Journal of Pharmaceutics and Biopharmaceutics*, 87(2):271–278, 2014.
- [25] A A Gowen, C Po O'Donnell, P J Cullen, G Downey, and J M Frias. Hyperspectral imaging—an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, 18(12):590–598, 2007.
- [26] Pshemak Maslak. Organic chemistry, vibrations: Methane. online course for organic chemistry at Penn State, 2010. accessed 16.01.2018.

## Bibliography

- [27] M Blanco, J Coello, H Iturriaga, S MasPOCH, and C De La Pezuela. Near-infrared spectroscopy in the pharmaceutical industry. *ANALYST-LONDON-SOCIETY OF PUBLIC ANALYSTS THEN ROYAL SOCIETY OF CHEMISTRY-*, 123:135R–150R, 1998.
- [28] G Reich. Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. *Advanced drug delivery reviews*, 57(8):1109–1143, 2005.
- [29] T De Beer, A Burggraeve, M Fonteyne, L Saerens, J P Remon, and C Vervaet. Near infrared and raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *International Journal of Pharmaceutics*, 417(1):32–47, 2011.
- [30] J Luypaert, D L Massart, and Y Vander Heyden. Near-infrared spectroscopy applications in pharmaceutical analysis. *Talanta*, 72(3):865–883, 2007.
- [31] O Scheibelhofer, D M Koller, P Kerschhaggl, and J G Khinast. Continuous powder flow monitoring via near-infrared hyperspectral imaging. In *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*, pages 748–753. IEEE, 2012.
- [32] I Aakre. A method for rapid localisation of hydrocarbon compounds on surfaces using chemical imaging and back-projection. Master's thesis, Institutt for kjemi, 2013.
- [33] L Eriksson, T Byrne, E Johansson, J Trygg, and C Vikström. *Multi-and megavariable data analysis basic principles and applications*. Umetrics Academy, 2013.
- [34] W Kessler. *Multivariate datenanalyse: für die pharma, bio-und Prozessanalytik*. John Wiley & Sons, 2007.
- [35] S Wold, M Sjöström, and L Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130, 2001.
- [36] G Palermo, P Piraino, and H-D Zucht. Performance of pls regression coefficients in selecting variables for each response of a multivariate pls for omics-type data. *Advances and applications in bioinformatics and chemistry: AABC*, 2:57, 2009.
- [37] H Abdi. Partial least squares regression and projection on latent structure regression (pls regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):97–106, 2010.
- [38] A Höskuldsson. Pls regression methods. *Journal of chemometrics*, 2(3):211–228, 1988.
- [39] S De Jong. Simpls: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, 18(3):251–263, 1993.
- [40] K H Esbensen and T T Lied. Principles of image cross-validation (icv): representative segmentation of image data structures. *Techniques and Applications of Hyperspectral Image Analysis*, 7:155–180, 2007.

## Bibliography

- [41] S Kim, P A Thiessen, E E Bolton, J Chen, G Fu, A Gindulyte, L Han, J He, S He, B A Shoemaker, et al. Pubchem substance and compound databases. *Nucleic acids research*, 44(D1):D1202–D1213, 2015.
- [42] E Glynn. Usaf 1951 and microcopy resolution test charts and pixel profiles, November 2002. accessed 29.12.2017.
- [43] A Rinnan, F van den Berg, and S B Engelsen. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10):1201–1222, 2009.
- [44] R W Schafer. What is a savitzky-golay filter?[lecture notes]. *IEEE Signal processing magazine*, 28(4):111–117, 2011.
- [45] B Perston. Perkinelmer ir data file import tools, version 1.0. <https://de.mathworks.com/matlabcentral/fileexchange/22736-perkinelmer-ir-data-file-import-tools>, January 2009. The MathWorks, Natick, MA, USA.