



LUND UNIVERSITY  
Faculty of Science

# SPECTROSCOPY AND RGB-COLORIMETRY FOR QUANTIFICATION OF PLANT PIGMENT AND FRUIT CONTENT IN FRUIT DRINKS

Daniele C. Rodman

Thesis submitted for the degree of Bachelor of Science  
Project duration: 7 months

Supervised by: Cecilia Jarlskog

Department of Physics  
Division of Atomic Physics  
March 2018



## **Abstract**

Collimated Transmission Spectroscopy (CTS) is a technique that is commonly used to study the optical absorption of different media. In this work CTS was used to measure the absorbance of pure raspberry juice in order to investigate how well the fruit content may be determined in a commercially available raspberry fruit drink. Another method, RGB colorimetry, which is a much more available technique that can be performed using a standard camera in e.g. a smartphone was also tested for the same purpose. The aim was to evaluate the viability and accuracy of CTS in quality monitoring of fruit drinks as compared to two differently sophisticated approaches to RGB colorimetry. CTS was indeed found to be the most accurate of the techniques, but the more sophisticated of the colorimetry methods, which used partial least squares regression to evaluate unknown concentrations was found to have some merit. The simpler of the colorimetry methods was found to be the most inaccurate, but it could potentially be used to make rough estimates of concentrations, and it has the advantage of being simple and available for e.g. educational or consumer purposes.

# Contents

List of Abbreviations	5
About this thesis	6
<b>1 Introduction</b>	<b>7</b>
1.1 Alternative techniques . . . . .	7
1.2 Aim . . . . .	8
<b>2 Material and methods</b>	<b>9</b>
2.1 CTS - Collimated Transmission Spectroscopy . . . . .	9
2.2 Colorimetry . . . . .	11
2.3 Experimental setup . . . . .	12
2.3.1 CTS setup . . . . .	12
2.3.2 Colorimetry setup . . . . .	14
2.4 Sample preparation . . . . .	14
2.5 Measuring the absorption of water . . . . .	14
2.5.1 Sample preparation and measurement . . . . .	15
2.5.2 Result . . . . .	15
2.6 Data analysis . . . . .	17
2.6.1 PLSR - Partial Least Squares Regression . . . . .	17
<b>3 Results</b>	<b>18</b>
3.1 An even more simplified colorimetry approach . . . . .	21
3.2 Data analysis . . . . .	22
3.3 Commercial fruit drink . . . . .	23
<b>4 Discussion and outlook</b>	<b>26</b>
<b>References</b>	<b>29</b>

## List of acronyms and abbreviations

NMR	Nuclear Magnetic Resonance
HPLC	High Performance Liquid Chromatography
CTS	Collimated Transmission Spectroscopy
PLSR	Partial Least Squares Regression
RGB	Red Green Blue, color representation
MEP	Mean Error of Prediction
RMSE	Root Mean Squared Error

## About this thesis

This work originated from a project proposal by Nina Reistad and Ann-Marie Pendrill. The proposal involved use of the mobile phone for spectroscopic analysis to assess quality of fruit and fruit drinks, with the aim to create an experiment for use in high-school physics education. Nina Reistad also provided access to the office space and the laboratory that I used in the work. Through Nina Reistad, I also had access to a computer, the collimated transmission setup, the spectrometer, and all laboratory supplies for the measurements, as well as supervision in the use. However, I chose to carry out the project with a different approach than in the original project description.

# 1 Introduction

Spectroscopic techniques are some of the most important and widely used tools in science today. The idea of using light to study properties of matter was first introduced by Joseph von Fraunhofer in the early 1800s [1], who invented the first spectroscope and discovered the atomic absorption lines in light emitted from the sun. The fundamental idea behind spectroscopy is to measure the intensity of different wavelength components of light, to obtain a spectrum. These spectra are characteristic of the light source in the case of emission spectra, and/or the medium which the light has interacted with on its way to the detector in the case of absorption spectra. Besides its usefulness in astronomy, spectroscopy is also commonly used in fields such as medicine, applied physics and chemistry to name a few, in order to probe properties such as concentration, composition or temperature of different materials and compounds.

This work is based on methods that are mainly used in the field of biophotonics, where the composition of organic tissues can be studied using various spectroscopic techniques. This way of doing compositional analysis can be applied to foods such as fruits or fruit drinks, and it has the potential to be useful in terms of quality assurance for these types of products, especially since fruits often contain pigments which can readily be detected and quantified using spectroscopic techniques in the visible wavelength range.

Ripe fruits and berries are often strongly colored in order for them to stand out from a green backdrop in nature. This increases their chances of being eaten by an animal, and in that way spread their seeds [2]. Blue, purple and red plant pigments commonly belong to a group of flavonoids called anthocyanins, which are responsible for the majority of the color absorption in e.g. raspberries, which will be studied in this thesis work. Anthocyanins have been the subject of a lot of research due to their antioxidative properties, which might have health benefits, such as reducing oxidative stress in the body.

In this work, CTS (Collimated Transmission Spectroscopy) will be applied to determine the concentration of pure raspberry juice, and consequently anthocyanins in commercial raspberry fruit drinks. A simpler method is to determine the the concentration colorimetrically using e.g. a mobile camera with an RGB analyzer. This method is readily available for educational or consumer purposes where advanced and expensive spectroscopic equipment may be inaccessible, and depending on the accuracy it may find other applications as well [3, 4, 5].

## 1.1 Alternative techniques

There are a few different techniques that are currently being used for quality monitoring and compositional analysis of foods and drinks. A brief overview of two important ones are given here:

- NMR (Nuclear Magnetic Resonance) techniques use rapidly altering magnetic fields which can interact with atomic nuclei, provided they have non-zero spin. The resonant frequencies of nuclei depend on the energy difference between adjacent magnetic sub-states, which can be shifted as an effect of the chemical environment of the atom. This chemical shift provides a very specific NMR signature of a particular molecule, which is why NMR can be used to identify different substances and determine their

concentration in a sample. The technique is sensitive to different isotopes, which can be useful in some applications as well [6].

NMR is a very accurate and sensitive technique which can detect very small deviations in the composition of compounds. NMR can for example distinguish between wines of different origin, being able to differentiate between grapes grown in different regions, even though they are of the same variety. This level of accuracy would be difficult to achieve with spectroscopic techniques. The downside of the NMR technique however, is that it is expensive and the technique is slow for large scale applications [7].

- HPLC (High Performance Liquid Chromatography) is used to separate components that are dissolved in a liquid known as the mobile phase. This is done by pumping the solution through a column of chromatographic packing material known as the stationary phase at high pressure. Different components will interact differently with both the solvent and the packing material, which can e.g be due to differences in polarity or size. The stronger the analytes are attracted to the liquid phase, the faster they will traverse the column, and the opposite is true for strong interaction with the stationary phase. HPLC can separate molecules at concentrations down to parts per trillion, which makes the technique useful in applications such as pharmacology, where high precision is necessary [8]. In order to identify the compounds as they exit the column, spectroscopic techniques are used, much like the ones that will be presented in this text, with the benefit of only having to deal with the spectrum of one analyte at a time.

Even though HPLC machines makes compositional analysis easier, they are quite expensive and slow, and they still rely on spectroscopic techniques. Another drawback with the technique is that the samples are inevitably destroyed as they are analyzed, while transmission spectroscopy generally does not affect the samples much. In conclusion, HPLC is primarily useful for high precision measurements, where the occurrence of specific molecules are of interest. HPLC has been used to investigate the occurrence of different antioxidants in red fruit juices [9], but for more general quality assurance and composition analysis and at larger scales, other methods are preferable.

## 1.2 Aim

The aim of this thesis is to investigate the viability and accuracy of CTS and two different methods of RGB colorimetry, when used to measure the fruit content in commercially available fruit drinks. To do this, the optical properties of raspberry juice will be measured, and the data will be used to determine its concentration in a commercial fruit drink. The detection limit will be tested, and the accuracy of the different techniques will be compared.

## 2 Material and methods

### 2.1 CTS - Collimated Transmission Spectroscopy

Spectroscopic techniques are used to measure the intensity of different wavelength components in emitted, reflected or transmitted light, and by doing so, deduce characteristics of the light source and/or the medium between the light source and the detector/spectrometer. Light can interact with matter in a variety of ways, depending on the wavelength of the light and the optical properties of the medium.

**Specular reflectance** occurs when the incident light is reflected off a smooth surface, such that the reflected beam is directional, i.e. the angle of the incident light is equal to the angle of the reflected light. If light is specularly reflected or transmitted through a surface depends on the wavelength of the light, the incident angle and the difference in refractive index at the intersection of the surface. In a transmission measurement some light will often be lost due to specular reflectance, e.g. off the surface of a cuvette (liquid container) in the case of studying liquids, or off the surface of a lens used in the setup.

**Diffuse reflectance** is non-directional, and arises when light scatters in random directions on a rough surface, or in a diffuse medium. When performing transmission spectroscopy, some light will likely be diffusely reflected off particles, or redirected through bubbles or other local areas of differing refractive index, and subsequently lost to the detector. In cases with very “cloudy” media, diffuse reflectance spectroscopy techniques are preferred, where only the diffusely reflected light is detected.

**Fluorescence** happens when light excites molecules or atoms in a medium, which then de-excite through intermediate steps, emitting characteristic wavelengths of light. Fluorescence can thus result in detected wavelengths that are longer than any of the wavelengths from the light source, or even increase the intensity of some wavelengths above 100% of the intensity from the light source. Fluorescence is non-directional, and therefore suitably measured in a similar way to diffuse reflectance.

**Absorption** occurs when a photon excites an atom, a molecule or a solid, i.e puts it into a higher energy state. In the case of atomic absorption, the state usually de-excites, re-emitting the photon in an arbitrary direction. In a molecule or a solid however there can be vibrational or rotational energy states, in which case the energy is converted into heat. Atomic energy states are discrete with a very narrow wavelength range, which is why an atomic absorption spectrum consist of lines, or equivalently very narrow peaks. Vibrational and rotational states however are generally more abundant and closely spaced, which results in seemingly continuous absorption spectra when the resolution is limited.

Absorption is the optical property that is of interest in this work, and it is what effects the apparent color of a medium. When certain wavelength components of light are absorbed, the other transmitted wavelengths are enriched, and it is these wavelengths that cause the perceived color of a medium or an object.

**Transmission.** When white light is shone through a semi-transparent liquid, all or some of the above interactions may affect the amount of light that is transmitted through the sample. By collimating the light, and by using reference measurements on cuvettes and media of similar refractive index to the analyte, the effect of specular reflectance can be reduced and compensated for. When performing measurements on clear samples without fluorescence,

the other effects can usually be neglected as well, and the transmission measurement can be approximated as a pure absorption measurement. The amount of light absorbed is related to the path-length  $d$ , and the absorption coefficient  $\mu_a(\lambda)$  of the medium, through Beer-Lambert's law:

$$I_T = I_0 e^{-\mu_a(\lambda)d}, \quad (1)$$

where  $I_0$  and  $I_T$  is the intensity of the light before and after transmission. The absorbance, or optical density, of a liquid is defined as:

$$A(\lambda) \equiv -\log_{10} \left( \frac{I_T}{I_0} \right), \quad (2)$$

which together with Eq. (1) gives the relation between the absorbance and the absorption coefficient:

$$A(\lambda) = \mu_a(\lambda) \cdot d \cdot \log_{10}(e). \quad (3)$$

$A(\lambda)$  is related to the concentration  $c$  and a medium specific extinction coefficient  $\varepsilon(\lambda)$ , so that:

$$A(\lambda) = \varepsilon(\lambda) \cdot c \cdot d. \quad (4)$$

The extinction coefficient  $\varepsilon$  allows for concentration determination of analytes of unknown concentration, provided that the path-length is known. The relation between absorbance and concentration is not always linear though. This is partly because intra-molecular interactions become increasingly more important with concentration, which can affect the absorptivity of the sample. Another non-linearity effect comes from the fact that the refractive index of the sample may change with concentration. This could lead to spread that affects the amount of detected light. Both of these effects can be neglected as long as the concentration is kept low. Other deviations from Beer-Lambert's law can be seen in wavelength ranges where the extinction coefficient changes rapidly, i.e where  $\frac{dA}{dc}$  is large. A wavelength at a broad absorbance peak should therefore be chosen when evaluating the extinction coefficient.

To measure an absorbance spectrum,  $I_T$  and  $I_0$  have to be obtained. In an ideal transmission measurement the  $I_0$  spectrum should be measured in a vacuum, so that there is no absorption or reflection between the light source and the detector. In the ideal scenario  $I_T$  is then measured under the exact same conditions, with the analyte as the only absorber, without any surfaces, like that of a cuvette wall, causing light to be lost to scattering, specular reflectance or absorption in the cuvette material. Realistically however a cuvette has to be used when the analyte is in liquid form, and in order to keep the conditions as similar as possible for the reference and the sample measurements, a blank is used for the reference spectrum. A blank should mimic the optical properties of the sample as closely as possible, and should therefore consist of an identical cuvette filled with the same solvent used for the sample. This way, the contributions of specular reflection, scattering and absorption by the cuvette walls cancel out.  $I_0$  is redefined to be the light transmitted through the blank [10], call it  $I_c$ , and what is actually measured is the relative transmittance of the sample compared to the blank,  $\left( \frac{I_T}{I_c} \right)$ . If the optical absorption of the solvent is known, it is possible to

recalculate the real absorbance of the sample from the relative absorbance, by subtracting the effect of the solvent from the blank,

$$I_0 = \frac{I_c}{e^{-\mu_s d}}, \quad (5)$$

where  $\mu_s$  is the absorption coefficient of the solvent.

## 2.2 Colorimetry

Digital imaging devices such as the CCD or CMOS sensors commonly found in digital cameras or smartphones, represent color in a similar way to the human eye. These sensors register color in terms of the intensity of the primary colors red, green and blue, so that each wavelength in the visible range (and to some extent in the IR) can be represented by an RGB vector. It should be noted however that while a wavelength can be represented by an RGB color, an RGB value does not necessarily correspond to a single wavelength. In order to assess the intensity of light at a specific wavelength using an RGB sensor, the spectral sensitivity of the sensor should be known, i.e the RGB vector representation of that wavelength. Figure 1 shows an example of the spectral sensitivity of a Nokia N900 camera phone, measured by D. Liu 2013 [11], although it should be noted that the sensitivity may vary between different kinds of sensors and manufacturers.

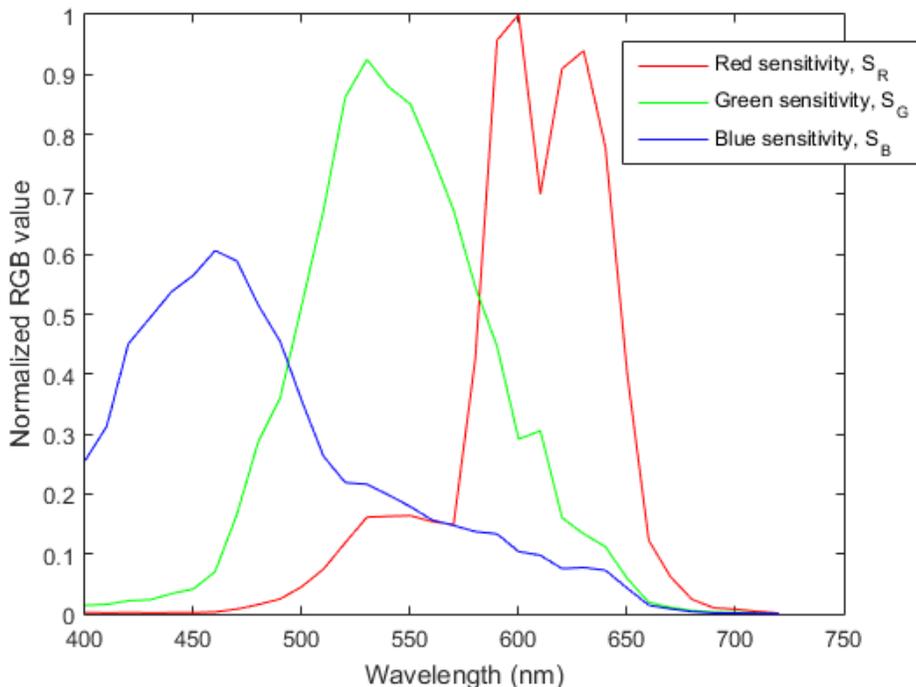


Figure 1: An example of the spectral sensitivity of a camera found on a smart phone [11].

Using spectral sensitivities such as in Figure 1, the intensity of a certain wavelength of light can be estimated, in analogy to Beer-Lamberts law [4]:

$$A(\lambda) = -\log \frac{S_R(\lambda) \cdot R_s + S_G(\lambda) \cdot G_s + S_B(\lambda) \cdot B_s}{S_R(\lambda) \cdot R_c + S_G(\lambda) \cdot G_c + S_B(\lambda) \cdot B_c}, \quad (6)$$

where the sensitivity coefficients  $S$  can be read from spectral sensitivities as those in Figure 1. Subscript  $s$  denotes the normalized R/G/B values of the sample, and  $c$  denotes the normalized values of the blank. The normalized values are obtained from the measured  $r/g/b$  values as:

$$R, G, B = \frac{r, g, b}{\sqrt{r^2 + g^2 + b^2}}. \quad (7)$$

Using this method, absorbance spectra can be obtained from an ordinary camera, provided that the spectral sensitivities are known. The resulting spectra can be treated in a similar manner as CTS spectra, see section 2.6.

RGB colorimetry is proposed in this work to be a simple alternative to CTS, suitable for education on a secondary level. An alternative to the full analysis suggested above is therefore desirable, especially since the spectral sensitivities are not always available for the camera that is used, and the process of acquiring them can be difficult without specialized equipment, such as a monochromator and a white light source. An alternative method is to simply observe a relation between the R, G and B values respectively and the concentration. Depending on which wavelengths are absorbed, one or more of the RGB values should be affected approximately linearly by the concentration, and the relative absorbance analogues would be [3]:

$$A_{R/G/B} = -\log_{10} \frac{R_s/G_s/B_s}{R_c/G_c/B_c}. \quad (8)$$

## 2.3 Experimental setup

### 2.3.1 CTS setup

The transmission measurements were made using a Tungsten-Halogen light source (Ocean Optics), producing white light. The light was directed to one of the optical ports on a cuvette holder (Thorlabs CVH100), through an optical fiber (Thorlabs FT038 mm, M41L02), and a collimating lens. The transmitted light was passed through a collecting lens, before entering another optical fiber of the same type to be analyzed in a spectrometer with a range of  $\sim 350$ - $1000$  nm, and a wavelength resolution of  $\sim 0.22$  nm. The spectra were recorded on a computer using Ocean Optics SpectraSuite software. The experimental setup is shown in Figure 2.

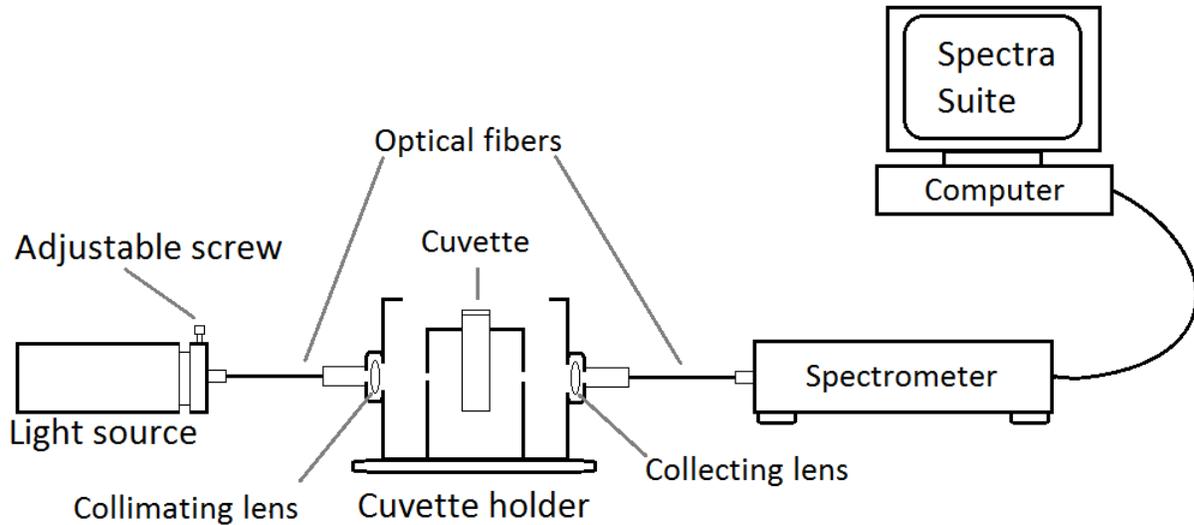


Figure 2: The CTS setup.

To avoid saturation of the spectrometer due to over exposure, the integration time was set to 18 ms in Spectra Suite. The integration time controls the exposure time of the spectrometer, and affects the total number of counts that is recorded in each wavelength bin. The integration time was kept as short as possible in order to achieve fast measurements, but still long enough to provide a signal close to the maximum, in order for the signal to noise ratio to be as high as possible. Consequently the light source was set to the maximum intensity. The maximum signal is limited by the detector and the software. In order to reduce noise, each spectrum was acquired 100 times and averaged. Each spectrum was then binned into integer nm intervals to reduce the effect of noise further. The edges of the spectra were trimmed due to an increased amount of noise near the shortest and longest wavelengths in the range, which was caused by the non-uniform intensity profile of the light source. For this reason most spectra presented in the following are from 440-900 nm.

In order to obtain accurate transmission spectra for the media under study, reference measurements were made with a blank consisting of a cuvette containing distilled water before and after every measurement. The light being absorbed by or reflected off the cuvette walls or absorbed by the water was recorded in calibration/reference spectra  $I_c$ , and used as a reference for subsequent measurements. So called dark noise, caused by e.g. thermal excitations in the detector or ambient light effects were reduced in a similar way by recording background spectra  $I_b$ . This was done by blocking the light source, and recording dark spectra before the measurements, to then be subtracted from subsequent measurements. The transmittance through a sample was then obtained from the measured signal  $I_t$  by:

$$T = \frac{I_t - I_b}{I_0 - I_b}, \quad (9)$$

where  $I_0$  was calculated from Eq. (5), using the absorption of water published by Nachabé et al. 2012 [12]. The absorbance was calculated in analogy with Eq. (2):

$$A = -\log_{10} T. \quad (10)$$

### 2.3.2 Colorimetry setup

The cuvettes containing the samples of diluted raspberry juice as well as a blank containing distilled water were lined up in front of a white screen, illuminated by standard fluorescent lamps. The setup was aligned so that the samples did not cast a shadow on the screen. A camera phone was placed in front of the samples along an edge, as to be able to slide parallel to the samples at a constant distance. A colorimetry application (ColorAssist, available from the App Store for iOS devices) for the phone was used to record the average RGB value of a 50x50 pixel area contained within the image of a cuvette, as in Figure 3. The RGB values of 40 samples and the blank were recorded.

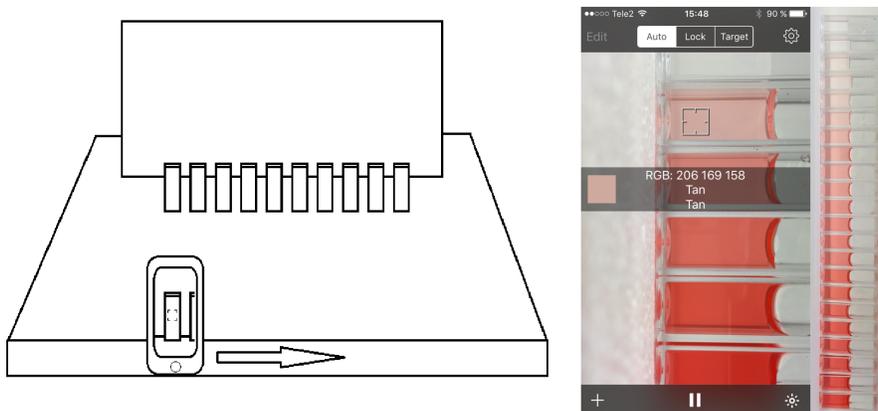


Figure 3: The colorimetry setup to the left, and an example of the colorimetry application in use to the right, as well as some of the samples lined up to the far right.

## 2.4 Sample preparation

Frozen raspberries were thawed, then mashed and strained to remove larger solids like seeds and most of the skin. The resulting purée was then filtered using a paper coffee filter to remove most of the remaining solids, as to ensure that the final product was clear enough for transmission measurements. Different concentrations of clarified raspberry juice in distilled water were then prepared in polystyrene cuvettes (VWR PS macro) with a 10 mm path length.

## 2.5 Measuring the absorption of water

In order to obtain the absorption coefficients of the solvent,  $\mu_s$  in Eq. (5), an absorption spectrum of distilled water was measured. For this measurement an empty cuvette was used as the blank, which caused a problem since air and water have different refractive indices. The interface between water and the cuvette walls has a different reflection coefficient than the interface between air and the cuvette walls. The issue can be solved by considering

that the transmission depends on the pathlength  $d$ . With two different pathlengths/cuvette thicknesses  $d_1$  and  $d_2$ , Eq. (1) can be written:

$$\frac{I_w(d_1) I_e(d_2)}{I_e(d_1) I_w(d_2)} = e^{\mu_a(d_1-d_2)}. \quad (11)$$

The subscripts  $w$  and  $e$  refer to cuvettes filled with water and empty cuvettes respectively. from Eq. (10) the absorption coefficient can be obtained as:

$$\mu_a = \frac{\ln \left( \frac{I_w(d_1) I_e(d_2)}{I_w(d_2) I_e(d_1)} \right)}{d_1 - d_2}. \quad (12)$$

This way the effect due to reflection is canceled out by division, since the difference in path length should not effect the reflection coefficient [13].

### 2.5.1 Sample preparation and measurement

Two cuvettes with funnel shaped, rectangular bottoms (Eppendorf Uvette<sup>®</sup>) were used, as they provided two different thicknesses, 2 mm and 10 mm for the same sample, by a 90° rotation. One cuvette was filled with distilled water, and the other remained empty. Measurements were carried out on both cuvettes, using both path lengths at 20°C. The final absorption spectrum was calculated according to Eq. (12).

### 2.5.2 Result

The left image in Figure 4 shows that the transmission through 2 mm of water is significantly reduced compared to that of 10 mm. This is clearly wrong and completely opposite of what was expected. Since water is optically more dense than air, it stands to reason that the transmission would decrease as air is replaced by water, and not the other way around. This ~ 29% gap was present for three different samples measured at three different occasions, and the origin of the effect could not be determined within the scope of this work. All of the other CTS and colorimetry measurements in this work were preformed with the same sized, 10 mm cuvettes. The effect which arises from a difference in path length, should therefore not affect any of the other results in this work.

By making the unfounded assumption that the effect simply shifted the spectra in relation to each other, an adjustment was made by adding 29% to the transmission of the 2 mm spectrum, as in the right image in Figure 4. Absorption coefficients could then be obtained which compared quite well to the absorption coefficients of water published by Nachabé et al. 2012 [12]. As is shown in Figure 5, the spectra match well above 700 nm, but at shorter wavelengths where the absorption of water is significantly lower [12], the arbitrariness and inaccuracy of the 29% adjustment makes itself apparent, and the spectra do not match at all. Since the measured absorption spectrum is the result of an arbitrary adjustment, and contains more noise than the one published by Nachabé et al. theirs is the one that will be used henceforth in this work, as  $\mu_s$  in Eq. (5), to calculate absorbance.

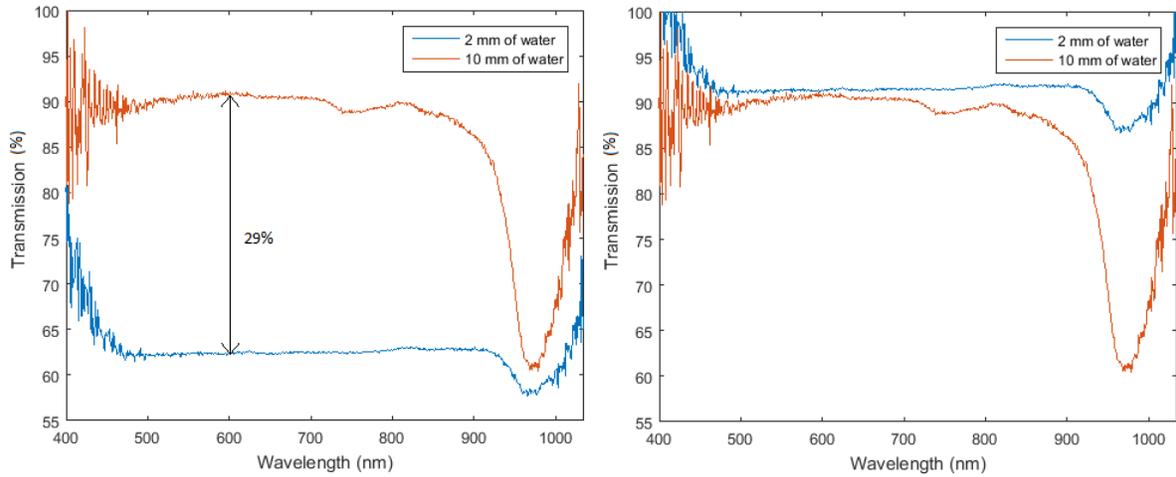


Figure 4: Measured transmission through 2 mm and 10 mm of distilled water to the left, and adjusted transmission to the right.

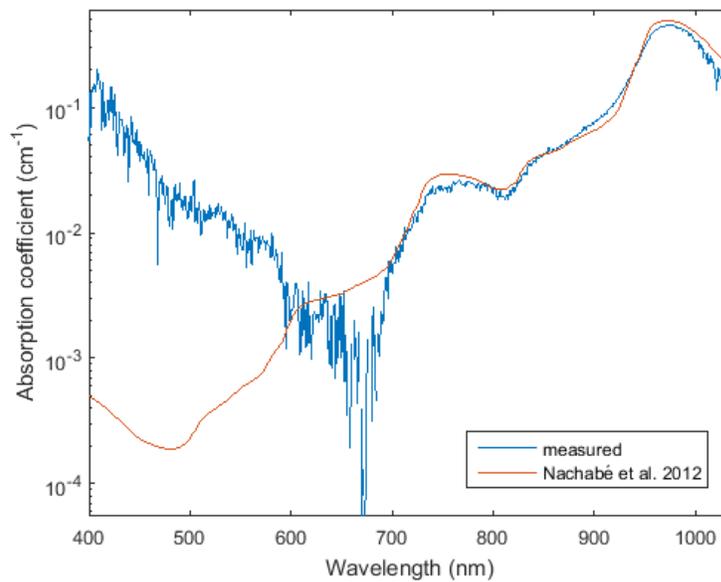


Figure 5: Measured absorption coefficients of water in the visible wavelength range. The spectrum is compared to one published by Nachabé et al. [12].

## 2.6 Data analysis

### 2.6.1 PLSR - Partial Least Squares Regression

A simple predictive model for determining the concentration based on the absorbance can be constructed by measuring the extinction coefficient at a suitable wavelength, provided that the relation between absorbance and concentration is linear. This is a rather limited method however, and it only utilizes a few data points, letting the majority go to waste. A common analysis tool is instead PLSR, which has the ability to predict a set of dependent variables, or responses, (concentration) from a very large set of independent variables, or predictors, such as the absorbance spectra, even though they are noisy, correlated or incomplete. The PLSR algorithm searches for a set of components that explain the variance in the predictor as well as the response variables. The more components that are used, the better the fitted response will match the data provided for the model. If too many components are used however, the model will be overdetermined, and the model will lose generality and will not handle new data very well [14].

### 3 Results

The transmittance (see Eq. (9)) measured by CTS, for 40 different concentrations between 0.25 and 10% of raspberry juice in water is presented in Figure 6. The corresponding absorbance, as in Eq. (10), is presented in Figure 7. The data is presented between 440-900 nm for both figures, and the gradual increase in concentration is indicated by the RGB colors of each sample, which were obtained during the RGB colorimetry measurements.

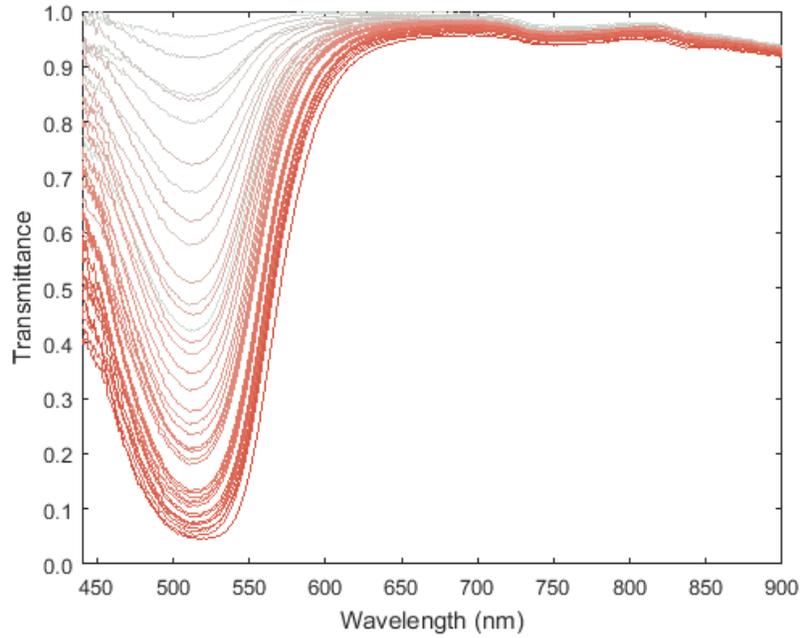


Figure 6: CTS measurement of transmittance through raspberry juice in water. Concentrations ranging from 0.25 to 10%.

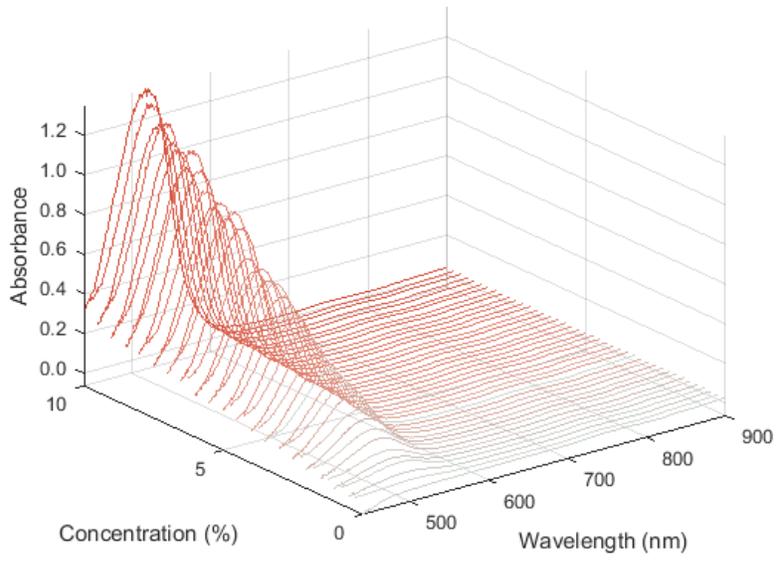


Figure 7: CTS measurement of absorbance through raspberry juice in water. Concentrations ranging from 0.25 to 10%.

The absorbance for the peak at 515 nm in Figure 7 is plotted against the concentration in Figure 8, together with a trend line with slope  $\varepsilon = 0.1270 \text{ cm}^{-1}/\%$  as in Eq. (4).  $R^2 = 0.9966$ . This indicates that Eq. (4) holds, at least at 515 nm.

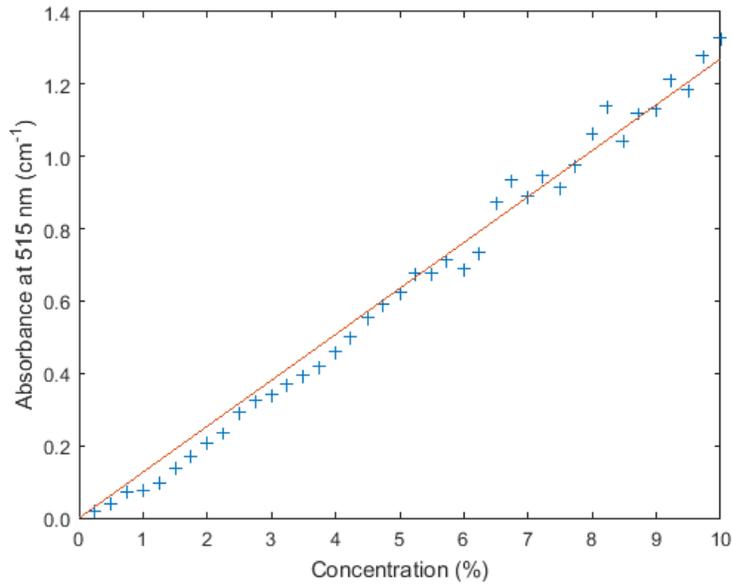


Figure 8: CTS-measured absorbance at 515 nm through raspberry juice in water. Concentrations ranging from 0.25 to 10%.

The color of the 40 different samples, as recorded by the camera are represented in normalized RGB space (as in Eq. (7)) in Figure 9, with the same color scheme as in Figure 6 and 7.

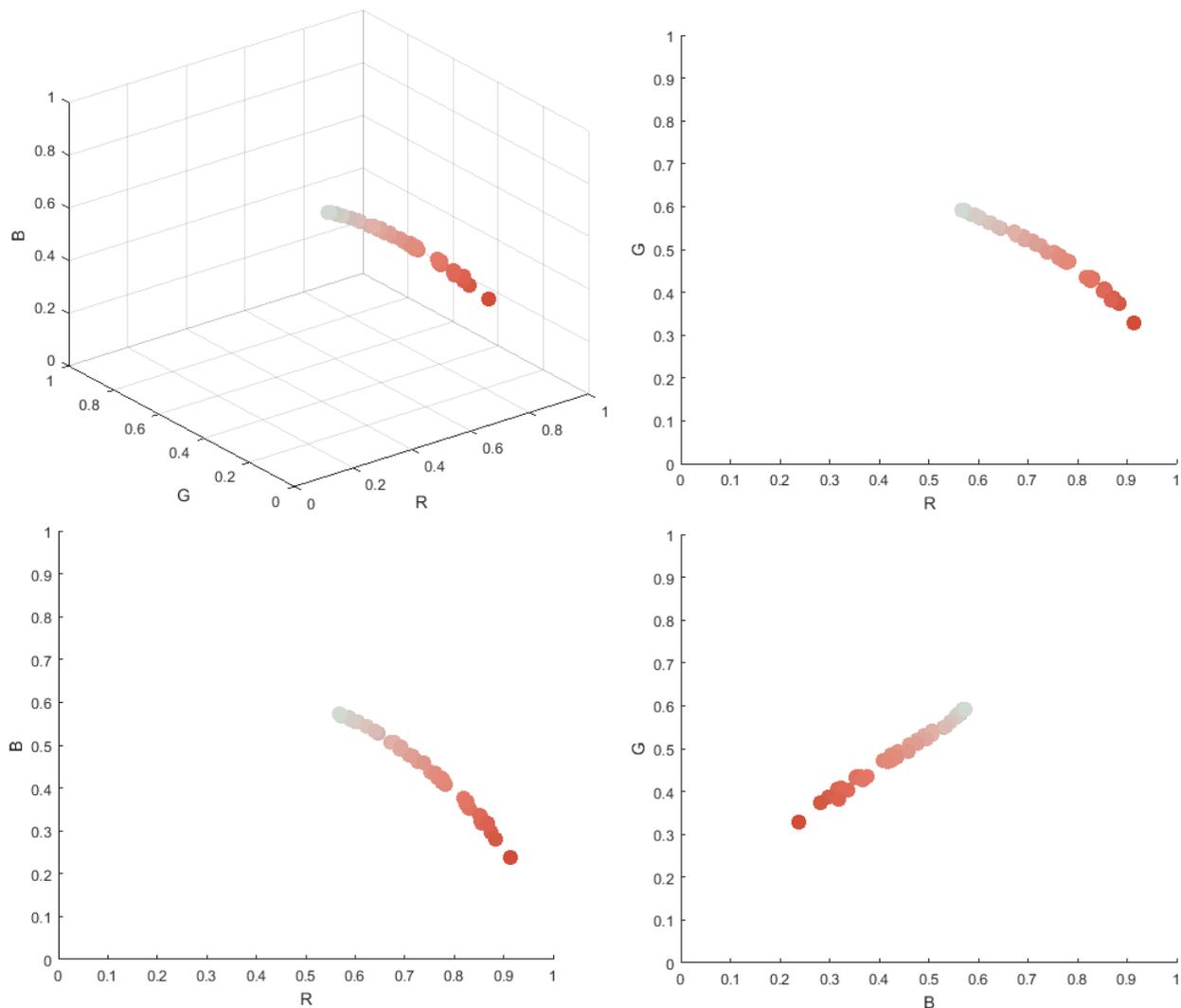


Figure 9: Normalized RGB values plotted in RGB space (top left), and in three different planes. The color of the data points match the averaged RGB color recorded by the camera.

The relative absorbance analogues were calculated as in Eq. (8), for all three RGB values, and plotted in Figure 10. The colorimetrically measured relative absorbance analogues are plotted similarly to the absorbance in Figure 7, as they will be treated analogously to the CTS-measured absorbance spectra, but with only three wavelengths represented by R, G and B. The relative absorbance analogues are used rather than the “pure” absorbance analogues for simplicity.

Some periodic behavior can be seen in the supposedly linear relations between absorbance and concentration in Figure 10. This is likely due to a measurement error, caused by performing the colorimetry measurements in two different series, under similar, but evidently

not identical circumstances. One series was performed with "odd"-numbered cuvettes and the other with "even"-numbered ones, which is why every other measurement point is slightly shifted in relation to the previous one.

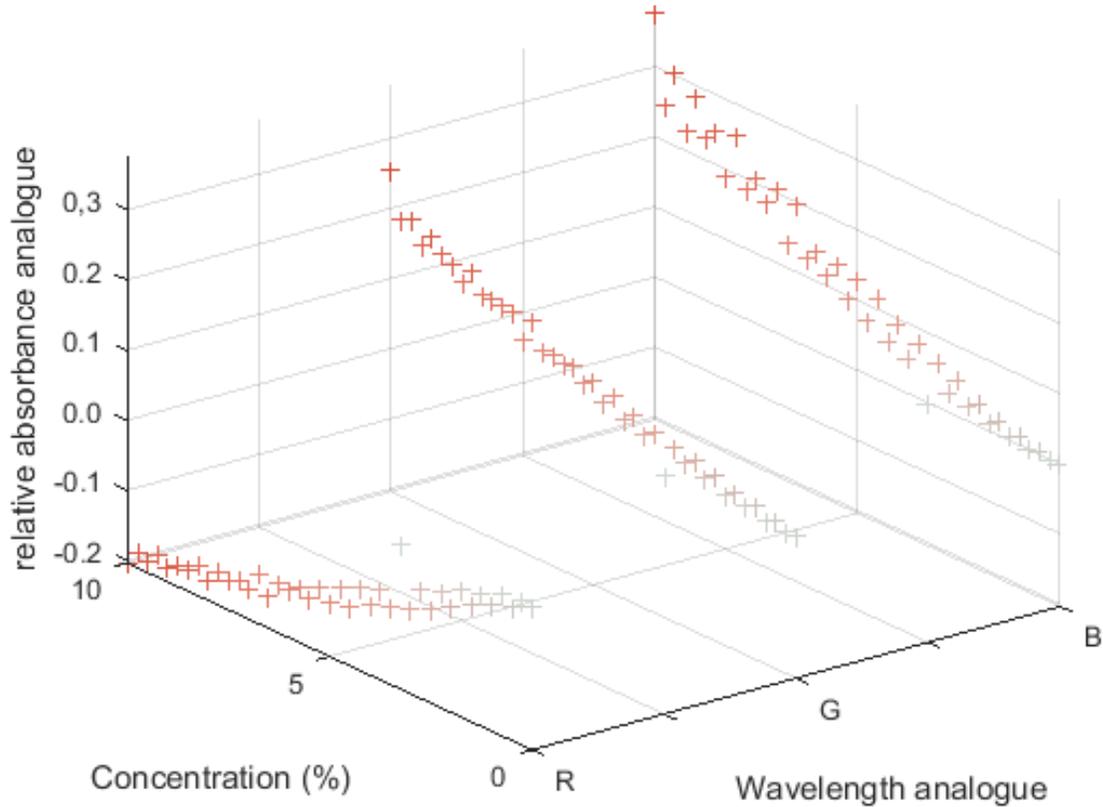


Figure 10: RGB-colorimetry data of the relative absorbance analogue through raspberry juice in water. Concentrations ranging from 0.25-10%

### 3.1 An even more simplified colorimetry approach

A very simplified protocol for determining concentrations colorimetrically has been suggested by [4], proposed to teach Beer-Lamberts law to high school and college students. In this method the RGB values are not normalized, and only the RGB value that is affected the most is used to calculate the relative absorbance analogue using Beer-Lamberts law. A linear relation between the relative absorbance analogue and the concentration is then used to predict unknown concentrations. In the case of raspberry juice, the absorbance is such that the blue value is affected the most. Figure 11 shows how the blue relative absorbance analogue is affected by the concentration. It also highlights the periodic feature caused by the two inconsistent measurement series. The fit was made using linear regression, with  $R^2 = 0.9208$  and root mean squared error  $RMSE = 0.853\%$ .

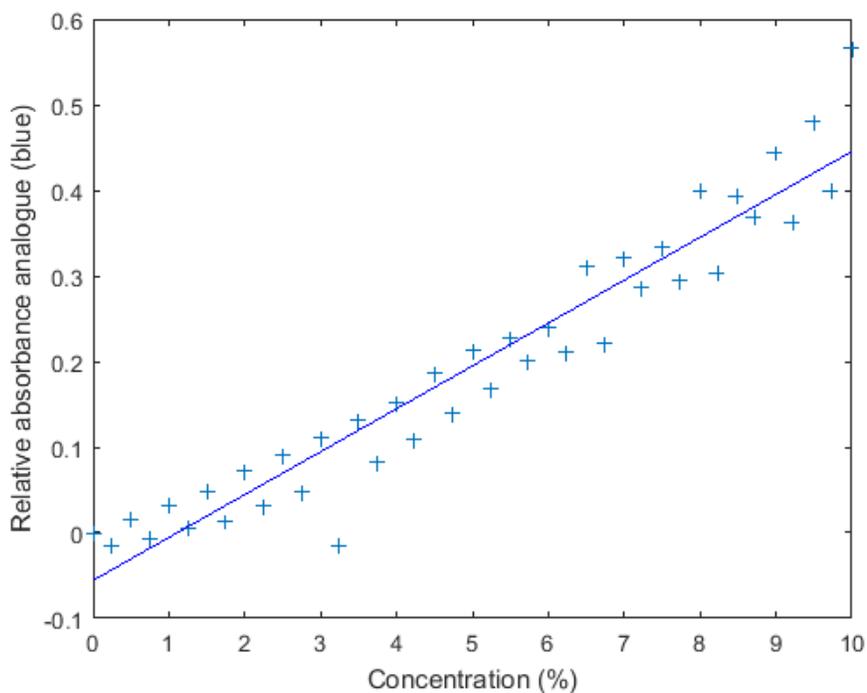


Figure 11: Scatter plot of the absorbance analogue calculated from the blue colorimetry values at 40 different concentrations, with a linear regression fit.

### 3.2 Data analysis

A PLSR analysis was performed on the CTS absorbance spectra in Figure 7, and one component was adequate to account for 99.17% of the variance in the response, i.e.  $R^2 = 0.9917$ . The mean error of prediction is  $MEP = 0.287\%$ .

A similar PLSR analysis was performed using all three values of the relative absorbance analogue from the RGB measurement in Figure 10. Three components accounted for 96.03% of the variance in the response, i.e.  $R^2 = 0.9603$ . In this case  $MEP = 0.703\%$ . Figure 12 shows the fitted concentration plotted against the actual concentration for both the CTS and the RGB measurements. This gives an indication of the goodness of the models.

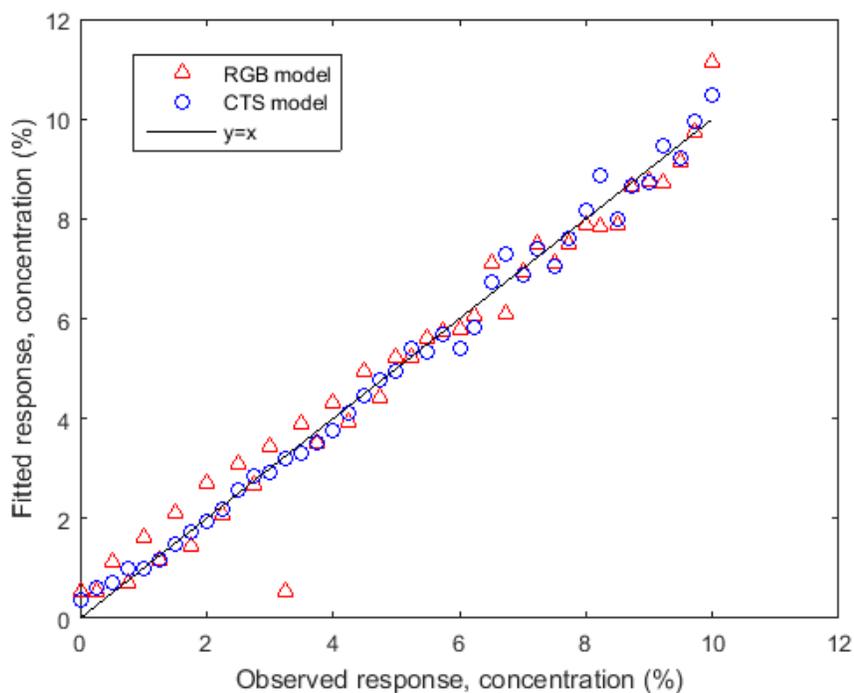


Figure 12: Fitted concentration of the PLSR models plotted against the actual concentration

For the simplified colorimetry approach the fit was made using linear regression, with  $R^2 = 0.9208$  and root mean squared error  $RMSE = 0.853\%$ .

### 3.3 Commercial fruit drink

Both the CTS and the RGB colorimetry methods were used to assess the fruit juice content in a commercially available raspberry fruit drink concentrate. The fruit drink contained juice only from raspberry, and no added coloring agents. With the recommended dilution 3:1, the fruit drink contains 9.3% raspberry juice according to the manufacturer, so the fruit content of the concentrate should be 37.2%. Both the CTS and the RGB methods were used on ten differently diluted samples, including five very dilute samples, to investigate the detection limits of the techniques. The dilutions and the results of the concentration determinations are presented in Table 1, and in Figure 13.

Table 1: Different concentrations of fruit drink concentrate, as well as the measured fruit content using CTS, combined RGB and “blue only” absorbance analogue values respectively.

Dilution (concentrate/water)	Fruit content (%)	CTS (%)	RGB(%)	B (%)
1/400	0.09	0.47	0.70	1.15
2/400	0.19	0.49	0.77	1.12
3/400	0.28	0.59	0.86	1.19
4/400	0.37	0.63	0.80	1.22
5/400	0.47	0.69	0.96	1.22
20/400	1.86	1.51	1.77	1.46
40/400	3.72	2.58	3.07	1.96
60/400	5.58	3.79	4.52	2.60
80/400	7.44	5.11	6.31	3.49
100/400	9.30	6.15	7.76	4.06

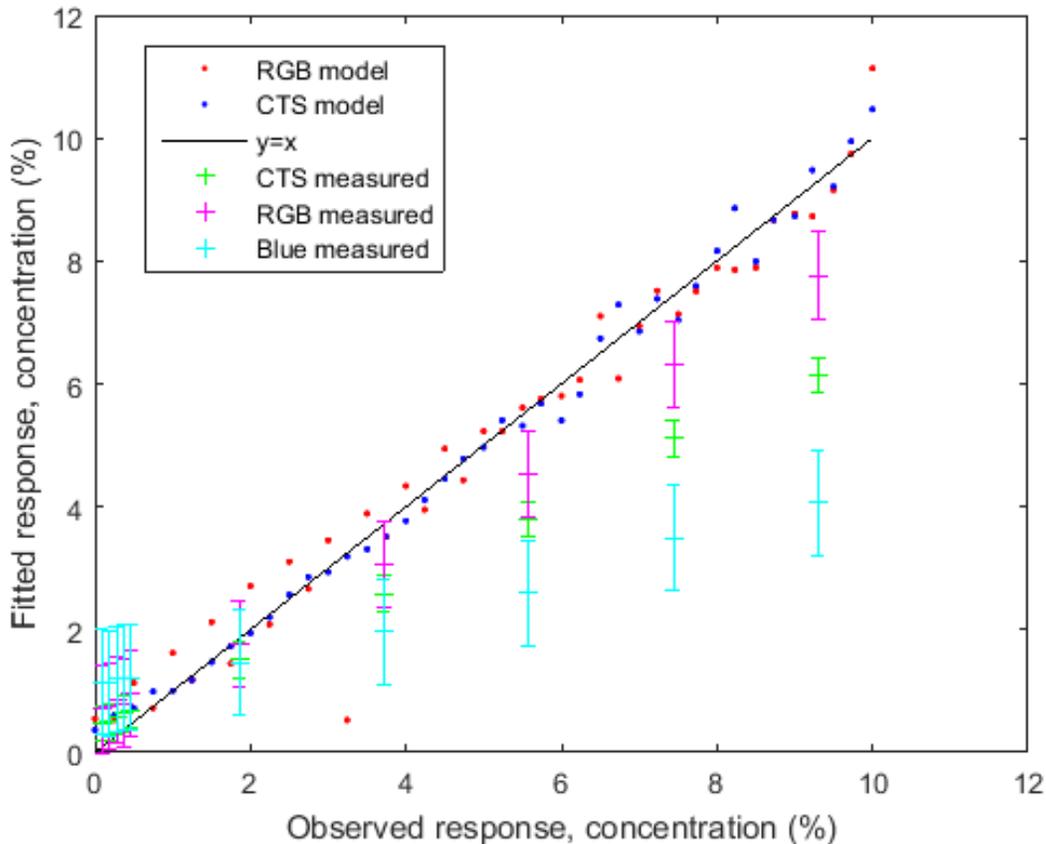


Figure 13: Predicted fruit content of the concentrate for ten different concentrations, and three different methods of concentration determination, CTS, combined normalized RGB analysis and “only blue” analysis.

Since the error margins are the same for different concentrations, they will scale differently

when the original concentration (fruit content of the concentrate) is calculated from the measurements, i.e. the samples with low concentration will produce larger errors when they are translated into the full concentrations. This is illustrated for all three methods in Figure 14.

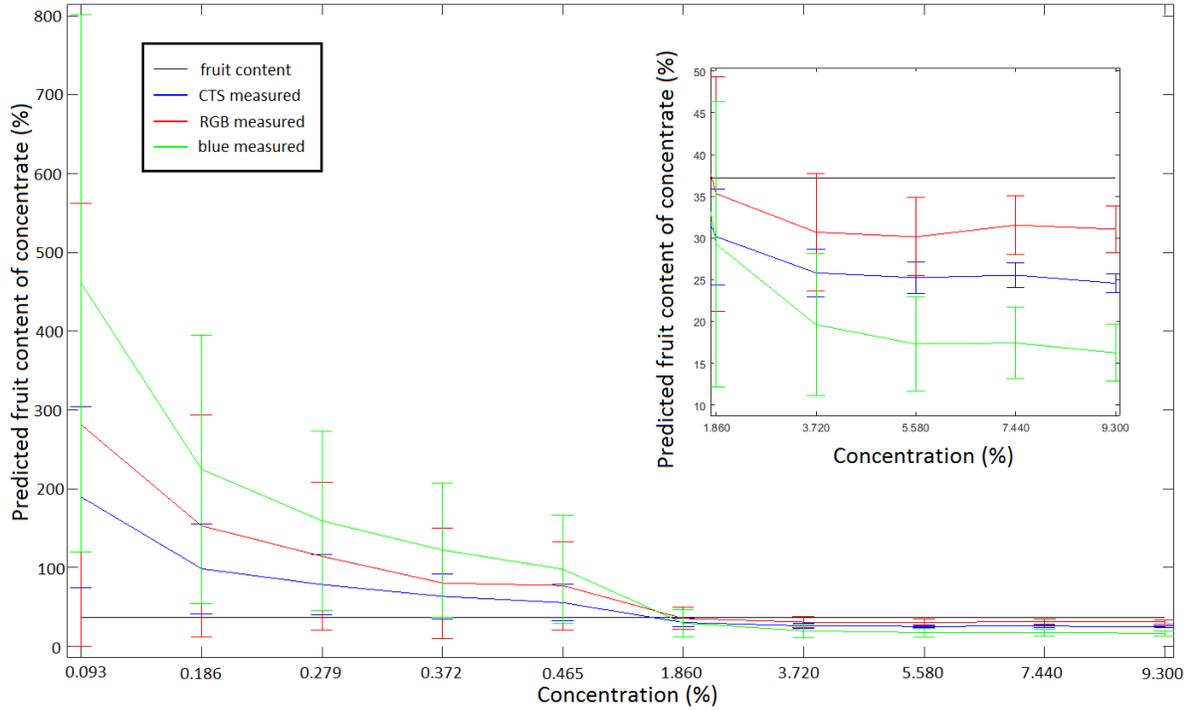


Figure 14: Predicted fruit content of the concentrate for ten different concentrations, and three different methods of concentration determination, CTS, combined normalized RGB analysis and “only blue” analysis. The inlay shows a magnified view of the five last samples. Observe that the concentration axis is not scaled properly, for increased readability.

## 4 Discussion and outlook

The results presented in Table 1, Figure 13 and Figure 14 all indicate that the lower concentrations tend to be overestimated, and the higher ones underestimated. This is observed for both the CTS technique, and for the two colorimetry methods. A comparison between the transmission spectra of pure diluted raspberry juice and the fruit drink reveal that they do have quite similar optical profiles as expected (see Figure 15). A non-linear effect seems to be more prominent for concentrations below  $\sim 5\%$  concentrate, while higher concentrations give more consistent results. A reason for this may be that the calibration data did not include such low concentrations, and the models therefore missed an important effect that only present for low concentrations. The physical reason for this behavior is not clear, since normally non-linear effects become more important for higher concentrations, due to e.g. changes in refractive index. The predictive power of the models outside of the concentrations used for their calibration is weak in this case, and efforts should always be made to include an exhaustive set of samples for the model calibration, as to not miss possible effects outside of the modeled region.

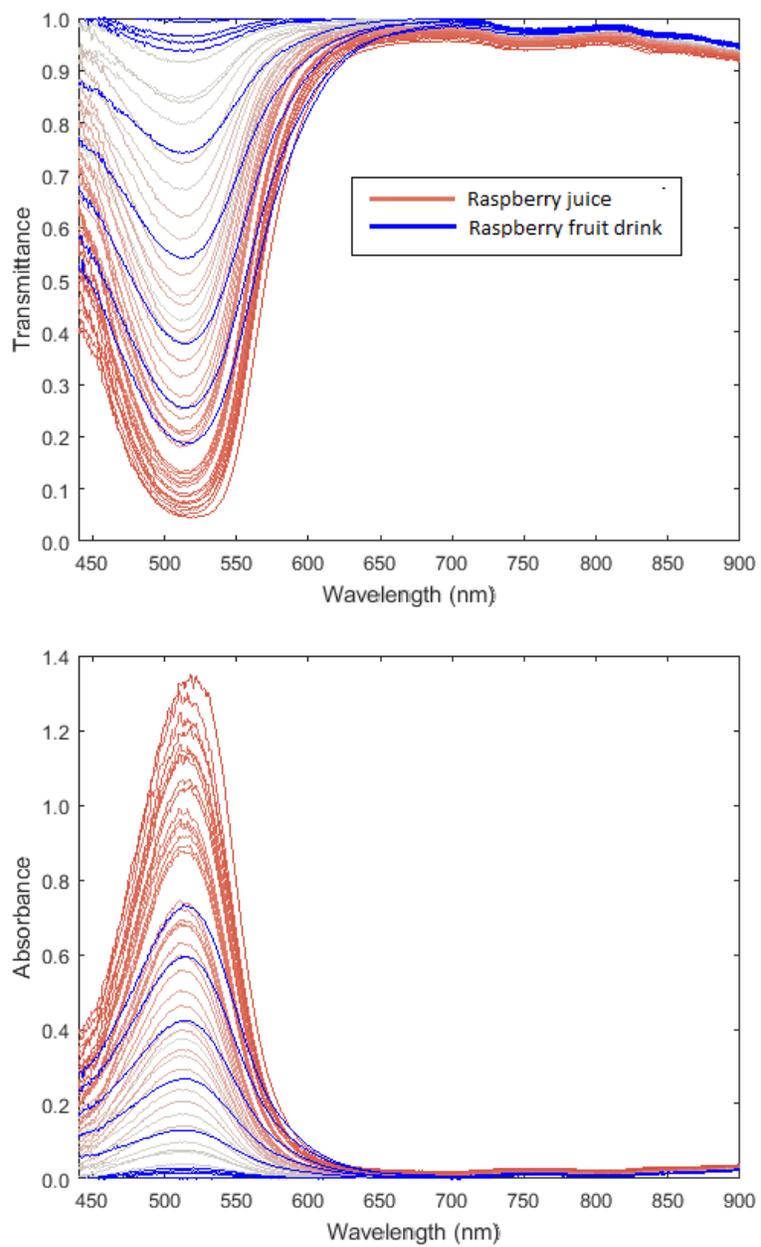


Figure 15: Transmittance and absorbance spectra of pure diluted raspberry juice from 0.25 to 10% (red), and diluted raspberry fruit drink from 1.9 to 9.3% claimed fruit content (blue).

The CTS method in combination with PLSR analysis should give the most reliable results, considering the number of data points included in the model, which is also reflected in the smaller error. It is also a tried and tested technique for these kinds of measurements as it has been used successfully to measure e.g. hemoglobin concentrations in blood with standard errors of prediction of only a few percent [15]. As for the colorimetry, the combined RGB analysis with PLSR gives larger errors than CTS, but smaller than the simplified “only blue” analysis, which was as expected.

In conclusion the sample that is tested should have a concentration within the range of concentrations used for creating the model. Also the scaling of the errors make samples which are very diluted more unreliable than samples that are not, in cases where the original undiluted concentration is to be estimated. The results above confirm that CTS is a more consistent and accurate technique than the two colorimetry methods in terms of reduced errors, but the results also suggest that RGB colorimetry can be used to give a rough estimate, using only a smartphone which is a readily available tool for many people. As for the fruit drink that was tested, it seemed to contain less anthocyanins than advertised, but the different methods gave inconclusive results, so not much can be said of the actual fruit content. Even though the combined RGB technique came closest to predicting the advertised fruit content, it may not be the better of the three techniques.

A reason for an unexpected measured fruit content could be that some anthocyanins have decayed/oxidized since the production of the fruit drink, or that the raspberries were harvested under different conditions than those used for the model. A more complete model could be made by including measurements of raspberry juice from different harvests and at different stages of aging. The model could also be expanded by including a wider range of concentrations, and more data points to reduce the MEP. It should also be noted that raspberry juice which only has one significant absorption peak may not be the best suited candidate for CTS, and that a compound with more absorption features could potentially provide a more accurate model.

To increase the flexibility and consistency of, and to further validate the CTS technique, it is important to be able to measure the absorbance of different solvents using the same equipment as for the principal measurements. This was attempted for water in section 2.5, but due to some unknown effect the results were not usable. Solving this issue and expanding to other solvents, e.g. ethanol, would be a natural next step for this work.

The colorimetry method turned out to be quite sensitive to changes in the environment. This is especially apparent in Figure 11, where every other measurement is shifted due to slightly different lighting conditions between two measurement series. Ideally a test chamber/cuvette holder should be constructed for the colorimetry measurements, similar to the one used in the CTS setup, where the light-source and ambient light is strictly controlled.

## References

- [1] Thorne A, Litzén U, and Johansson S. Spectrophysics: Principles and applications. *Lund: Media-Tryck*, 2007.
- [2] Lee D. Nature's palette: The science of plant color. *University of Chicago Press*, 2007.
- [3] Kuntzleman T. S and Jacobson E. C. Teaching beer's law and absorption spectrophotometry with a smart phone: A substantially simplified protocol. *J. Chem. Educ*, 93(7):1249–1252, 2016.
- [4] Confessor M. R Moraes E. P and Gasparotto Luiz H. S. Integrating mobile phones into science teaching to help students develop a procedure to evaluate the corrosion rate of iron in simulated seawater. *J. Chem. Educ*, 92(10):1696–1699, 2015.
- [5] Huang H. W and Zhang Y. Flame colour characterization in the visible and infrared spectrum using a digital camera and image processing. *Meas. Sci. Technol.*, 19(8):085406, 2008.
- [6] Krane K. S. Introductory nuclear physics. *Wiley*, 1988.
- [7] Bruker. Nmr foodscreener - sgf profiling module for the analysis of juices. Available from: <https://www.bruker.com/products/mr/nmr/food-screener/juice-profiling/overview.html>, 2016.
- [8] The Linde Group. High performance liquid chromatography (hplc). Available from: [http://hiq.linde-gas.com/en/analytical\\_methods/liquid\\_chromatography/high\\_performance\\_liquid\\_chromatography.html](http://hiq.linde-gas.com/en/analytical_methods/liquid_chromatography/high_performance_liquid_chromatography.html), 2016.
- [9] Obón J.M, Díaz-García M.C, and Castellar M.R. Red fruit juice quality and authenticity control by hplc. *Journal of Food Composition and Analysis*, 24(6):760–771, 2011.
- [10] Harvey D. Analytical chemistry 2.0. available from: [https://www.researchgate.net/publication/50927915\\_Analytical\\_Chemistry\\_20](https://www.researchgate.net/publication/50927915_Analytical_Chemistry_20), 2012.
- [11] Liu D. Space of spectral sensitivity functions for digital color cameras. *Rochester Institute of Technology, Center for Imaging Science*, available from: <http://www.cis.rit.edu/~dxl5849/projects/camspec/>, 2013.
- [12] Nachabé R. et al. Validation of interventional fiber optic spectroscopy with mr spectroscopy, mas-nmr spectroscopy, high-performance thin-layer chromatography, and histopathology for accurate hepatic fat quantification. *Investigative Radiology*, 47(4):209–216, 2012.
- [13] S. Kedenburg, M Vieweg, T Gissibl, and H. Giessen. Linear refractive index and absorption measurements of nonlinear optical liquids in the visible and near-infrared spectral region. *Optical Materials Express*, 2(11):1588–1611, 2012.
- [14] Wold S. Sjöström M. and Eriksson L. Pls-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130, 2001.

- [15] Kim Y.J. et al. Data preprocessing and partial least squares regression analysis for reagentless determination of hemoglobin concentrations using conventional and total transmission spectroscopy. *Journal of Biomedical Optics*, 6(2):177–182, 2001.