Linearization of Spectral Reflectance

Report Modelling week - Canon
Group 5

Milo Goolkate, Luuk Hofmeijer, Tuur Willio, Tijmen van der Zwaag

November 2024

1 Problem Statement

To support the ongoing advancements in appliances and applications within the printing industry, we seek to deepen our understanding of spectral reflectance. In the following report, we consider printers that use a CMYK ink set, which contains cyan, magenta, yellow and a key plate, which is commonly just black ink. This is a standard system used to print the desired colours of an image [SB17]. Various colour configurations are evaluated, using these inks, by measuring their spectral reflectance over different wavelengths. This spectral reflectance is a measure of how much light gets reflected by the paper[WC16]. Using different types of print results in a distinct spectral reflectance measure for the same CYMK value. The goal is to find a mapping F of the CYMK ink values to a corresponding spectral reflectance, that is,

$$F : C \times Y \times M \times K \longrightarrow R$$

here R is the spectral reflectance. More specifically, we are looking for a transformation T such that $T(R)$ is a linear function of CMYK values. We do this by finding the inverse mapping F^{-1} from R to the CYMK values. A conventional way, and we will see reliable way, to achieve this is by applying the $T = -\log(R)$ transformation on the spectral reflectance. In this report, we investigate the situation in which this $-\log$ transformation does not perform well. For these situations, we introduce other transformations which perform better than the $-\log$ transform. An additional goal is being able to distinguish between paper types in a mathematical model, and more specifically between coated and uncoated papers.

1.1 Data

The data provided consists of measurements of the spectral reflectance on five different paper types and different combinations of CMYK ink values, ranging from 0 to 100 in steps of around 12.5 for all four different ink colours. Some combination of colours results in too much ink on a paper which means the paper might not be able to quickly absorb all the ink. These combinations are called saturated and behave differently from the other data points. To avoid complications, these data points have been removed. Spectral reflectance ranges from 0 to 1, where a reflectance of 0 means no light gets reflected, and a reflectance of 1 means all the incident light gets reflected. However, data might spill over one. We will not go further into this behaviour and assume it as if it were one. The spectral reflectance is measured by testing different wavelengths. These wavelengths range from 380 to 730 in steps of 10 nm. Data is provided in the following format:

The data was provided by Canon Production Printing. Their goal, with a model as described

CMYK_C	CMYK_M	CMYK_Y	CMYK_K	SPECTRAL_NM380	SPECTRAL_NM390	...
83.5300	50.2000	50.2000	25.1000	0.0491	0.0491	
0.0000	60.0000	60.0000	37.6500	0.0397	0.0397	
85.8800	85.8800	42.7500	12.5500	0.0650	0.0650	
...						

Table 1: The table shows how the received data is structured. The first four columns represent the amount of ink applied to the paper of the colours cyan, magenta, yellow and black. These are values between 0 and 100, with 0 being no ink and 100 the maximal amount. After this follow 36 columns of spectral reflectance measurements for different wavelengths.

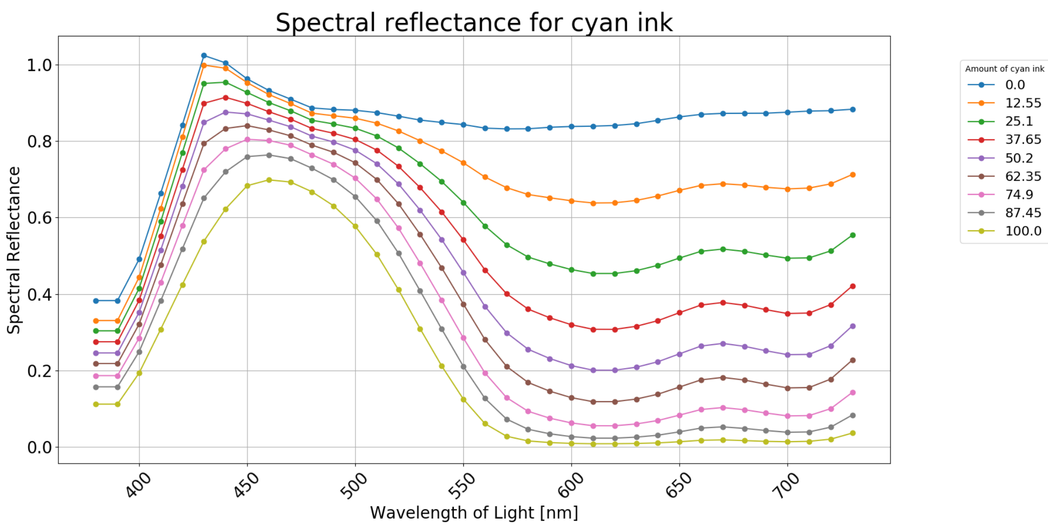


Figure 1: Spectral reflectance across wavelengths for different Cyan values, with the others colours fixed to zero.

above, is to find values for the amounts of ink that are required to produce a given spectral reflectance on a specific type of substrate. The example plot in Figure 1 is a first graph, which aims to visualise the data to get a sense of the behaviour of the data.

2 Single colour printing

To simplify the problem, we start by focusing on printing with a single colour. To print with a single colour we fix all but one ink amount of the colours to zero. The non-fixed colours ink value is what we vary. To get a better insight into the behaviour of the ink for certain wavelengths, we plot the spectral reflectance against the amount of ink for each wavelength. One such plot is shown in Figure 2. In this section we explore the model that we deduced from the plot in Figure 2 and discuss other models we have explored. But before we dive into the models, we introduce a weighted mean to be able to quantify and compare the different models.

2.1 Weighted mean

In order find out how linear the data is, we apply a linear regression model. Doing this yields an r^2 value for each wavelength. The r^2 value, also known as the coefficient of determination [BF10]. Let y_i be the value of the spectral reflectance, and f_i the value predicted by some model, for all

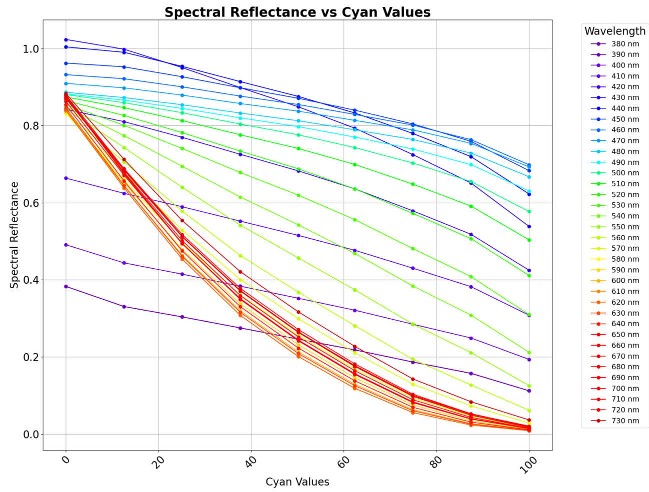


Figure 2: The plot shows the reflectance of the printed paper as a function of different amounts of ink that are applied to the paper. Each line corresponds to a different wavelength of light. Axes are flipped in comparison to Figure 1.

$i = 1, .., n$ points. Then we define the r^2 for a wavelength λ in the following way

$$r_\lambda^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i-1}^{n}(y_i - \bar{y})^2}. \tag{1}$$

Where the mean of the observed data is given by

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Note that an r^2 value of 1 would imply the model can predict the observations perfectly. Since the r^2 value is obtained by fitting a line through the data, the closer the r^2 value is to 1, the more linear the data is. Later in the report, we will also encounter negative values for r^2. Negative r^2 value implies that the model is a very poor fit for the data. Furthermore, since we apply a transformation per wavelength, this will also yield a r^2 value per wavelength.

But we would like to encapsulate the performance of the model in one number, not in a number per wavelength. To find one value that can function as a score for the model, one could simply take the mean value of these r^2 values. However, during our exploration of the data we noticed that non-linearity for a certain wavelength may be less consequential compared to other wavelengths. Consider for example the spectral reflectance of yellow ink as shown in Figure 3. One can see that for wavelengths 380 nm up to 480 nm the different amounts of ink impact the spectral reflectance a lot more than at wavelengths larger than 520 nm, since the data points are further away from each other in the first part. Therefore, having linearity in this first part matters more than in the later part. Thus, to compare the methods it will not be enough to take the mean of the r^2 values for all wavelengths. Instead, at wavelength λ we define

$$\Delta_\lambda = \max_{\text{amount of ink}}(\text{spectral reflectance}) - \min_{\text{amount of ink}}(\text{spectral reflectance}).$$

Then, for a certain model we define its score s as

$$s = \frac{1}{|\Lambda|}\sum_{\lambda\in\Lambda}(1 - r_\lambda^2)\Delta_\lambda$$

where Λ is the set of wavelengths considered and r_λ^2 is the r^2 value of the linear regression on the (transformed) data at wavelength λ, as defined in Equation 1. A score zero for s implies thus that all r_λ^2 are one, and we will have a perfect linear model, i.e. we want s as close to zero as possible.

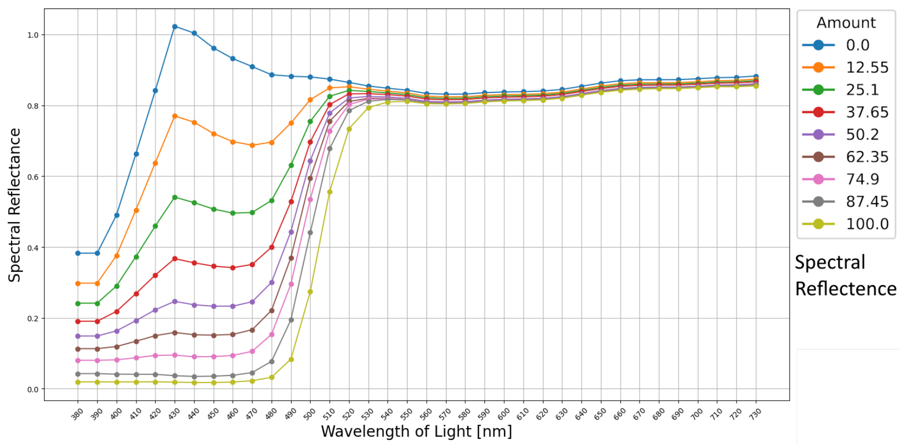


Figure 3: The reflectance is shown as a function of the wavelength of the light. Each line corresponds to a different amount of yellow ink.

2.2 Leaf model

When we looked at Figure 2, we noticed that there seemed to be three kinds of curves. Curves that seem to follow a e^{-x} function, curves that seem to follow a $1-e^{x}$ function and curves that are already linear. These three curves resemble a leaf as shown in Figure 4. This is the motivation for the name *the leaf model*. The existing model, which uses a $-\log(y)$ transformation, works well for

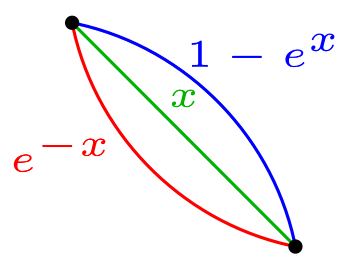


Figure 4: The characteristic curves that build up the leaf model.

the curves that follow the e^{-x} function, as this transform makes these curves linear. However, it does not help for the other types of curves. Our idea was to test the transform that corresponds to each part of the leaf in Figure 4 and finally use the transform best suited for each wavelength. Note that the transform is the inverse of the function that the curve follows. Putting it in mathematical terms we get,

$$x(y) = ay + b\log(1-y) - c\log(y),$$
$$x(y) : \lambda \to \{0,1\}$$
$$\text{where } (a,b,c) = (1,0,0) \vee (a,b,c) = (0,1,0) \vee (a,b,c) = (0,0,1)$$

Consider r_l^2, r_u^2 and r_b^2 to be the r^2 value corresponding to the linear, upper and lower transform respectively. Then a, b and c are chosen by

$$a := \begin{cases} 1 & \text{if } r_l^2 \geq r_u^2 \wedge r_l^2 \geq r_b^2, \\ 0 & \text{otherwise,} \end{cases}$$

$$b := \begin{cases} 1 & \text{if } r_u^2 > r_l^2 \wedge r_u^2 > r_b^2, \\ 0 & \text{otherwise,} \end{cases}$$

$$c := \begin{cases} 1 & \text{if } r_b^2 \geq r_u^2 \wedge r_b^2 > r_l^2, \\ 0 & \text{otherwise.} \end{cases}$$

However, since the spectral reflectance can be greater than 1, we have to alter the transform $\log(1-y)$. We considered the transform $\log(\alpha - y)$, for $\alpha \in [1, 2]$, for which we observed that for different inks the optimal value of α lies around 1.15. This, however, does not consider multicolour printing.

Figure 5 shows the r^2 values considering each of the three transformations for the leaf model for each wavelength. We see that the $-\log(y)$ model performs well overall. The performance seems to have a dip for the wavelengths in the interval $[420, 520]$. The $\log(\alpha - y)$ model performs contrary to the $\log(y)$ model. It performs really well in the interval $[420, 520]$, where the performance of the $-\log(y)$ model was less. Lastly the linear model is shown. This model performs the least consistent of the 3 models. But it performs well in the sections around the best performance of the $\log(\alpha - y)$. The bottom panel of Figure 5 shows the leaf model, it takes the best performing model of the 3 models shown in the upper panel. Note that the y-axis is different from the upper panel. The weighted mean of the r^2 is shown in Table 2.

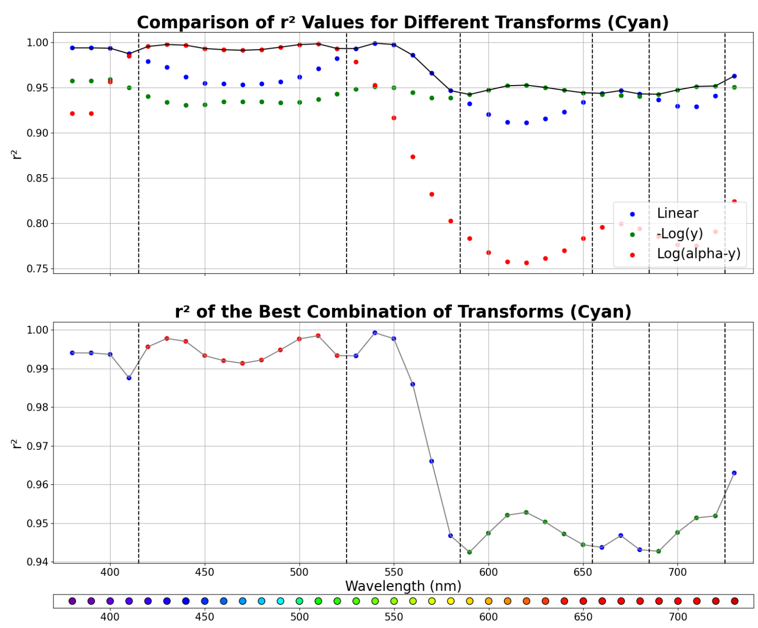


Figure 5: Performance of the leaf model for Cyan. Note the different scales on the y-axis. The bottom panel takes the best performing model for each wavelength. Which model is chosen is indicated by the colour.

We can then apply the leaf model to the other colours as well. However, as can be seen in Figure 6, the black ink behaves almost exclusively exponentially. We therefore just apply the already used $-\log$ transform. The leaf will not improve much further upon this colour.

Method	Weighted mean
linear	0.0299
$-\log(y)$	0.0332
$\log(1.15-y)$	0.0929
leaf	0.0213

Table 2: Weighted mean as introduced in subsection 2.1 per method for cyan. Values closer to 0 are better, with 0 being a perfect model.

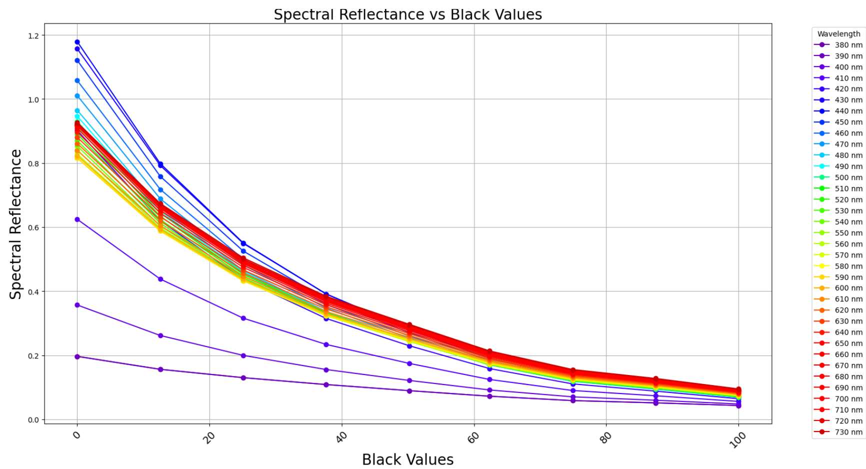


Figure 6: Spectral reflectance as a function of different amounts of black ink that are applied to the paper. Other colours fixed to zero.

2.3 Other models

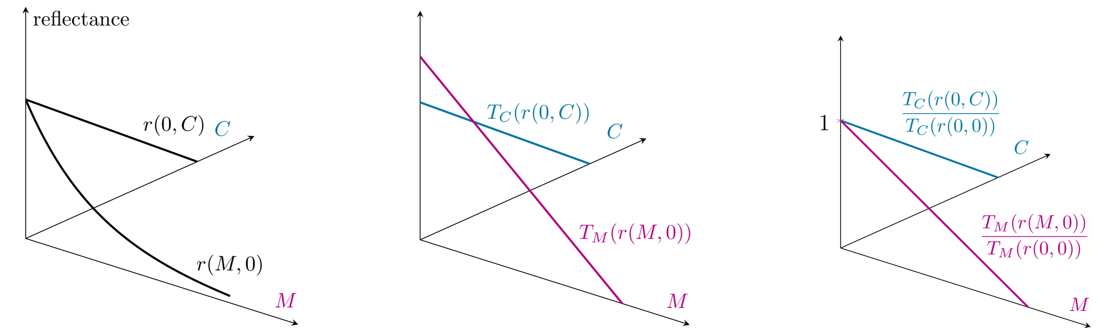
Besides the leaf model we also investigated ways to fit the curves with other family of functions. For example, we tried to fit quadratic functions to both types of curves, that is, $r = (ax+b)^2 + c$. Here r is the reflectance and x the colour value. This would yield the transformation $\sqrt{r-c}$ to obtain a linear model. However, in practice this gave rise to some practical problems such as negative values under the root. Other suggested models include $r = \frac{1}{ax}$, with transformation $\frac{1}{r}$, but these methods performed worse than the leaf model in terms of linearization. In the end, considering the limited time for this project, we decided to select the best performing method at the time, namely the leaf model, as the main model to further investigate.

3 Multicolour printing

Now that a decent model for the linearization of single inks has been developed, it is time to look at creating a model where multiple inks can be combined. The method that was already provided to us, simply applying $-\log(x)$ to every measured spectral reflectance x, is tested first. Furthermore a new method is developed that attempts to extend leaf model in multiple dimensions.

3.1 Two-colour leaf model

The two-colour leaf model takes two coordinates, say the amount of cyan ink C and the amount of magenta ink M, and uses the separate leaf model transformations T_C and T_M for cyan and magenta respectively. Note that these transformations can be one of the three parts of the leaf model, but also the combination of all three which would also make it a function of the wavelength. The reflectance is now a function of C and M, so we write $r(C, M)$ for this reflectance. This model is illustrated in Figure 7. Figure 7a shows the original data, without any transformation.



(a) Original reflectance data

(b) Two different surfaces appear when applying T_C or T_M

(c) The transformations are normalized such that they meet at 1 when no ink is applied.

Figure 7: The reflectance is plotted on the vertical axis as a function of the amounts of cyan and magenta ink, while a 2D transformation is built on it.

When applying the different transformation to the colours, we will get different linear results, as illustrated in 7b. To handle this, we rescale these models to meet in the point $(0, 0, 1)$ on the z-axis. This point makes sense in theory, because no ink implies a blank substrate, assumed a type of white, which gives the best reflection. Note that this is in actuality really dependent on the type of paper, however we will not go deeper into different types of paper.
Lastly, we would like all plots to be normalized between 0 and 1. This is helpful since it becomes easier to find a new model for how the inks interact, when all of these two-colour plots have a similar shape. This is currently not the case, since a lower-leaf curve becomes a line with a positive slope when the logarithmic transformation is applied. This is fixed by taking the transformation Lastly, it becomes easier to find a new model for how the inks interact, when all of these two-colour plots have a similar shape. This is not the case currently, since a lower-leaf curve becomes a line with a positive slope when the logarithmic transformation is applied. Therefore, we would like all plots to be normalized between 0 and 1. This is fixed by taking the transformation

$$T(C, M) = \frac{C}{C+M}\left[\frac{T_C(r(C,M) - T_C(r(0,M))}{T_C(r(0,0)) - T_C(r(0,M))}\right] + \frac{M}{C+M}\left[\frac{T_M(r(C,M)) - T_M(r(C,0))}{T_M(r(0,0)) - T_M(r(C,0))}\right].$$

Note that this kind of transformation can also easily be extended to even more colours. We can again then check the r^2 of this model for two colours and compare it to the original data and the original transform $-\log$. In Figure 8 the r^2 of the original and the $-\log$ transformation are shown. This figure shows that this transformation already greatly improves the linearity as compared to the original data. In Figure 9, the r^2 of the leaf model in two dimensions is plotted. As can be seen in this last plot, it performs well only for wavelengths below 540 nm. Remember that we introduced a weighted mean metric for this reason. Yellow, for example, gives barely any difference in spectral reflectance from around wavelength 520 nm to the end of the visible spectrum. We will thus ignore any linearity in this last part and also in our model. Hence, an improvement in the linearity metric for wavelengths between 380 nm and 500 nm indicates the two-colour leaf model is performing better than the simple $-\log$ transformation in those parts of the spectrum that are import for this ink colour. Accepting this, it is clear that the leaf model in two dimensions also improves upon the original transformation. We can also plot the r^2 for other colour combinations, as can be seen in Appendix A.

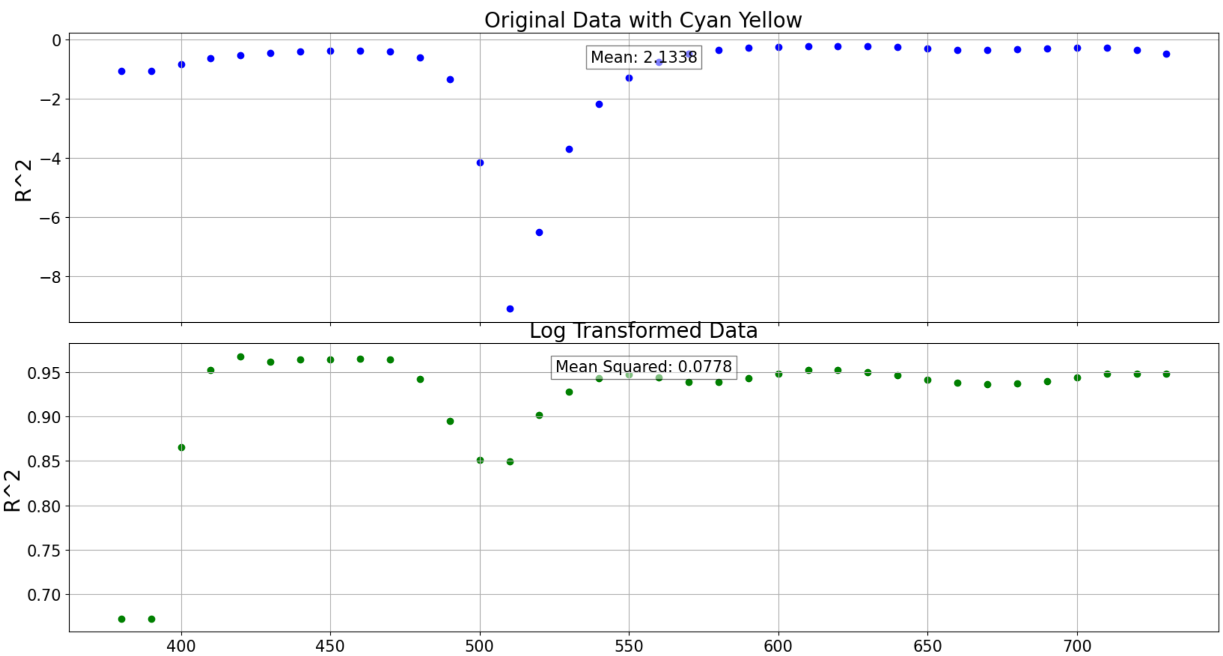


Figure 8: Cyan & Yellow original data and $-\log$ transform

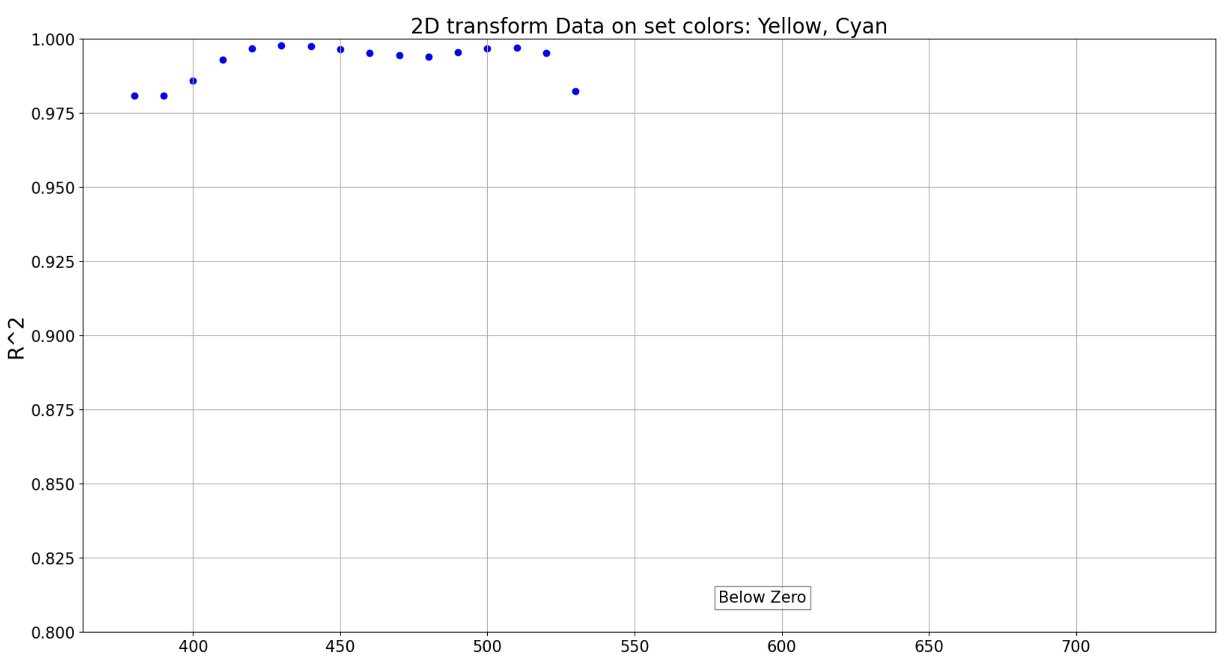


Figure 9: Cyan & Yellow r^2 leaf model

4 Conclusion

We are interested in creating a linear interpolation model to approximate spectral reflectance. However, this reflectance is a nonlinear function of the CMYK values and wavelength. In this project, we modelled the spectral reflectance in order to find a transformation such that the transformed reflectance is linear with respect to CMYK. Our approach was to identify the reflectance domains for which the function had the same curvature. Using this approach, a different transformation could be applied to each of these domains to obtain a linear model. The linearity of the result after transformation was quantified using the standard and weighted r^2 metric. Our proposed leaf model is an improvement on the existing $-\log$ transform, a standard in this field, in terms of linearisation in one dimension. For the colours cyan, magenta, and yellow, our transformation result was more linear than the $-\log$ transform. The spectral reflectance for Black was already very exponential and, therefore, $-\log$ was already a good choice for the transformation.
Having improved upon the one-dimensional model, we tried to extrapolate our knowledge to two dimensions. We did this by taking a normalised combination of our one-dimensional model. With the end of the project approaching, we could not develop this idea much further. However, we believe that this idea, when worked out, is easily implemented into all dimensions, that is all the CYMK values. We have improved upon the existing model, but in general $-\log$ is not a bad choice of transform.

5 Future work

There are many alternative approaches and unexplored ideas that could be investigated in a future project.

Most importantly, one should explore multiple ways of fitting the one-dimensional model to obtain transformations that could fit the data better than our proposed leaf model. A better insight in one dimension might also give better insight in multiple dimensions.

Secondly, we would have further worked out our idea for extending the leaf model from a single colour to multiple colours. This, as mentioned, is similar to the implementation in two dimensions. Other possible methods include the approach of modelling multiple dimensions from the beginning, instead of using one-dimensional models, and using those to extend into more dimensions. Alternative ways to extend from one dimension would also be of interest.
Our proposed leaf model also requires more testing and a performance analysis, compared to other models, as described above.

Lastly, while we investigated the influence of different papers on our leaf model on a surface level, a larger analysis is required to get a better insight on the performance of our model and the behaviour of different types of paper.

A Additional error plots for the two-colour leaf model

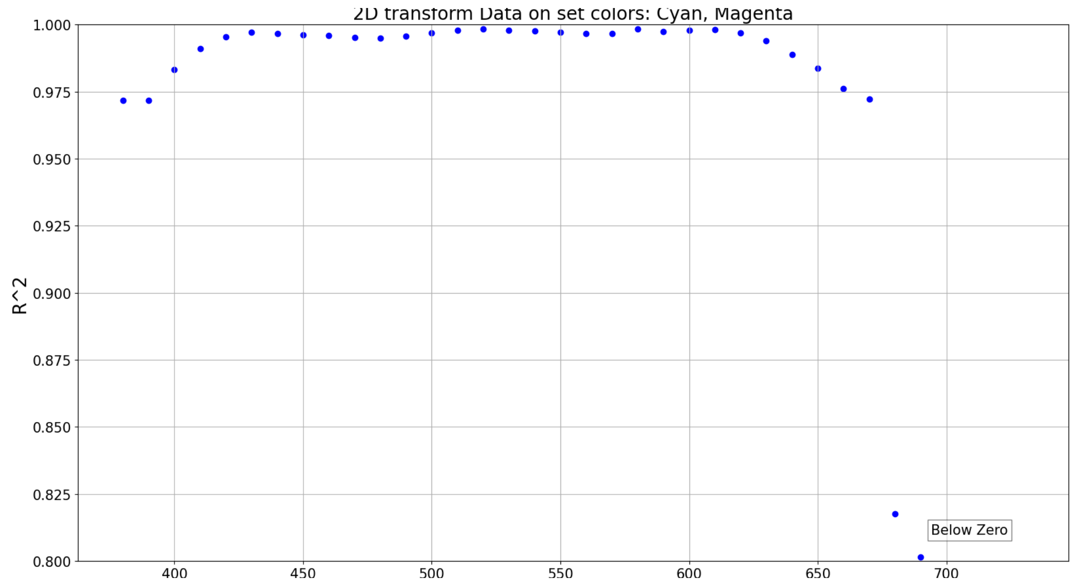


Figure 10: Cyan & Magenta r^2 leaf model

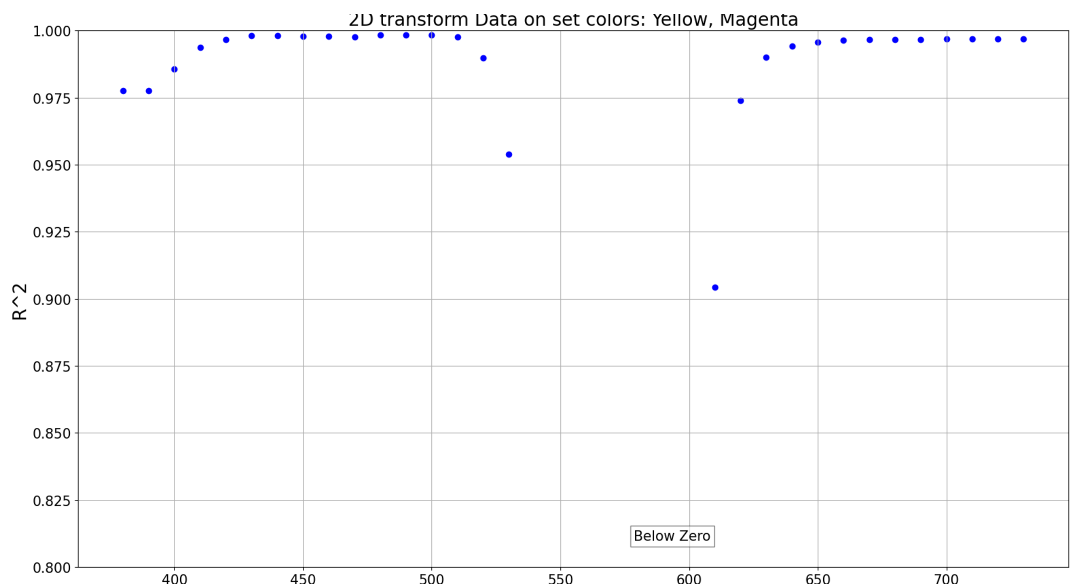


Figure 11: Cyan & Magenta r^2 leaf model

References

[BF10] N. H. Bingham and John M. Fry. *Regression: Linear models in statistics*. Springer, 2010.

[SB17] Gaurav Sharma and Raja Bala. *Digital Color Imaging Handbook*. CRC Press, Dec. 2017. ISBN: 9781420041484.

[WC16] Stephen Westland and Vien Cheung. "CMYK systems". In: Jan. 2016. DOI: 10.1007/978-3-319-14346-0_13.