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. 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. Using different types of paper could also result in different results for this measure. The goal is to find a mapping of the CYMK ink values to a corresponding spectral reflectance.

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

here R is the spectral reflectance. More specifically, we interpolate experimentally found data and use a transform T to map it to R.

$$M : (c, y, m, k) \longmapsto T(linearInterpolation(Data)(c, y, m, k))$$

This transform T serves to make the spectral reflectance more linear w.r.t. the CYMK values. We will focus on the latter, transform T. We do this by finding an inverse mapping T^{-1} from R to the CYMK values,

$$T^{-1}(R) \approx linearInterpolation(CMYK\_Data)$$

A conventional way, and we will see reliable way, to achieve this is by making use of a $-\log(y)$ transform on the data, that is the spectral reflectance. An extra 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:

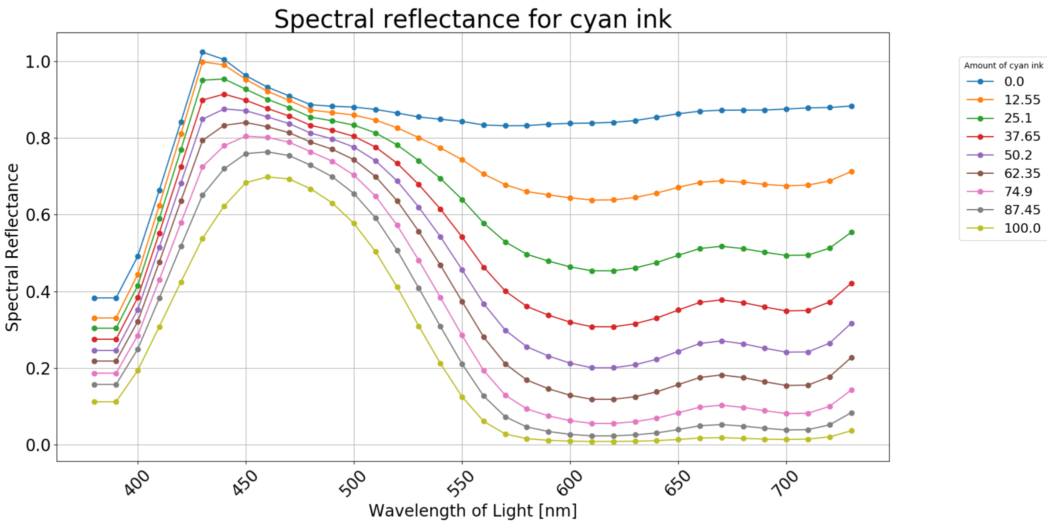


Figure 1: Spectral reflectance across wavelengths for different Cyan values, with the others colours fixed to zero.

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	
...						

The example plot in Figure 1 is a first graph, visualizing the data, to get a sense of the data behaviour.

2 Single colour printing

To simplify the problem, we start by focusing on printing with a single colour. 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. To then 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 a non-linearity for a certain wavelength may be less consequential than that for another. 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

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

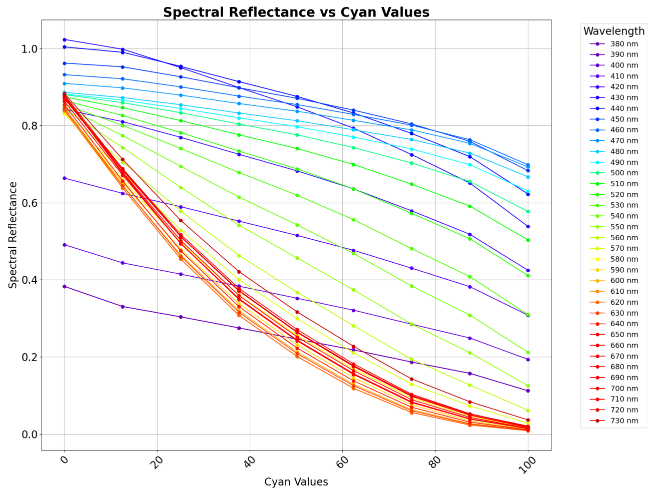


Figure 2: The plot shows the reflectance of the printed paper as a function of different amounts of ink that are applied to the paper. Each line corresponds to a different wavelength of light.

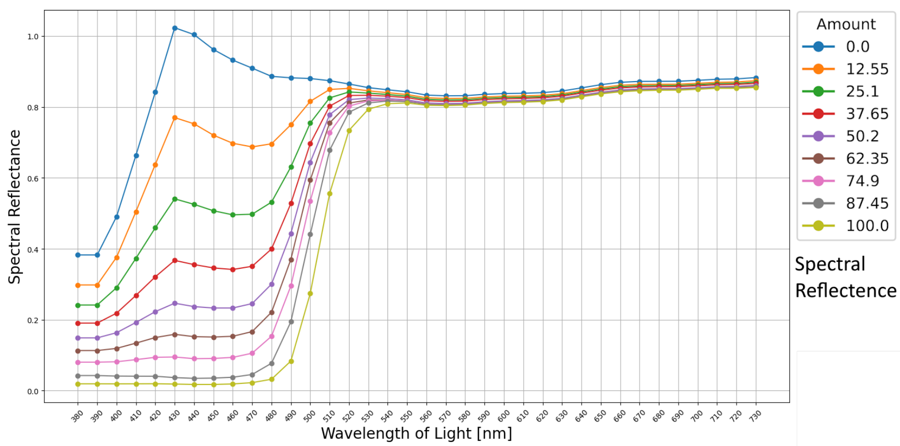


Figure 3: The reflectance is shown as a function of the wavelength of the light. Each line corresponds to a different amount of yellow ink.

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. 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.

2.2 Leaf model

When we looked at the leaf plots, 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 the curves that follow the e^{-x} function, as this transform makes these curves linear. However, it

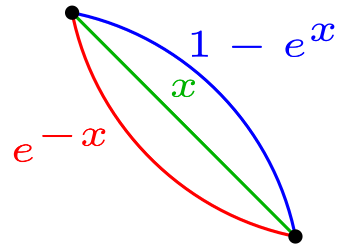


Figure 4: The characteristic curves that build up the leaf model.

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,

$$\begin{aligned}
&x = ay + b\log(1-y) - c\log(y),\\
&\text{with } a, b, c : \Lambda \to \{0, 1\},\\
&\text{s.t. } (a+b+c)(\lambda) = 1, \quad \forall \lambda \in \Lambda.
\end{aligned}$$

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

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

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 lies around 1.15. This, however, does not consider multicolour printing.

When we apply our method, just the $-\log(y)$, linear or $\log(1.15 - y)$ transforms and plot the r^2 values, we get the plot shown in Figure 5. The weighted mean of the r^2 is shown in Table 1. We see that the $\log(1.15 - y)$ performs well where the linear and $-\log(y)$ transform perform worse. Furthermore, the leaf model performs well in general. Note that on the bottom graph, in Table 1, the y-axis is rescaled and the plot is really the combination of methods as shown in the top graph.

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

Table 1: Weighted mean per method for cyan.

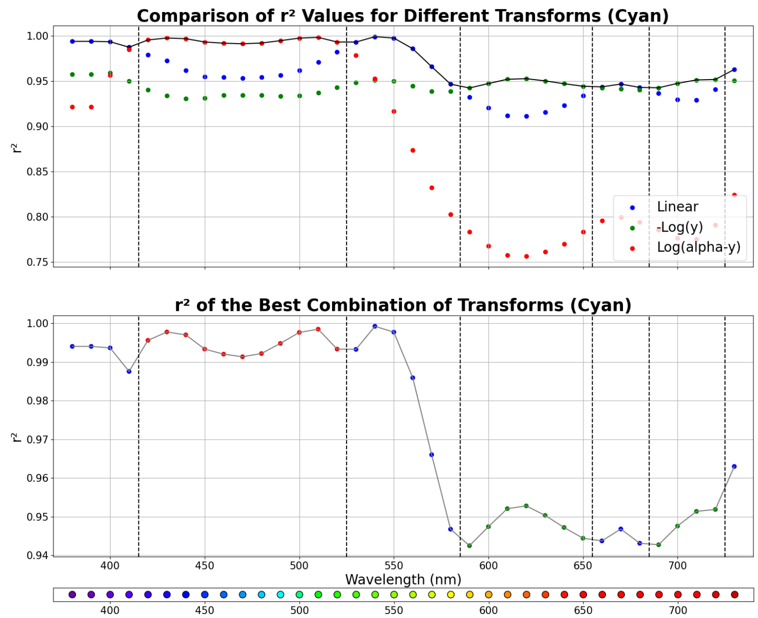


Figure 5: Performance of the leaf model for Cyan.

2.3 Other models

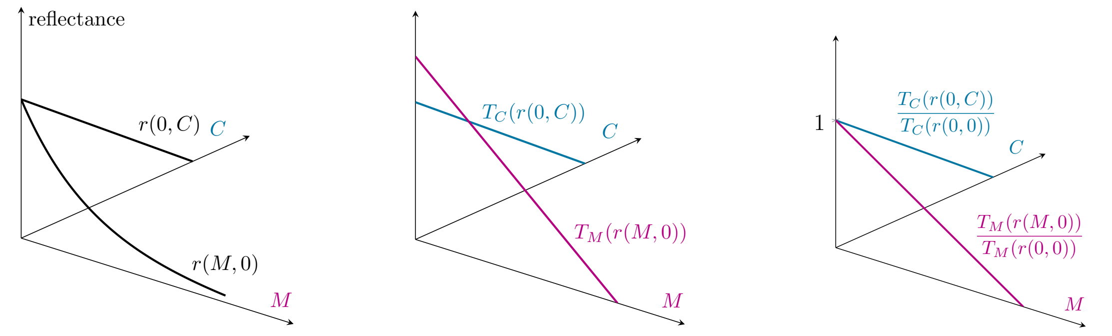
Besides the leaf model we also investigated ways to fit the curves with other family of functions. For example, we tried to fit quadratic functions to both types of curves, that is, $r = (ax + b)^2 + c$. Here r is the reflectance and x the colour value. This would yield the transformation $\sqrt{r - c}$ 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. 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 6. Figure 6a shows the original data, without any transformation. When applying the different transformation to the colours, we will get different linear results, as illustrated in 6b. 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.



(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 6: 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.

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 7 the r^2 of the original and the $-\log$ transformation are shown. In Figure 8, the r^2 of the leaf model in two dimensions is plotted. As can be seen in this last plot, it performs well on only one side of the spectrum. 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, the results only being good before this point in wavelength is justified. 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.

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

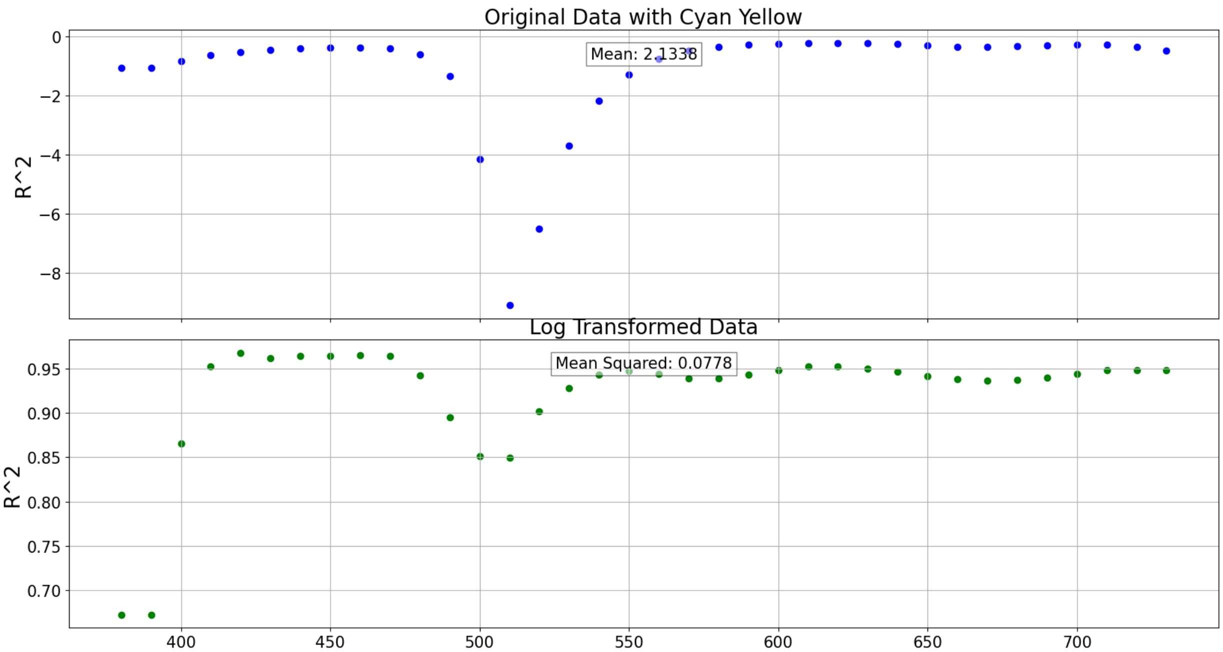


Figure 7: Cyan & Yellow original data en $-$log transform

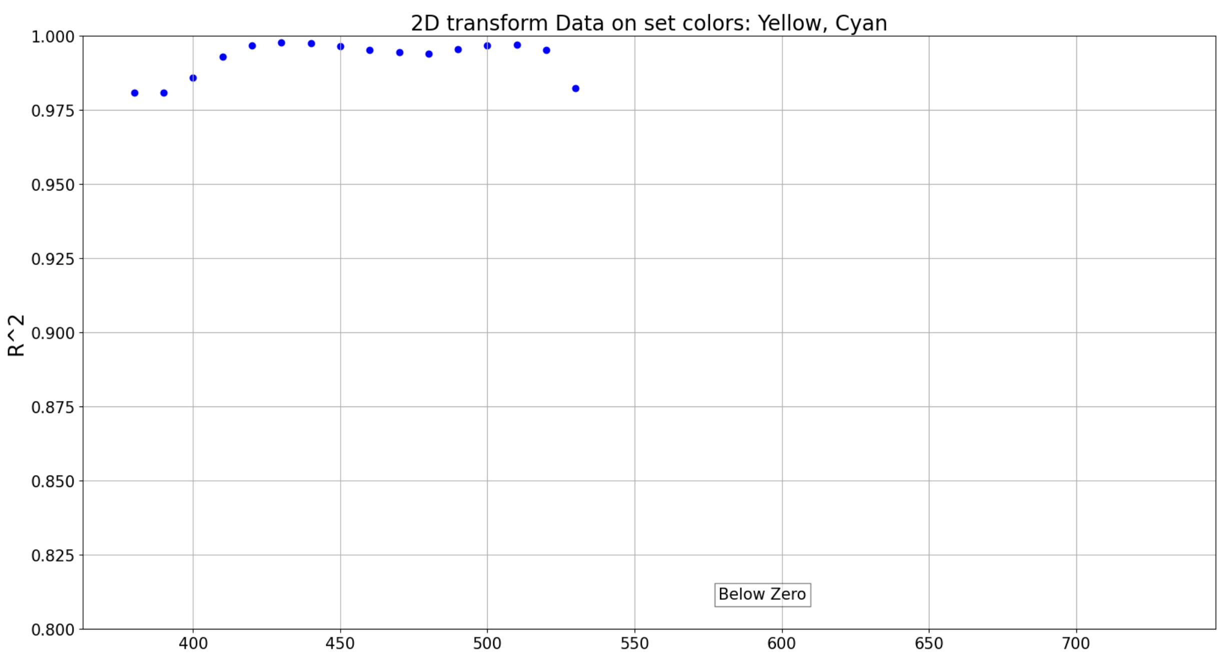


Figure 8: Cyan & Yellow r^2 leaf model

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 behaved 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 overall is the $-\log$ not a bad choice of transform.

5 Future work

Due to limited time assigned to this project, 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. Moreover, with additional time, 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 including the approach of modelling multiple dimensions from the beginning, or alternative ways to extend from one dimension would be of interest. Through this approach, we could look further into the performance of our dimension extension and compare that to the performance of other suggested approaches. 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

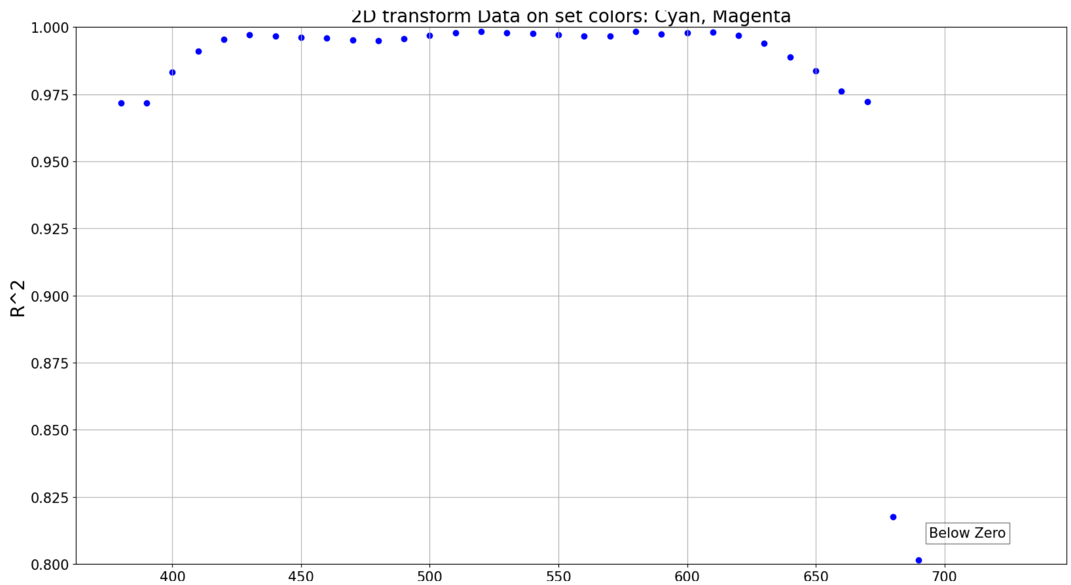


Figure 9: Cyan & Magenta r^2 leaf model

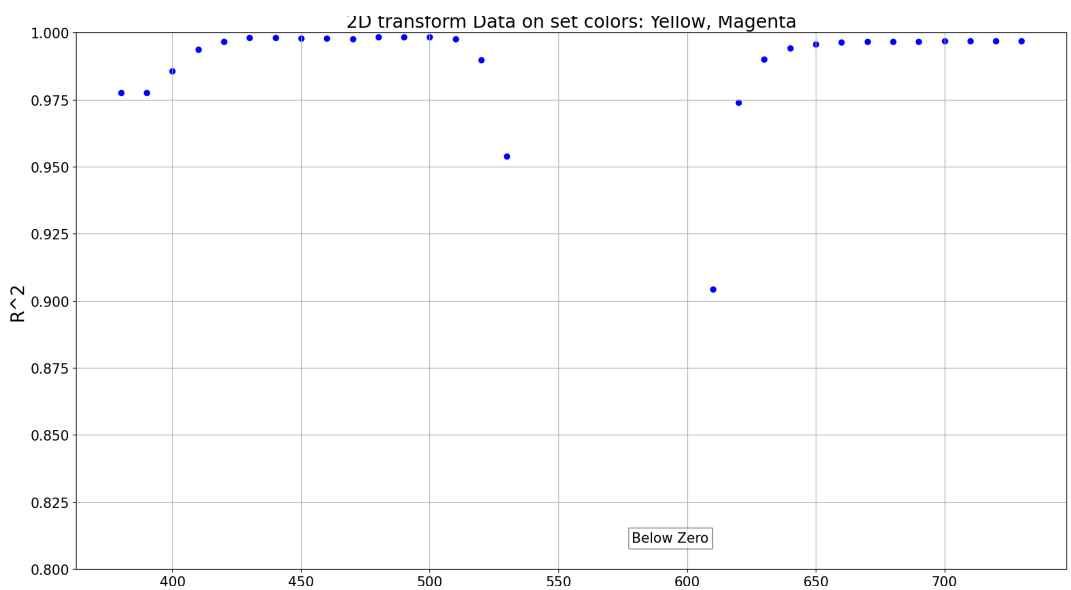


Figure 10: Cyan & Magenta r^2 leaf model