



## A numerical solution for correct color rendering of narrowband emitters in applied astrophotography.

Markus A. R. Langlotz

In deep sky photography, astronomers are frequently facing the problem, how to correctly apply a color toning to narrowband emitters, such as  $H\alpha$  or  $OIII$ . For that reason, different images of the same object often show totally different colors. This does, for sure, not add to the scientific correctness of images and often leads to discussions between astrophotographers about a correct color reproduction. As those emitters are monochromatic, color calibration on stars with known spectral classes – which is based on continuum light sources - does not affect the nebulae in a correct way. This paper shall provide a methodology to overcome that situation.

### 1. Standard color metrics

All visible colors can be described using the CIE1931 color space, a U-shaped locus of vectors representing all possible colors from full saturation (monochromatic stimuli) to completely desaturated white. The latter point is represented by the so called equivalence point  $E$ , whereas all monochromatic stimuli can be found on the curved contour surrounding the U. The straight line, locating the base of the CIE1931 color space, is called purple boundary. Colors on this line can not be represented by single wavelengths. In this model, all colors can be described as numerical triples  $(x;y;z)$ . As by definition  $z=1-x-y$ ,  $z$  is typically omitted and a color can be described by the doublet  $(x;y)$ . As for the equivalence point  $E$ , all components have to be equal,  $E$  translates to  $(0.33;0.33)$ . The CIE color space and the corresponding wavelengths can be found in figure 1.

**Note:** Full saturated (monochromatic) colors are located on the u-shaped outline of the CIE1931 color space. Given a line from the equivalence point stretching to the outline of the CIE1931 color space, this line represents the same color (hue) with a saturation increasing from zero (equivalence point, pure white) to 100% at the outer border (monochromatic wavelength).

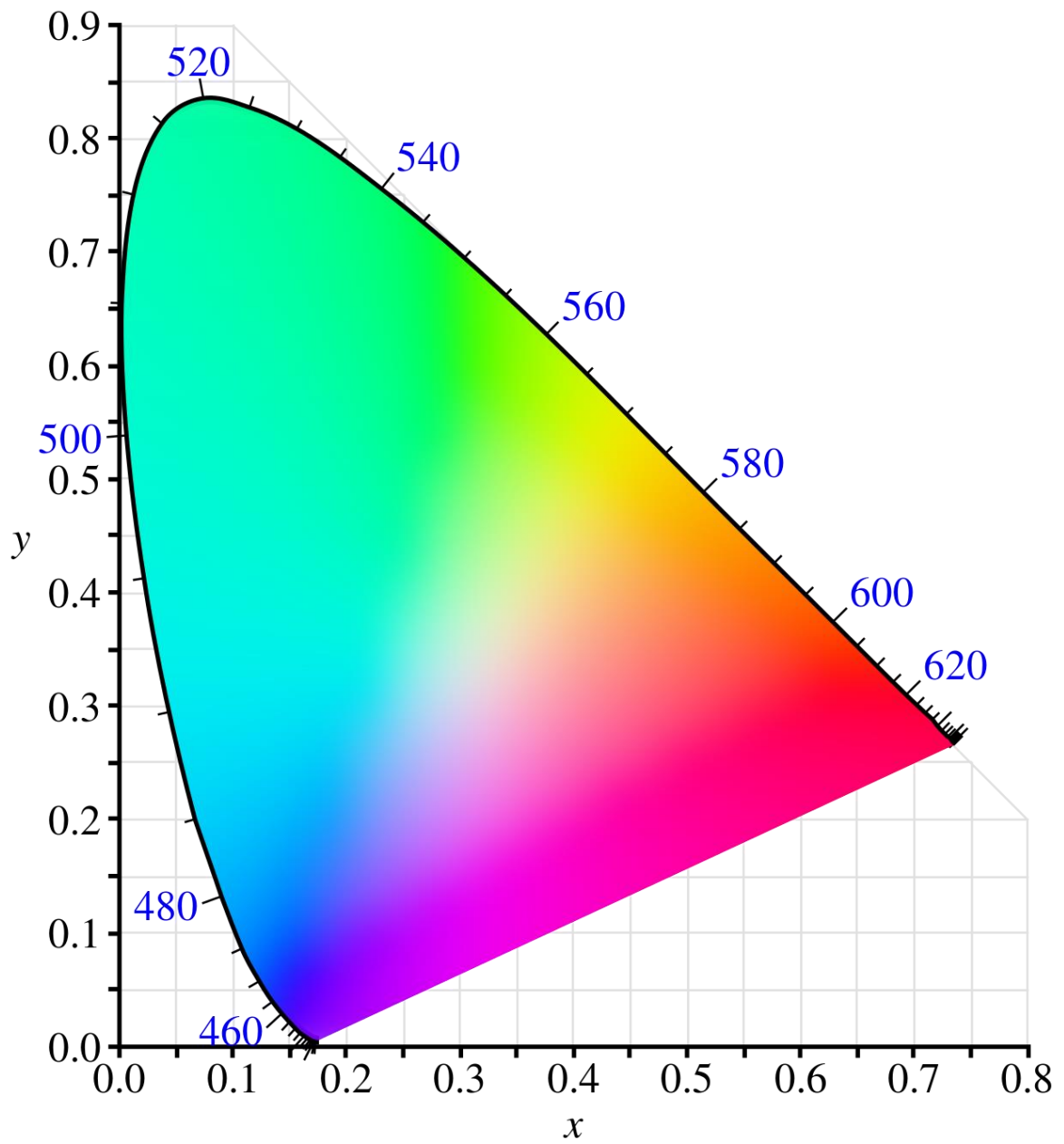


Fig. 1: The CIE 1931 color space

## 1. Additive color mixing

Given two colors  $(x_1; y_1)$  and  $(x_2; y_2)$ , their additive combination lies on the straight line between those two color coordinates, the exact position being dependent on the intensity of either of the two stimuli. Based on that principle, three color stimuli spanning a triangle, can be used for representing either color within that triangle. This is the basic working principle of color displays with three primary stimuli for red (R), green (G) and blue (B). The larger this triangle is, the larger the color gamut of the respective display will be. A similar principle is applicable for the subtractive color mixing in printing devices.

### 1.1. Common color gamuts for displays and monitors

This article concentrates on a correct color representation for narrowband emitters on displays (e. g. computer monitors). If color management will be applied in the post processing workflow, then it will be assured, that also on prints, all colors will be rendered exactly in the same way as they had been perceived on the monitor. Table 1 shows the two most commonly used color gamuts sRGB and AdobeRGB.

Color Gamut	Color coordinate R	Color coordinate G	Color coordinate B	White Point
sRGB	(0,64;0,33)	(0,30;0,60)	(0,15;0,06)	(0,3127;0,3290)
Adobe RGB	(0,64;0,33)	(0,21;0,71)	(0,15;0,06)	(0,3127;0,3290)

Table 1: Commonly used color spaces

The white point is an essential metric, as it defines the color temperature of a pure white shown on the respective monitor. The white point for typical applications as sRGB or AdobeRGB is defined by a color temperature of sunlight, which translates to 6500K. This is also referred to as D65. As can be seen, the only difference between the two mentioned color spaces is a wider extension in the green area for AdobeRGB, whereas the red and blue primaries are exactly the same.

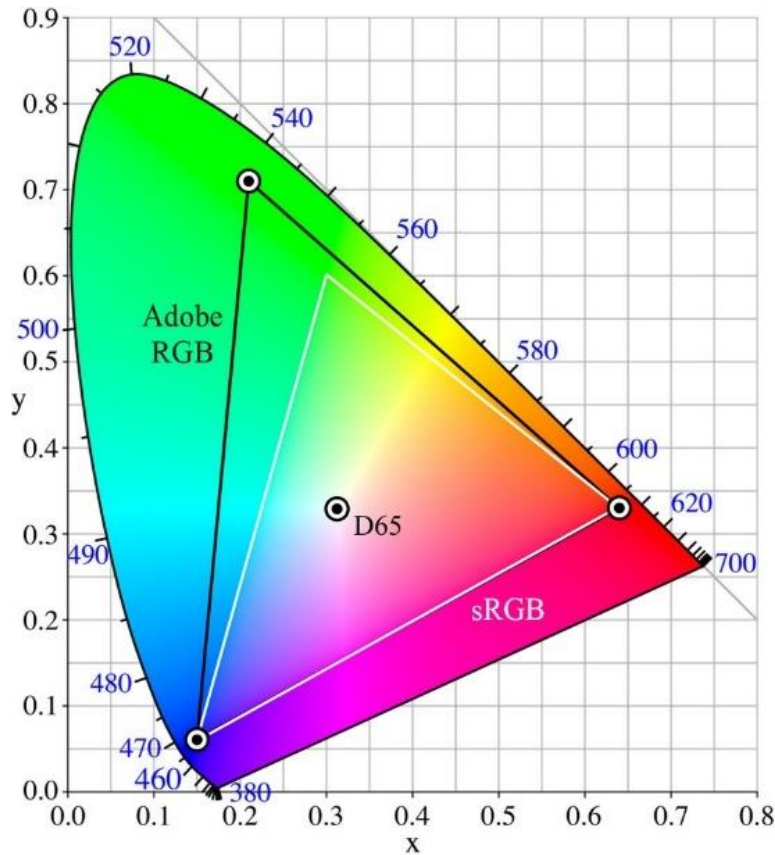


Fig. 2: sRGB and AdobeRGB color spaces

As can be seen in figure 2, both, the sRGB and AdobeRGB, color spaces do not touch the outline of the CIE1931 color space. This is due to the fact, that colors in imaging devices are generated by dye based color filters and therefore the red, green and blue primaries are by nature not fully saturated. You can further see, that on a line from red, orange, yellow to green, a display can show dedicated wavelengths by a quite good approximation, whereas between green and blue, purple and red, the colors are far away from monochromatic colors and therefore a very vague approximation. Unfortunately, most emission lines of astronomical interest lie exactly on these two borders of the color triangles of modern display systems.

**Note:** As all colors in nature are created by reflection of white (sun)light on colored surfaces and objects, typically there are no very saturated colors present in real life. That's why the sRGB and AdobeRGB color gamuts are a very good representation of what can typically be seen. In astronomy, however, there are narrow emission lines, which – in typically workflows – are more or less colorized to taste. To avoid this, this paper shall show an exact way, how monochromatic colors can best be approximated on common imaging devices.

## 2. Color coordinates and (RGB)-Values

The first (and most unintuitive) transformation we have to do, is to find a correlation between (RGB)-Values (as are well known in the image processing world) and CIE1931 color coordinates (which represent the real world color).

Let's assume, a display will be filled with a rectangle of the color (R;G;B). Then we have to calculate the color coordinate (x;y) of the light emitted by this rectangle. We are using the following color transformation:

$$(X; Y; Z) = \underline{A}_{ij}(R;G;B) \quad \text{Eqn. (1)}$$

Where  $\underline{A}_{ij}$  is the transformation matrix describing the transition from RGB to XYZ. The components of the matrix  $\underline{A}_{ij}$  can be calculated using the primary stimuli  $(X_r; Y_r)$ ,  $(X_g; Y_g)$ ,  $(X_b; Y_b)$  of the underlying imaging device's color gamut.

For the red primary, we get the following equation:

$$(X; Y; Z)_{red} = \underline{A}_{ij}(1; 0; 0)$$

This is because the pure red primary is shown, when a color with (R=1; G=0; B=0) is displayed. Expanding above equation results in the following:

$$(X; Y; Z)_{red} = \underline{A}_{1j} \quad \text{Eqn. (2)}$$

Explicit notation for all three primaries, using Eqn. (2) gives the following system of equations:

$$\begin{array}{lll} A_{11} = X_{red}; & A_{12} = Y_{red}; & A_{13} = 1 - X_{red} - Y_{red}; \\ A_{21} = X_{green}; & A_{22} = Y_{green}; & A_{23} = 1 - X_{green} - Y_{green}; \\ A_{31} = X_{blue}; & A_{32} = Y_{blue}; & A_{33} = 1 - X_{blue} - Y_{blue}; \end{array}$$

Eqn. (3)

The next step is to convert from (X;Y;Z) color coordinates to normed (x;y;z) color coordinates, as we typically find in the CIE1931 color space.

Given

$$(x; y; z) = \frac{(X; Y; Z)}{X + Y + Z}$$

results in

$$x + y + z = 1 \quad (\text{Eqn. 4})$$

**Note:** Equation (4) shows, that the color coordinate  $z$  is a redundant value and therefore can typically be omitted, as  $z = 1 - x - y$ .

A tuple of  $(x; y; z)$ -Values is a scaled  $(X; Y; Z)$  tuple with a scaling factor of  $k=(X+Y+Z)$ . By multiplication, the color coordinate does not change:

$$x(X; Y; Z) = x(kX; kY; kZ) = \frac{kX}{kX+kY+kZ} = X/(X + Y + Z) \quad (\text{Eqn. 5})$$

Remark: Equation (5) also applies for the  $y$  and  $z$  components, the latter of which can be omitted due to redundancy reasons.

From expanding Eqn. (1) directly follows, that all line totals of the transformation matrix are proportional to the color coordinates of the white point. This is the white point of the underlying color gamut, i. e. sRGB or AdobeRGB. From Eqn. (5) follows, that the columns of the transformation matrix are proportional to the primaries, with typically different factors in each column:

$$\underline{A}_{ij} = k_i(x_i; y_i; 1-x_i-y_i), \quad i=1(1)3$$

From a brightness standardization point of view, the middle line total of the transformation matrix has to be equal to a value of 1.

Reference: C.A. Bournman, Digital Image Processing – January 12, 2015, page 18. [1]

Hence we have an explicitly solvable set of equations for the three constants  $k_i$  :

Let  $(x_{red}; y_{red})$ ,  $(x_{green}; y_{green})$ ,  $(x_{blue}; y_{blue})$  be the primary stimuli of a given display system and let  $(x_w; y_w)$  be the white point of that display.

This gives the resulting equation for  $k_1$ ,  $k_2$  and  $k_3$ :

$$1/y_w(x_w; y_w; 1-x_w-y_w) = \underline{A}_{1j}(k_1; k_2; k_3)$$

Herein the transformation matrix is defined as follows:

$$\begin{array}{lll}
 A_{11} = x_{red}; & A_{12} = y_{red}; & A_{13} = 1 - x_{red} - y_{red}; \\
 A_{21} = x_{green}; & A_{22} = y_{green}; & A_{23} = 1 - x_{green} - y_{green}; \\
 A_{31} = x_{blue}; & A_{32} = y_{blue}; & A_{33} = 1 - x_{blue} - y_{blue};
 \end{array}$$

The final transformation matrix  $\underline{B}_{1j}$  is given by multiplication:

$$\underline{B}_{1j} = \underline{A}_{1j} \underline{C}_{1j}, \text{ with}$$

$$c_{ij} = 0 \text{ for } i \neq j$$

$$c_{ij} = k_i \text{ for } i=j \quad (\text{Eqn. 6})$$

Herewith the final transformation between color coordinate and (R;G;B)-tripels:

$$(x;y;z) = \underline{B}_{1j} (R; G; B)_\gamma \quad (\text{Eqn. 7a})$$

Here, we take into account, that the linear RGB-stimuli have to be corrected with the gamma characteristic of the underlying display:

$$(R;G;B) \rightarrow (R;G;B)_\gamma \text{ with}$$

$$R_\gamma = R^\gamma$$

$$G_\gamma = G^\gamma$$

$$B_\gamma = B^\gamma \quad (\text{Eqn. 7b})$$

For current display and monitor applications, gamma equals a value of 2.2. In Apple systems, a gamma of 1.8 has also been usual in former times. In this equation, R,G and B are taken as the fraction of the maximum R, G and B levels, this maximum being typically 255 (0xFF).

As mentioned in chapter 1.1, the typical color gamuts used in display and monitor systems are sRGB and AdobeRGB, resulting in different color transformation matrices for both types of display.

Remark: Comparable methodology is also used in the color profiling process of monitors, where (R;G;B) triples are transformed from one gamut into another.

Bringing all values into Eqn. (7a) leads to the following final color transformation equations:

- (i) Color coordinates to (R;G;B) values into sRGB:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0,4124 & 0,3576 & 0,1805 \\ 0,2126 & 0,7152 & 0,0722 \\ 0,0193 & 0,1192 & 0,9505 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad \text{Eqn. (8a)}$$

- (ii) Color coordinates to (R;G;B) values into Adobe RGB:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0,5767 & 0,1856 & 0,1882 \\ 0,2973 & 0,6274 & 0,0753 \\ 0,0270 & 0,0707 & 0,9913 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad \text{Eqn. (8b)}$$

With these and their inverse transformations, both color representation schemes (x;y;z) and (R;G;B) can be converted into each other.

### 3. Color representation of narrowband emitters

As mentioned in chapter 1 (see Note), all points with a certain wavelength are on the line from the equivalence point E to the wavelength at the border of the U-shaped border of the CIE1931 color space being more and more desaturated the nearer they are located to E. This means, that a given wavelength can be in a most saturated manner be displayed with the color coordinate (x;y), where the line from E to this wavelength intersects the outline of the triangular color gamut border, the so called color triangle. As can be seen in Fig. 2, turquoise colors like the OIII emission line can only be shown with a quite low saturation. Reds longer than approximately 620nm can only be represented via the purple line, meaning red with a small addition of blue.

Remark: Any monochromatic source can be represented by an (R;G;B) triple, where at least one of the components must be zero. If the wavelength to be represented is that of one of the three primary stimuli, only the value of this particular stimulus is greater than zero, the other two components being zero.

Calculating these points in the (x;x)-color space is quite easy as it is simply an intersection of two straights, one defined by E and the color coordinate of the respective wavelength, the other defined by the two loci of the adjacent primary stimuli. As we are only considering color coordinates, but not brightness, we can omit dealing with gamma.



## 3.1. Color coordinates of narrowband emitters

Matching the straights of the narrowband emitter and the border of the color gamut leads to the following loci in the (x;y)-color plane:

Emission line	H- $\alpha$	H- $\beta$	OIII	SII
Wavelength	656,28nm	486,13nm	501nm / 496nm	672nm
CIE-1931-color locus	(0,729;0,271)	(0,064;0,2175)	(0,020;0,438)	(0,732;0,268)
( $x_\lambda$ ; $y_\lambda$ ) in sRGB	(0,577;0,295)	(0,211;0,281)	(0,235;0,366)	(0,574;0,294)
( $x_\lambda$ ; $y_\lambda$ ) in AdobeRGB	(0,577;0,295)	(0,169;0,263)	(0,180;0,385)	(0,574;0,294)

Table 2: narrowabd emitters on real displays

## 3.2. RGB values for narrowband emitters

The color gamuts ( $x_\lambda$ ;  $y_\lambda$ ) shown in table 2 can be converted to (R;G;B) triples using equantions 8a (for sRGB) and 8b (for AdobeRGB). The final results can be seen in the following table:

Emission line	H- $\alpha$	H- $\beta$	OIII	SII
wavelength	656,28nm	486,13nm	501nm / 496nm	672nm
Color in sRGB	R=255 G=000 B=082	R=000 G=216 B=255	R=000 G=255 B=225	R=255 G=000 B=083
Color in AdobeRGB	R=255 G=000 B=096	R=000 G=206 B=255	R=000 G=255 B=220	R=255 G=000 B=097

Tsable 3: Color values for narrowabd emitters

Remark: Those (R;G;B) triples are the brightest colors, that can be shown on an adobeRGB or sRGB device, as one of the color channels is reaching a value of 255=0xFF. Showing weaker colors (less bright) can be done by multiplying a complete (R;G;B) tripel by a factor  $\mu < 1$ :

$$(R;G;B) \rightarrow (\mu R; \mu G; \mu B)$$

The latter is a very important correlation: multiplying an (R;G;B) triple with a constant factor does only change the brightness, but NOT the color. An intuitive example is the gray scales from a pure black (0;0;0) to pure white (255;255;255) with all grey values

$$(gr;gr;gr), gr=1(1)254$$

In practical use, to get a correctly colored narrowband image, the greyscale image of an astronomical narrowband object has to be multiplied (colored) with the respective (R;G;B) triple from Table 3.

## 4. Conclusion

This paper shows the correct color representation of narrowband emitters on common display and print devices. Color space transformations can be used to exactly define, how a certain monochromatic wavelength has to be best represented on a technical device.

A very important point, however, is, that commonly used methods for colorimetric color calibration based on star colors only work very well with continuum emitters like galaxies or star clusters, as the latter adjust images by applying a correct white balance. As narrowband emitters are highly saturated and located at the borders of the color triangle, these methods can never work with pure colors. For example, H alpha only affects the red channel. However, pure reds on a display or printed have typically a dominant wavelength of 620nm. All reds longer than 620nm would therefore be shown as the same 620nm red, the so called primary stimulus. The same applies for all monochromatic sources of light: They can not be generated by white balancing but have to be numerically assigned by calculation.

## 5. References

- [1] C.A. Bournman, Digital Image Processing – January 12, 2015, page 18.