https://doi.org/10.2352/CIC.2024.32.1.22

Mean color, probably

Ján Morovič, Peter Morovič, HP Inc.

Abstract

Colors are characterized by their appearance attributes such as brightness, colorfulness and hue, or lightness, chroma and hue when considered relative to an adapted white. Consequently, they can be represented by coordinates in three-dimensional spaces like those of CIELAB or CIECAM02 appearance predictors. Therefore, the color of a stimulus is customarily associated with a point in 3D. While this is appropriate when dealing with theoretical quantities, in practice the color space coordinates of a stimulus are subject to variations both in the stimulus itself and in its measurements. This tends to lead to multiple measurements being the basis of identifying the color of a stimulus by averaging them and in some cases excluding outliers. This, however, obscures the fundamental variability of color data. In this paper an alternative, probabilistic approach to dealing with color data will be introduced and applied to the computation of color difference, color gamuts and gamut inclusion, yielding distributions of answers instead of only single values.

Introduction

When asking the question of what the color of an object or light source is, the expectation is to receive a single answer, when all relevant conditions are well defined. An example here could be to say that a certain print, under specific lighting, viewing and measurement conditions has certain colorimetry, as expressed in CIE XYZ, CIE LAB or other color spaces. While this may be unproblematic in a theoretical context, when it comes to practical use cases, the stimulus–colorimetry relationship is anything but this simple.

For a start, the great majority of stimuli vary either because of variations in the environment they are viewed in (e.g., a print viewed even in a light-box is subject to variations of the light source's output over time, with changes in ambient temperature, etc.) or because of variations in the devices used to bring them about (e.g., a display's output is also subject to changes over time). A further level of variation arises when stimuli are generated in response to inputs to an imaging system. E.g., making a print repeatedly in response to a certain CMYK input to a system, even if that system has not been intentionally changed between copies, will result in varying physical properties.

Added to the unavoidable variability of stimuli is also a variability of color measurement devices. Repeated measurements of a stimulus yield variations in the measured quantities since measurement instruments, like imaging devices, are subject both to systematic drifts, changes due to changing environmental conditions and random fluctuations.

The upshot of such variability is that the question of what the colorimetry of a stimulus is is hard to answer with a single response. Repeatedly measuring the varying states of a stimulus results in varying measurements instead of a single one.

A solution to such unwanted multiplicity is to take a set of measurements and derive from it a single set of colorimetries for a stimulus. This can be done by, e.g., computing the mean of such multiple measurements (e.g., Hunt and Pointer, 2011; Ly *et al.*, 2020), or some other central tendency of the data, such as their median, when distributions are not normal. The removal of some of "outliers" from the set of measurements before computing their central tendency is also a strategy that aims at greater robustness (e.g., ASTM, 2008).

Such solutions to dealing with the variability of stimulus colorimetry are certainly good if the objective is to obtain stable, single colorimetries per stimulus. Their use is widespread in color science and engineering and has clear advantages versus using single measurements of stimuli.

They do, however, have an important shortcoming, which is that they obscure variability and present a view of what colorimetries and their relationships are like. In the context of evaluating color differences, e.g., of a color match or of the color stability of an imaging system, they suggest a single state being the case while in reality such differences form distributions. The same applies to the question of what ranges or gamuts a set of color has, which also is not best expressed by a single boundary or a single volume but by their distributions.

The following sections will set out an exploration of how color could be dealt with in a probabilistic way that preserves and expresses the variability of colorimetry in providing answers to questions about stimulus color, color difference and color gamuts, including the indication of which stimuli are enclosed by a gamut and which are not. Answers will be in the form of distributions, their central tendencies and percentage ranges derived from them.

From means to all pairs

A typical approach to quantifying the differences between two contexts is to first attempt their robust characterizations and to then compute differences between them.

Let's start with a simple scenario of wanting to quantify the difference between two imaging systems for a given input to them. E.g., in the case of printing, the question may be, how different is the color output of two printers, with their respective inks, substrates and settings, for a given CMYK input. One way to answer such a question would be to take that CMYK input, obtain multiple prints for it and measure each of those prints multiple times. Let's refer to the number of printed repetitions on system 1 as p_1 and the number of measurement repetitions of each of those prints as m_1 and let's use analogous nomenclature for system 2.

The color difference between the mean outputs of the two systems can then be expressed as follows:

$$D_{1,2} = \Delta E\left(\frac{1}{p_1 m_1}\sum_{i_1=1}^{p_1}\sum_{j_1=1}^{m_1} C_{1(i_1,j_1)}, \frac{1}{p_1 m_1}\sum_{i_2=1}^{p_2}\sum_{j_2=1}^{m_2} C_{2(i_2,j_2)}\right) \qquad (1)$$

where $\Delta E()$ is a color difference equation and C_k is a color measurement from system k. In other words, the $p_1 \times m_1 + p_2 \times m_2$ measurements yield a single color difference, the color difference between the respective mean colors of the two systems. Fig. 1 shows an example of this approach for a case where $p_1 = p_2 = 2$ and $m_1 = m_2 = 3$, i.e., where two prints are made of the same CMYK input on each system and each print is measured three times.

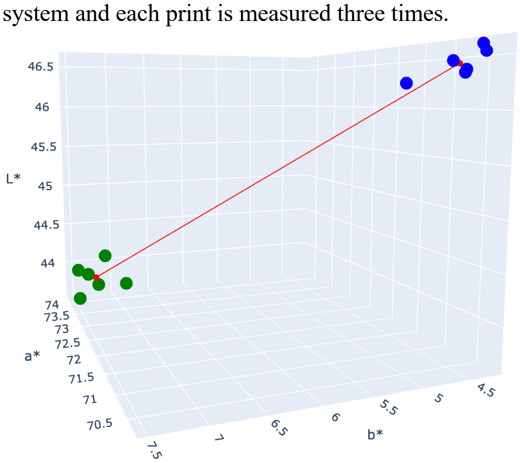

Figure 1. *Measurements of the output of two systems for the same CMYK input, their means and the color difference between those means.*

The data for the example in Fig. 1 is shown in Tab. 1 and the ΔE76 between the two mean colors is 5.15, with the ΔE00 (ISO/CIE, 2022) being 2.81.

Table 1: Example data visualized in Fig. 1 for print (P) 1 and two and measurements (M) 1, 2 and 3 on systems 1 and 2.

P	M	System 1			System 2		
		L*	a*	b*	L*	a*	b*
1	1	43.75	73.95	6.88	46.38	70.66	4.27
	2	43.75	74.02	7.22	46.25	70.89	4.78
	3	43.64	73.85	7.48	46.35	70.64	4.30
2	1	44.07	74.24	7.09	46.54	70.43	4.24
	2	43.95	74.04	7.47	46.45	70.55	4.51
	3	43.91	73.95	7.36	46.59	70.28	4.40
Mean		43.85	74.01	7.25	46.43	70.58	4.42

While computing mean colors and then color differences between them is a good way to make the indication of color difference more robust to variations both in printing and measurement, doing so obscures those variations, which are an inextricable part of imaging and color measurement too.

A further shortcoming of such an approach is also that the means do not actually correspond to any printed and measurement colorimetry (see the small red dots in Fig. 1 which indicate the means of the two sets of measurements, shown as larger green or blue dots).

The approach to characterizing this same kind of data proposed here is one that characterizes the relationship between such two sets of measurements not with a single color difference value, but with a distribution of such values. Instead of computing a single color difference on the basis of $p_1 \times m_1 + p_2 \times m_2$ measurements, all $p_1 \times m_1 \times p_2 \times m_2$ pair differences are computed as follows:

$$D_{1,2(i_1,j_1,i_2,j_2)} = \Delta E\left(C_{1(i_1,j_1)}, C_{2(i_2,j_2)}\right) \quad (2)$$

In the case of our toy example, this yields 2x3x2x3=36 ΔEs, as shown in Fig. 2, with pair values shown in Tab. 2 and their distributions visualized in Fig. 3.

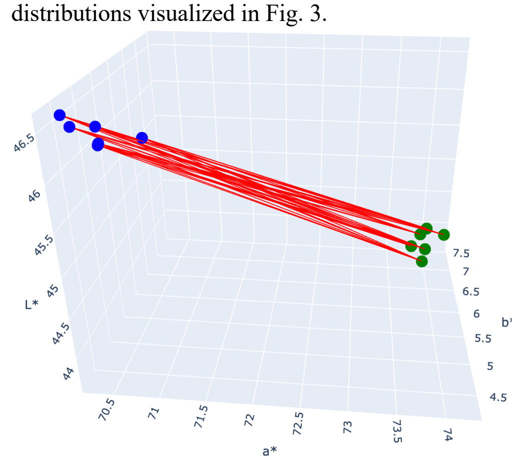

Figure 2. *Measurements of the output of two systems for the same CMYK input and all their pair color differences.*

Table 2: Pair color differences in ΔE76 and ΔE00.

ΔE76		S2					
		p_1m_1	p_1m_2	p_1m_3	p_2m_1	p_2m_2	p_2m_3
S1	p_1m_1	4.96	4.47	4.94	5.21	4.95	5.26
	p_1m_2	5.19	4.69	5.17	5.44	5.16	5.48
	p_1m_3	5.29	4.78	5.27	5.53	5.25	5.56
	p_2m_1	5.11	4.62	5.10	5.36	5.10	5.41
	p_2m_2	5.25	4.74	5.23	5.49	5.22	5.53
	p_2m_3	5.14	4.63	5.12	5.39	5.11	5.42

ΔE00		S2					
		p_1m_1	p_1m_2	p_1m_3	p_2m_1	p_2m_2	p_2m_3
S1	p_1m_1	2.81	2.61	2.78	2.96	2.84	2.99
	p_1m_2	2.87	2.66	2.84	3.02	2.89	3.05
	p_1m_3	2.99	2.78	2.96	3.14	3.01	3.17
	p_2m_1	2.60	2.39	2.57	2.75	2.63	2.78
	p_2m_2	2.75	2.53	2.72	2.90	2.77	2.92
	p_2m_3	2.76	2.54	2.73	2.91	2.78	2.93

Looking at the 36 pair color differences in Tab. 2 and their distribution in Fig. 3 already gives a sense of the more varied, probabilistic nature of how the output of the two systems relates. Instead of simply stating that their mean colorimetries are 2.81 ΔE00 apart, it is more representative to say that, while the mean of their pair differences is 2.81 ΔE00, it has an interquartile range of 2.73 to 2.94 ΔE00 and a 95 percent range of 2.52 to 3.15 ΔE00. In other words, while the difference is 2.81 on average, half the time it will be between 2.73 to 2.94 and 95 percent of the time it will be in the 2.52 to 3.15 range.

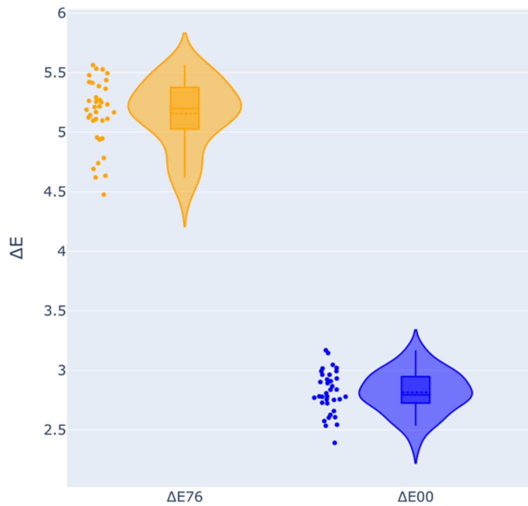

Figure 3. *Distribution of pair differences in ΔE76 and ΔE00.*

Expressing the differences between the outputs of these two systems for a single input probabilistically also enables a probabilistic way of thinking about relationships versus thresholds. E.g., if an acceptability threshold for some color reproduction use case were set at 3.0 ΔE00, then, instead of considering that the difference between our systems is below threshold (based on a difference between means of 2.81), it can now be stated that their differences would be not exceed the 3.0 ΔE00 threshold in 87 percent of cases (since 3.0 is at the 87th percentile of the pair color difference distribution).

Let's next consider the impact of moving from comparing means to evaluating the color differences between all corresponding pairs on the descriptive statistics of whole sets of system inputs. Taking the set of 1485 CMYKs of the ECI2002 chart (ISO, 2006) that samples the full device color space, we obtain the statistics shown in Tab. 3. As can be seen, the inclusion of the pair ΔEs barely changes the central tendency of differences and even the effect on the 95th percentile is mild, with it being principally minima and maxima that are more significantly affected.

Table 3: Descriptive statistics of color differences between printed outputs corresponding to 1485 device color inputs to two systems.

	ΔE76		ΔE00	
	Mean	Pair	Mean	Pair
Minimum	**0.43**	**0.15**	**0.35**	**0.12**
Median	12.17	12.16	7.93	7.92
Mean	13.01	13.02	8.42	8.43
95th percentile	24.42	24.49	15.12	15.25
Maximum	**35.70**	**36.78**	**22.71**	**24.54**

As a consequence, there is little benefit from computing all pair differences when the aim is to understand the overall distribution of differences over a large set, while for an individual case they provide a relevant sense of variability and allow for a probabilistic understanding of tolerances.

Even for large sets the computation of pair color differences per member does have value in that it allows for a characterisation of the distribution of variability across the set. Fig. 4 shows the histogram of 95th percentile ranges over the 1485 CMYK inputs and it can be seen both that they have significant medians (1.00 in ΔE76 and 0.68 in ΔE00) and that they vary substantially from color to color, with a 95 percent ranges of 0.46–2.83 in ΔE76 and 0.29–1.71 in ΔE00.

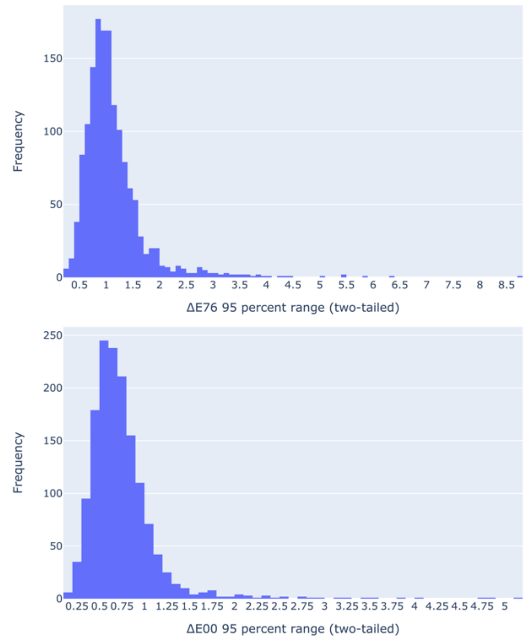

Figure 4. *Histograms of pair difference 95th percent ranges for 1485 ECI2002 CMYK inputs to two printing systems in ΔE76 (top) and ΔE00 (bottom).*

In other words, only reporting summary statistics of the differences between the colorimetries obtained from the two systems does give a sense of the distribution of their magnitudes but supplementing them with the statistics of the 95 percent ranges also gives a sense of the variability of those differences.

And back to means, assuming normality

An alternative to computing pair differences is to characterize the distributions of colorimetries corresponding to the cases being compared and to then compute differences between those distributions and the confidence intervals of differences on the basis of per-distribution statistics. Assuming that colorimetries are samples from a normal distribution, their means ($\bar{C}$) and variances (*Var(C)*) can be computed as follows:

$$\bar{C} = \frac{1}{p_1 m_1} \sum_{i_1=1}^{p_1} \sum_{j_1=1}^{m_1} C_{(i_1,j_1)} \quad (3)$$

$$Var(C) = \frac{1}{p_1 m_1} \sum_{i_1=1}^{p_1} \sum_{j_1=1}^{m_1} \left(C_{(i_1,j_1)} - \bar{C} \right)^2 \quad (4)$$

The difference of two normal distributions can then be characterized by the difference of their means and their pooled variance (Dodge, 2008):

$$\Delta\bar{C}_{1,2} = |\bar{C}_1 - \bar{C}_2| \quad (5)$$

$$Var_p(C_{1,2}) = \frac{(n_1-1)Var(C_1) + (n_2-1)Var(C_2)}{n_1+n_2-2} \quad (6)$$

where n_1 and n_2 are the sample sizes of the two sets of colorimetries.

Returning to the toy data set from the previous section, the colorimetries from Tab. 1 yield the following normal statistics (Tab. 3).

Table 4: Normal statistics of example colorimetries from Tab. 1.

	System 1			System 2		
	L*	a*	b*	L*	a*	b*
$\bar{C}$	43.85	74.01	7.25	46.43	70.58	4.42
Var(C)	0.02	0.02	0.05	0.02	0.03	0.04

	System 1 – System 2			
	L*	a*	b*	L2 norm
$\Delta\bar{C}_{1,2}$	2.58	3.43	2.83	5.15
$Var_p(C_{1,2})$	0.02	0.03	0.04	0.05

As can be seen, the L2 norm of the per-dimension differences between sample means of 5.15 is their ΔE76 color difference. The standard deviation of this norm can then be obtained by taking the square root of the L2 norm of per-dimension pooled variances of 0.05, yielding 0.22. Finally, the two-tailed 95 percent range is ±1.96 times this standard deviation, resulting in a range of 4.71 to 5.59. This compares with the 95 percent range computed directly from the pair differences of the two samples (Tab. 2) of 4.60 to 5.54, which in this particular case is a reasonable prediction of the pair differences derived from the normal statistics of the two samples.

Looking at the relationships between these 95 percent ranges for ΔE76 (which can be predicted directly from normal statistics) between direct pair difference computation and prediction from per-set mean and variance, Fig. 5 shows that the latter are a good approximation of the former with an R^2 coefficient of determination of 0.837. At the same time, the individual sample sets of 36 pair differences per CMYK input do not pass the Kolmogorov-Smirnov (K-S) test (Smirnov, 1939) for normality at the 95% confidence level.

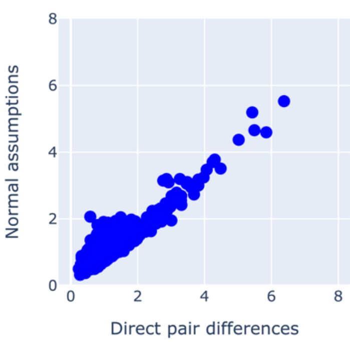

Figure 5. *95 percent ranges for 1485 ECI2002 CMYK inputs predicted from normal statistics versus computed directly from pair differences.*

To test whether the failure of normality is related to the limited scale of the toy data set, a more extensive set was compiled, consisting of printing three copies of the same ECI2002 chart at three different moments in time and measuring each of them three times – i.e., a data set where 27 colorimetries were available on two systems in correspondence to the same CMYK input. Computing pair differences then yielded 729 ΔEs per CMYK. In this case the 95 percent ranges had a coefficient of determination of 0.692 and again the K-S test for normality failed as for the much smaller, toy data set. Nonetheless, predicting 95 percent ranges from per-sample means and variances can be considered to be a useful approximation.

Looking more closely at the first member of the large set, Fig. 6 shows both the colorimetries from the two systems and the distribution of their pair differences. As can be seen, there are clusters among the colorimetries, which also correspond to clusters in the difference histograms that are not consistent with a unimodal, normal distribution.

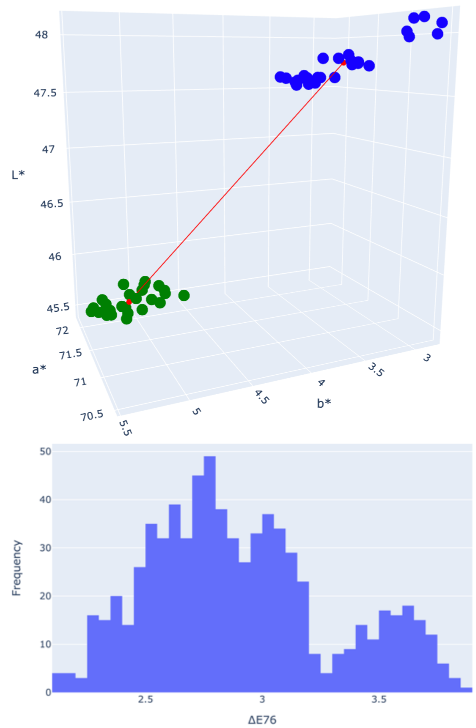

Figure 6. *Colorimetries of first member of large dataset from two systems (top) and the histogram of their pair differences (bottom).*

Probabilistic gamut volume

Given the variability of colorimetric data, their volume – or the color gamut – also results in variation and instead of having a single volume, reporting a range or distribution is required.

The aforementioned set of 27 prints and measurements of the same CMYK data on two systems can first be analysed as individual sets of measurements of the 1485 samples of the ECI chart. This yields the results shown in Table 5, where gamut volumes are computed following the ISO specification (ISO, 2022) for discrete data, using alpha-shapes with a constant value of alpha=40.

Table 5: Color gamut volumes (LAB) – statistics over 27 sets of prints and measurements

	System 1	System 2
Min	443,749	425,165
Mean	447,449	430,418
Median	447,930	430,705
Max	452,179	433,896

This would result in a 1.9% and 2.1% range of variability between the smallest and largest volumes for System 1 and System 2 respectively.

However, for each of the 1485 samples, each of the 27 measurements is a valid possibility, hence any combination of picking one of 27 samples for each of the patches is a valid representation of colorimetry for the full chart. While evaluating all such possibilities is prohibitive (with a total of 27^{1485} possible charts), the maximum volume can be had easily by computing the volume of all 27 data sets pooled into one. For the mean volume, the volume of the mean measurements can be computed. Instead, the smallest volume can be estimated by computing the volume of all those samples that are closest to an interior point such as [50, 0, 0]. Since strict convexity cannot be assumed, this may not yield the exact minimum gamut. On the other hand, the gamuts considered here are close to convex, so this simple approach will yield a good approximation. Likewise other possible charts can be constructed by randomly selecting one of the 27 possible measurements for each of the 1485 patches and repeating such selection many times. Generating over 1M such possibilities results in a sampling of these possibilities and as before, the Kolmogorov-Smirnov test for normality fails at a 95% confidence level.

Tab. 6.a shows statistics of additional valid color gamut data points for each system, such as the gamut of the mean colorimetry, the near-minimum and maximum volumes.

Table 6.a: Color gamut volumes (LAB) – near-min, mean, maximum

	System 1	System 2
Near-Min	440,563	422,489
Mean color gamut	447,729	429,911
All max	455,941	438,377

Taking the above data points into account, System 1 has a gamut volume range of 440,563 (near-minimum) to 455,941 (maximum) LAB units, resulting in a range of 3.5%, while System 2 has a gamut volume range of 422,489 (near-minimum) to 438,377 (maximum) LAB units, corresponding to a 3.8% range. Fig. 7 shows the near-minimum and maximum gamuts for System 2. What can be seen in Fig. 7 is that the differences are well distributed in a*b* terms, while in L*b* terms there is a bias in differences towards the darker colors. This is not unexpected since measurement noise may be larger for darker samples.

In Tab. 6.b all data is taken together, also including 1M random choices as well as all other possibilities to express valid gamuts. Note how the mean of the random samples is different again, compared to the mean color gamut and the color gamut of mean colorimetries. Further note how considering the 95% range, the gamut volumes only vary 0.4% and 0.5% for System 1 and System 2 respective, assuming normality, or 0.3% to 0.4% without the normal assumption. This is to be expected since sampling all possibilities will be biased towards the mean and therefore even 1M random samples (an insignificant fraction of all choices) do not represent the variation sufficiently well.

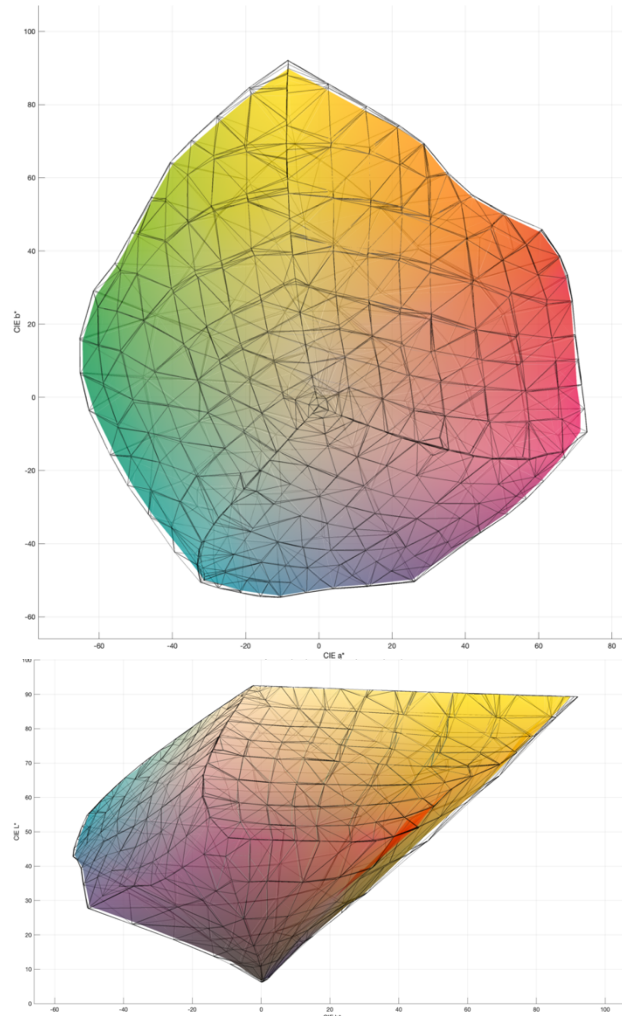

Figure 7. *Near-minimum (solid color) vs maximum (black mesh) color gamut in CIE a*b* view (top) and CIE L*b* view (bottom), showing the largest variation of gamut volumes and the distribution of differences for System 2 (corresponding to a 3.8% range).*

Table 6.b: Color gamut volumes (LAB) including 1M random gamut volumes

	System 1	System 2
Mean (1M samples)	448,093	430,521
5% Conf. Interval	447,250	429,511
95% Conf. Interval	448,937	431,532
5^{th} %tile	447,392	429,677
95^{th} %tile	448,807	431,372

Note also how the mean gamut (i.e. the average over 27 gamut volumes) in Tab. 5 differs from the mean color gamut (i.e. the color gamut of the mean colorimetries) in Table 6.a as well as the mean over the 1M samples in Table 6.b, even though all are quite close.

In summary, instead of thinking that a given system has a gamut volume of e.g. 447kLAB units (the mean gamut from Table 5, System 1), a probabilistic way that takes variation into account is to report 441kLAB to 456kLAB as the gamut volume range of System 1.

Probabilistic gamut coverage

The normal approximation of multiple colorimetries per color set member also lends itself to being extended to computing gamut coverages in a probabilistic way. This can be done by tracking not only mean colorimetries per set member, but also their variances and by characterizing a gamut boundary not only in terms of a triangulated surface of mean colorimetries but also by variance over such a surface.

This can be done as follows, assuming within-triangle variance convexity:

1. Compute means and variances of per-member colorimetries.
2. Construct a color gamut based on mean colorimetries. Here the alpha shapes method (Edelsbrunner and Mücke, 1994; Cholewo and Love, 1998) will be used, with an alpha value of 40. This yields a triangulated surface.
3. Given a test color *T*, with its mean colorimetry and variance, find *P* – the nearest point to *T* on the gamut hull from step 2.
4. Compute the barycentric weights of P within the gamut hull triangle that contains it.
5. Apply the same barycentric weights, derived from colorimetry, to the per-vertex variances to obtain the variances corresponding to P.
6. Given P and T, with their respective variances, compute the distance between P and T and the pooled variance of that distance using Eq. 5 and 6. The result is a distance distribution from which, e.g., 95% or interquartile ranges can be computed.

Fig. 8 illustrates the computation of pooled variance for a point on a gamut surface. Applying this approach to the extensive data sets referred to before, it is possible to derive the mean + variance gamut hull of one data set and then compute not only the proportion of in-gamut colors from the other data set based on mean colorimetries, but also obtaining a range of coverages at different ranges within the per-member distributions, such as the 95% range. Fig. 9 then shows the gamut boundary of one of the two systems where the norm of per-dimension variances is shown at each vertex. It can be seen both how variances differ from color to color and that their values can be used to predict variance across the full gamut hull surface.

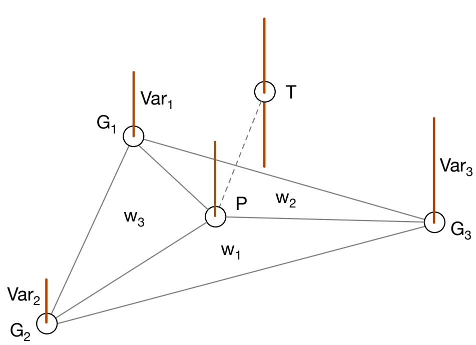

Figure 8. *Barycentric convex combination of gamut hull vertex G variances, resulting in variance at closest point on gamut P to test color T. Variances shown only in one dimension for illustrative purposes.*

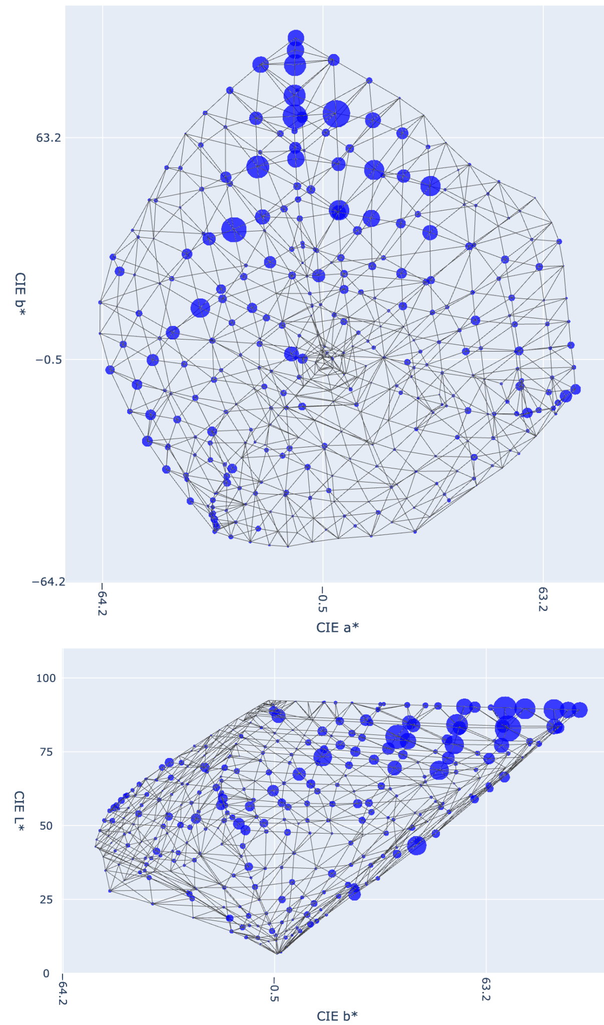

Figure 9. *The gamut boundary of one of the extensive data sets, showing the norm of per-dimension variances at each vertex. For ease of illustration, circle radii indicating variance are 50x the actual variance value.*

Taking a member of the System 2 data set as a test color, with CIE LAB values of [22.1, -7.8, -22.1], variances of [0.096, 0.015, 0.041] and an L2 norm of 0.11, we can find the nearest point to it on the mean gamut of System 1 at a distance of 1.37 ΔE76, with CIE LAB coordinates of [23.2, -7.7, -21.3], a pooled variance based on the enclosing surface triangle's vertex variances of [0.085, 0.031, 0.103] and L2 norm of 0.14. Given the variances, we can determine a range of distances between the test color and the gamut with a 95% range of 0.70 to 2.04 ΔE76. In other words, instead of thinking that the out-of-gamut distance is 1.37, a probabilistic treatment of the question predicts that the distance will be in the 0.70–2.04 ΔE76 range 95% of the time. Another way to express this relationship is to say that the probability of the test color being no more than 1.0 ΔE76 out-of-gamut is 16% in this case and that the probability of it being no more than 1.5 ΔE76 is 64%.

Applying the same process to the full 1485-member System 2 mean data set and evaluating each member's signed distance to the System 1 mean gamut yields a probabilistic view of gamut coverage. Instead of resulting in a single coverage percentage based on mean data of 79.46%, a 95% range can also be obtained, which in this case is between 72.67% and 89.95%. Computing distances based on means and obtaining their variances based on pooled test color and closest gamut hull color variances results in distance ranges as shown in Fig. 10, where negative values correspond to interior and positive values to exterior points relative to a gamut hull. In other words, a negative value indicates the distance from an interior point to the nearest point on the hull that contains it (i.e., large negative values mean that a point is deep inside a hull), while a positive value represents the distance between an exterior point to the nearest point on a hull.

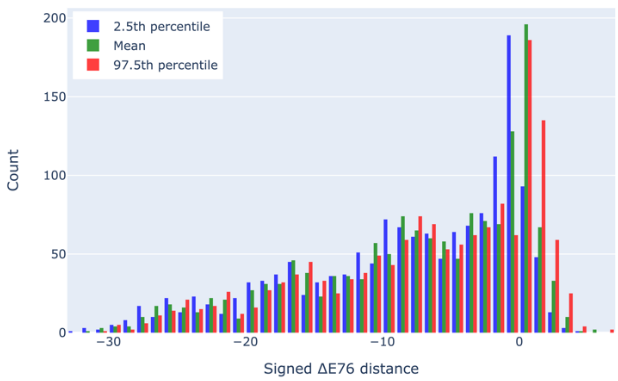

Figure 10. *Signed test color to gamut hull distances for 2.5th, 50th and 97.5th percentiles.*

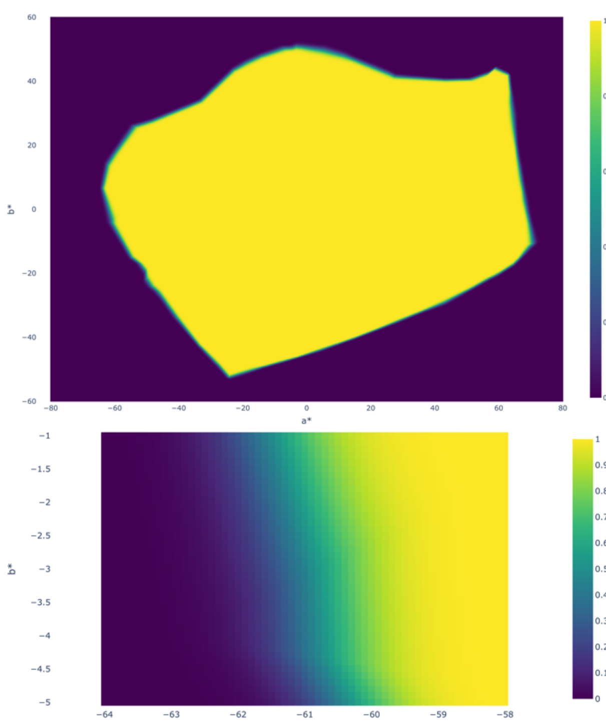

Figure 11. *Probability of inclusion in System 2 gamut of color distributions with mean L*=50. (Top) full gamut view, (bottom) zoomed in boundary region.*

In effect, dealing with color probabilistically also means a shift from considering whether a color is inside a boundary to asking what the probability is that a color's distribution of possible values is inside a boundary, also defined by its probabilities. To illustrate this, Fig. 11 shows the likelihood of a color with its mean L* at 50 and with a variance that matches the mean variance of System 2 being inside the mean gamut of System 2, given the boundary's variances.

Conclusions

While the color of a stimulus can theoretically be expressed by a single set of appearance attribute values, such as its brightness, colorfulness and hue, in practice any measured data about a stimulus forms a distribution. As a consequence, the question about how far two colors are in some color space is not one about the distance between two points, but about the distance between two distributions. By extension, other color-related entities, derived from multiple color stimuli, also become statistical. In the case of color gamuts, a 2D boundary in 3D becomes a probability map where any color coordinate can have inclusion probabilities continuously ranging from zero to one.

The atomic entity of a color stimulus therefore is not a set of coordinates but a set of descriptive statistics. This paper illustrates how such a move impacts three common color analyses of ascertaining the color difference between two stimuli or imaging system outputs for two inputs, of determining the range of colors that a system can reproduce, and of determining the inclusion of a stimulus in a color gamut. Using a statistical approach then yields ranges of values at given confidence levels or probabilities of the answer being above or below a given threshold, which more fully captures the nature of color analysis based on measured data and for systems with variable output.

While it is not always feasible to have the amount of data available that was used in this analysis, having a system characterization done once extensively and deriving variability thresholds as described would then allow to use these to interpret single data points or few repetitions and still handle color properties in a probabilistic way.

Future work could consider how the statistical approach to color stimulus specification introduced here can be applied in other contexts, e.g., that of psychophysics and psychometrics, and to other computations performed over a set of color stimuli, e.g., modelling, optimizations and machine learning.

Acknowledgements

The authors would like to thank their colleagues at HP Inc. for their support, especially Marc Isal and Alan Lobban.

References

[1] Y. Dodge, The Concise Encyclopedia of Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-32833-1_323 (2008)

[2] N. V. Smirnov, Estimate of deviation between empirical distribution functions in two independent samples. (Russian). Bull. Moscow Univ. **2**(2), 3–16 (6.1, 6.2) (1939).

[3] International Organization for Standardization. (2006). *Graphic technology. Input data for characterization of 4-colour process printing. Part 2: Expanded data set* (ISO Standard No. 12642-2:2006). https://www.iso.org/standard/39371.html.

[4] ISO/CIE 11664-6:2022, *Colorimetry — Part 6: CIEDE2000 colour-difference formula*, CIE, Vienna. (2022).
[5] T. J. Cholewo and S. Love, "Gamut boundary determination using alpha-shapes", in Proceedings of IS&T and SID's 7th Color Imaging Conference (IS&T, Springfield, VA, 1998), pp. 200–204.
[6] H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes", ACM Trans. Graphics 13, 43–72 (1994).
[7] B. C. K. Ly, E. B. Dyer, J. L. Feig, A. L. Chien, S. Del Bino, Research Techniques Made Simple: Cutaneous Colorimetry: A Reliable Technique for Objective Skin Color Measurement, Journal of Investigative Dermatology, 140(1):3-12.e1 (2020). https://doi.org/10.1016/j.jid.2019.11.003.
[8] R. W. G. Hunt, M. R. Pointer, Measuring Colour, 4th Edition, John Wiley & Sons, Ltd., pp. 198 (2011).
[9] ASTM Document E1345, *Standard practice for reducing the effect of variability of color measurement by use of multiple measurements*, ASTM International, West Conshohocken, PA, USA (2008).
[10] International Organization for Standardization. (2022). *Graphic technology — Image quality evaluation methods for printed matter. Part 11: Colour gamut analysis* (ISO Standard No. ISO/TS 18621-11). https://www.iso.org/standard/81505.html.

Author Biography

Ján Morovič received his Ph.D. in color science from the University of Derby (UK) in 1998, where he then worked as a lecturer. Since 2003 he has been at Hewlett-Packard in Barcelona as a senior color scientist and later master technologist. He has also served as the director of CIE Division 8 on Image Technology and Wiley and Sons have published his 'Color Gamut Mapping' book. He is the author of over 120 papers and has filed 180+ US patents (143 granted).

Peter Morovič received his Ph.D. in computer science from the University of East Anglia (UK) in 2002 and holds a B.Sc. in theoretical computer science from Comenius University (Slovakia). He has been a senior color and imaging scientist at HP Inc. since 2007, has published 65+ scientific articles and has 180+ US patents filed (142 granted) to date. His interests include color science, image processing, color vision, computational photography, computational geometry.