COLOR DIFFERENCE EQUATIONS AND THEIR ASSESSMENT

Scott Millward
sxm@rit.edu

KEYWORDS

color difference, visual assessment, color measurement

ABSTRACT

In 1976, the International Commission on Illumination, CIE, defined a new color space called CIELAB. It was created to be a visually uniform color space. At the same time the color difference equation ΔE_{ab} was developed to communicate color tolerances. However, CIELAB is not truly visually uniform, making colors having the same ΔE_{ab} magnitude in different regions of the color space appear of different magnitude. Instead of developing a new color space, the color science community has developed several other color difference equations that use higher order mathematics to give more or less weight to CIELAB values in different areas of the color space, resulting in color difference equations that better correlate with visually perceived differences.

This research uses ten reference Pantone color samples, and four other Pantone colors, which are distanced about 6 ΔE_{ab} around each reference color. A paired comparison test was conducted so that the perceived color differences between the reference color and the four sample colors could be ranked. Five color difference equations (ΔE_{ab}, ΔE_{94}, ΔE_{00}, ΔE_{CMC}, and ΔE_{DIN99}) were evaluated to determine which best correlates with the perceived color difference. The results show (1) that only four out of ten paired comparison tests had significant agreement among all 10 observers; (2) the ΔE_{ab} equation did a good job in predicting color differences for near neutrals; and (3) there is no clear winner for a color difference equation that outperforms the rest.

1. INTRODUCTION

CIELAB color space was intended to represent color by numbers in a visually uniform way. The difference between two colors can be calculated easily using an equation developed by the CIE in 1976 called ΔE_{ab}. This equation calculates the linear (Euclidian) distance between two points in the L*a*b* 3D space. Even though, L*a*b* is not truly visually uniform, it is the standard color space used by the graphic communications industry.

The human visual system is more sensitive to different kinds of changes and perceives these differences in different magnitudes even though they may have the same calculated difference. For instance, the eye is more sensitive to changes in neutrals than in high-chroma colors having the same lightness. In addition, the eye is more sensitive to changes in chroma than changes in lightness for neutrals, but this is not so for yellows. So the same vector distance may not be perceptually the same for all colors.

Many ΔE equations have been created to try to calculate the perceived difference between two colors by giving more tolerance to areas that the eye is not as sensitive to. This research will try to identify which of five equations best describes the perceived color difference in areas of interest.

2. METHODOLOGY

This study is a small-scale experimentation of a possible way of evaluating color difference equations. Only ten areas of interest were chosen due to the limitation of using pre-printed Pantone colors. This section explains how the areas of interest were chosen, preparation of patches for evaluation, and the evaluation procedure.

2.1 CHOOSING AREAS OF INTEREST

There are data sets, e.g., RIT-DuPont and others with equal visual differences, that have been developed and used by researchers to evaluate the performance of different color difference equation. This research is organized to identify color patches with equal, but noticeable color differences as the first step in the assessment of color difference equations.

Using colorimetric values from measurements taken from a Pantone swatch book, calculate the ΔE_{ab} of all possible combinations. Determine which Pantone colors have at least four other Pantone colors with a ΔE_{ab} of 6.0 (±0.5). An automated script was written to generate this list for this study. From this list, select ten patches as areas of interest in different hues, saturations, and lightnesses, and mark as References R1 to R10. Select four Pantone colors with a color difference of 6.0 ΔE_{ab} (±0.5) for each reference patch and mark them as Samples S1 to S4.

2.2 PAIRED COMPARISON EVALUATION

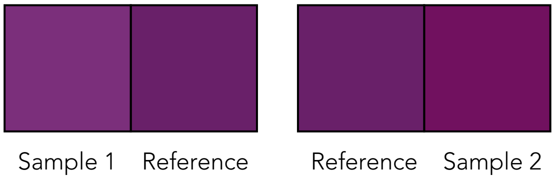

Figure 1: Paired comparison setup.

Prepare references and samples by cutting squares from a Pantone Solid Chip book and mounting them flush on mounting board. Place two reference tiles side-by-side separated by about four inches on the gray table inside a light booth (D50 lighting was used). Place two samples against the outside edges of the references (Figure 1). Ask the observer “Which pair demonstrates the smaller color difference?” Continue switching pairs of samples for each combination and recording the observer’s response.

3. RESULTS AND ANALYSIS

Results are reported by selected areas of interest and paired comparison evaluation, followed by analyses of the five color difference equations that were used.

3.1 SELECTED AREAS OF INTEREST

Figure 2 shows the Pantone swatches that were used as reference and sample patches. The Pantone Library is an example of a large database of colors with preprinted samples. It does not consist of evenly distributed colors and therefore does not include a large selection of patches that meet the requirements for this test. Most of the viable patches are in the L* 75 to 85 range and are not representative of a typical color gamut. An effort was made to select areas of interest with different hues and saturations.

Reference 1 (R1) represents a saturated red hue and Reference 2 (R2) is a chromatic blue. Reference 3 (R3) is a neutral gray that will show colorcasts and changes in lightness. Reference 4 (R4) is a less chromatic pink than R1; Reference 5 (R5) is a mint green. References 6 (R6) and 7

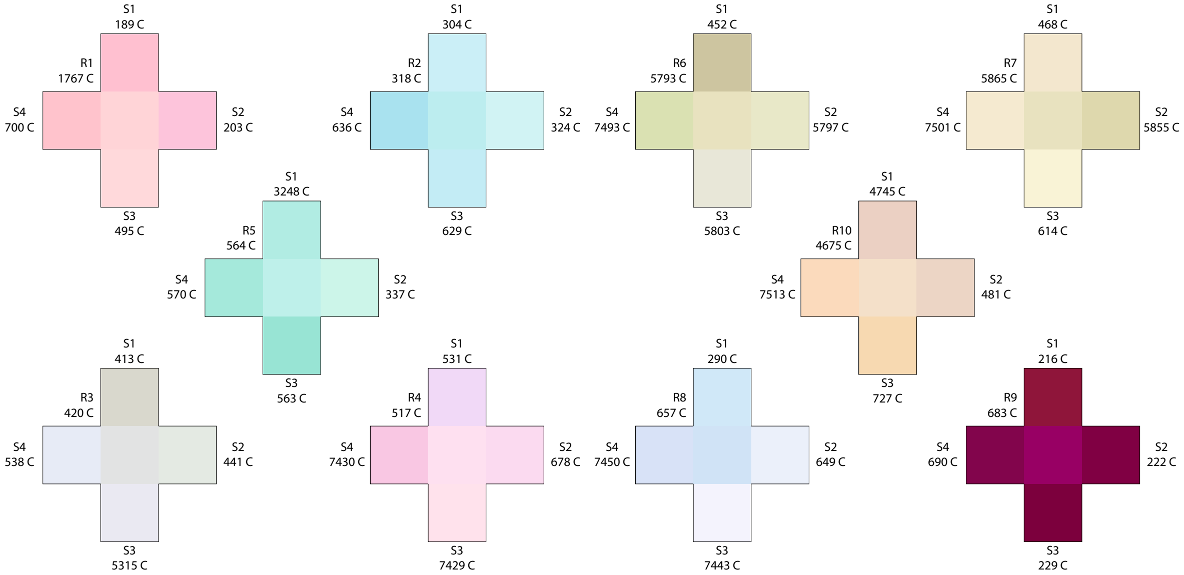

Figure 2. Selected Reference and Sample Pantone swatches. Note: This figure was printed with CMYK and may not represent the actual Pantone colors listed. An effort was made to maintain a similar perceptual difference.

(R7) are both low chromatic yellows with Reference 7 being less chromatic. All of the surrounding samples are more or less chromatic or have slight hue or lightness shifts. Reference 8 (R8) is a pale blue color that will show hue shifts in color. Reference 9 (R9) is a dark maroon color. Reference 10 (R10) is a flesh tone area and represents one of the memory colors.

3.2 OBSERVERS

Ten observers were chosen from within RIT's College of Imaging Arts and Sciences building. The ages of the observers ranged from 18 to 30 years old. Eight participants were male and two were female. The group consisted of observers from India, China, Denmark, and North America for cultural diversity. Five observers were graduate students in the School of Print Media with interest and a high degree of experience in color theory. Three observers were undergraduate students who have some experience with color and images because they were enrolled in programs within the College of Imaging Arts and Sciences. The tenth observer was a visitor to the school with little experience with judging color. None of the observers were aware if they had any level of color deficiency when asked.

3.3 PAIRED COMPARISON EVALUATION

Each observer was asked to identify the sample that demonstrated less color difference to the reference in each pair. These rankings reflect the sample that shows the smallest to largest difference to the reference. Only four of the ten areas of interest showed a significant agreement among judges (R1, R3, R4, R6), which indicates for many of the areas of interest, it is difficult to choose between two samples that are approximately the same difference, just in different ways. There were relatively few triads found in all of the observations, indicating that the judges are fairly consistent within themselves. These triads were excluded from the analysis.

In light red Reference 1, there was a significant agreement among the judges, with a coefficient of concordance of 0.9. The darker and slightly less saturated sample (R1S4 – Pantone 700C) was viewed to be the least different of all the samples and the much more saturated sample

(R1S1 – Pantone 189C) was viewed to be the most different. These two sample patches differed mostly in the a* where the less saturated patch matched closer than the more saturated color.

In the light, chromatic blue Reference 2, the sample that was viewed as the least different (R2S2 – Pantone 324C) did not change in lightness very much but was more yellow than the other samples. Sample 4 (R2S4 – Pantone 636C), which was deemed to have the largest change, was darker than the other patches and hue shifted to be more purple. This would indicate that we see more change in lightness than chromatic changes. However there was a low agreement among judges, with a coefficient of concordance of 0.32, so this set may not be an accurate measure.

The neutral gray set (Reference 3) had the highest agreement among judges, with a coefficient of concordance of 0.95. The slightly darker and more yellow sample (R3S1 – Pantone 413C) was viewed to be the most like the reference. The sample that was viewed as the worst match (R3S3 – Pantone 5315C) was much lighter and redder than the reference even though the third most different sample (R3S4 – Pantone 538C) was lighter. Neutral colors are more susceptible to changes in hue and lightness because it does not take much deviation to get away from the neutral axis and become noticeable.

In low chromatic pink Reference 4, the patch that was observed to be the least different (R4S2 – Pantone 678C) was the patch with the lowest hue difference. It was darker and less chromatic but there was no hue shift, indicating that this is important. The largest difference was viewed in the patch that was greatly darker and more saturated (R4S4 – Pantone 7430C). This patch actually had a significantly higher ΔE than the others so this is to be expected. This set also had a high coefficient of concordance of 0.90.

In high chromatic mint green Reference 5, the smallest difference was seen in the patch that was significantly lighter but approximately the same chromaticity (R5S2 – Pantone 337C). The largest difference (R5S1 – Pantone 3248C) was much more chromatic than the reference indicating that saturation is the influence in this area. However, there was little agreement among judges, with a coefficient of concordance of 0.35, so opinions varied.

In low chromatic yellow Reference set 6, the lowest difference (R6S2 – Pantone 5797C) was seen in the lighter and more chromatic patch and the most (R6S1 – Pantone 452C) was seen in the much darker patch. Again this indicates sensitivity to lightness over other factors. The agreement among judges was high with a coefficient of concordance of 0.90.

Reference 7, another low chromatic yellow, was difficult to judge according to the coefficient of concordance of 0.66. The sample that was judged as the least different (R7S2 – Pantone 5855C) actually had the largest change in lightness (darker) and was much more chromatic than the others. It was also the patch that had the smallest hue shift along with the second least different (R7S3 – Pantone 614C), which also had a large lightness difference (lighter) and was also more chromatic. This indicates that a hue shift is very important rather than lightness and chromaticity. Because the patches ranked first and second were both very similar in their changes, just in different degrees, this would explain the lower agreement among judges. The worst patch (R7S1 – Pantone 468C) had only slight changes in lightness and chromaticity but had a significant hue shift.

Reference 8, a pale blue, also had a low agreement among judges with a coefficient of concordance of 0.44. The patch that showed the least difference (R8S2 – Pantone 649C) was lighter and significantly less chromatic but had a very small hue shift. The worst patch (R8S1 – Pantone 290C) had little change in saturation and was a little darker but the major change was in the hue. Again this indicates that a hue shift is most noticeable.

Reference 9, a dark maroon, also had low agreement among judges with a coefficient of concordance of 0.60. The observers stated that this was a hard set to judge because the difference of each pairing seemed so similar. Both the observed lowest (R9S2 – Pantone 222C) and second lowest (R9S3 – Pantone 229C) patches had a small hue difference and the least different was darker the change in saturation was lower than the second. The worst patch was significantly less chromatic and much darker than the reference.

The 10th reference set, the flesh tones, did not have a statistically high agreement among judges (coefficient of concordance of 0.73) but agreement was close. In this case the chromaticity was a factor in choosing the least different (R10S2 – Pantone 481C) since the second least different (R10S3 – Pantone 727C) had less of a hue shift and both were darker. The worst sample (R10S4 – Pantone 7513C) and second worst sample (R10S1 – Pantone 4745C) both had large hue shifts.

While not all judges could agree, it seems that, in lighter colors, a hue shift is most important, seconded by lightness and chromaticity. Because this study was limited to patches in the lighter areas of a typical color gamut, the results can only be attributed to these areas. The limited number of observers also makes the data susceptible to bias. A more finely tuned study would require a more comprehensive selection of areas of interest that would represent all hues, saturations, and lightnesses of a color gamut as well as a larger observer base.

3.4 COLOR DIFFERENCE EQUATIONS

Color difference equations are designed to quantify the color differences as perceived by the human visual system. The paired comparison test above sets guidelines as to how people perceive color difference in the areas of interest. Next, the color difference equations are used to quantify the perceived differences and compare them to the guideline.

Each of the five color difference equations used in this study tries to more accurately match the visual difference seen by the human visual system than the traditional ΔE_{ab} equation. Some work better than others in different areas. For instance if we compared two colors with L*a*b* values of 50; 48; 73 and 48; 47; 60 the ΔE_{ab} would be 13.19, which is considered to be very poor and unacceptable. The same patches when considered using other formulas produce very different results: ΔE_{94}=4.20, ΔE_{00}=4.91, ΔE_{CMC}=6.84, ΔE_{DIN99}=3.54. ΔE_{94}, ΔE_{00}, and ΔE_{DIN99} predict that the patches are different but may be visually acceptable. ΔE_{CMC} was getting to the point of being unacceptable but was half of the ΔE_{ab} equation.

The color difference equations are identified and shown below with explanations based on the results and analyses of findings.

3.4.A CIE1976 (ΔE_{ab})

As discussed this is the Euclidian distance between two points in a 3D space. This would work fine if the L*a*b* color space were visually uniform, but it is not. This equation is mathematically easy but does not generally correlate with a visual difference.

$$\Delta E_{ab} = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2}$$

(ISO/DIS 13655, 1996)

3.4.B CIE1994 (ΔE_{94})

In 1994 the CIE made an attempt to correct for the visually non-uniformity of L*a*b* by weighting lightness, chroma, and hue in different proportions (Habekost, 2007). The math is not overly complicated and correlates better to the visual difference.

$$\Delta E_{94} = \sqrt{\left(\frac{\Delta L}{K_L S_L}\right)^2 + \left(\frac{\Delta C}{K_C S_C}\right)^2 + \left(\frac{\Delta H}{K_H S_H}\right)^2}$$

(Hunt, 2004)

3.4.C CIE2000 (ΔE_{00})

ΔE_{00} was an attempt to improve upon the 1994 equation by adding more weighting factors depending on the hue angle of the color (Habekost, 2007). This is the most complicated color difference equation mathematically but does tend to correlate better to the visual difference. There is some question about the data that was used to create this equation but it seems to work (Granger, 2008).

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{K_L S_L}\right)^2 + \left(\frac{\Delta C'}{K_C S_C}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2 + R_T\left(\frac{\Delta C'}{K_C S_C}\right)\left(\frac{\Delta H'}{K_H S_H}\right)}$$

(Hunt, 2004)

3.4.D COLOR MEASUREMENT COMMITTEE (ΔE_{CMC})

This equation was not created by the CIE but by the Color Measurement Committee (of the Society of Dyers and Colourists of Great Britain) and is used primarily in the textile industry. Again, there is weighting placed on the lightness, chroma, and hue of the colors (Habekost, 2007). It is similar to the ΔE_{94} equation but is slightly more complicated.

$$\Delta E_{CMC} = \sqrt{(\Delta L / lS_L)^2 + (\Delta C / cS_C)^2 + (\Delta H / S_H)^2}$$

(Hunt, 2004)

3.4.E DIN99 (ΔE_{DIN99})

DIN99 is a German standard not well known in North America. This equation warps the actual color space to a more uniform model before calculating the difference. This unique method makes the math relatively simple; in fact, the color difference equation is the same as ΔE_{ab} after the color space is warped.

$$\Delta E_{DIN\,99} = \sqrt{\Delta L_{99}{}^2 + \Delta a_{99}{}^2 + \Delta b_{99}{}^2}$$

(DIN 6176, 2001)

4. DISCUSSION AND CONCLUSION

As discussed earlier, the different color difference equations give weightings in different parts of the color space to better match the differences seen by the human eye. This means that different areas of the color space will show difference more than others and different factors of the color difference are more perceivable than others. Table 1 shows two examples of how visual ranking and calculated rankings agree. The agreement increased if calculated rank matched the visual rank of a specific sample.

Table 1. Examples of agreement between visual ranking and calculated rankings. Red indicates match.

Sample	Visual Rank	ΔE_{ab} Rank	ΔE_{94} Rank	ΔE_{00} Rank	ΔE_{CMC} Rank	ΔE_{DIN99} Rank
Reference 3 (R3)						
R3S1	1	1	1	1	1	1
R3S2	2	2	2	2	2	2
R3S4	3	4	4	4	4	4
R3S3	4	3	3	3	3	3
Agreement		**2**	**2**	**2**	**2**	**2**
Reference 9 (R9)						
R9S2	1	1	2	2	2	2
R9S3	2	2	1	1	1	1
R9S4	3	3	4	4	3	3
R9S1	4	4	3	3	4	4
Agreement		**4**	**0**	**0**	**2**	**2**

Figure 3 shows the agreement between the observers' ranking, of least different patches to greatest difference, to the calculated difference of each color difference formula. In some cases none of the calculated rankings matched the observed rankings and resulted in a zero agreement.

In R1, all five of the equations predicted the smallest differences as compared to the visual observations, but did poorly ranking the rest. In R2 ΔE_{DIN99} predicted the two least differences and ΔE_{94} predicted the worst two, but there was little agreement between observers. In R3 all of the equations predicted the two least different patches but switched the more different patches indicating that neutrals can be calculated by any equation. Using R4 the ΔE_{94} and ΔE_{DIN99} predicted the entire ranking correctly and the others predicted only the least and worst patches. This set is a good indicator of the validity of each equation. In R5 all but ΔE_{94} predicted the least different and ΔE_{00} and ΔE_{CMC} the third smallest, however there was little agreement between the judges on this set. Using R6 all of the equations agreed with the visual rankings with the exception of ΔE_{94}, which swapped the two least different. R7 was not agreed upon between judges very well but ΔE_{00} and ΔE_{CMC} predicted the least different, ΔE_{00} predicted the second least different and ΔE_{DIN99} predicted the third different. ΔE_{ab} was way off in R8 not predicting any and actually transposing the least and most different samples. The other four equations predicted the least different and ΔE_{CMC} predicted them all. Using R9 ΔE_{ab} agreed with all the visual rankings and ΔE_{CMC} and ΔE_{DIN99} with the two worst patches. In R10 all but ΔE_{CMC} predicted the least different and ΔE_{00} agreed completely.

As seen in Figure 3, the equation that agrees most with the visual observations is ΔE_{DIN99} followed by ΔE_{CMC}, ΔE_{00}, ΔE_{ab}, and finally ΔE_{94}. This means that ΔE_{DIN99} or ΔE_{CMC} are most likely to provide a color difference factor that most closely matches the difference perceived by the human visual system. Since ΔE_{CMC} and ΔE_{00} are very complex formulas, ΔE_{DIN99} may be a valid choice for everyday use.

Other studies on the assessment of color difference equations have said that ΔE_{00} quantify small perceived color differences more accurately than other equations (Luo et al., 2004; Johnson, 2006; Habekost, 2007). While others hold ΔE_{CMC} to more consistently correlate with perceived differences (Habekost, 2008). These studies use various methods, color sample base and observer group sizes, which will vary their final conclusions.

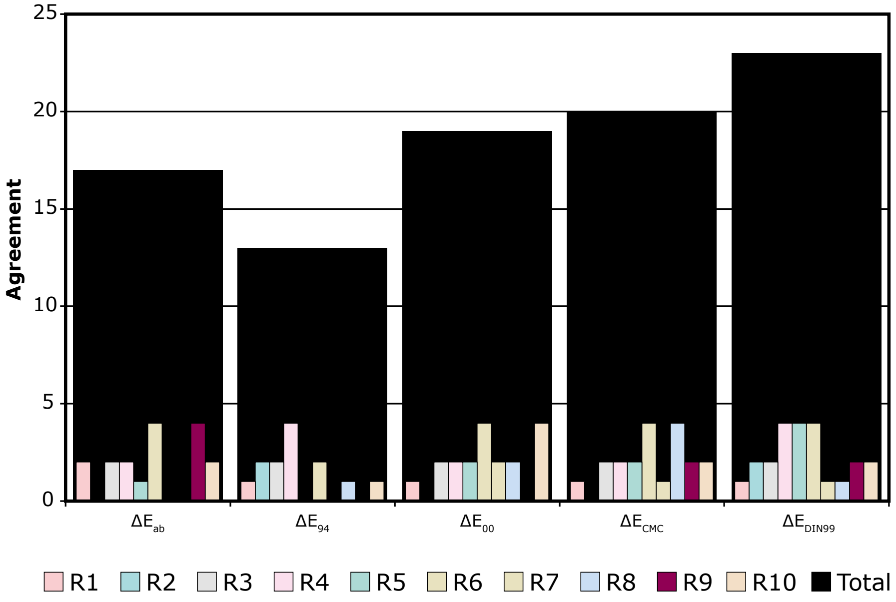

Figure 3. Agreement between observer ranking and color difference equation rankings. Note: This figure was printed with CMYK and may not represent the actual Pantone colors listed. An effort was made to maintain a similar perceptual difference.

5. FUTURE RESEARCH AND LIMITATIONS

In this research, the assessment of color difference equations is based color on samples selected from an existing Pantone color swatchbook with noticeable visual difference of around 6 ΔE_{ab}. Exact measured color differences between samples and their reference are not critical because these color difference pairs are judged visually to form a visual scale. These visually derived color difference scales are used to evaluate the performance of five color difference equations.

Color has three dimensions. It is difficult to tell if one of the three attributes carries more influence in the visual color assessment than the other two. A possible improvement of the experiment is to limit color samples with similar lightness values by producing color patches varying in hue and chroma only using ICC color management.

This study does not look at the scale of the color difference, only the rank. Further study is needed to see how accurately and how uniformly these equations perform in placing a usable scale of difference on two colors.

A greater number of observers than this study sampled would also be necessary to average out the inherent personal bias of two equally different sample patches.

6. ACKNOWLEDGMENTS

I would first like to thank Professor Robert Chung for his encouragement for this study and guidance with the evaluation format. I would also like to thank Dr. Martin Habekost of Ryerson University for his help in interpreting the different equations. Lastly, I would like to extend my gratitude to the entire *Test Targets 9.0* team. Without all of their hard work this publication would not be possible.

7. REFERENCES

DIN 6176: Colorimetric calculation of color differences with the DIN99 Formula (2001). Berlin, Germany: German Institute for Standardization.

Granger, E. (2008). A comparison of color difference data and formulas. *2008 TAGA Proceedings,* pp. 191-202.

Habekost, M. (2007). Color difference equations and the human eye, *2007 TAGA Proceedings,* pp. 291-310.

Habekost, M. (2008). Evaluation of digital proofs using the GRACoL dataset and various color difference equations, *2008 TAGA Proceedings,* pp. 392-415.

Hunt, R. W.G. (2004). *The Reproduction of Colour.* Chichester, England: John Wiley & Sons Ltd.

ISO/DIS 13655:1996 Graphic technology — Spectral measurement and colorimetric computation for graphic arts images. Geneva, Switzerland: International Organization for Standardization.

Johnson, A., Green, P. (2006). The Color Difference Formula CIEDE2000 and its Performance with a Graphic Arts Data Set, TAGA Journal of Graphic Technology, Vol. 2, pp. 59-71.

Luo, M.R, Minchew, C., Kenyon, P. and Cui, G. (2004). Verification of CIEDE2000 using industrial data, *AIC Color and Paints, Interim Meeting of the International Colour Association, Proceedings,* pp. 97–102.