Evaluation of Performance of Several Color-Difference Formulae Using a New NCSU Black Experimental Dataset

Renzo Shamey[1], Juan Lin[1], Weethima Sawatwarakul[1], Gang Fang[1]. [1]Color Science Laboratory, TECS Department, North Carolina State University, Raleigh, NC 27695-8301 USA;

Abstract

The objectives of this study were to develop a specific visual dataset comprising samples with low lightness (L range from 11 to 19), covering near neutral black substrates that varied in hue and chroma, and testing the performance of major color difference formulae currently in use as well as more recent CIECAM02 color difference formulae including CAM02-LCD, CAM02-SCD as well as CAM02-UCD models based on black samples. The dataset comprised 50 dyed black fabrics with a distribution of small color differences from 0 to 5. The visual color difference between each sample and the standard was assessed by 20 observers in three separate sittings with an interval of at least 24 hours between trials using an AATCC standard gray scale and a total of 3000 assessments were obtained. A third-degree polynomial equation was used to convert gray scale ratings to visual differences. The Standard Residual Sum of Squares (STRESS) was employed to evaluate the performance of thirteen color difference formulae based on visual results. Based on the analysis of STRESS index results the CAM02-SCD color difference equation performed better than other equations, however, all equations performed poorly in this region of the color space.*

Introduction

Establishing an accurate relationship between visual assessments of the perceived differences between two color patches and various models that predict the average perceived magnitude of such differences is of importance in the development of color difference models. The CIE color difference formulas are based on color patches that are assessed under controlled viewing and illumination conditions and form specific visual datasets. These models have been widely accepted and applied in industry. Nevertheless, there are differences among formulas in terms of their overall performance, with each newer model attempting to overcome the shortcomings of the previous versions. It is important to determine the performance of the recommended color difference equations, and the corresponding computation of differences in specific regions of the color space to minimize prediction errors and improve results. Over the past 40 years, several models have been developed to correlate color appearance with color difference data. The most recent CIE recommended color difference formula for general use is CIEDE2000 which was developed to improve prediction of differences especially in the neutral region as well as for blue colors. More recent developments include those based on color appearance models such as, CIECAM02, as a revision of CIECAM97s model. These models were developed to improve the accuracy of predictions by employing surround conditions and incorporating chromatic adaptation transforms. The CIECAM02 model was adopted by the CIE [1]. Li et al. [2] found out that CIECAM02 outperformed CIECAM97s in all spaces, and concluded that CIECAM02 may be used as a universal color model for all colorimetric applications. In a previous work, we reported that DIN99d equation outperformed 11 other equations using NCSU-B1 low chroma blue dataset as well as NCSU-2 general dataset. This evaluation was based on minimizing the standardized residual sum of squares (STRESS) index [3]. CIECAM02 equations did not perform as well and moreover the performance of CIEDE2000 was not significantly better than some of the other equations examined.

In order to investigate the performance of color difference formulae in the neutral region, an experimental visual dataset was developed to supplement the study and contribute to the development of a uniform color space for industrial color difference assessments. The distribution of samples in the NCSU-BK (black) dataset is displayed in three-dimensional coordinates in Figure 1. The sample shown in red represents the Standard or the center of the dataset.

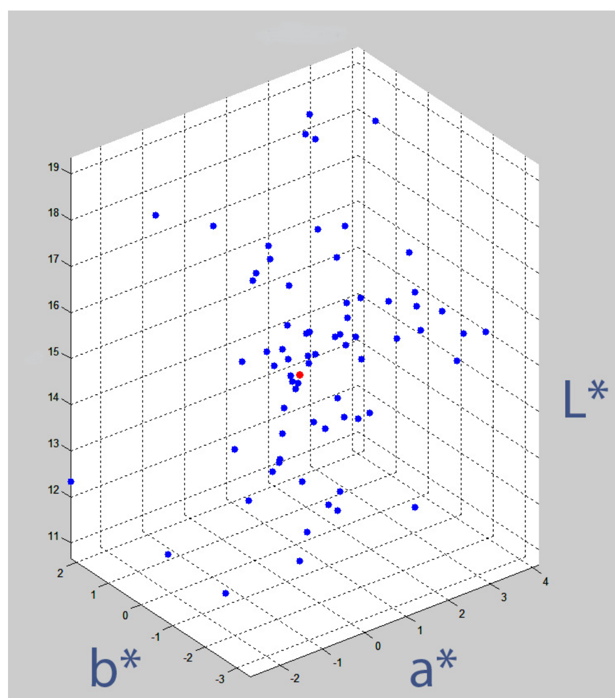

***Figure 1**. Distribution of Samples Representing NCSU-BK dataset in CIELAB Color Space.*

Experimental

Sample Preparation

Several hundred black textile samples were prepared, out of which fifty (50) samples were selected based on their distribution around a selected standard ($L^* = 14.98, a^* = 0.37, b^* = -0.92$ and X, Y, Z of 1.822, 1.905 and 2.152 respectively) for visual assessments. The black fabrics were cut into 2×2 inch square dimensions. Samples were orientated in the same direction for visual assessments to maximize visual uniformity and minimize variations in the visual assessment conditions. The colorimetric attributes of samples, $L^*a^*b^*C^*h$ as well as XYZ were measured with a DataColor SF600X bench top reflectance spectrophotometer using a large area view aperture, illuminant D65, CIE 1964 supplementary colorimetric observer (10°) and with specular light excluded and UV included. During the colorimetric measurements, each single sample was folded into 4 layers to ensure opacity and measured a total of 3 times and results were averaged. Samples were rotated 90° after each reading to reduce measurement variability due to fabric structure, directionality of yarns as well as the non-uniformity of dyeings. The 50 sample pairs used in the NCSU-BK dataset had an average ΔE_{ab}^* of 2.35, with a range from 0.29 to 5.11. Figure 2 shows the ΔE_{ab}^* histogram, based on the results of colorimetric measurements, of all black samples used in the NCSU-BK dataset.

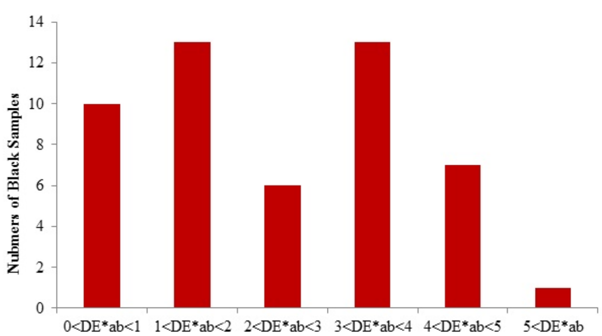

Figure 2. *Histogram of NCSU-BK samples in the dataset as a function of their CIELAB color differences.*

Visual Assessments

For the development of the NCSU-BK visual dataset a panel of 20 naive observers (8M and 12F) with an average age of 23 ranging from 20 to 43 was employed to assess the perceptual differences amongst black samples. During the visual assessment procedure, the experimenter as well as each subject wore a mid-gray laboratory coat and a pair of grey gloves in order to minimize color variability of the surround in the course of the experiment and to avoid damaging samples. All observers were color normal according to the Neitz test for color vision [4].

A SpectraLight III calibrated viewing both from X-Rite illuminated with filtered tungsten blubs simulating illuminant D65 at a color temperature of 6489K and an illumination intensity of 1400 lux inside the chamber was used. All extraneous light sources were excluded during the assessment. Before the experiment, each subject viewed the empty illuminated viewing booth for 2 minutes to adapt to the light source during which time the experiment was explained. Each observer assessed NCSU-BK samples three times in separate sittings with a gap of at least 24 hours between trials. Each black sample was displayed next to the black standard with no gap between the samples and the AATCC Standard Gray Scale for Color Change [5] was used to determine the pair's color difference by comparing to contrast pair differences shown in the scale directly beneath them. A custom stand painted in neutral gray, to match the interior of the viewing booth, was used to display samples. The distance from the eye to the sample was kept at the arms length in order to simulate an approximately 10° field of view. The visual assessment procedure including approximately 45/0 illumination/viewing geometry is shown in Figure 3.

Assessment of visual results indicates that while all observers involved in the study were color normal, one observer exhibited outlier behavior. It is debatable whether any color normal observer(s) should be removed from visual data and there are proponents and opponents to this approach. In this study we considered data from all observers, however, in a subsequent manuscript the results will be reexamined in view of the reduced number of observers.

In addition, in the analysis of samples we found one sample that exhibited the unusual behavior of being close to standard with high visual difference, while two samples had the unusual behavior of being far from standard with small visual difference. These samples were examined in greater detail and re-measured but no particular causes were obtained and thus the samples were kept in the dataset.

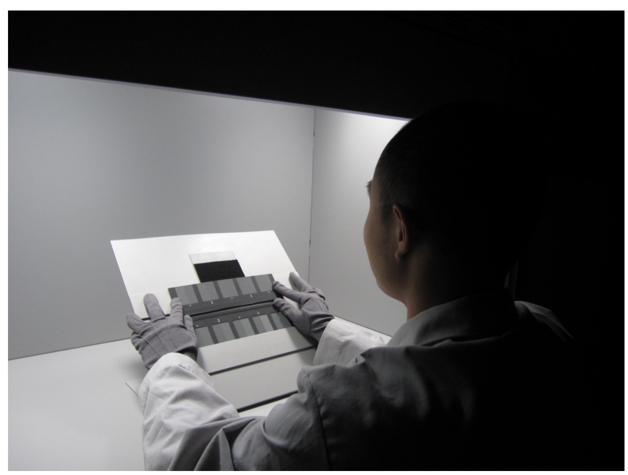

Figure 3. *Viewing/Illumination Geometry Employed for the Assessment of Samples in the NCSU-BK Dataset.*

Color Difference Equations Examined

The performance of eight color difference formulae was examined in the work reported here. These included CIELAB, CMC and CIE94 color difference formulae [8] as models established prior to the recommendation of CIEDE2000, and CIEDE2000 as the currently recommended CIE formula for general use [9], and DIN99 and BFD as two models that have shown good performance in previous studies [3, 7], and finally the CIECAM02 formulae, including CAM02-SCD, CAM02-LCD and CAM02-UCD [10] which are developed based on CIECAM97s model. Similar to conditions used in our previous studies [3], the parameters used in the CAM02's formulae were selected according to the experimental conditions, i.e. LA = 89.1cd/m2, Yb = 44.4, c = 0.69, Nc =1.0 and F = 1.0.

Optimization of K_L

The parametric factor k_L (or l) was optimized for each formula by minimizing STRESS. Results are shown in Table 1. The STRESS values were thus calculated for the black dataset at $k_L = 1$, k_L=2 as well as optimized k_L for each formula as shown in Table 2.

Table 2 shows that the STRESS values based on the optimized k_L are considerably less than those for k_L =1 and 2. Results shown in Table 2 also indicate that the performance of BFD is better in comparison to CIE94, CMC and CIE 2000 for k_L = 1 and also for the optimized k_L. As expected, in general, the performance of the color difference equations improves based on the optimized K_L values.

In addition, the STRESS values calculated for three CIECAM02 models, that is, CIECAM02-LCD, CIECAM02-SCD and CIECAM02-UCS, show significant improvement compared to other models examined. The lowest STRESS value among all equations was for the CIECAM02-SCD which indicates the equation based on the color appearance model for small color differences performs well in this region of the color space.

Table 1. Optimized K_L values for various color difference formulae using NCSU-BK dataset

	CMC	CIE 94	CIEDE 2000	BFD	CAM LCD	CAM SCD	CAM UCS
K_L	2.16	1.63	0.82	1.23	1.25	1.34	1.30

Table 2. Summary of STRESS values at K_L=1, K_L=2 and Optimized K_L for various color difference formulae using NCSU-BK dataset

	K_L = 1	K_L = 2	K_L = Opt
CMC	49.02	44.55	44.51
CIE94	46.51	44.83	44.54
CIEDE2000	44.46	47.59	44.15
BFD	44.11	45.05	43.76
CAM-LCD	42.86		41.10
CAM-SCD	40.87		40.84
CAM-UCS	41.42		40.94

While parametric factors like kL are related to viewing conditions (e.g. texture) and are necessary in textured objects, such as those used in this study, weighting functions should be employed to improve the overall performance of colour difference formulas in the “black” region of colour space. This will be the subject of additional work in a subsequent manuscript.

Assessment of Reduced Models

One of the main reasons for the development of a new color difference formula is to improve the performance of the previous model(s) in regions of a given color space where results are not considered satisfactory. In comparison to the simple DE_{ab}* color difference equation several parameters have been added over the years to enhance the performance of the equation(s) by taking into account several factors, such as hue super importance, surface characteristics of the objects being measured and the location of the object in the color space. Since the color space region comprising very dark objects has been an area where equations have not performed well, it was decided to test the performance of reduced models to determine the effectiveness of additional parameters. Results are shown in Table 3. Based on the results reported in Table 3 it can be concluded that:

1. The STRESS values based on the optimized k_L (or l) for each of the CMC, CIEDE94, CIEDE2000 and BFD formulas and their reduced forms are less than those based on k_L (or l) of 1 and 2. This confirms that k_L (or l) of 2 should not be assumed to always generate optimal results for all equations and that the use of optimized k_L (or l) values in these equations results in better performances, albeit in this case only slightly.

2. In the case of the NCSU-BK dataset, the most important correction to the CMC formula was found to be the chroma correction. STRESS increases by nearly 2.3 units when this correction is suppressed (S_C=1). The remaining two corrections to the CMC equation are also effective in that STRESS generally increases when they are removed. Compared with l =1, STRESS is reduced by a few units when l is set to 2 which shows the adjustment of this parametric factor is also effective for low chroma and dark textile samples, albeit not sufficiently.

3. In the case of the CIEDE94 color difference equation, suppressing lightness correction does not affect the STRESS value for k_L of 1, 2 or optimized k_L. The STRESS values for all suppressed cases are very close to that of the full model indicating that the correction functions are not very effective based on the visual dataset reported in this study.

4. In the case of the CIEDE2000 model, the S_L term for this visual dataset does not seem to be effective. The CIEDE2000 assumes L*=50 while in this study L*~72 was used. Testing a new S_L function in CIEDE2000 according to this relevant difference in the experiment could potentially be useful and will be examined in a subsequent manuscript.

The rotation correction did not affect the performance of the model and the STRESS values were approximately similar when the corrections were suppressed for the NCSU-BK samples. The rotation parameter is incorporated to account mostly for the irregularities in the blue region and this result is in line with expectations. No significant effects were found when the hue term was removed from the CIEDE2000 for black samples. As mentioned earlier, it has been shown that hue influences subjects' perception of blackness and thus this factor needs to be modified for near neutral dark samples. The chroma factor influences results to some extent and the STRESS values are increased when S_C is suppressed. The dataset is comprised of neutral samples and thus this term was expected to affect results significantly. Surprisingly the G correction term did not seem to influence the performance of the color difference equation based on the STRESS values obtained. The G term is designed to adjust the magnitude of differences for low chroma samples, as in this case, however, the performance of the model was not affected when G was removed. This implies that several of the existing parameters are not effective in improving the performance of the model for black appearing objects.

5. No change in the STRESS values, in comparison to the full BFD model, was noted when the lightness correction was removed. For the BFD equation suppressing the hue correction also does not affect the STRESS values significantly. However, the chroma correction played a role in improving the performance of the equation especially at the optimized *l*. However, again the G correction term did not influence the performance of the BFD color difference equation based on the STRESS values obtained. There are significant similarities between BFD and CIEDE2000 color difference models and thus results are in line with those observed for the CIEDE2000 equation.

Table 3. Comparison of STRESS values for data from NCSU-BK dataset for CMC, CIE94, CIE2000 and corresponding reduced models.

	K_L=1	K_L=2	Optimization
CMC-Full Model	49.02	44.55	44.51
CMC-Lightness	44.59	46.38	46.74
CMC-Chroma	51.30	45.84	45.58
CMC-Hue	49.93	44.56	44.39
CIE94-Full Model	46.51	44.83	44.54
CIE94-Lightness	46.51	44.83	44.54
CIE94-Chroma	46.25	44.85	44.48
CIE94-Hue	46.49	44.87	44.56
CIE2000-Full Model	44.46	47.59	44.15
CIE2000-Lightness	44.67	47.59	45.76
CIE2000-Chroma	45.49	49.11	44.83
CIE2000-Hue	44.75	47.93	44.38
CIE2000-Rot.term	44.46	47.59	44.15
CIE2000-G term	44.15	45.67	44.90
BFD-Full Model	44.11	45.05	43.76
BFD-Lightness	44.11	45.05	43.76
BFD-Chroma	46.85	46.22	45.93
BFD-Hue	45.54	44.90	43.77
BFD-Rot. term	44.11	45.05	43.77
BFD-G term	44.08	44.95	43.70

Assessment of Statistical Significance of Differences amongst Tested Formulae

The significance of differences in performance among equations can be tested using various statistical measures. One of the recent approaches is based on a statistical F test using the STRESS function. In Tables 4 and 5, the light green cells denote an insignificant improvement in performance for the formula shown in the row compared to the one given in the column. Light red cells indicate that the performance of the formula shown in the row is insignificantly *worse* than that given in the column and dark red cells signify the deterioration in performance is <u>significant</u> at 95% confidence level. As explained earlier, based on the STRESS results shown in Table 2 CIECAM equations outperformed all other equations examined by up to 9 units when compared to CMC(1:1) and on average 3-4 STRESS units when compared to other equations. Results in Table 2 show that in general equations perform better at k_L (or *l*) of 2 compared to k_L (or *l*) of 1. Detailed examination of results shows that all equations outperform the DIN99d equation, although the only significant difference was for CAMSCD. Otherwise statistically significant differences were not found among any of the equations examined in this study. Results at k_L of 1 also indicate that the performance of the CIEDE2000 compared to CMC, CIE94 and DIN99d equations was better although not significantly. The difference in performance of the CIEDE2000 (2:1:1) equation compared to other equations was found to be worse though the deterioration was also insignificant.

Results shown in Table 4 indicate that at the optimized k_L or *l* values the performances of BFD and CIEDE2000 color difference formulas in comparison to most of the other equations examined (with the exception of CIECAM02 based models) are improved, while again the improvement was statistically insignificant. Again DIN99d performed the worst and the difference was statistically significant compared to CAM-SCD and CAM-UCS models. Otherwise, there were no significant differences among any of the equations examined at the optimized k_L values.

Conclusions

For the NCSU-BK dataset, preliminary results based on the standardized residual sum of squares (STRESS) index show that at both k_L=1, k_L=2 and optimized k_L, CAM02-SCD outperforms all other equations tested in this study.

Absent of CIECAM equations, at k_L or *l* of 1 the BFD equation performed slightly better than the CMC, CIE94 and CIEDE2000 formulas, but the improvement was statistically insignificant. At k_L or *l* of 2 the CMC, CIE94, and BFD equations performed approximately similarly, and CIEDE2000 performed slightly worse than other equations.

Suppression of the chroma correction factor in the BFD and CIEDE2000 formulas lowered the performance of these models. In contrast, the lightness term was not found to be an effective correction for these models for the black dataset examined. Rotation and G correction terms also did not affect the STRESS values. The effect of the hue term on the performance of the models was not large. Previous studies have shown that bluish blacks appear blacker. This implies that several of the existing parameters are not effective in improving the performance of the model for black appearing objects. In general the STRESS values for NCSU-BK dataset are high indicating that accurate prediction of color in the low chroma neutral region is still a major challenge and that further modifications to the models are needed.

Three samples and one observer exhibited outlier behavior. Since we did not detect a major reason for the removal of this color normal observer and since the repeated measurements of the three samples did not change their unusual behavior we included results from all observers and samples in this study. In a future paper, visual results from the outlier observer will be removed from the dataset and the performance of equations in absence of 3 outlier samples will be reexamined to determine their effect on the performance of various equations.

In addition, we will examine a modification of weighting functions in various equations to determine the root causes for the poor performance of various equations in this region of the color space. Moreover, we will determine unit difference contour for neutral region to design appropriate weighting functions in the "black" region.

Acknowledgments

The authors thank all observers who took part in the visual assessments in this study.

Note

The visual data can be obtained by contacting the lead author at **rshamey@ncsu.edu**

References

[1] CIE. A Color Appearance Model for Color Management Systems: CIECAM02, CIE Publ. 159, 2004.
[2] Li C, Luo MR, Cui G. Color Difference Evaluation using Color Appearance Models. The 11th Color Imaging Conference, IS&T and SID, Scottsdale, Arizona, 13-15, 2002, 51-54.
[3] Shamey R, Hinks D, Melgosa M, Luo R, Cui GH, Huertas R, Cardenans L, Lee SG, Evaluation of Performance of Twelve Color-Difference Formulae Using Two NCSU Experimental Datasets, CGIV, 2010.
[4] Neitz J., Manual: Neitz Test for Color Vision, Western Psychological Services, Torrance, Calif, USA, 2001.
[5] AATCC EP 1, Gray Scale for Color Change, AATCC Technical Manual, 2004, 378-379.
[6] AATCC EP 9, Visual Assessment of Color Difference of Textiles, AATCC Technical Manual 82, 2007, 394-396.
[7] Lee SG, Shamey R, Hinks D, Jasper W, Development of a Comprehensive Visual Dataset Based on a CIE Blue Color Center: Assessment of Color Difference Formulae Using Various Statistical Methods, Color Res. Appl. 36,1, 2011, pp27-41.
[8] CIE 15:2004, 3rd Edition (CIE Central Bureau), 2004.
[9] CIE, Improvement to Industrial Color-Difference Evaluation, CIE Publ. 142 (CIE Central Bureau), 2001.
[10] Luo MR, Cui G and Li C, Uniform Color Spaces Based on CIECAM02 Color Appearance Model, Color Res. Appl. 31, 2006, 320-330.

Author Biography

Renzo Shamey is the Polymer and Color Chemistry Program Director in the TECS Department at North Carolina State University (NCSU). He obtained a Bachelor of Engineering in 1989 followed by MSc (1992) and PhD (1997) from the Colour Chemistry Department of Leeds University, England. He was a faculty member at Heriot-Watt University from 1998 to 2003 and since then has been at North Carolina State University where he currently serves as professor. He is active in the CIE Division 1 committees 55, 76, 77 and Chairs the AATCC RA36 Color Measurement Test Methods Committee.

Table 4. Significance test using NCSU-BK dataset for optimized k_L (or l) (N=50, α = 0.05, Fc = 0.62, 1/Fc = 1.61)

	k_L (or l)	CIELAB	CMC	CIE94	BFD	CIE2000	DIN99d	CAM02-LCD	CAM02-SCD	CAM02-UCS
		-	2.16	1.63	1.23	0.82	-	1.25	1.34	1.30
CIELAB	-	1	1.079	1.077	1.116	1.096	0.787	1.265	1.281	1.275
CMC	2.16	0.927	1	0.999	1.035	1.016	0.730	1.173	1.188	1.182
CIE94	1.63	0.928	1.001	1	1.036	1.018	0.731	1.174	1.189	1.184
BFD	1.23	0.896	0.967	0.965	1	0.982	0.705	1.134	1.148	1.143
CIE2000	0.82	0.912	0.984	0.983	1.018	1	0.718	1.154	1.169	1.163
DIN99d	-	1.270	1.370	1.368	1.417	1.393	1	1.607	1.627	1.619
CAM02-LCD	1.25	0.790	0.853	0.851	0.882	0.867	0.622	1	1.013	1.008
CAM02-SCD	1.34	0.780	0.842	0.841	0.871	0.856	0.614	0.987	1	0.995
CAM02-UCS	1.30	0.784	0.846	0.845	0.875	0.860	0.617	0.992	1.005	1

Table 5. Significance test for color difference equations based on the NCSU-BK dataset (N=50, α = 0.05, Fc = 0.62, 1/Fc = 1.61)

		CIELAB	CMC		CIE94		BFD		CIE2000		DIN99d	CAM-LCD	CAM-SCD	CAM-UCS
	k_L (or l)	-	1	2	1	2	1	2	1	2	-	0.77	1.24	1
CIELAB	-	1	0.889	1.077	0.988	1.063	1.098	1.053	1.081	0.944	0.787	1.163	1.279	1.246
CMC	1	1.124	1	1.211	1.111	1.196	1.235	1.184	1.216	1.061	0.885	1.308	1.439	1.401
	2	0.929	0.826	1	0.917	0.988	1.020	0.978	1.004	0.876	0.731	1.080	1.188	1.157
CIE94	1	1.012	0.900	1.090	1	1.076	1.112	1.066	1.094	0.955	0.797	1.178	1.295	1.261
	2	0.940	0.836	1.013	0.929	1	1.033	0.990	1.017	0.887	0.740	1.094	1.203	1.171
BFD	1	0.910	0.810	0.980	0.899	0.968	1	0.959	0.984	0.859	0.717	1.059	1.165	1.134
	2	0.950	0.845	1.023	0.938	1.010	1.043	1	1.027	0.896	0.748	1.105	1.215	1.183
CIE2000	1	0.925	0.823	0.996	0.914	0.984	1.016	0.974	1	0.873	0.728	1.076	1.183	1.152
	2	1.060	0.943	1.141	1.047	1.127	1.164	1.116	1.146	1	0.834	1.233	1.356	1.320
DIN99d	-	1.270	1.130	1.368	1.255	1.351	1.395	1.337	1.373	1.199	1	1.478	1.625	1.582
CAM-LCD	0.77	0.860	0.764	0.926	0.849	0.914	0.944	0.905	0.929	0.811	0.677	1	1.100	1.071
CAM-SCD	1.24	0.782	0.695	0.842	0.772	0.831	0.858	0.823	0.845	0.738	0.615	0.909	1	0.974
CAM-UCS	1	0.803	0.714	0.864	0.793	0.854	0.882	0.845	0.868	0.758	0.632	0.934	1.027	1