How do Major Color-Difference Formulae Perform in the High Chroma Blue Region?

Renzo Shamey, Renbo Cao, Syed Shahzad Hussain Zaidy, Juan Lin, Seung Geol Lee
Polymer & Color Chemistry Program, TECS Department, North Carolina State University, Raleigh, NC 27695-8301, USA

Abstract

The objectives of this work were to develop a comprehensive visual dataset, NCSU-B2, around the CIE high chroma blue color centre and to use the new dataset as well as the low chroma blue dataset, NCUS-B1[1,2], to test the performance of the major color difference formulae in this region of color space using the standardized residual sum of squares (STRESS) index, correlation coefficient (r) and Spearman Rank coefficient (ρ) as performance metrics. The visual differences between the 65 samples and the color center were assessed by 16 observers under highly controlled viewing and illuminaiton conditions, using AATCC Gray Scale for Color Change, in three separate sittings and a total of 3120 assessments were obtained. The results showed that CIEDE2000 exhibited relatively better performance for the NCSU-B2 dataset in comparison to other equations examined. However, none of the tested equations gave a satisfactorily low STRESS value, or high correlation coefficient (r) and Spearman Rank coefficient (ρ) values. For the combined blue datasets (NCSU-B1 and B2), CAM02-UCS, CAM02-SCD, DIN99d and CIEDE2000 showed the best performance.

Keywords: color difference, visual dataset, CIEDE2000, statistical significance, performance, reduced-models

Introduction

The ultimate goal of research in the field of color difference equations is to develop a single-number shade pass/fail equation for evaluating objects. In surface color industries including plastics, paper and textiles typically small to medium color differences are used. The ideal model is intended to represent the average visual assessment of two stimuli and predict the magnitude of such differences accurately. Over the last several decades, a large number of color difference equations, such as CIELAB, CMC (*l:c*), BFD (*l:c*), CIE94 [3-6], etc. have been proposed. These are based on several visual datasets generated under various experimental conditions and employing different substrates as well as different evaluation methods. However, these formulas show large discrepancies in their performance and suffer from certain shortcomings. Specifically large errors in predicting chromatic differences for blue, dark and neutral colors are noted.

The newest CIE recommended color difference equation is CIEDE2000. This formula was optimized against five independent sets of perceptual color difference data [7, 8], mainly aiming to address the problems in predicting blue, dark and neutral colors. More specifically, in this equation, a rotation term is used to improve the performance of the model for blue colors, and a *G* factor is used to rescale the a^* scale in the CIELAB space to improve the performance of the model for near neutral colors. In an effort to continually improve the performance of color difference equations it is important to test the performance of CIEDE2000 using independent visual datasets [9].

With this in mind, the objectives of this work were to generate a comprehensive visual dataset around high chroma blue color center, denoted NCSU-B2, and compare the performance of CIEDE2000 to other color difference equations. The reference color-difference formulas tested were: CIELAB, CMC, BFD and CIE94 color difference formulas [4-6], established models prior to the recommendation of CIEDE2000; and a group of formulas including DIN99d [3], and those based on CIECAM02 color appearance model, i.e. CAM02-SCD and CAM02-UCS[11], which were developed after CIEDE2000.

Experimental Procedure

Description and Measurement of Samples

A total of 65 samples and a standard with attributes corresponding to a CIE recommended high chroma blue center (L^*:34, a^*:7, b^*:-44)[10] were produced on 100% knitted polyester fabrics using commercial disperse dyes stable to light and weathering, and a conventional dyeing method. Fabrics were precision cut into 2 × 2 inch dimensions for visual assessment after dyeing. Their distribution around the center is shown in Figure 1. The selection of samples ensured division into seven subsets, with samples selected to have color differences mostly due to lightness, chroma, hue, and their combinations as shown in Table I.

Table I Relative contribution of Lightness, Chroma and Hue to the overall color difference

Subset	Avg.			No
	$\Delta L^{*2}/\Delta E^{*2}$	$\Delta C^{*2}/\Delta E^{*2}$	$\Delta H^{*2}/\Delta E^{*2}$	
1	79.68	0.81	19.50	15
2	0.62	91.71	7.69	15
3	25.85	11.00	63.15	15
4	44.54	53.91	1.55	5
5	8.85	49.86	41.29	5
6	53.82	1.20	44.98	5
7	33.30	22.55	44.15	5

The knit structure was oriented during the preparation of the mounted samples to ensure maximum visual uniformity of all samples. The CIE illuminant D65 and CIE 1964 Supplementary Colorimetric Observer were used for all colorimetric calculations. The reflectance of all samples was measured with a Datacolor SF600X spectrophotometer using a large area view aperture (30mm). UV light was excluded and specular light was included. Each sample was folded into 4 layers to ensure opacity and was measured a total of 4 times and averaged. Samples were rotated 90° and repositioned after each reading to reduce measurement variability due to fabric construction, directionality of yarns, and non-uniformity in dyeing. The 65 sample pairs had an average ΔE^*_{ab} of 2.59, with a range of 0.23-4.59. Details of the NCSU-B1 dataset have been previously reported [1, 2]. Figure 2 shows the histogram of all samples representing the NCSU-B2 dataset.

Visual Assessment Methodology

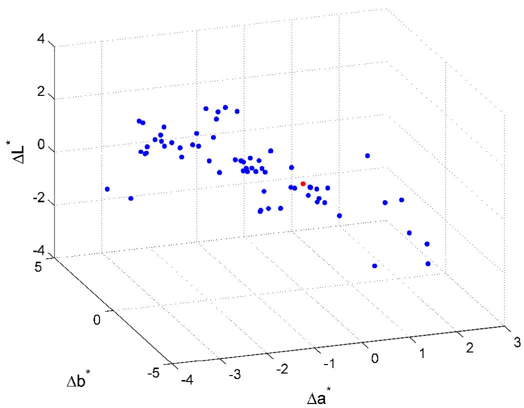

Figure 1 *The distribution of 65 samples on CIEL*a*b* color space (Red circle is the location of the standard).*

For the methodology described, a custom sample stand, based on a 45/0 degree illumination-viewing geometry, was manufactured and painted neutral gray (Munsell N7.25). The stand was used to house the standard and test samples as well as a gray scale pair, as shown in Figure 3.

The observers ranged in age from 20 to 42 years old. All observers had normal color vision according to the Ishihara confusion plates [13] and the Neitz test for color vision [14]. Most of the observers were naïve for the purposes of the experiment with 0 to 1 year prior experience in color assessments.

Each observer assessed samples three times with at least 24 hours gap between assessments. During the assessment subjects wore a mid-gray laboratory coat and a pair of gray gloves to minimize color variability of the surround during the course of the experiment and to prevent damaging the samples. The samples were placed by the experimenter who also wore a mid-gray laboratory coat. At the beginning of the experiment, the subject viewed the empty illuminated viewing booth for 2 minutes to adapt to the light source; during which time the experiment was explained.

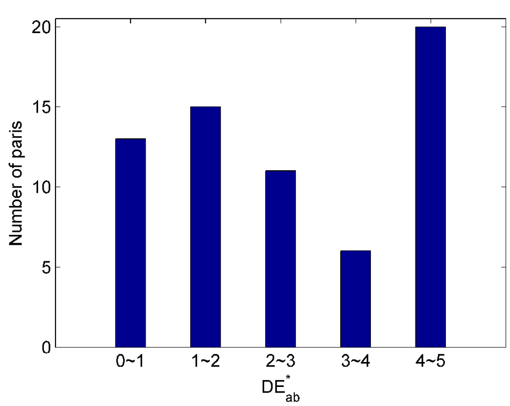

Figure 2 *Histogram of the 65 high chroma blue samples as a function of CIELAB color differences.*

The AATCC Gray Scale for Color Change [12] was placed directly below the standard and sample pair. During the assessment, observers gave a rating for the visual difference (ΔV) between each sample and standard pair according to AATCC Gray Scale contrast ratings from 1 to 5. The visual assessments were conducted under well controlled viewing and illumination conditions using a calibrated SpectraLight III (X-Rite) viewing booth equipped with a filtered incandescent daylight simulator. All extraneous light sources were excluded during the assessments. The viewing/illumination geometry, including the gray scale used is shown in Figure 3.

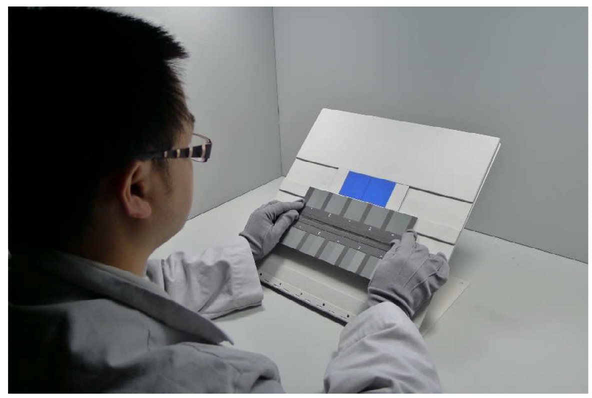

Figure 3 *Visual assessment involving 45/0 illumination viewing geometry, and a custom sample stand painted in neutral gray that housed the standard and test samples; as well as the AATCC gray scale used in this study.*

Conversion of Gray Scale Ratings to Visual Difference

A third degree polynomial equation that converted gray scale ratings based on the AATCC standard gray scale to visual differences (ΔV) is shown in Equation 1. The R^2 value for the polynomial fitting is 0.998.

$$\Delta V = -0.21\times G^3 + 2.684\times G^2 - 12.84\times G + 23.5 \quad (1)$$

The *STRESS* index, correlation coefficient (r) and Spearman rank coefficient (ρ) were used to evaluate the performance of color difference equations [1, 15, 16].

Results and Discussion

The contribution of each component to visual difference

The relative contribution of each component (ΔL^*, ΔC^* and ΔH^*) to overall color difference was plotted against the visual difference (ΔV) for subsets 1, 2 and 3, as shown in Figure 4. The percent of each average relative contribution (C_r) to one unit of visual difference (ΔV) was: 25.86 for ΔL^*, 43.97 for ΔC^*, and 27.06 for ΔH^*. Thus larger chroma variations are needed to results in the same magnitude of visual difference when compared to hue and lightness differences. Thus chroma had the least influence to the total visual difference of the samples produced; while lightness and hue differences contributed more in terms of the perceptual differences. In addition, a large variation existed in assessing lightness and chroma differences, in that the hue C_r for several samples was about 55%, while their visual difference, ΔV, ranged from 0.78 to 4.78.

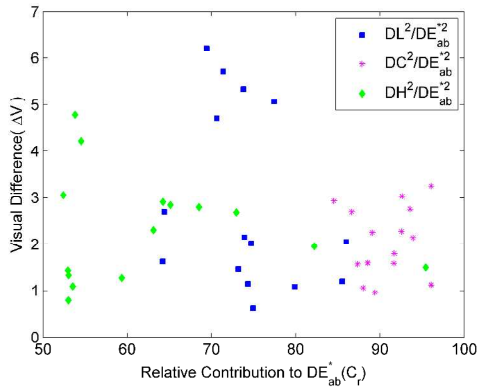

***Figure 4** Correlations between the relative contribution to DE^*_{ab} and Visual Difference (ΔV).*

Evaluation of the performance of color difference formulae

In this work, 7 color difference formulae with different parametric factors, resulting in 13 different combinations, were examined. The *STRESS* values, correlation coefficient (*r*) and Spearman Rank coefficient (*ρ*) for each formula are shown in Table II. The parameters used in the CAM02 formulas were selected according to the experimental conditions employed in NCSU experiments, i.e. L_A =89.1 cd/m^2; Y_b = 44.4; c = 0.69; N_c = 1.0; and F = 1.0. The scatter plots between visual difference ΔV and the computed color difference (ΔE) for various models are shown in Figure 5.

TABLE II Summary of the results of STRESS, correlation coefficient (r) and Spearman Rank coefficient (ρ) between ΔV and ΔE for various color difference formulae.

	STRESS	r	ρ
CIELAB	47.77	0.50	0.51
CIEDE94(1:1)	40.17	0.67	0.60
CIEDE94(2:1)	40.97	0.63	0.62
CIEDE2000(1:1:1)	42.13	0.69	0.55
CIEDE2000(2:1:1)	38.56	0.72	0.63
CMC(1:1)	39.93	0.67	0.61
CMC(2:1)	41.54	0.62	0.61
BFD(1:1)	40.22	0.66	0.62
BFD(2:1)	42.06	0.61	0.60
DIN 99	41.44	0.66	0.57
DIN 99d	39.77	0.67	0.61
CIECAM02UCS	40.18	0.68	0.60
CIECAM02SCD	40.01	0.68	0.60

From Table II, it can be seen that CIEDE2000 (2:1:1) gave the best performance based on all three metrics. The results according to *STRESS* values, correlation coefficient (*r*) and Spearman Rank coefficient (*ρ*) agree in most cases. These observations can also be confirmed by Figure 5.

A critical aspect of color difference modeling is to determine whether a color difference formula gives a better

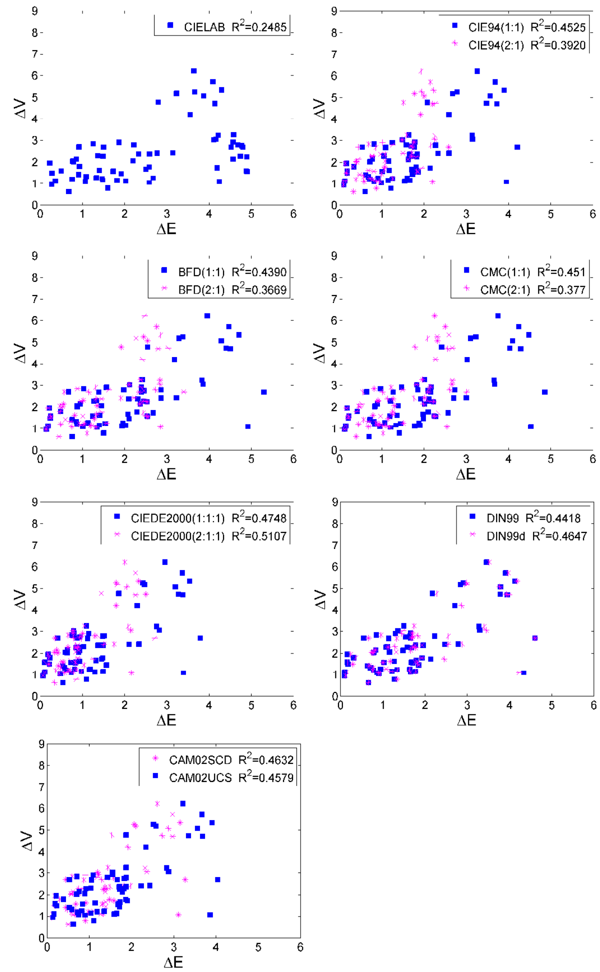

***Figure 5** Correlations between the color differences calculated for different equations.*

performance than another color difference equation. In this work, the *F*-test using the *STRESS* function was employed to calculate the significance of variation between two formulae. Firstly, the high chroma blue dataset was tested (the critical F_c = 0.610). Then, the combination of NCSU-B1 and NCSU-B2 datasets was examined (the F_c = 0.708). In comparisons if calculated F values are smaller than F_c, Model A is better than Model B; in addition when F values are between F_c and 1 the difference between Model A and Model B is insignificant. Results are shown in Tables IV and V respectively.

In Table IV, cells in orange indicate that the formula given in the column performs slightly better than the formula given in the row. In Table V, cells in blue indicate that the formula shown in the column is *significantly* better than the formula given in the row and cells in green indicate the opposite. The results shown in Tables IV indicate that CIEDE2000 (2:1:1) performed slightly better than other color difference models. However, no color difference equation was found to exhibit significant improved performance over another. Again, it seems that all color difference models produce average performance in the high chroma blue region and no equation generates sufficiently low *STRESS* or high correlation values.

For the combined low and high chroma blue samples, as shown in Table V, CAM02-UCS, CAM02-SCD, CIEDE2000, BFD, CMC(2:1) and CIEDE94(2:1) showed significantly better results than CIELAB. The performance of CIEDE2000, CAM02-UCS, CAM02-SCD and DIN99d were statistically significantly better than that of DIN99. DIN99d, CAM02-UCS and CAM02-SCD performed slightly better than CIEDE2000. An examination of the four major datasets used to derive the CIEDE2000 color difference model shows that while these datasets contain visual data for the low chroma blue region samples in the high chorma blue color center are not represented. This may explain the relatively poor performance of equations in this region.

In addition, no statistically significant difference in performance was found for CIEDE2000 at different k_L values, and similar observations were found for CIEDE94, CMC and BFD equations. While parametric factors are adjusted for different viewing conditions, e.g. for samples containing texture, an optimization of these factors for each formula does not seem to affect their performance significantly.

Evaluation of Reduced CIEDE2000 (1:1:1), CIEDE94 (1:1:1), and CMC (1:1) Models

STRESS and correlation coefficient (*r*) values were computed for CIEDE2000 (1:1:1), CIEDE94 (1:1:1), and CMC (1:1) as well as their reduced models, as shown in Table V. In each of the reduced models, one of the correction functions was removed from the equation. Since these function serve to improve the performance of the color difference formulas, removing these factors was expected to deteriorate the performance of the given model.

In the case of the CIEDE2000, the findings show that:

1- The chroma correction for the CIELAB is the most important correction in the blue region of the CIELAB color space based on both *STRESS* and correlation coefficient(*r*) and this is in agreement with the results reported previously[1,15], i.e. the *STRESS* value for high chroma blue was increased by as much as 7 units.

2- This is followed by the rotation and lightness adjustment, however, the models that include these two corrections only show a slight change in their performance when the functions are removed.

3- For the high chroma blue colors tested in this study, *STRESS* and correlation coefficient (*r*) metrics exhibit different results. STRESS indicates that the inclusion of the rotation term, which was specifically designed for blue colors, deteriorates the performance of the model for the high chroma samples. However, according to *(r)*, the model attains a worse fit when R_T is removed.

With regard to CIEDE94, *STRESS* and correlation coefficient (*r*) metrics indicate that:

1- The chroma adjustment function is the most significant function in improving the performance of the model.

2- CIEDE94 without the hue correction function performed slightly better than its full model for all three datasets.

CMC (*l:c*) also showed the same trends for chroma adjustment based on the two metrics and chroma correction function was found to be the most important.

Overall, *STRESS* and correlation coefficient (*r*) metrics showed inferior performance for the high chroma blue dataset compared to those for the NCSU-B1 (low chroma) as well as the four datasets used to develop CIEDE2000 [1,15]. The new dataset comprises a large set of samples populating a region of the space where only a few, if any, samples were present in the testing and development of previous models. It seems that color difference models need further refinement to produce sufficient performance in the high chroma blue region. Also, in some cases STRESS and correlation coefficient (*r*) gave inconsistent results for the performance of the CIEDE2000 for the high chroma blue samples. This suggests that a suitable metric is needed for the evaluation of the models.

TABLE III The STRESS values and correlation coefficient (r) for CIEDE2000 (1:1:1), CIEDE94, CMC (1:1) and their reduced models.

		NCSU-B1		NCSU-B2		Combined	
		STRESS	r	STRESS	r	STRESS	*r*
CIEDE2000 (1:1:1)	Full Model	21.22	0.91	42.13	0.71	33.71	0.76
	-Rot. Term	25.39	0.84	38.88	0.68	34.00	0.74
	-Lightness	22.63	0.89	42.18	0.67	36.42	0.75
	-Chroma	33.40	0.79	49.62	0.50	42.24	0.63
	-Hue	20.42	0.93	42.01	0.69	33.54	0.79
	-G	21.46	0.91	42.14	0.69	34.00	0.78
CIEDE94 (1:1)	Full Model	28.91	0.79	40.17	0.67	37.29	0.69
	-Chroma	40.94	0.65	50.94	0.45	46.78	0.53
	-Hue	25.39	0.84	39.20	0.68	35.44	0.72
CMC(1:1)	Full Model	27.41	0.81	39.93	0.67	36.55	0.70
	-Lightness	26.74	0.82	39.61	0.67	36.82	0.70
	-Chroma	37.23	0.70	47.85	0.50	43.66	0.58
	-Hue	25.78	0.83	39.25	0.68	36.02	0.71

Conclusions

For the high chroma blue dataset, the CIEDE2000 shows slightly better performance than other equations evaluated in this study based on *STRESS*, correlation coefficient (*r*) and Spearman Rank coefficient (ρ) metrics, but no equation showed a statistically significant improvement compared with others. For the combined low chroma and high chroma blue datasets, CAM02-UCS, CAM02-SCD, CIEDE2000 and DIN99d gave a better performance based on the *F*-test using the *STRESS* function. The optimization of k_L or l, k_C or c and k_H, did not change the performance of models significantly.

For the reduced CIEDE2000, CIEDE94 and CMC models, chroma correction was found to be the most important parameter. The least effective factor for the CIEDE2000 was the rotational term based on the *STRESS* function.

The metrics used indicate that an accurate prediction of color in the high chroma blue region remains a challenge. The weighting functions should be reexamined to determine whether a significant improvement in the performance of equations can be obtained.

Acknowledgments

The authors thank all color assessors who participated in the study.

References

[1]. Shamey R., Lee S.G., Hinks D, Jasper W "Performance of Recent Color Difference Equations around a CIE Blue Color Center" 4th European Conference on Colour in Graphics, Imaging, and Vision, 10th International Symposium on Multispectral Colour Science, 7-11 June (2008) Terrassa, Spain.

[2]. Lee S.G., "Assessment of Metrics in Color Spaces". Master's Thesis (North Carolina State University, 2007).

[3]. Luo M.R., Cui G., Li C., "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. 31, 320-330 (2006).

[4]. Luo M.R. "Development of colour-difference formulae." Rev. Prog. Coloration, 32, 28, (2002).

[5]. Luo M.R.; Rigg B. "BFD (l:c) colour difference formula, Part I-Development of the formula". J. Soc. Dyers Col. 103, 86, (1987).

[6]. Kuehni R.G., Color Space and Its Divisions: Color Order from Antiquity to the Present; John Wiley & Sons. Inc: New York, NY, (2003); pp 204-270.

[7]. Luo M.R., Cui G., Rigg B., "The development of the CIE colour-difference formula: CIEDE2000," Color Res. Appl., 26, 340-350 (2001).

[8]. CIE Publication 159:2004, 3rd Edition (CIE Central Bureau 2004).

[9]. Melgosa M., "Request for existing experimental datasets on color differences," Color Res. Appl. 32, 159 (2007).

[10]. Witt K., "Cie guidelines for coordinated future work on industrial colour-difference evaluation," Color Research & Application 20, 399-403 (1995).

[11]. Luo M.R., Cui G., Li C., "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. 31, 320-330 (2006).

[12]. AATCC, Evaluation Procedure 9-2002, "Visual assessment of color difference of textiles," in AATCC Technical Manual 82, (AATCC, 2007). pp. 394-396.

[13]. American Optical Company, Pseudo-Isochromatic Plates for Testing Color Perception, Beck Engraving Company, Inc.: New York, PA.

[14]. Neitz J., Manual: Neitz Test of Color Vision (Western Psychological Services, 2001).

[15]. Melgosa M., Huertas R., Berns R.S., "Performance of recent advanced color-difference formulas using the standardized residual sum of square," J. Opt. Soc. Am. A, 25, 1828-1834, 2008).

[16]. Kirchner E., Dekker N. "Performance measures of color-difference equations: correlation coefficient versus standardized residual sum of squares," J. Opt. Soc. Am. A, 25, 1841-1848, 2011).

Author Biography

Renzo Shamey is the director of Polymer and Color Chemistry Program at North Carolina State University. He obtained a Bachelor of Engineering in 1989 followed by MSc (1992) and PhD (1997) from the Colour Chemistry Department of Leeds University in England. After serving as a faculty member at Heriot-Watt University (Edinburgh) from 1998 to 2003 he joined North Carolina State University in 2003 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 IV Significance test of NCSU-B2 dataset (N=65, 1/Fc=1.639, Fc=0.610)

F_values	CIELAB	CIEDE94		CIEDE2000		CMC(l:c)		BFD(l:c)		DIN 99	DIN 99d	CAMUCS	CAMSCD
		1:1:1	1:1:1	1:1:1	2:1:1	1:1	2:1	1:1	2:1				
CIELAB	1.000	1.414	1.359	1.286	1.535	1.431	1.322	1.411	1.290	1.329	1.443	1.413	1.426
CIEDE94(1:1)	0.707	1.000	0.961	0.909	1.085	1.012	0.935	0.998	0.912	0.940	1.020	1.000	1.008
CIEDE94(2:1)	0.736	1.040	1.000	0.946	1.129	1.053	0.973	1.038	0.949	0.977	1.061	1.040	1.049
CIEDE2000(1:1:1)	0.778	1.100	1.057	1.000	1.194	1.113	1.029	1.097	1.003	1.034	1.122	1.099	1.109
CIEDE2000(2:1:1)	0.652	0.921	0.886	0.838	1.000	0.933	0.862	0.919	0.840	0.866	0.940	0.921	0.929
CMC(1:1)	0.699	0.988	0.950	0.898	1.072	1.000	0.924	0.986	0.901	0.928	1.008	0.988	0.996
CMC(2:1)	0.756	1.069	1.028	0.972	1.161	1.082	1.000	1.067	0.975	1.005	1.091	1.069	1.078
BFD(1:1)	0.709	1.002	0.964	0.911	1.088	1.015	0.937	1.000	0.914	0.942	1.023	1.002	1.011
BFD(2:1)	0.775	1.096	1.054	0.997	1.190	1.110	1.025	1.094	1.000	1.030	1.118	1.096	1.105
DIN 99	0.753	1.064	1.023	0.968	1.155	1.077	0.995	1.062	0.971	1.000	1.086	1.064	1.073
DIN 99d	0.693	0.980	0.942	0.891	1.064	0.992	0.917	0.978	0.894	0.921	1.000	0.980	0.988
CAMUCS	0.707	1.000	0.962	0.910	1.086	1.013	0.936	0.998	0.913	0.940	1.021	1.000	1.009
CAMSCD	0.701	0.992	0.954	0.902	1.077	1.004	0.928	0.990	0.905	0.932	1.012	0.992	1.000

TABLE V Significance test for the combined NCSU-B1 and NCSU-B2 datasets (N=131, 1/Fc=1.412, Fc=0.708)

F_values	CIELAB	CIEDE94		CIEDE2000		CMC(l:c)		BFD(l:c)		DIN 99	DIN 99d	CAMUCS	CAMSCD
		1:1:1	1:1:1	1:1:1	2:1:1	1:1	2:1	1:1	2:1				
CIELAB	1.000	1.313	1.463	1.600	1.712	1.367	1.439	1.634	1.544	1.202	1.805	1.767	1.826
CIEDE94(1:1)	0.762	1.000	1.114	1.218	1.304	1.041	1.096	1.245	1.176	0.915	1.375	1.346	1.390
CIEDE94(2:1)	0.684	0.898	1.000	1.093	1.170	0.935	0.984	1.117	1.055	0.822	1.234	1.208	1.248
CIEDE2000(1:1:1)	0.625	0.821	0.915	1.000	1.070	0.855	0.900	1.022	0.965	0.751	1.128	1.105	1.141
CIEDE2000(2:1:1)	0.584	0.767	0.855	0.934	1.000	0.799	0.840	0.955	0.902	0.702	1.054	1.032	1.067
CMC(1:1)	0.731	0.960	1.070	1.170	1.252	1.000	1.052	1.195	1.129	0.879	1.320	1.293	1.335
CMC(2:1)	0.695	0.913	1.017	1.112	1.190	0.950	1.000	1.136	1.073	0.835	1.254	1.228	1.269
BFD(1:1)	0.612	0.803	0.895	0.979	1.047	0.836	0.880	1.000	0.945	0.735	1.104	1.081	1.117
BFD(2:1)	0.648	0.851	0.948	1.036	1.109	0.886	0.932	1.059	1.000	0.779	1.169	1.145	1.183
DIN 99	0.832	1.093	1.217	1.331	1.424	1.138	1.197	1.360	1.284	1.000	1.502	1.470	1.519
DIN 99d	0.566	0.743	0.828	0.905	0.969	0.774	0.814	0.925	0.874	0.680	1.000	1.000	1.033
CAMUCS	0.566	0.743	0.828	0.905	0.969	0.774	0.814	0.925	0.874	0.680	1.021	1.000	1.033
CAMSCD	0.548	0.719	0.801	0.876	0.938	0.749	0.788	0.895	0.846	0.658	0.989	0.968	1.000