Research Article | Vol. 29, No. 5 / 1 March 2021 / Optics Express 7778

Optics EXPRESS

Testing uniform colour spaces using colour differences of a wide colour gamut

QIANG XU,[1] BAIYUE ZHAO,[1] GUIHUA CUI,[2] AND MING RONNIER LUO[1,*]

[1] *State Key Laboratory of Modern Optical Instrument, Zhejiang University, Hangzhou, China*
[2] *School of Physics & Electronics Information Engineering, Wenzhou University, Wenzhou, China*
[*] *m.r.luo@zju.edu.cn*

Abstract: An experimental dataset, *WCG*, was assembled. The set includes 416 pairs of samples that surround 28 colour centres and covers a wide colour gamut. The data were used to test the performance of seven colour-difference models, *CIELAB*, *CIEDE2000*, *CAM16-UCS*, *DIN99d,* OSA_{GP}, and IC_TC_P, $J_za_zb_z$. Colour discrimination ellipses were also fitted to compare the uniformity of the colour spaces. Different versions of the models were derived to improve the fit to the data, including parametric factors, k_L, k_C, and a power factor. It was found that the k_L optimised *CAM16-UCS, DIN99d,* OSA_{GP} models significantly outperformed the other colour models. In addition, the magnitude of the colour difference had an impact on visual assessment.

1. Introduction

With the increasing popularity of high dynamic range (HDR) and wide colour gamut (WCG) TVs and displays, the traditional technology is facing challenge, i.e., the 8-bits/channel display controllers following the sRGB specification with a 6500 *K* white point at luminance of about 30 cd/m^2. The new configuration uses 10 to 12-bits per channel, e.g., quantum dot or OLED technology to provide much wider gamuts [1], and to increase the luminance to the range of 300 to 1000 cd/m^2. One of the challenges is that the existing colour models including colour-difference formulae and associated uniform colour spaces can still give satisfactory performance on colour control. Note those models were derived to fit experimental datasets using surface colours or sRGB type displays. They covered a relatively small colour gamut and low dynamic range. This may give imprecise estimation for HDR and WCG displays. Although new spaces, such as IC_TC_P [1] and $J_za_zb_z$ [2] were developed to be used for these applications, there is a lack of robust experimental data to verify their performance.

Let's give a brief overview of the development of colour difference models. Our goal is to achieve a uniform colour space (UCS) which provides a single tolerance for the evaluation of colour differences across entire colour regions. The real progress has been made in 1976 since the recommendation of *CIELAB* and *CIELUV* colour spaces by the CIE, (International Commission on Illumination) [3]. Between 1976-2001, much efforts have been made to improve the accuracy of *CIELAB* space in predicting medium size colour difference magnitudes, i.e., less than 5 *CIELAB* units [4–6]. Robust datasets were produced in this period including BFD [5,7], RIT-DuPont [8], Leeds [6] and Witt [9]. Later, they were merged to form a combined visual dataset (*COMBVD*). Various colour difference formulae [6–8] were developed from these datasets. In 2001, one of these, *CIEDE2000* colour-difference formula [4], was recommended by the CIE for industrial applications. Later, it was published as the ISO/CIE standard colour-difference formula [10]. The formula has outperformed the previously recommended colour-difference formulae [4] using the *COMBVD*. Note the above colour-difference formulae were modifications of *CIELAB* and they do not have their own associated UCSs. In the same period, the *COMBVD* was used to develop new UCSs including *DIN99d* [11], OSA_{GP} [12] and *CAM02-UCS* [13]. The latter was based on the *CIECAM02* colour appearance model [14]. Further experiments

#413985 https://doi.org/10.1364/OE.413985
Journal © 2021 Received 3 Nov 2020; revised 12 Feb 2021; accepted 15 Feb 2021; published 26 Feb 2021

showed that *CAM02-UCS* gave the best performance when tested using new datasets [15,16]. It gave very similar performance to that of *CIEDE2000* in predicting the *COMBVD*. *CAM02-UCS* first predicts the colour appearance correlates under different viewing conditions (including illuminant, luminance level, neutral background luminance factor, and surround condition), and then uses these correlates to estimate the visual colour differences. It has been well received in different applications such as the CIE colour fidelity index, R_f [17], the IES-TM30 [18,19] colour gamut index, R_g, and chroma shift index, $R_{cs,hj}$, and the Gao *et al.* colour appearance model for unrelated colours [20]. The latest colour appearance model, *CAM16*, and its associated UCS, *CAM16-UCS* [21], overcomes some problems with the chromatic adaptation and cone response transforms in *CIECAM02*. The performance of the new model is almost identical to that of its predecessor and its structure is also simpler. Both *CAM16* and *CAM16-UCS* are expected to become new CIE recommendations in 2021. Note that the *COMBVD* has had varying names in different publications, including COM-corrected dataset [16], and COM [13].

This paper describes two experiments carried out using a single wide-colour-gamut display. Experiment 1 covered the most saturated colour regions close to the border of the colour gamut [22]. It included a total of 192 pairs of stimuli that surrounded 12 colour centres. The colour difference of each pair was assessed by a panel of 18 observers. Experiment 2 included 224 pairs surrounding 16 colour centres, with assessments by 20 observers. These colours were selected to fill in the gap between the most saturated colour regions used in Experiment 1 and the less saturated colours found in the *COMBVD*. A new dataset, *WCG*, was formed by combining the results from Experiments 1 and 2. Thus the combined *COMBVD* and *WCG* dataset cover both less and highly saturated colour regions, respectively. Note that the results from Experiment 1 have been previously reported [22]. The data used however, were based on the target colorimetric values; the present work uses the actual measured data.

The goals of the present study were *(i)* to produce a dataset to cover a wide colour gamut, and *(ii)* to test the performance of colour models using the new *WCG* dataset.

2. Experimental

2.1. Display

Both experiments were conducted in a darkened room on an NEC PA302W display, with a size of 30 inches and a resolution of 2560 × 1600 pixels. The display peak white had a correlated colour temperature of 6500 *K* and a luminance of 310 cd/m^2. Its colour performance was detailed examined. It had a one ΔE_{00} on spatial uniformity of the display. It was evaluated by dividing the display into 3 by 3 segments and the mean colour difference calculated between the centre. The Gain-Offset-Gamma (GOG) model [23] was used to characterize the display and had an average predictive accuracy of 0.42 ΔE_{00} units over the 416 samples used in the present experiment, with a standard deviation of 0.21 ΔE_{00} units. In addition, each colour was measured before, during and after the experimental period and the MCDM (mean colour difference from the mean) was 0.39 ΔE_{00} units, with a standard deviation of 0.12 ΔE_{00} units. All the results were measured using a Konica Minolta CS2000A tele-spectroradiometer and colorimetric values were calculated using the CIE 1964 standard colorimetric observer, or 10° observer, with the chromaticity of the peak white of display. The above performances are indicative of the high quality of the display, making it suitable to perform vision experiments.

2.2. Stimuli

Figure 1(a) shows the colour centres from the *COMBVD* (black dots) plotted in the *CIELAB a*b** plane, together the colour centres from the present Experiment 1 (blue dots) and Experiment 2 (red dots). Figure 1(b) shows the colour centres for Experiments 1 and 2 plotted in CIE 1964 chromaticity diagram. In addition, the sRGB, DCI-P3 colour gamuts, and that of the actual

Optics EXPRESS

display used in the experiment, are plotted in this figure, showing that the selected colours covered most of the saturated colour region. Table 1 lists the *CIELAB* specification of each colour centre, calculated using the chromaticity of the display peak white, [x_{10}=0.3152, y_{10}=0.3279], together with their associated colour names. Repeated colour centres in Experiment 2, the symbol of '-0', was added to their names, e.g., CIE-Grey-0. Note that the 5 colour centres i.e., grey, red, blue, magenta and cyan-green (see Table 1), were investigated in both experiments. The first three are included in the 5 colour centres recommended by CIE [24] for further colour-difference research. Results using these centres were widely reported in the literature, for example by Witt [9], RIT-DuPont [8], Cheung and Rigg [25], Cui and Luo [26], and Mirjalili *et al.* [27]. The first three datasets are included in *COMBVD*. The results from the 3 centres were used to adjust the different datasets to have the same visual scale in the *COMBVD*. Note that in the late stage, there were more guidelines proposed for coordinated work on colour difference studies [28,29].

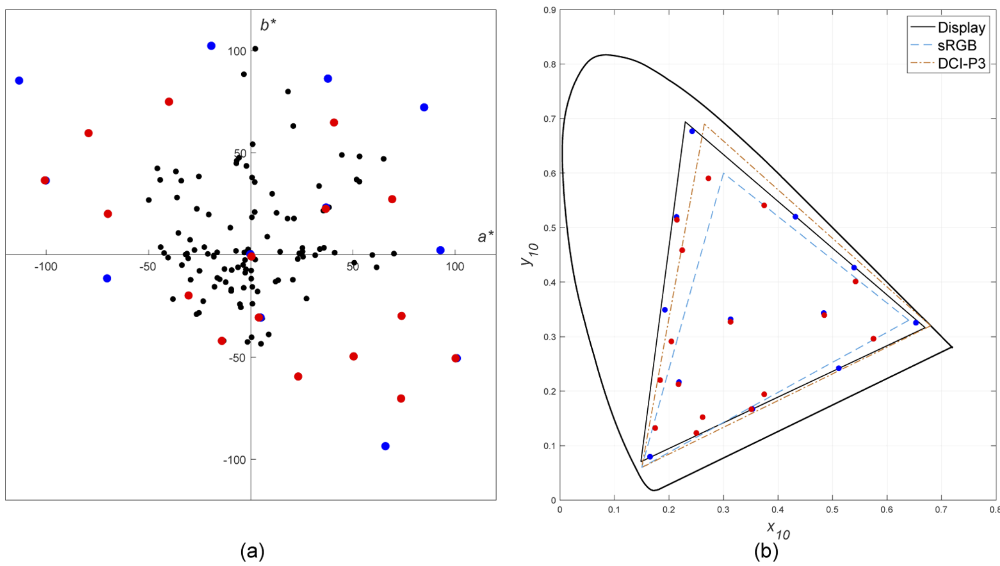


Fig. 1. (a) The colour centres from the *COMBVD* (black dots), the Experiment 1 set (blue dots) and the Experiment 2 set (red dots). (b) The distribution of the same colour centres in the CIE 1964 chromaticity diagram. The triangles represent the sRGB, DCI-P3 and display gamut primaries.

The distributions of samples around each colour centre in $\Delta a^*\Delta b^*$ plane is shown in Figs. 2(a) and 2(b) for Experiments 1 and 2, respectively. In Experiment 1, Fig. 2(a), the pairs had two levels of colour-difference magnitude of, 3 or 6 *CIELAB* units. It can be seen that the two levels of colour difference cover five directions from 0° to180° at an interval of 45° in the $\Delta a^*\Delta b^*$ plane. The other six pairs covered the ΔL^* axis, and 45° in the $\Delta a^*\Delta L^*$ plane and 45° in the $\Delta b^*\Delta L^*$ plane, again with each of two magnitudes. In total, 192 pairs of samples (12 centres × 16 pairs) were prepared. Experiment 1 was also designed to verify the earlier results of Mirjalili *et al.* [27] showing that colour-difference magnitudes have an impact on perceived colour difference. In Experiment 2, Fig. 2(b), all the pairs had a difference of 3 *CIELAB* units from 0° to180° at an interval of 18°. The other 3 pairs included the ΔL^* axis, 45° in the $\Delta a^*\Delta L^*$ plane and 45° in $\Delta b^*\Delta L^*$ plane. In total, 224 pairs of samples (16 centres × 14 pairs) were prepared in Experiment 2. For both experiments, sample pairs at the grey centre were repeatedly assessed to evaluate intra-observer variation. The above samples were chosen for visual assessments with a goal to produce reliable colour discrimination ellipses. From earlier studies, a number of rules had been learned: 1) to have a good coverage of samples against colour centre [25], 2) to apply a symmetry

Optics EXPRESS

Table 1. The *CIELAB* colour specification of each colour centre calculated using the chromaticity of the display peak white and the 1964 standard colorimetric observer. [a]

		Colour Centre	L^*	a^*	b^*	C^*_{ab}	h_{ab} (°)
Exp 1	1	Max. Red	53.1	84.8	72.2	111.4	40.4
	2	Orange	70.1	37.7	86.3	94.1	66.4
	3	Yellow	90.1	-19.3	102.3	104.1	100.7
	4	Max. Green	79.1	-113.3	85.3	141.8	143.0
	5	Cyan-green	80.1	-100.3	36.4	106.7	160.1
	6	Cyan	83.1	-70.3	-11.5	71.3	189.3
	7	Max. Blue	35.1	65.9	-93.6	114.4	305.1
	8	Magenta	60.1	100.8	-50.5	112.7	333.4
	9	Pink	54.1	92.8	2.3	92.8	1.4
	10	CIE-Red	44.1	36.8	23.3	43.5	32.3
	11	CIE-Blue	36.1	4.8	-30.7	31.1	279.0
	12	CIE-Gray	62.1	-0.2	0.4	0.4	122.3
Exp 2	1	CIE-Grey-0	61.9	0.3	-0.8	0.8	290.9
	2	CIE-Red-0	44.4	36.7	22.5	43.0	31.6
	3	Red-1	45.4	69.2	27.3	74.4	21.5
	4	Orange-1	60.2	40.7	64.8	76.5	57.8
	5	Green-1	75.0	-40.0	74.9	85.0	118.1
	6	Green-2	70.0	-79.4	59.6	99.2	143.1
	7	Green-3	65.1	-69.9	20.1	72.7	164.0
	8	Cyan-1	55.0	-30.5	-19.9	36.4	213.1
	9	Cyan-2	53.5	-14.2	-42.1	44.4	251.4
	10	CIE-Blue-0	36.3	3.9	-30.6	30.9	277.3
	11	Blue-1	35.3	23.2	-59.5	63.9	291.3
	12	Blue-2	40.4	50.3	-49.7	70.7	315.4
	13	Blue-3	50.4	73.8	-29.8	79.6	338.0
	14	Blue-4	45.2	73.5	-70.3	101.7	316.3
	15	Magenta-0	62.5	100.2	-50.6	112.2	333.2
	16	Cyan-green-0	79.9	-100.8	36.5	107.2	160.1

[a] Note that those named with '-0' at the end mean the colour centres studied in Experiments 1 and 2.

rule [30], e.g., in Figs. 2(a) and 2(b), the Δa^* values of each sample are unchanged but $+\Delta b^*$ values are changed to $-\Delta b^*$ for the colour centres having samples outside of the colour gamut (the pattern of distribution will thus be inverted), 3) to choose less samples along the ΔL^* direction and in the $\Delta a^*\Delta L^*$ or $\Delta b^*\Delta L^*$ planes because little tilting of the discrimination ellipsoids was found in earlier studies [25,30].

2.3. Visual assessment

Figure 3 shows the experiment interface on the display. The sample pair were displayed in the centre of the screen with no separation, and the background was set to a mid-grey (L^* equal to 43.6). The colour difference of a red test pair was assessed against the grey scale pairs shown at the top of the display. The grey-scale method has been widely used for assessing colour-differences. For the industrial applications, colour fastness is an important quality control property for all surface products. Although the method was first adopted as an ISO standard in

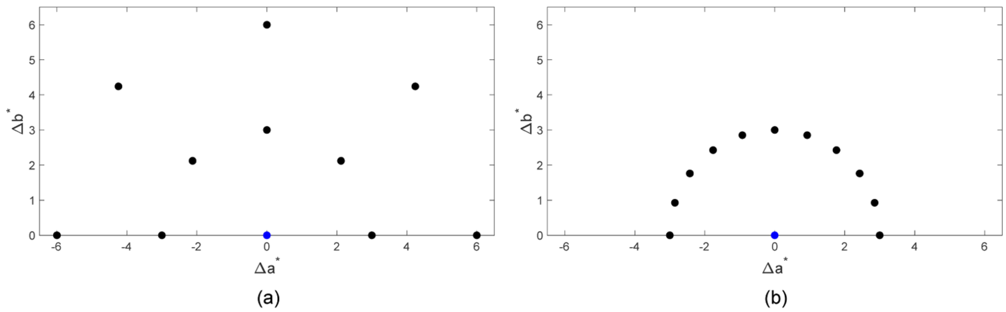


Fig. 2. Distribution of samples surrounding a centre in *CIELAB* $\Delta a^* \Delta b^*$ plane for (a) Experiment 1 and (b) Experiment 2.

the textile industry for assessing different types of fastness including colour change, staining and light fastness [31], it has been extended to the other industries including coatings, plastics and printing. This results in the publication of ASTM standards, including D1729 [32], and E3040 [33]. They give detail procedure to visually assess the total colour difference, including lightness, chroma and hue differences, of a pair of samples via the grey-scale pairs. For the academic research, the method was first utilised in 1986 [34] and it was found to be robust in comparison with other psychophysical methods and the different methods produce similar results [25,30]. It has been extensively used in various studies and their results showed a good observer consistency [15,26,27,34,35].

Fig. 3. The display showing the 5-step grey scale (top), and the colour difference pair (centre) and the slider for the observer to record the colour difference (bottom).

The grey scale consisted of five grey-scale samples [31]. Table 2 shows the ISO standard *CIELAB* lightness values, L^*_{STD}, of the individual samples *GS-1* to *GS-5* (having a^*_{STD} and b^*_{STD} are equal to zero), and the reproduced samples used in the experiment (see L^*_M, a^*_M, b^*_M, values in Table 2). It can be seen that the measured lightness values, L^*_M, agreed well with the standard values, L^*_{STD}, i.e., with lightness differences of less than 0.60, and a^* and b^* differences less than 0.5.

The 5 grey scale pairs were constructed between the standard (*GS-1*) and each of *GS-1* to *GS -5* samples. Figure 3 shows the 5 pairs on the top of the screen. Each pair should only exhibit a lightness difference, having target values, $\Delta E^*_{ab,S}$ of 0, 1.5, 3.0, 6.0 and 12.0 units respectively

Optics EXPRESS

Table 2. Colour specifications of the grey scale samples and the colour-difference pairs.

Grey scale number (*GS*)	L^*_{STD}	L^*_M	a^*_M	b^*_M	Grey scale pairs	$\Delta E^*_{ab,S}$	$\Delta E^*_{ab,E}$	$\Delta E^*_{ab,P}$
GS-1	40.0	40.6	-0.3	0.1	1 (1-1)	0.0	0.0	0.2
GS-2	41.5	41.9	-0.2	-0.5	2 (1-2)	1.5	1.4	1.2
GS-3	43.0	43.6	-0.4	0.2	3 (1-3)	3.0	3.0	3.0
GS-4	46.0	46.6	-0.2	-0.4	4 (1-4)	6.0	6.0	6.2
GS-5	52.0	52.3	-0.2	-0.4	5 (1-5)	12.0	11.7	11.7

(Table 2, column 7). The actual differences achieved, $\Delta E^*_{ab,E}$, are given in Table 2, column 8 and these values are considered acceptably close to the standard values.

Equation (1) was used to scale the visual judgements in terms of *GS* grade values, given by the observers, to visual colour difference values, ΔV. The coefficients in Eq. (1) were obtained by minimising the difference between the values of $\Delta E^*_{ab,E}$ and the *GS* scale values to give predicted colour differences, $\Delta E^*_{ab,P}$. These predicted values of colour difference also agreed with values of $\Delta E^*_{ab,E}$ within differences of 0.2, Table 2, column 9.

$$\Delta V = 0.7999e^{0.5567GS} - 1.2359 \tag{1}$$

The experiment was conducted in a dark room. Observers were seated approximately 60 *cm* in front the display for which the test pair subtended about 8°. The height of the chair was adjusted to maintain the viewing/illumination geometry of 0° : 0°. Observers were required to adapt to the viewing conditions for one minute prior to each observing session. Subsequently, observers viewed the sample pairs in a different random order. The observer was asked to scale the colour difference using a mouse to control the slider below the test pair on the display. This meant that the scale was not limited to the integer values 1 to 5. After scaling the test pair, the observer clicked ‘Next’ to move to the next sample pair, or ‘Previous’ to go back to a previous sample pair.

In Experiments 1 and 2, 18 and 20 observers took part respectively, with ages in the range 22 to 25 years (mean 23 years, standard deviation 0.74), half were male and half female. All the observers were the students at Zhejiang University, and all had normal colour vision according to the Ishihara colour vision test. Sample pairs in the grey colour centre were repeated to evaluate of intra-observer variation.

3. Results and discussion

3.1. Observer variation

The standard residual sum of squares (*STRESS*) metric [16,36] was used to evaluate the intra-observer and inter-observer variation, Eq. (2). The percent *STRESS* has values between 0 and 100 and for perfect agreement between two sets of data, will be equal to zero. A higher *STRESS* represents a poorer agreement between two sets of data.

$$STRESS = 100\sqrt{\frac{\sum (F\Delta E_i - \Delta V_i)^2}{\sum \Delta V_i^2}}, \tag{2}$$

with $F = \frac{\sum \Delta E_i \Delta V_i}{\sum \Delta E_i^2}$, where F is a scaling factor to adjust ΔV and ΔE to be on the same scale.

For the intra-observer variation, the average *STRESS* values for the 18 and 20 observers, based on assessment of pairs of the CIE-Grey centre, were 24 and 21 for Experiments 1 and 2 respectively. The inter-observer variation for Experiments 1 and 2 were 42 and 41 *STRESS* units, respectively. These values indicate that the results of the observers for these two experiments were consistent. These results were somewhat larger than those in the other studies, e.g., about

Optics EXPRESS

35 units [15] using lower chroma surface colours than the present study. Figures 4(a) and 4(b) show the *STRESS* values for the inter-observer variations for each centre in Experiments 1 and 2 respectively, together with the MEAN and TOTAL inter-observer variations, plus the intra-observer variation. Note that the MEAN was the mean of the *STRESS* values of individual centres, and the TOTAL was calculated for all colour centres. It can be seen that the variations were very similar in both experiments and, as expected, the TOTAL is larger than the MEAN. Note that as the pairs in the CIE-grey centre were assessed repeatedly to represent intra-observer variation, its *STRESS* value gave very similar performance to the MEAM in both experiments. So, the intra-observer variation based on the pairs from the grey centre is a good representation of all the data, i.e., it does not cause the great difference between inter- and intra- observer variations.

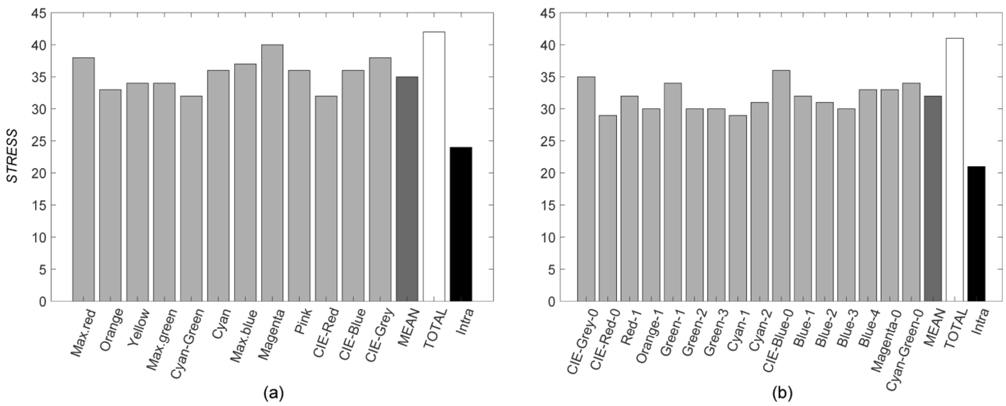


Fig. 4. The inter-observer variations, in *STRESS* units, for each colour centre, the mean and the total, together with intra-observer variations for Experiment 1 (a) and Experiment 2 (b).

3.2. Performance of the colour models

3.2.1. Using the STRESS measure

The present two datasets were used to test the performance of several colour models, including the *CIEDE2000* colour-difference formula, and the *CIELAB, CAM16-UCS,* IC_TC_P, $J_za_zb_z$, *DIN99d* and OSA_{GP} UCSs. Note that the parameters for *CAM16-UCS* (adapting luminance (L_a), luminance factor of the neutral background (Y_b) and the surround) were set at 42.06, 13.56 and 'dim', respectively. Generally, the Euclidean distance of a pair of samples in an UCS represents the colour difference. As mentioned earlier, *CIEDE2000* colour difference formula does not have an associated colour space. *CIELAB* was chosen because it is the most widely used CIE recommended UCS. The *CIEDE2000*, *CAM16-UCS* and $J_za_zb_z$ models were developed by fitting the *COMBVD*. The IC_TC_P and $J_za_zb_z$ colour spaces were specifically designed for HDR and WCG applications. Note that IC_TC_P has two versions to calculate colour difference [1,37], and the latter version given in Eq. (3) is investigated in the present study.

$$\Delta IC_TC_P = 720\sqrt{(\Delta I)^2 + 0.25(\Delta C_T)^2 + (\Delta C_P)^2}. \tag{3}$$

Two methods were used to evaluate colour models' performance, by the *STRESS* values between the visual differences (ΔV) and predicted differences (ΔE), and by comparing the global and local uniformities based on chromatic discrimination ellipses.

The performance of the models was first assessed in terms of the *STRESS* values for all pairs in a dataset. Table 3 summarises the results of each colour model for each colour centre, together with their MEAN and TOTAL sets. The MEAN result was calculated by averaging the *STRESS* values from all colour centres and the TOTAL result was calculated by combining the results for all colour centres. Finally, Experiments1 and 2 datasets were combined and the test TOTAL and MEAN from all models were also given in Table 3.

Table 3. The performance of the seven colour-difference models in terms of *STRESS*.[a]

	Colour Centre	*CIELAB*	*CIEDE2000*	*CAM16-UCS*	ΔIC_TC_P	$J_za_zb_z$	*DIN99d*	OSA_{GP}
Exp 1	Max. Red	67	38	38	43	43	35	48
	Orange	70	53	53	48	48	44	62
	Yellow	63	49	44	40	40	37	51
	Max. Green	73	33	32	40	48	35	56
	Cyan-green	67	37	39	40	51	37	58
	Cyan	62	48	43	37	49	35	60
	Max. Blue	46	26	30	28	39	32	35
	Magenta	64	26	40	51	56	34	50
	Pink	67	32	39	46	56	36	51
	CIE-Red	58	38	43	42	45	36	51
	CIE-Blue	44	31	32	20	33	35	33
	CIE-Gray	37	45	40	21	27	44	47
	MEAN	60	**38**	**39**	**38**	45	**37**	50
	TOTAL	62	42	**39**	46	47	**39**	50
Exp 2	CIE-Grey -0	59	65	63	41	51	65	67
	CIE-Red -0	68	53	56	57	60	52	62
	Red-1	73	47	52	64	64	47	61
	Orange -1	77	60	64	64	61	53	69
	Green -1	73	49	50	54	56	46	53
	Green -2	79	49	51	59	66	54	64
	Green -3	73	46	49	47	62	49	63
	Cyan -1	68	52	54	48	62	53	61
	Cyan -2	69	50	54	49	62	56	59
	CIE-Blue -0	61	52	52	40	54	53	54
	Blue -1	62	53	52	46	57	55	55
	Blue -2	63	36	49	53	58	44	54
	Blue -3	75	38	56	68	70	51	60
	Blue -4	61	20	42	51	52	34	47
	Magenta -0	73	35	52	65	66	45	61
	Cyan-green -0	78	48	51	53	66	48	68
	MEAN	69	**47**	53	54	60	50	60
	TOTAL	70	**57**	**57**	**56**	60	**58**	64
Combined	MEAN	65	**43**	47	47	54	**45**	56
	TOTAL	65	**47**	**46**	49	51	**45**	55

[a]Note: For the MEAN and TOTAL results, the *STRESS* values for the best performed models were in bold and underlined.

Table 3 lists the performance, in terms of the *STRESS* value, of the colour difference models as applied to the data from Experiments 1 and 2. The MEAN results for Experiment 1 data, showed that *DIN99d* performed the best, followed by *CIEDE2000*, ΔIC_TC_P and *CAM16-UCS*, then $J_za_zb_z$, OSA_{GP}, and *CIELAB* the worst. The order of merit was changed for Experiment 2 MEAN

Optics EXPRESS

results, *CIEDE2000* performed the best, followed by *DIN99d*, *CAM16-UCS*, ΔIC_TC_P, $J_za_zb_z$, OSA_{GP}, and *CIELAB* the worst. Comparing the TOTAL results for both experiments (Combined dataset), *CAM16-UCS* and *DIN99d* outperformed the others, followed by *CIEDE2000*, ΔIC_TC_P, $J_za_zb_z$, OSA_{GP}, with *CIELAB* the worst. The TOTAL results can be considered to include all sample pairs from the different centres and thus represent the overall performance of each model. It is encouraging that *CIEDE2000*, *CAM16-UCS* and *DIN99d*, which were derived to fit the *COMBVD*, performed the best for both MEAN and TOTAL results. The order of merits for each model is quite similar between the *COMBVD* as reported in [16] and the present *WCG* dataset in Table 2. However, the predictive accuracy of 47 *STRESS* units for the TOTAL results, comparing with that of 28 *STRESS* units for the *COMBVD* in [16]. This seems to imply that the colour differences behave quite differently between the low and high chroma regions respectively.

Comparing the results from Experiments 1 and 2, all models performed better for Experiment 1 results than Experient 2 results. This could be due to the sampling of the colour centres, the sample distribution of each centre, the colour difference magnitudes, or the number of lightness and chromatic differences in the dataset in question. This will be clarified in Section 3.2.4.

From the above analysis, it can be concluded that the results of the two experiments agreed well in terms of the ranks of the models as given by the MEAN and TOTAL results in Table 3. For this reason, it was decided to merge the results from Experiments 1 and 2 to form a combined *WCG* dataset that included 416 pairs of colours.

3.2.2. Plotting chromatic discrimination ellipses

Chromatic discrimination ellipses for the 23 colour centres were fitted in *CIELAB*, *CAM16-UCS*, IC_TC_P, $J_za_zb_z$, *DIN99d*, and OSA_{GP}, colour space respectively. Note that the 5 repeated colour centres were merged in the two experiments to fit individual ellipses. A colour-difference ellipse is given by Eq. (4):

$$\Delta E^2 = g_{11}\Delta a^{*2} + g_{12}\Delta a^*\Delta b^* + g_{22}\Delta b^{*2} \tag{4}$$

where coefficients g_{11} to g_{22} are optimised to give the lowest *STRESS* between the calculated colour difference using the ellipse equation and the visual data. Setting ΔL^* to zero allows the $\Delta a^*\Delta b^*$ plane chromatic discrimination ellipse to be calculated.

When fitting the ellipses from the *COMBVD*, they are called *COMBVD* ellipses hereafter. Each ellipse equation in *CIELAB* under CIE D65 and 10-degree observer was used to extract samples from an ellipse at 10° interval from 0 to 180°. These values were then transformed to *XYZ* tristimulus values and converted to the designated space, e.g., *CAM16-UCS*. Finally, an ellipse was fitted from the data.

When plotting the ellipses for the *COMBVD* and *WCG* dataset in one colour space, it is desirable to ensure the ellipses have the same visual scale. A scaling factor was calculated between the sizes of ellipses of the *COMBVD* and *WCG* dataset using the 3 CIE centres as noted earlier. It was found necessary to reduce the size of the 23 *WCG* ellipses by a factor of 2 to be correctly scaled to those of the 126 *COMBVD* ellipses.

Figures 5(a)–5(f) show all the ellipses plotted in the chromatic plane of *CIELAB*, *CAM16-UCS*, IC_TC_P, $J_za_zb_z$, *DIN99d* and OSA_{GP}, respectively. For a perfectly UCS, all ellipses should be circles with equal radius. It can be seen that, for *CIELAB*, Fig. 4(a), the *COMBVD* ellipses plotted in *CIELAB* were smallest close to the neutral point, and they progressively increased in size with increasing chroma. Also, most ellipses are orientated towards the origin, the exception being those in the blue-purple region, which are relatively long and thin. This has been noted in the literature [4,6–10] and is caused by the poor uniformity in the blue-purple region, e.g., especially shown by hue linearity data for hue angles in the range from 260° to 300° [38]. The new *WCG* ellipses, followed a similar trend to the *COMBVD* ellipses, except that the ellipses for the high chroma purple and magenta regions are very small. Also, the new ellipses in the orange region do not orientate along the chroma axis and rotate towards 90°. The IC_TC_P space

Optics EXPRESS

showed a similar pattern to that of *CIELAB*, i.e., the size of the ellipses increases with increasing chroma. However, unlike the other spaces, the number of colour centres in each quadrant was more varied, i.e., many centres are crowded in the second and third quadrants, or the negative C_P region, than in the other regions.

CAM16-UCS gave the best performance, i.e., the shapes of individual ellipses are close to being equal-sized circles. Also, the number of colour centres in each quadrant is evenly distributed. The ellipse pattern in $J_z a_z b_z$ space is like that of *CAM16-UCS* (i.e., all ellipses are close to circles), but the trend of an increase in the size of the ellipse as chroma increases still can be discerned.

A quantitative method, developed by Huang *et al.* [15], was used to compare the performance of the models using the ellipse parameters, the semi-major axis (A), the semi-minor axis (B) and orientation angle (θ). The performance can be divided into two: local and global uniformity. The local uniformity describes the shape of ellipses in terms of their closeness to a circle, i.e., to have A/B equal to unity. The second measure concerns global uniformity where all ellipses should be of equal size, i.e., the value of the area (πAB) should be constant.

Table 4 shows the performance of the spaces in terms of the local and global uniformity. The local uniformity is measured by calculating root mean square error (*RMSE*) multiply by 100%, Eq. (5), between the ratios of semi axes (A/B) and that of circle ($A/B = 1$). The global uniformity is measured by calculating the coefficient of variation (*CV*), Eq. (6), between the size (S) of each ellipse and the average of all ellipses ($\bar{S}$). The results clearly show that for local uniformity, OSA_{GP} performed the best, followed by $J_z a_z b_z$, and then *CAM16-UCS*, *CIELAB*, *DIN99d* and $IC_T C_P$ the worst. For global uniformity, OSA_{GP} outperformed other UCSs, followed by *DIN99d*, *CAM16-UCS* and then $J_z a_z b_z$, $IC_T C_P$, and *CIELAB* the worst. Overall, OSA_{GP} performed the best in this test. Note that the present results based on ellipses are somewhat disagreed with those from the previous *STRESS* study, i.e., OSA_{GP} clearly performed the best here for both the local and global uniformity, but it is not so obvious for the TOTAL and MEAN results in Table 3.

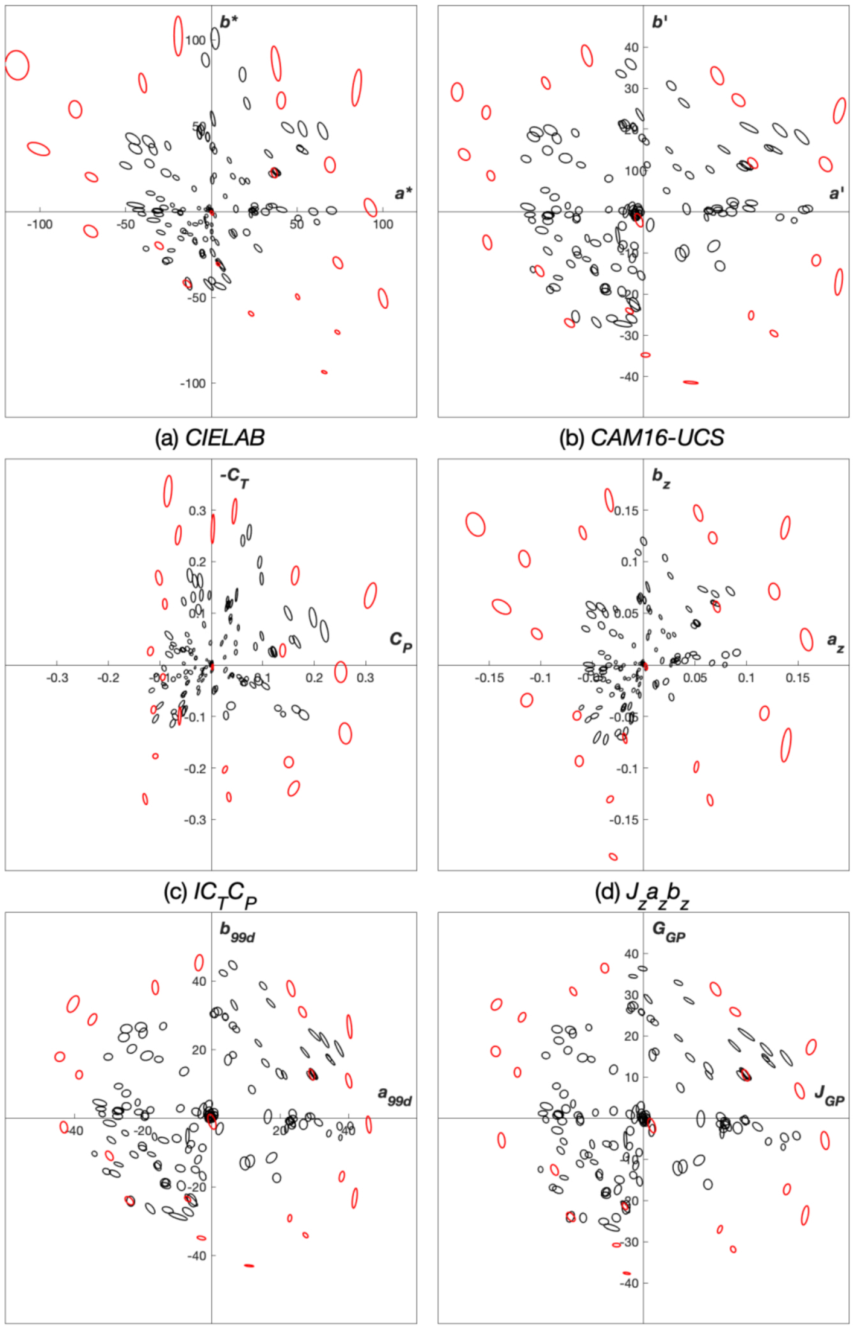
$$Local = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{A_i}{B_i}-1\right)^2} \times 100\%, \tag{5}$$

$$Global = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i-\bar{S})^2}}{\bar{S}} \times 100\%. \tag{6}$$

Table 4. Local and global uniformity of chromatic discrimination ellipses. Values are the *RMSE* between the ratio of the semi-major and semi-minor axes (local), and the Coefficient of Variation between the size of the ellipses and the average over all ellipses (global).

Colour Spaces	*Local (%)*	*Global (%)*
CIELAB	161	100
CAM16-UCS	160	54
$IC_T C_P$	292	78
$J_z a_z b_z$	133	71
DIN99d	198	52
OSA_{GP}	**117**	**49**

Optics EXPRESS



(a) *CIELAB* (b) *CAM16-UCS*

(c) IC_TC_P (d) $J_za_zb_z$

(e) *DIN99d* (f) OSA_{GP}

Fig. 5. Chromatic discrimination ellipses plotted in different colour spaces: (a) *CIELAB*; (b) *CAM16-UCS*; (c) IC_TC_P; (d) $J_za_zb_z$; (e) *DIN99d*; (f) OSA_{GP}.

Optics EXPRESS

3.2.3. Model performance

In this section, each colour model was improved for predicting the *WCG* dataset by introducing parametric factors, k_L, k_C and γ, Eq. (7).

$$\Delta E = \sqrt{\left(\frac{\Delta L}{k_L}\right)^2 + \left(\frac{\Delta C}{k_C}\right)^2 + (\Delta H)^2 + R_T\left(\frac{\Delta C}{k_C}\right)(\Delta H)}^{\,\gamma}, \tag{7}$$

where k_L and k_C are lightness and chroma factors, respectively; $\boldsymbol{\gamma}$ is a power factor; R_T is a rotation factor (only used in the *CIEDE2000* formula). Different sets of factors were optimised by minimizing the *STRESS* value between the visual data, $\boldsymbol{\Delta V}$, and the corresponding predicted values, $\boldsymbol{\Delta E}$. The results are listed in Table 5.

Table 5. The performance of each model, without and after optimization using parametric factors, in *STRESS* units.

Colour Centre	*CIELAB*	*CIEDE2000*	*CAM16-UCS*	ΔIC_TC_P	$J_za_zb_z$	*DIN99d*	OSA_{GP}
original	65	47	46	49	51	45	55
k_L	31	28	23	28	28	24	26
k_C	59	43	42	40	47	42	52
γ	64	46	43	48	48	45	55
k_L, k_C	31	27	23	28	28	24	26
k_L, γ	31	27	23	28	28	24	26
k_L, k_C, γ	31	27	23	27	28	24	26
k_L value	0.18	0.35	0.34	0.30	0.29	0.35	0.25

Table 5 shows that for all the models, the introduction of the parametric factors had a systematic effect, and each factor gave a different degree of improvement. For the optimised k_L models, the improvement was approximately 300%. Other versions of the models that included various combinations of k_L, k_C or/and γ factors showed little improvement compared to the k_L optimised models. Thus, it can be concluded that only the k_L factor had a strong impact on the colour-difference models when evaluating the *WCG* dataset. Table 5 also shows the values of the optimised k_L factor for each model, the mean value is 0.30, indicating that the perceived lightness difference is approximately 300% more obvious than the perceived chromatic difference in a sample pair.

Another study carried out by Melgosa *et al.* [39] to test the AUDI2000 color-difference formula, there was a small dependence of the weighting function for lightness with chroma indicating that high chroma values (approximately equivalent to high color gamut centres in the present study) reduce perceived lightness difference. This trend was not found in this study.

The F-test [15] was used to test the differences between models. For two given colour-difference formulae, the F value can be calculated using Eq. (8).

$$F = \frac{STRESS^2_{DE1}}{STRESS^2_{DE2}}. \tag{8}$$

For the present data, F_C was 0.82 with a 95% confidence level. There is significant difference between the two colour-difference formulae when $F< F_C$ or $F> 1/F_C$. Tables 6 and 7 show the F values between a pair of models for the original and optimised k_L models, respectively. An underlined value indicates that the model at the head of the column significantly outperformed that at the beginning of the row. Tables 6 and 7 showed that regardless of the original or optimised k_L models, *CAM16-UCS* and *DIN99d* performed the best amongst all the models tested. And *CIELAB* always performed the worst. The other models, i.e., *CIEDE2000*, ΔIC_TC_P, $J_za_zb_z$ and OSA_{GP}, gave similar performance (no significant difference between them).

Optics EXPRESS

Table 6. *F*-test values for all possible combinations of colour models (*Fc* = 0.82, *1/Fc* = 1.21) – without optimising k_L. *F values underlined are statistically significant.*

	CIELAB	*CIEDE2000*	*CAM16-UCS*	ΔIC_TC_P	$J_za_zb_z$	*DIN99d*	OSA_{GP}
CIELAB	–	**1.91**	**2.00**	**1.76**	**1.62**	**2.09**	**1.40**
CIEDE2000	**0.52**	–	1.04	0.92	0.85	1.09	**0.73**
CAM16-UCS	**0.50**	0.96	–	0.88	**0.81**	1.04	**0.70**
$\boldsymbol{\Delta IC_TC_P}$	**0.57**	1.09	1.13	–	0.92	1.19	**0.79**
$\boldsymbol{J_za_zb_z}$	**0.62**	1.18	**1.23**	1.08	–	**1.28**	0.86
DIN99d	**0.48**	0.92	0.96	0.84	**0.78**	–	**0.67**
$\boldsymbol{OSA_{GP}}$	**0.72**	**1.37**	**1.43**	**1.26**	1.16	**1.49**	–

Table 7. *F*-test values for all possible combinations of colour models (*Fc* = 0.82, *1/Fc* = 1.21) – after optimising k_L. *F values underlined are statistically significant.*

	CIELAB	*CIEDE2000*	*CAM16-UCS*	ΔIC_TC_P	$J_za_zb_z$	*DIN99d*	OSA_{GP}
CIELAB	–	**1.23**	**1.82**	**1.23**	**1.23**	**1.67**	**1.42**
CIEDE2000	**0.82**	–	**1.48**	1.00	1.00	**1.36**	1.16
CAM16-UCS	**0.55**	**0.67**	**–**	**0.67**	**0.67**	0.92	**0.78**
$\boldsymbol{\Delta IC_TC_P}$	**0.82**	1.00	**1.48**	–	1.00	**1.36**	1.16
$\boldsymbol{J_za_zb_z}$	**0.82**	1.00	**1.48**	1.00	–	**1.36**	1.16
DIN99d	**0.62**	**0.73**	1.09	**0.73**	**0.73**	–	0.85
$\boldsymbol{OSA_{GP}}$	**0.67**	0.86	**1.28**	0.86	0.86	1.17	–

3.2.4. Parametric effect

Note that the models tested (*CIEDE2000, CAM16-UCS,* $J_za_zb_z$*, DIN99d and* OSA_{GP}) were derived to fit *COMBVD*. They all had k_L factor of unity to give the best fit to the data. For the *WCG* dataset, it was found that by the introduction of k_L factor, each original model's performance improved greatly. This section attempts to clarify why the inclusion of an optimization of the factor controlling the lightness variable is important. This discrepancy was found to be due to the magnitudes of the colour differences and media difference. For the former, all pairs had a $\Delta E^*{}_{ab}$ value of 3 in Experiment 2. Experiment 1 had $\Delta E^*{}_{ab}$ values to be either 3 or 6. Table 8 will reveal the discrepancy, where the *STRESS* values for the different models have been divided into two Groups. The two Groups summarise the performance, in terms of *STRESS* value, of the original and optimised k_L models tested using three Pairs of datasets: Pair 1 to include the full set from Experiments 1 and 2; Pair 2 to consist of $\Delta E^*{}_{ab} = 3$ and $\Delta E^*{}_{ab} = 6$ subsets from Experiment 1; Pair 3 to comprise subsets of Experiments 1 and 2 having identical pairs. And the optimised k_L values for each model are also given in the bracket of Group 2.

From Table 8, some conclusions can be drawn.

1) Comparison between the full set from Experiments 1 and 2 of the original models (see Pair 1 of Group 1), the results showed that all models performed better when fitting Experiment 1 than Experiment 2 data by a mean *STRESS* unit of 14. Conversely, all models performed only slightly when fitting the Experiment 2 results as opposed to the Experiment 1 results (see Pair 1 of Group 2), by a mean *STRESS* unit of 4, for the optimised k_L models. This fmerit

2) The k_L value for Experiment 1 was consistently larger than that for Experiment 2, by a factor of about 1.5. This implies that the difference between Experiments 1 and 2 is mainly caused by the magnitude of the colour difference, $\Delta E^*{}_{ab}$ values of 3 and 6, and the $\Delta E^*{}_{ab}$ value of 3 for Experiments 1 and 2, respectively.

Optics EXPRESS

Table 8. The performance, in terms of *STRESS* values, of Group 1 (original) and Group 2 (the optimised k_L) models tested using three pairs of datasets (Pair 1 includes full sets for Experiments 1 and 2; Pair 2 includes $\Delta E^*_{ab} = 3$ and $\Delta E^*_{ab} = 6$ subsets of Experiment 1; Pair 3 includes subsets of Experiments 1 and 2 having identical pairs). Note that the optimised k_L values are given in the bracket of Group 2 models.

Group 1. Original model										
Pair of comparison		No. of Pairs	Magnitude	*CIELAB*	*CIEDE2000*	*CAM16-UCS*	ΔIC_TC_P	$J_za_zb_z$	*DIN99d*	OSA_{GP}
Pair 1	Exp 2	224	Full	70	57	57	56	60	58	64
	Exp 1	192	Full	62	42	39	46	47	39	50
Pair 2	Exp 1	96	$\Delta E^*_{ab} = 3$	65	47	45	51	51	45	53
	Exp 1	96	$\Delta E^*_{ab} = 6$	61	41	38	44	45	37	49
Pair 3	Exp 2	26	$\Delta E^*_{ab} = 3$	61	50	49	44	51	51	55
	Exp 1	26	$\Delta E^*_{ab} = 6$	58	45	44	43	48	46	51
Group 2. Optimised k_L model										
Pair of comparison		No. of Pairs	Magnitude	*CIELAB*	*CIEDE2000*	*CAM16-UCS*	ΔIC_TC_P	$J_za_zb_z$	*DIN99d*	OSA_{GP}
Pair 1	Exp 2	224	Full	27 (0.15)	25 (0.27)	19 (0.25)	22 (0.25)	24 (0.21)	22 (0.25)	22 (0.18)
	Exp 1	192	Full	32 (0.19)	27 (0.41)	22 (0.41)	30 (0.32)	28 (0.34)	23 (0.42)	26 (0.29)
Pair 2	Exp 1	96	$\Delta E^*_{ab} = 3$	31 (0.12)	29 (0.30)	24 (0.29)	28 (0.18)	29 (0.25)	27 (0.31)	24 (0.21)
	Exp 1	96	$\Delta E^*_{ab} = 6$	32 (0.21)	27 (0.44)	21 (0.44)	30 (0.36)	27 (0.36)	22 (0.45)	26 (0.31)
Pair 3	Exp 2	26	$\Delta E^*_{ab} = 3$	24 (0.18)	19 (0.27)	15 (0.26)	23 (0.31)	22 (0.22)	17 (0.25)	15 (0.20)
	Exp 1	26	$\Delta E^*_{ab} = 6$	23 (0.11)	23 (0.26)	19 (0.24)	24 (0.18)	21 (0.17)	21 (0.24)	18 (0.19)

3) Comparing the performance of the models between the two subsets in Experiment 1 ($\Delta E^*_{ab} = 3$ and $\Delta E^*_{ab} = 6$, see Pair 2 of Group 1), the results showed a mean difference of 6 *STRESS* units for the original models. And the difference is even smaller for the optimised k_L models (a mean *STRESS* unit of 2, see Pair 2 of Group 2). The k_L values, given in the bracket of Group 2, clearly showed that a larger colour difference set ($\Delta E^*_{ab} = 6$) had a higher value than those of small colour different set ($\Delta E^*_{ab} = 3$) by a factor of 1.5. This agrees with the finding the earlier results of Mirjalili *et al.* [27].

4) Comparison of the performance of the models between the two experiments using identical 26 pairs of samples, including the lightness and chromatic differences, showed that the original models (Pair 3 of Group 1) gave a similar performance for all models, as was found for the optimised k_L models (Pair 3 of Group 2). This indicates good repeatability between Experiments 1 and 2. The discrepancy of each model's performance between Experiments 1 and 2 is caused by the colour difference magnitudes.

Finally, it was also found that the present results indicate higher k_L values than those of Mirjalili *et al.* [27]. Note that the latter data were obtained using printed surface colours with no separation between a sample pair. The present experiments had no separation between the sample pair but were displayed on a monitor. This suggests that the difference could be caused by the media used, i.e., display vs. surface stimuli. Mirjalili *et al.* [27] proposed a simple linear equation [see Eq. (9)] to calculate the lightness parametric factor to fit their results for colour models tested.

$$\Delta E_{NS} = \sqrt{\left(\frac{\Delta L}{D_L}\right)^2 + \Delta C^2 + \Delta H^2}, \tag{9}$$

where $D_L = a\Delta E + b$, and a and b coefficients for each model are given in Table 9. Note that the a and b values for ΔIC_TC_p, $J_za_zb_z$, *DIN99d* and OSA_{GP} were newly optmised, or did not

Optics EXPRESS

reported in [27]. Equation (9) reflects a visual phenomenon, i.e., for a larger colour difference pair, a clearer dividing line (or separation) will appear than that of small colour difference. This will result in a larger D_L value, or a smaller computed lightness difference for a larger than for a smaller colour difference pair.

Table 9. Coefficients *a* and *b* from the Mirjalili *et al.* and present ΔE_{NS} formula.

	CIELAB	*CIEDE2000*	*CAM16-UCS*	ΔIC_TC_P	$J_za_zb_z$	*DIN99d*	OSA_{GP}
a	0.047	0.079	0.072	0.033	25.305	0.065	0.077
b	0.22	0.27	0.27	0.31	0.26	0.25	0.25
s_L	0.41	0.66	0.65	0.44	0.53	0.71	0.58
STRESS	34	30	28	31	32	29	29

Equation (9) is rewritten as Eq. (10) to consider the media effect.

$$\Delta E = \sqrt{\left(\frac{\Delta L}{s_L D_L}\right)^2 + \left(\frac{\Delta C}{k_C}\right)^2 + (\Delta H)^2}^{\,\gamma} \tag{10}$$

where s_L is optimised by each model to fit the present data. Values are given in Table 9 together with the associated *STRESS* values. Note that for each model, the *STRESS* values in Table 9 are only slightly worse than those in Table 8 with the optimised k_L values differing by about 4 units. This implies that the D_L function is robust to predict the colour difference magnitude effect, and the s_L factor, with a mean value of 0.57 for all models, indicating a media effect between the surface and display colours, i.e., a surface colour pair will exhibit 175% smaller perceived lightness difference than that when exactly reproduced the colorimetric value of pair on a display.

4. Conclusions

An experimental dataset, *WCG*, was accumulated to investigate colour differences covering a wide colour gamut of a DCI-P3 colour gamut display. 416 colour-difference pairs were assembled to cover the colour regions outside those of the surface colour dataset, *COMBVD*, used to develop the *CIEDE2000* colour-difference formula. One colour-difference formula and six UCSs were tested using the *WCG* dataset, and chromatic discrimination ellipses for each colour centre were fitted. Good agreement was found between the ellipses calculated from the *COMBVD* and the present data. All models' performances in *STRESS* unit were improved by the introduction of a factor, k_L, to correct a parametric effect, i.e., that lightness differences appear to be about 300% the chromatic differences in a colour difference pair. However, it was revealed colour difference magnitude and media to have big influence on the present experimental data having no separation between pairs of samples. The performance of the k_L optimised *CAM16-UCS*, *DIN99d* and OSA_{GP} models significantly outperformed the other colour models. The recently developed UCSs for HDR and WCG applications (IC_TC_P and $J_za_zb_z$) did not perform well.

Funding. National Natural Science Foundation of China (61775190).

Acknowledgment. The authors would like to thank Prof. Michael Pointer for his technical advice and his help to prepare the manuscript.

Disclosures. The authors declare no conflicts of interest.

References

1. Dolby, "ICtCp White Paper," https://www.dolby.com/us/en/technologies/dolby-vision/ICtCp-white-paper.pdf.
2. M. Safdar, G. Cui, Y. J. Kim, and M. R. Luo, "Perceptually uniform color space for image signals including high dynamic range and wide gamut," Opt. Express **25**(13), 15131–15151 (2017).
3. . "Colourimetry," CIE 015: 2018.

Optics EXPRESS

4. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," Color Res. Appl. **26**(5), 340–350 (2001).
5. M. R. Luo and B. Rigg, "BFD (l:c) colour-difference formula Part 2-Performance of the formula," J. Soc. Dyers Colour. **103**(3), 126–132 (2008).
6. D. H. Kim and J. H. Nobbs, "New weighting functions for the weighted CIELAB colour difference formula," in *Proceedings of the AIC* (1997), pp. 446–449.
7. M. R. Luo and B. Rigg, "BFD (l:c) colour-difference formula Part 1-Development of the formula," J. Soc. Dyers Colour. **103**(2), 86–94 (2008).
8. R. S. Berns, D. H. Alman, L. Reniff, G. D. Snyder, and M. R. Balonon-Rosen, "Visual determination of suprathreshold color-difference tolerances using probit analysis," Color Res. Appl. **16**(5), 297–316 (1991).
9. K. Witt, "Geometric relations between scales of small colour differences," Color Res. Appl. **24**(2), 78–92 (1999).
10. . "Colorimetry – Part 6: CIEDE2000 colour-difference formula," ISO/CIE 11664-6:2014(E).
11. G. H. Cui, M. R. Luo, B. Rigg, G. Roesler, and K. Witt, "Uniform colour spaces based on the DIN99 colour-difference formula," Color Res. Appl. **27**(4), 282–290 (2002).
12. C. Oleari, M. Melgosa, and R. Huertas, "Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space," J. Opt. Soc. Am. A **26**(1), 121–134 (2009).
13. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. **31**(4), 320–330 (2006).
14. . "A colour appearance model for colour management systems: CIECAM02," CIE 159: 2004.
15. M. Huang, H. Liu, G. Cui, and M. R. Luo, "Testing uniform colour spaces and colour-difference formulae using printed samples," Color Res. Appl. **37**(5), 326–335 (2012).
16. . "Recommended Method for Evaluating the Performance of Colour-Difference Formulae," CIE 217:2016.
17. . "CIE 2017 COLOUR FIDELITY INDEX FOR ACCURATE SCIENTIFIC USE," CIE 224:2017.
18. . "TM-30-15 IES Method for Evaluating Light Source Rendition," IES.
19. A. David, P. T. Fini, K. W. Houser, Y. Ohno, M. P. Royer, K. A. G. Smet, M. Wei, and L. Whitehead, "Development of the IES method for evaluating the color rendition of light sources," Opt. Express **23**(12), 15888–15906 (2015).
20. C. L. Gao, C. J. , Luo, and M. R. Pointer, "CAM20u: An extension of CAM16 for predicting colour appearance for unrelated colour," Accepted for publication by Color Res. Appl. (2021).
21. C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. H. Brill, and M. Pointer, "Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS," Color Res. Appl. **42**(6), 703–718 (2017).
22. B. Zhao, Q. Xu, and M. R. Luo, "Color difference evaluation for wide-color-gamut displays," J. Opt. Soc. Am. A **37**(8), 1257–1265 (2020).
23. R. S. Berns, "Methods for characterizing CRT displays," Displays **16**(4), 173–182 (1996).
24. A. Robertson, "CIE guidelines for coordinated research on color-difference evaluation," Color Res. Appl. **3**(3), 149–151 (1978).
25. M. Cheung and B. Rigg, "Colour-difference ellipsoids for five CIE colour centres," Color Res. Appl. **11**(3), 185–195 (1986).
26. G. H. Cui, M. R. Luo, B. Rigg, and W. Li, "Colour-difference evaluation using CRT colours. Part I: Data gathering and testing colour difference formulae," Color Res. Appl. **26**(5), 394–402 (2001).
27. F. Mirjalili, M. R. Luo, G. Cui, and J. Morovic, "Color-difference formula for evaluating color pairs with no separation: ΔE NS," J. Opt. Soc. Am. A **36**(5), 789–799 (2019).
28. T. Maier, "CIE guidelines for coordinated future work on industrial colour-difference evaluation," Color Res. Appl. **20**(6), 399–403 (1995).
29. M. Melgosa, "Request for existing experimental datasets on color differences," Color Res. Appl. **32**(2), 159 (2007).
30. S. S. Guan and M. R. Luo, "Investigation of Parametric Effects Using Large Colour Differences," Color Res. Appl. **24**(5), 356–368 (1999).
31. "Textiles - Tests for colour fastness - Part A02: Grey scale for assessing change in colour," ISO 105-A02.
32. "ASTM D1729-16 Standard Practice for Visual Appraisal of Colors and Color Differences of Diffusely-Illuminated Opaque Materials," ASTM International.
33. "ASTM E3040-18 Standard Practice for Evaluation of Instrumental Color Difference with a Gray Scale," ASTM International.
34. M. R. Luo and B. Rigg, "Chromaticity-discrimination ellipses for surface colours," Color Res. Appl. **11**(1), 25–42 (1986).
35. S. S. Guan and M. R. Luo, "Investigation of parametric effects using small colour differences," Color Res. Appl. **24**(5), 331–343 (1999).
36. P. A. Garcia, R. Huertas, M. Melgosa, and G. Cui, "Measurement of the relationship between perceived and computed color differences," J. Opt. Soc. Am. A **24**(7), 1823–1829 (2007).
37. E. Pieri and J. Pytlarz, "Hitting the Mark—A new color difference metric for HDR and WCG imagery," SMPTE Mot. Imag. J. **127**(3), 18–25 (2018).
38. B. Zhao and M. R. Luo, "Hue linearity of color spaces for wide color gamut and high dynamic range media," J. Opt. Soc. Am. A **37**(5), 865–875 (2020).
39. M. Melgosa, J. Martínez-García, L. Gómez-Robledo, E. Perales, F. M. Martínez-Verdú, and T. Dauser, "Measuring color differences in automotive samples with lightness flop: A test of the AUDI2000 color-difference formula," Opt. Express **22**(3), 3458–3467 (2014).