Prediction of Munsell Appearance Scales Using Various Color-Appearance Models

David R. Wyble,* Mark D. Fairchild

Munsell Color Science Laboratory, Rochester Institute of Technology,
54 Lomb Memorial Dr., Rochester, New York 14623-5605

Received 1 April 1999; accepted 10 July 1999

Abstract: The chromaticities of the Munsell Renotation Dataset were applied to eight color-appearance models. Models used were: CIELAB, Hunt, Nayatani, RLAB, LLAB, CIECAM97s, ZLAB, and IPT. Models were used to predict three appearance correlates of lightness, chroma, and hue. Model output of these appearance correlates were evaluated for their uniformity, in light of the constant perceptual nature of the Munsell Renotation data. Some background is provided on the experimental derivation of the Renotation Data, including the specific tasks performed by observers to evaluate a sample hue leaf for chroma uniformity. No particular model excelled at all metrics. In general, as might be expected, models derived from the Munsell System performed well. However, this was not universally the case, and some results, such as hue spacing and linearity, show interesting similarities between all models regardless of their derivation. © 2000 John Wiley & Sons, Inc. Col Res Appl, 25, 132–144, 2000

Key words: color appearance; color-appearance models; Munsell System

INTRODUCTION

Color-appearance models are used in many important applications related to color reproduction and, more generally, the prediction of color whenever source and destination viewing conditions are not identical. Examples of this abound in daily life: simultaneously viewing one image on a monitor and another nearby in hardcopy form; viewing the same image under light sources of different colors; viewing an object indoors and also outside under full sunlight. To predict the color of the objects accurately in these examples, a color-appearance model is required. Modern color-appearance models should, therefore, be able to account for changes in illumination, surround, observer state of adaptation, and, in some cases, media changes. This definition is slightly relaxed for the purposes of this article, so simpler models such as CIELAB can be included in the analysis.

This study compares several modern color-appearance models with respect to their ability to predict uniformly the dimensions (appearance scales) of the Munsell Renotation Data,[1] hereafter referred to as the Munsell data. Input to all models is the chromaticities of the Munsell data, and is more fully described below. As used in this work, these data are useful because of the uniformity achieved from an extensive visual evaluation performed to better adjust the chromaticities of the original Munsell system. Given that the conditions of the visual experiment were carefully controlled and reported, it should be possible to configure and apply color-appearance models successfully to these chromaticities. To the extent that the conditions of the original viewing conditions can be duplicated in the form of model input parameters, accurate models should properly predict the lightness, chroma, and hue scales of the Munsell data.

MUNSELL RENOTATION DATA

Input data for all model predictions are the chromaticities specified by Newhall, Nickerson, and Judd in their 1943 article.[1] Their study made a detailed investigation of minor non-uniformities in the 1929 *Munsell Book of Color*.[2] Observers viewed various sets of Munsell patches and judged how they should be adjusted to properly align the colors in relation to the surrounding patches. To adjust the colors, observers performed two different forms of ratio scaling. In Newhall's notation,[3] these were designated R, and R'. For R scaling, the observer judged what factor was required to scale a test color so that it would equal the reference color

* Correspondence to: David R. Wyble, Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science, 54 Lomb Memorial Dr., Rochester, NY 14623–5605 (e-mail: wyble@cis.rit.edu)

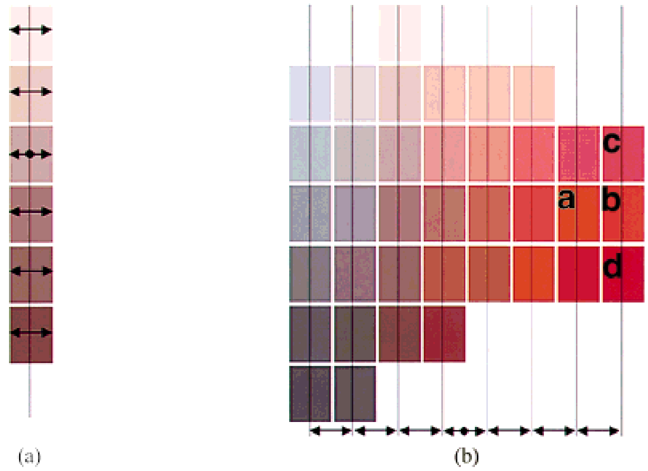


FIG. 1. Samples of the scaling experiment described in Ref. 1: (a) represents the chroma adjustment for a single hue/chroma combination. The dot shows the hypothetical average selected by the observer. The chromas of the other colors are adjusted relative to this average chroma, as indicated by the arrows. (b) represents the chroma adjustment for the entire sample hue leaf. The dot shows the hypothetical average chroma difference selected by the observer. The chroma difference between the other constant chroma lines are adjusted relative to this average chroma difference, indicated by the arrows. Of the steps listed in the text, steps 1–4 use (a); steps 5–7 use (b).

in some attribute (e.g., value, hue, or chroma). For R' scaling, the observer judged what factor was required to scale a test color difference to equal a reference color difference. Here, "color difference" is the difference along only one color dimension. Note that each dimension was always scaled separately. For example, the procedure of judging a single hue leaf for chroma adjustment is given below. Shown in Fig. 1(b), a hue leaf is a group of colors of constant hue and varying chroma and lightness. To perform the complete scaling experiment on a single hue leaf, an observer performs the following:

1. Select a line of constant chroma at this hue. Place the mask over the surrounding patches, exposing only patches of the selected chroma. See Fig. 1(a).
2. Examine all the patches in this vertical column, and determine which one best represents the average chroma. This is the reference chroma used in step 3, denoted by the black dot on the arrow near the center of Fig. 1(a).
3. Using the R method, judge the factor required to scale each of the other patches to make them equal the reference chroma.
4. Repeat steps 1–3 until all the constant chroma lines have been judged for this hue.
5. Remove the mask, exposing all the constant chroma lines at this hue. See Fig. 1(b).
6. Select one pair of adjacent constant chroma lines as representative of the average chroma difference among all the lines. This is the reference color difference used in step 7, denoted by the black dot on the arrow in Fig. 1(b).
7. Using the R' method, judge the factor required to scale the color difference between each of the other pairs of adjacent lines to equal to the reference color difference.

In a similar fashion, the other dimensions of Munsell space were judged and subsequently adjusted. In each case, small sets of patches were viewed for local adjustments, and then larger areas were exposed to make large-scale adjustments of, for example, entire hue leaves or circles of constant chroma. In most cases, a blend of the R and R' scaling methods was used in the final chromaticity renotation.

These renotation data have been used in many studies throughout the subsequent decades, and are useful in color-appearance work when they are properly applied. Critical to the understanding of Munsell renotation data is the fact that observers in the original study scaled only a single dimension of color with each set of visual

TABLE I. Input parameters for color-appearance models. For Hunt94, RLAB, LLAB, ZLAB, and CIECAM97s, parameters for dim and dark surround are also included. Alternative surround parameters were used only for lightness evaluation.

CIELAB	ZLAB	RLAB	LLAB
X_n 98.074 Y_n 100.0 Z_n 118.232	X_w 98.074 Y_w 100.0 Z_w 118.232 exps 0.345 L 400.0 D 1.0	X_n 98.074 Y_n 100.0 Z_n 118.232 — 1/2.3 L 400 D 1.0	X_0 98.074 Y_0 100.0 Z_0 118.232 D 1.0 F_S 3.0 F_L 1.0 F_C 1.0 Y_b 20.0
	Dim surround exps 0.295	Dim surround Sigmas 1/2.9	Dim surround D 0.7 F_S 3.5
	Dark surround exps 0.2625	Dark surround Sigmas 1/3.5	Dark surround D 0.7 F_S 4.0

Hunt94	Nay95	CIECAM97s	IPT[a]
X_W 98.074 Y_w 100.0 Z_w 118.232 L_A 400.0 CCT 6774 N_c 1.0 N_b 75.0 Y_b 20.0 D 1.0	X_n 98.074 Y_n 100.0 Z_n 118.232 L_0 400.0 E_{0r} 1000.0	X_w 98.074 Y_w 100.0 Z_w 118.232 L_A 400 C 0.69 N_c 1.0 F_{LL} 1.0 F 1.0 Y_b 20.0 D 1.0	X_n 98.074 Y_n 100.0 Z_n 118.232 L 400.0 D 1.0
Dim surround N_c 0.95 N_b 25.0		Dim surround c 0.59 N_c 1.1 F 0.9	
Dark surround N_c 0.9 N_b 10.0		Dark surround c 0.525 N_c 0.8 F 0.9	

[a] As mentioned in the text, IPT accepts only D65 tristimulus values. The notation used in Ref. 12 is, therefore, X_{D65}, Y_{D65}, and Z_{D65}.

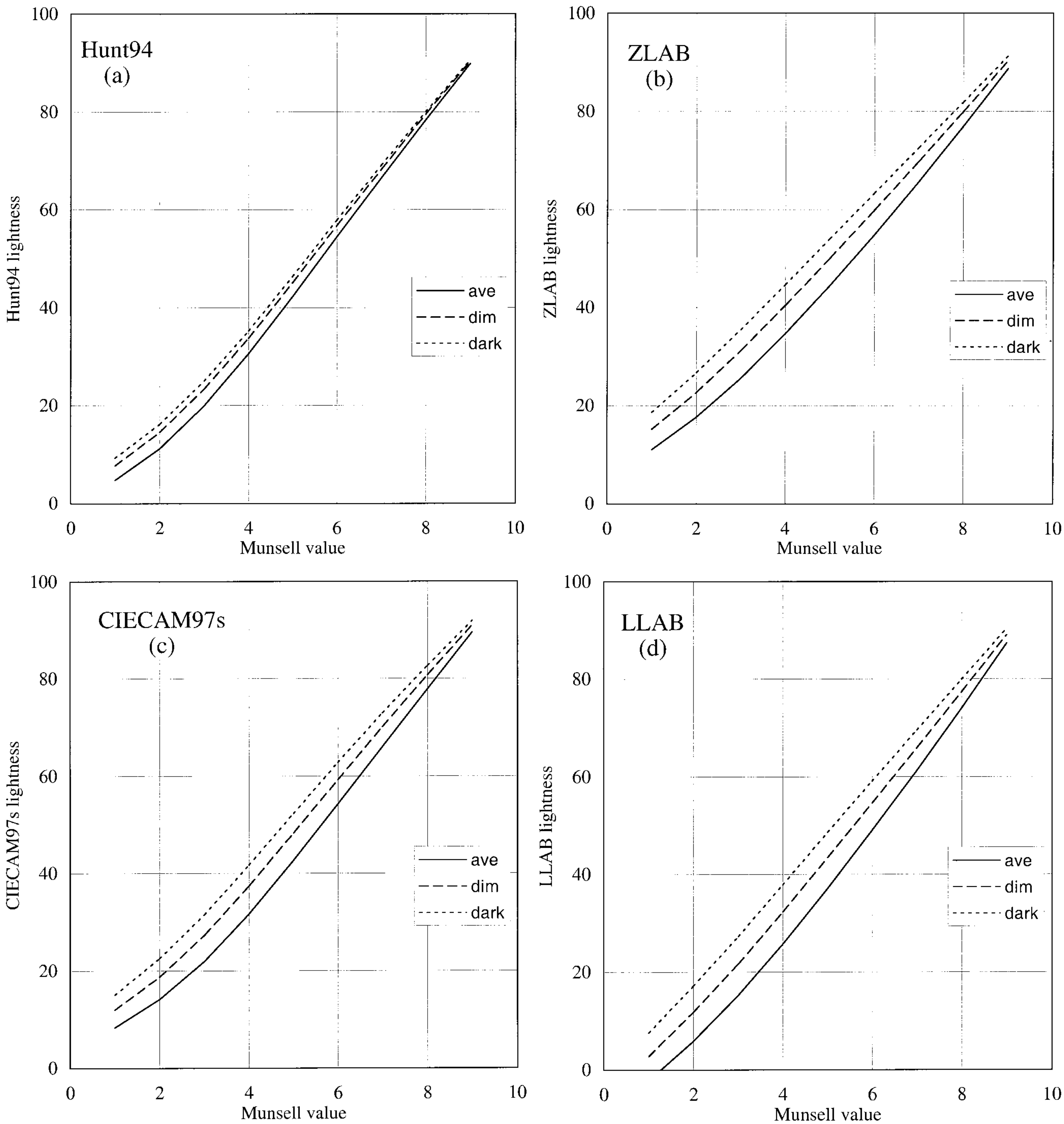


FIG. 2. Model lightness performance. Predicted lightness is plotted against Munsell value for all models: (a) Hunt94; (b) ZLAB; (c) CIECAM97s; (d) LLAB; (e) RLAB; (f) CIELAB; (g) IPT; (h) Nay95. Table II lists relevant statistics for lightness performance. (a)–(e) also show predictions for dim and dark surrounds. Parameters for alternative surround adjustments are listed in Table I.

judgments. Hence, it is never appropriate to expect uniform perceptual difference between two pairs of Munsell patches, if the patches vary in more than one dimension. For example, in Fig. 1(b), the color difference between the pair of colors marked (b) and (c) should be equal to that of the pair (b) and (d). They each represent one value step, with unchanged hue and chroma. However, the pair of colors (a) and (d) cannot be expected to have the same color difference as the pair (a) and (c). Comparing these colors requires changing two color attributes, chroma and value, which is inconsistent with the original scaling experiments. To retain consistency, all the metrics reported in this work rely on the variation of only one dimension at a time.

Another important issue to understand is that the chromaticities listed in the original article are extrapolated to

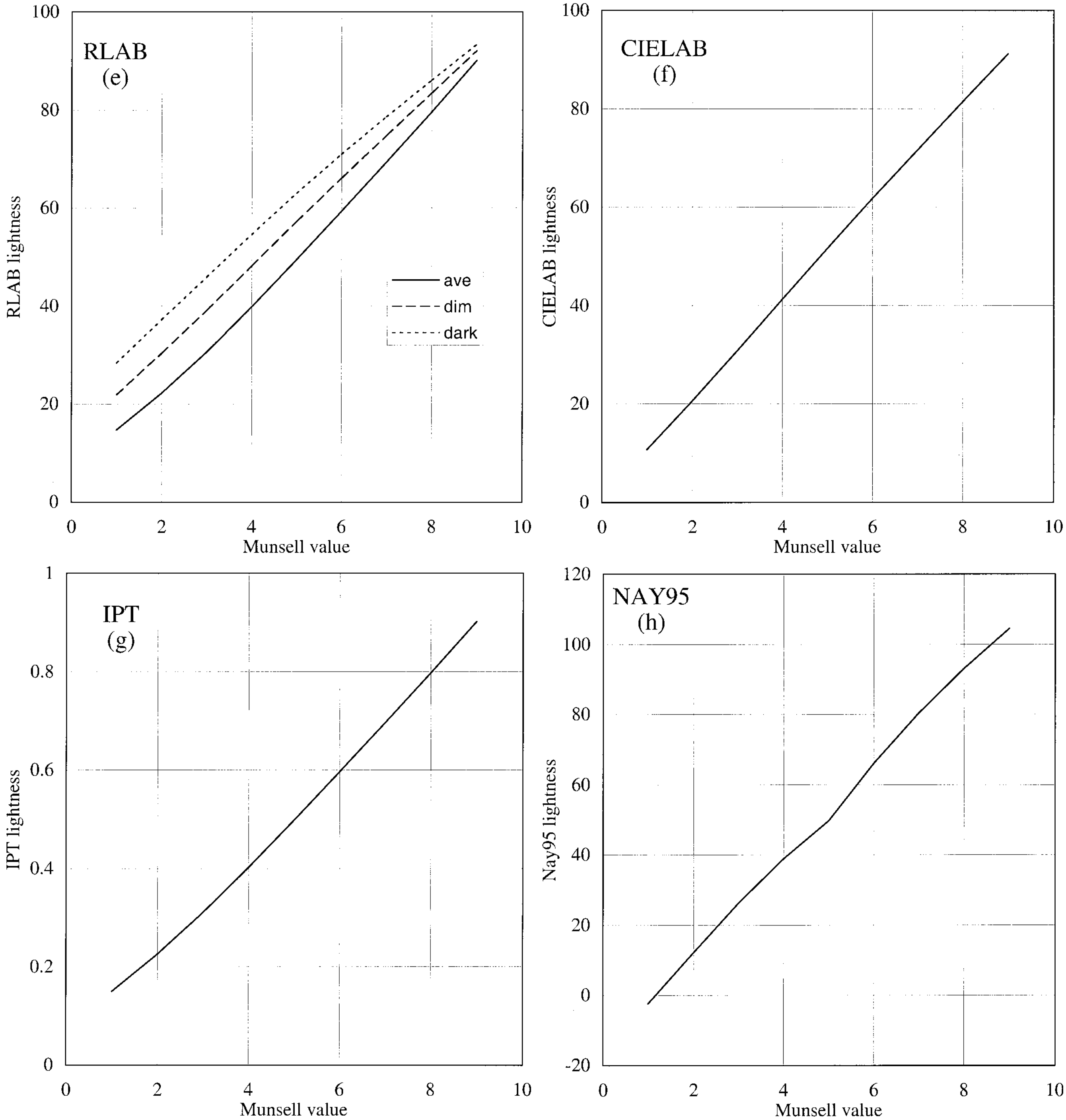


FIG. 2. (*Continued*)

the MacAdam Limits.[4] This means that the chromaticity values far exceed those of the 1929 *Munsell Book of Color*, from which the samples were taken for the scaling experiments. To separate the performance of the color-appearance models from the methods of extrapolation, the colors used for this article were limited to those found in the 1929 *Munsell Book of Color*. Note that even this choice includes colors not actually used as physical samples from the original scaling experiments, because not all 40 hues were used. However, no unreasonable colors were used here, in the sense that no chroma extrapolation was needed.

Subsequent studies have been performed on the Munsell System. One such study was made at NBS (National Bureau of Standards, now NIST, National Institute of Standards and Technology) in 1967.[5] The emphasis of this work was clearly related color difference equations, and therefore its relevance here is minimal. However, it represents the culmination of many years of effort, and we felt its mention was warranted.

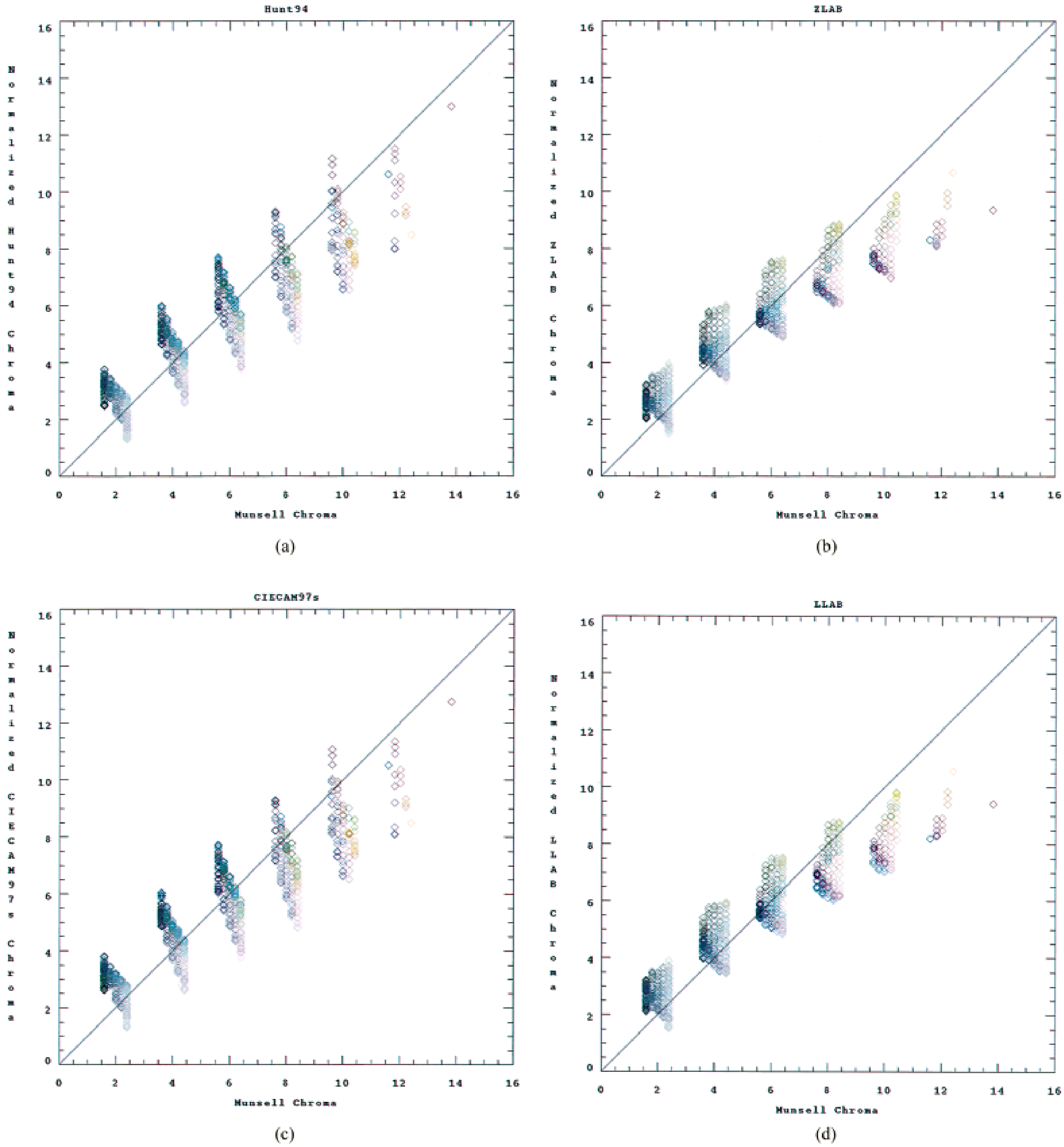


FIG. 3. Model chroma performance. Normalized model chroma is plotted against Munsell chroma for all models: (a) Hunt94; (b) ZLAB; (c) CIECAM97s; (d) LLAB; (e) RLAB; (f) CIELAB; (g) IPT; (h) Nay95. The solid line is drawn at unity, and does not indicate any type of fitting. Data points are shaded to approximate the color of the input Munsell color. The abscissa was slightly adjusted to prevent occlusion of data points. The input chroma is actually only even integers between 2–14, inclusive. The data clustered over chroma 2, for example, are all in fact calculated from an input color of *exactly* chroma 2. Vertical bands within each cluster represent Munsell Value. From left to right, values are: 2 or 3; 4; 5; 6; and 7 or 8.

PROCEDURE

The evaluation of color-appearance models took the form of a computer simulation. Input data were the renotation chromaticity coordinates and the model-specific parameters for viewing conditions. These are outlined for each individual model in Table I. Parameters were chosen to consistently and appropriately represent the viewing conditions recommended for Munsell samples: daylight (Illuminant C) and average surround. Multiple surround luminance was used for the lightness evaluation, but no direct inferences were made with respect to uniformity of model lightness. (More

TABLE II. Results from regression of model lightness vs. Munsell value. Only linearity was of interest, so no slope or intercept results are reported.

Model	R^2 of linear fit
CIELAB	0.999
ZLAB	0.994
RLAB	0.998
LLAB	0.995
Hunt94	0.994
Nay95	0.999
CIECAM97s	0.992
IPT	0.998

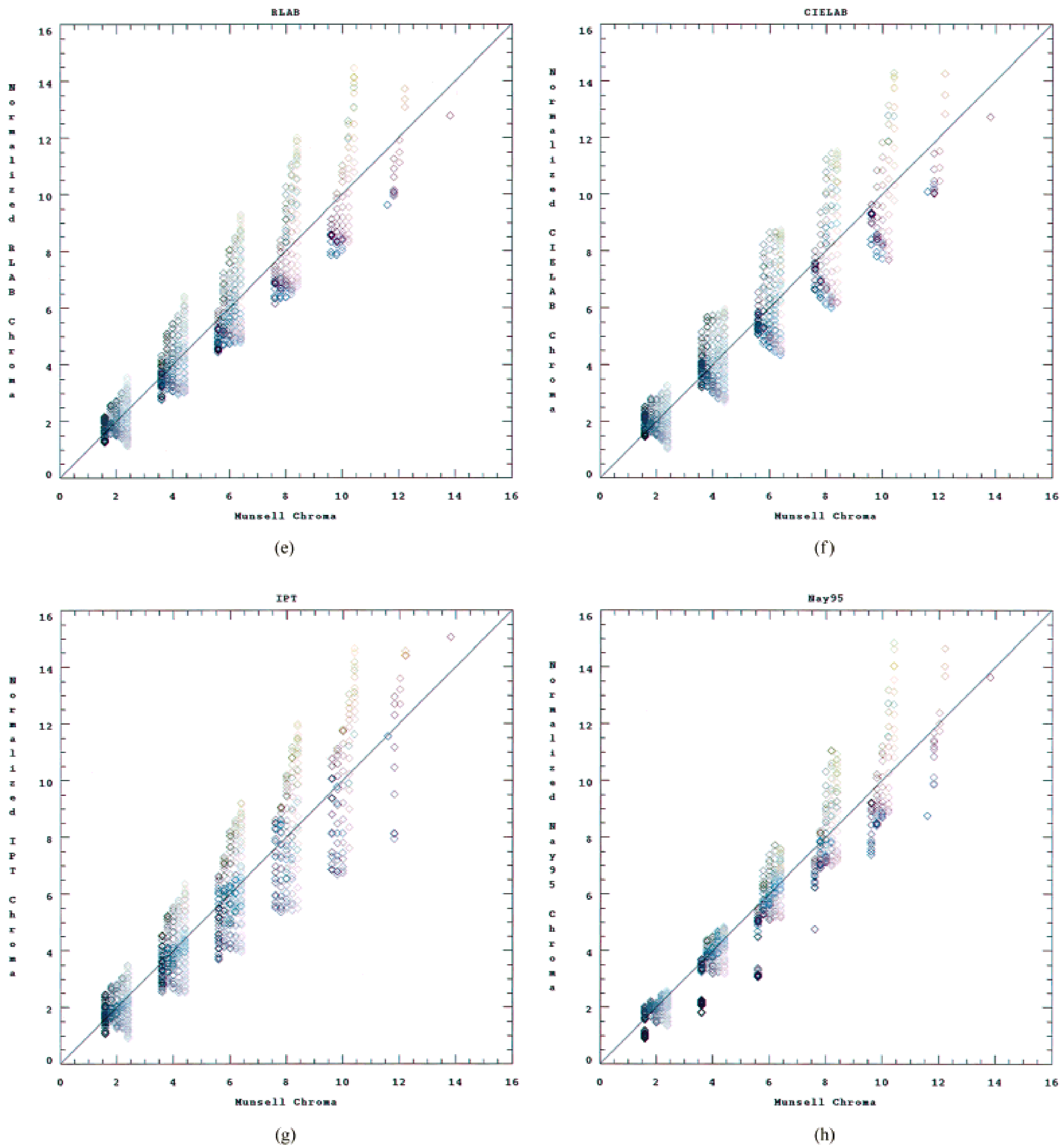
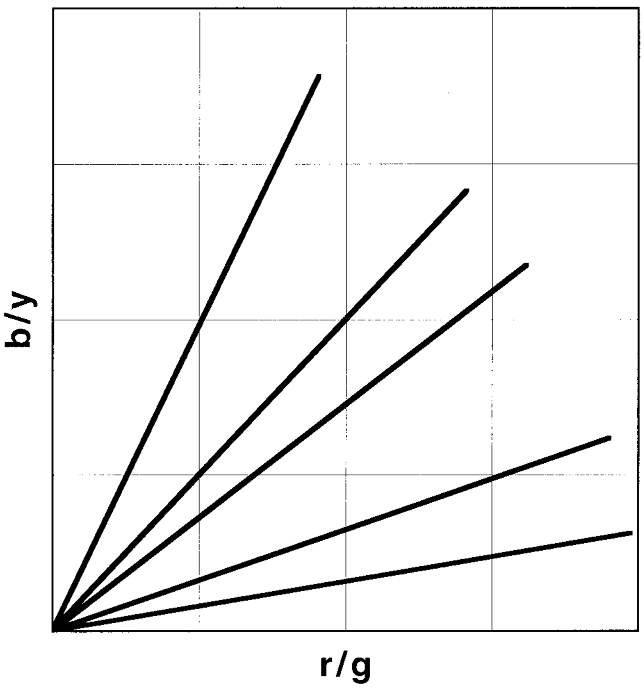


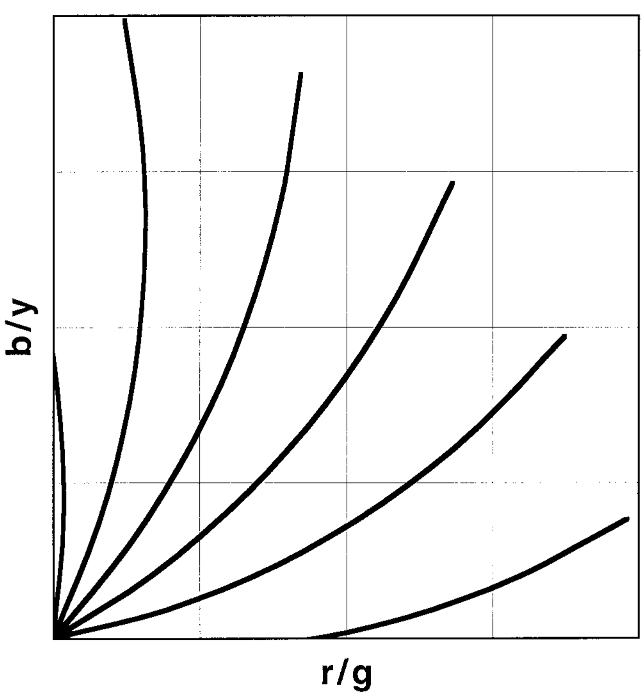
FIG. 3. (*Continued*)

TABLE III. Results from linear regression of normalized model chroma vs. Munsell Chroma. Bolded intercept values are not statistically significant (e.g., zero). Bolded slope values indicate that a slope of one lies within the 95% confidence intervals.

	β_0 (intercept)		β_1 (slope)			
Model	Value	*p*-value	Value	*p*-value	Lower 95%	Upper 95%
CIELAB	**0.04**	0.58	**0.98**	0.0	0.954	1.00
ZLAB	1.56	0.0	0.69	0.0	0.67	0.70
RLAB	**0.03**	0.61	**0.98**	0.0	0.96	1.01
LLAB	1.56	0.0	0.69	0.0	0.67	0.70
Hunt94	1.56	0.0	0.69	0.0	0.67	0.70
Nay95	**−0.23**	0.0	**1.02**	0.0	1.00	1.04
CIECAM97s	1.52	0.0	0.70	0.0	0.68	0.72
IPT	**−0.08**	0.35	**1.02**	0.0	0.99	1.05



(a)



(b)

FIG. 4. Sample hue spacing and linearity: (a) shows a model with perfect hue linearity but poor spacing. (b) shows a model with perfect spacing, but poor hue linearity. Both metrics are important for good hue prediction.

on the lightness conditions is explained below.) The color-appearance models used were: CIELAB,[6] Hunt,[7] Nayatani,[8] RLAB,[9] LLAB,[10] CIECAM97s,[11] ZLAB,[12] and IPT.[13] (Throughout this article these models are referred to as CIELAB, Hunt94, Nay95, RLAB, LLAB, CIECAM97s, ZLAB, and IPT, respectively. Numbers after Hunt and Nayatani indicate the year of publication of the specific model applied here.) The first five of these are well known, and thoroughly described in the past;[12] interested readers should check the references for a more detailed description of these models. CIECAM97s was proposed by the CIE in late 1997. ZLAB is a reduced form of CIECAM97s, proposed by Fairchild. It is most useful over a more limited set of viewing conditions than CIECAM97s. IPT was proposed in 1998 by Ebner and Fairchild. IPT was specifically designed to predict constant perceived hue.

The selection of any model for this study required that it predict the relative perceptual color attributes of lightness, chroma, and hue. Most of the above models predict more than just these three attributes, but limiting predictions to these allowed the use of CIELAB without compromising proper testing of the models with the Munsell renotation data. Models must also accept Illuminant C tristimulus values as input colors. This requirement is met for all models except IPT, which assumes input colors are D65 tristimulus values. For IPT, the RLAB chromatic adaptation model was used to transform the input from Illuminant C to D65 values. Since C and D65 are very close in chromaticity, it is not expected that prepending this transform onto the IPT model would have any significant detrimental effect.

The input conditions of all models are listed in Table I. Every attempt was made to properly configure the input parameters to match the recommended viewing conditions for viewing Munsell samples. In general, this simply means viewing object colors under Illuminant C with an average surround. For a complete description of these parameters, check the respective references for the models. For the lightness evaluation, five models were also used to predict the lightness of Munsell neutrals with dim and dark surrounds. Parameters for this adjustment are also shown in Table I for the five models that account for surround: Hunt94, RLAB, LLAB, ZLAB, and CIECAM97s.

It is important to recognize that we fully expect models derived from the Munsell data to perform very well in these evaluations. These models are CIELAB, and its derivative RLAB, and Nay95. We do not desire to simply verify that these models do indeed predict the data to which they were fit; however, the relative paucity of good color-appearance data require that some input data be selected. Models fit to the LUTCHI[13,14] data (LLAB, Hunt94, CIECAM97s, and its derivative ZLAB) do not predict all features of the Munsell renotation data. As such aspects of these models are uncovered, they are mentioned not to imply fault in the models, but rather to explore differences between the appearance data used to create the models. It is also possible that these evaluations might expose areas where both datasets are in general disagreement with all models. Perhaps experiments can be devised that avoid any such problems when creating future appearance data.

All the simulations were carried out using IDL™ (Inter-

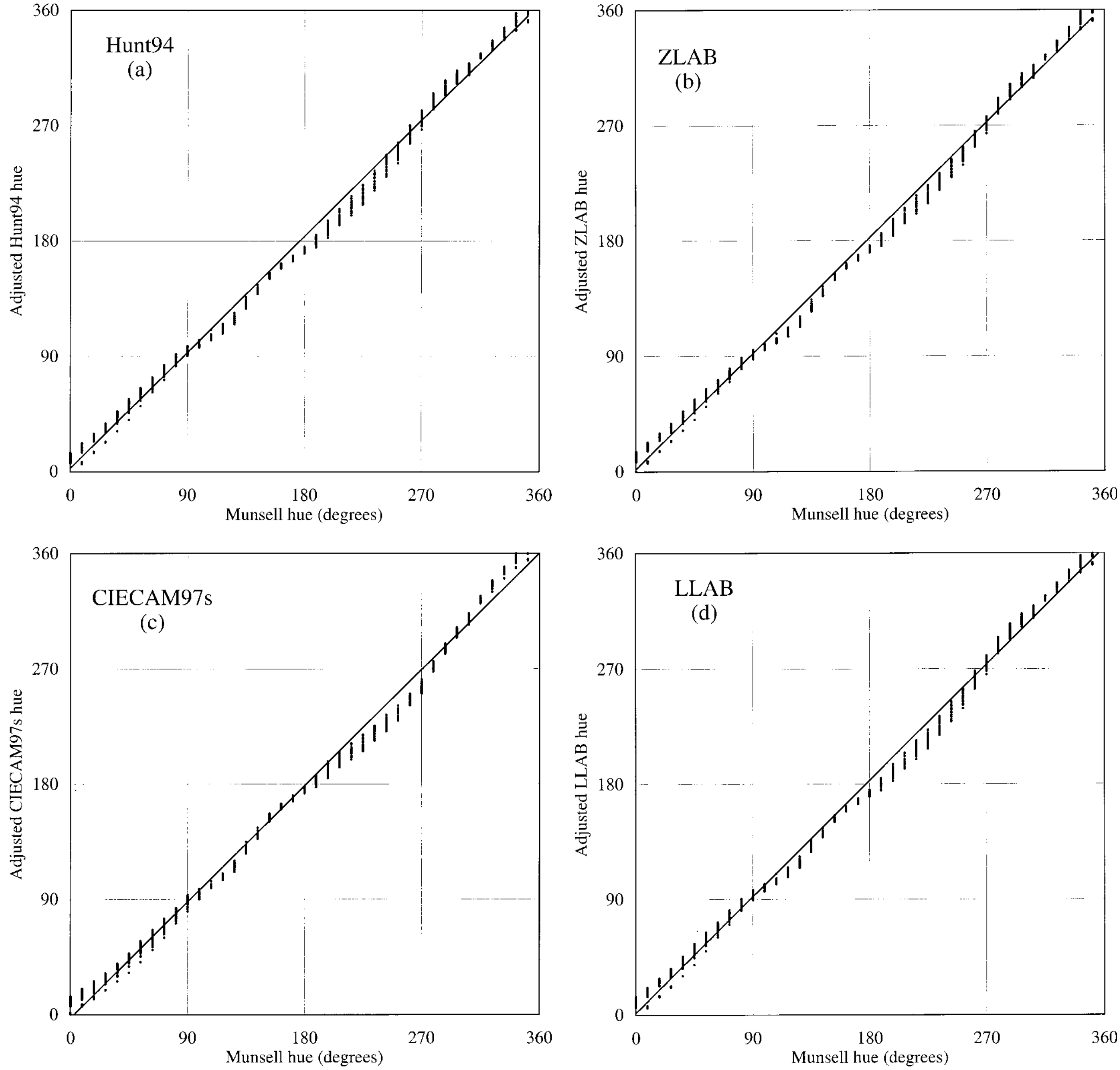


FIG. 5. Model hue performance. Adjusted hue is plotted against Munsell Hue for all models: (a) Hunt94; (b) ZLAB; (c) CIECAM97s; (d) LLAB; (e) RLAB; (f) CIELAB; (g) IPT; (h) Nay95. Adjusted hue is described in the text. The range of ordinate data for a given input hue represents the various predicted hues for an entire hue leaf. Predicted hue is not constant for constant input hue, because predicted hue also depends on input chroma and value. The solid line represents a linear fit. Systematic trends in the hue data for all models cannot be accounted for with the global linear fit. These trends are explored in Tables V and VI. (*Continued on next page*).

active Data Language). The code can be downloaded at www.cis.rit.edu/fairchild/CAM.html. On the same page, there are also Microsoft Excel™ worksheets with sample calculations from these models. The Munsell renotation data can be downloaded from www.cis.rit.edu/mcsl/online/.

RESULTS AND DISCUSSION

Given the reasonable division of the model predictions into lightness, chroma, and hue dimensions, the discussion focuses on each of these separately. Various metrics are reported, which allow quantitative comparison of the performance of the models. It should be noted that no attempt is made to select the "overall best" model. This is because, as described above, the three color dimensions in the Munsell Space cannot be appropriately combined. Remember that, in the original scaling experiments, observers adjusted each dimension of color separately. Hence, it would be inappropriate to make the assumption that the value, chroma, and hue performance can be united into an overall performance metric.

Another important point regarding the statistics per-

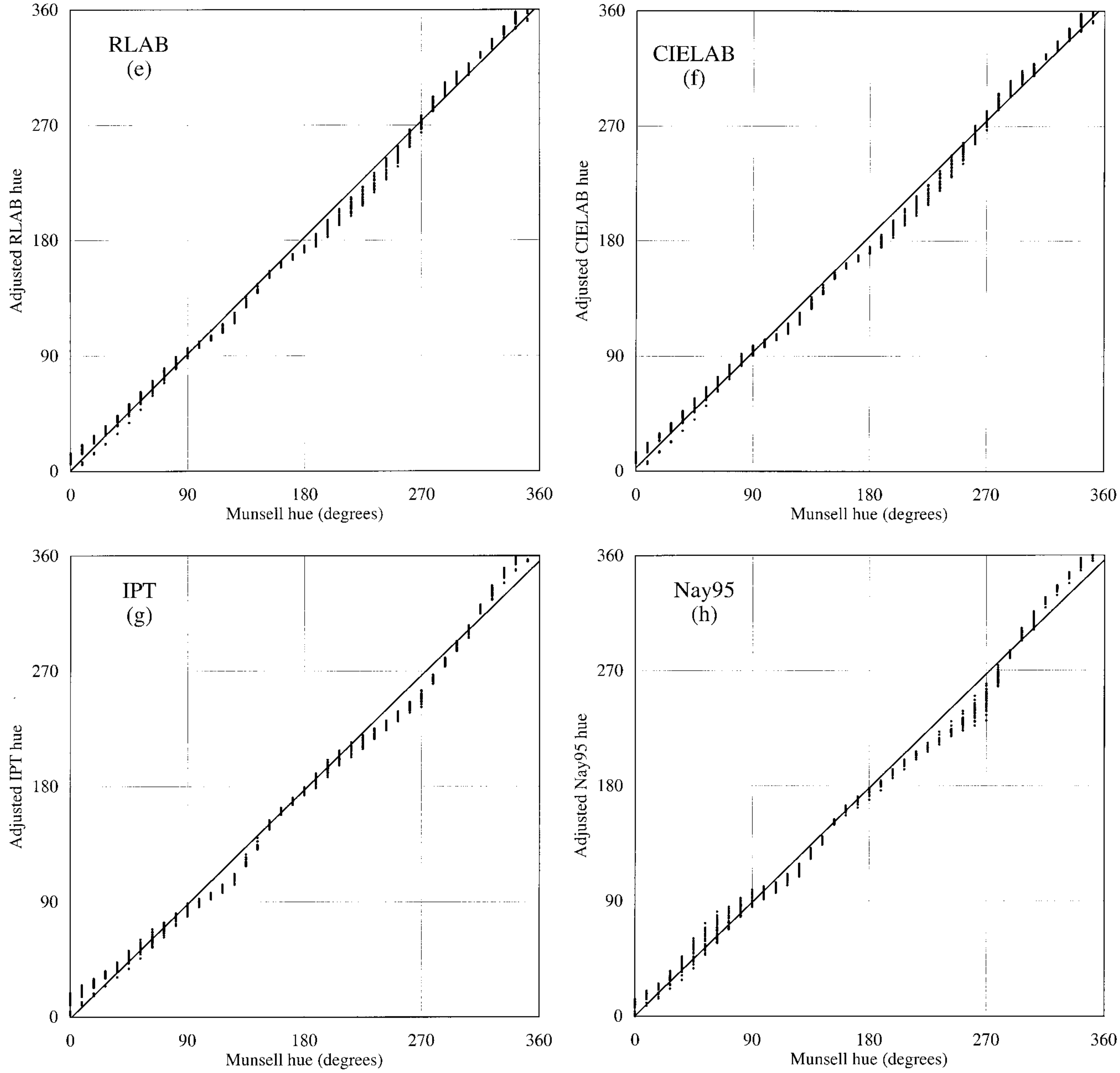


FIG. 5. (*Continued*)

formed is the assumption that the data are normally distributed. While no specific test was performed to verify normality, none of the data appear to be skewed in any way, indicating that this assumption is not met.

Lightness Linearity

Figure 2(a)–(h) shows model lightness for all models. The charts show model lightness plotted against Munsell value. A good lightness scale should be linear with Munsell value. To quantify the linearity, a linear fit was done on each set of model lightness data. Since the various models do not necessarily have the same absolute scales for lightness, no particular slope is desired for good performance. Only the linearity of lightness with respect to Munsell value is important. Therefore, only the correlation coefficient R^2 is reported for these model fits. Table II shows these values; all models performed quite well. Figure 2(a)–(h) also shows lightness performance for dim, and dark surrounds for models that comprehend surround luminance. These results are presented for completeness only; published results on the Munsell scaling experiment do not include the effects of varying surround on the Munsell value scale, so no quantitative statistics were performed on the dim and dark surround predictions.

Overall Chroma Performance

Figure 3(a)–(h) shows chroma performance for all models. The solid line is drawn with unity slope. The data points

TABLE IV. Results from the linear fit of adjusted model hue vs. Munsell Hue. As in Table III, good slopes are those for which unity lies within their 95% confidence intervals. Bolded slopes have one inside their 95% confidence limits. Even though only ZLAB is bolded, all the models produce a slope very close to unity. R^2 values for all models are in excess of 0.99. Note that intercept results are not shown because of the nature of hue as an interval scale.

	β_1 (slope)			
Model	Value	*p*-value	Lower 95%	Upper 95%
CIELAB	1.01	0.00	1.002	1.010
ZLAB	**1.00**	0.00	0.997	1.007
RLAB	1.01	0.00	1.006	1.014
LLAB	1.01	0.00	1.003	1.011
Hunt94	1.01	0.00	1.002	1.013
Nay95	0.989	0.00	0.983	0.995
CIECAM97s	1.01	0.00	1.002	1.012
IPT	0.991	0.00	0.985	0.997

are color coded to approximate the input Munsell color. The points are clustered about Munsell chromas of 2, 4, 6, 8, 10, and 12. This does not indicate that the input chroma was varied in this manner. The slight variation in the *x*-dimension was imposed to allow more points to be visible, and not hidden behind other data. The charts show normalized model chroma plotted against Munsell Chroma. Normalized chroma is needed to transform all model chromas to the same scale. This was done using Eq. (1):

$$C_{model,norm} = (C_M C_{model})/C_{model,ave}. \quad (1)$$

Here, C_M (Munsell Chroma) is always 6, C_{model} is model chroma, and $C_{model,ave}$ is the average model chroma of the 40 colors at *chroma* = 6, *value* = 5.* In Fig. 3, perfect model chroma would all lie precisely on the line of unity slope. This is clearly not the case, and it can be seen that, in general, the chroma for yellow colors is overpredicted (above the line) and that of blues is underpredicted (below the line).

A linear fit was done to these normalized data, and statistics were calculated to determine if the intercept is significantly different from zero, and if the value of one lies inside the 95% confidence intervals for the slope. Table III lists the intercept and probability (*p*-value) that the intercepts are significant for all the models. Also in Table III are slope, *p*-value, and upper and lower 95% confidence limits. Traditionally, authors publish statistical significance as a Boolean value, essentially whether or not the *p*-value exceeds a selected α. (The usual choices for a are 0.05 or 0.01, for 95% and 99% confidence, respectively.) By providing the *p*-value itself, readers are free to make their own judgment as to the significance of the coefficients. The *p*-value should be interpreted as the maximum choice of α for which the null hypothesis can be rejected, and, therefore, the intercept can be said to be something other than zero. Since we desire an intercept of zero, good models have high *p*-values for the intercept.

The bolded intercept values in Table III are those that are not statistically significant for any reasonable choice of α. For these models (CIELAB, RLAB, Nay95, and IPT), it can be said that the linear fit goes through the origin. The same four models also had slopes of unity within their 95% confidence intervals. Both of these are desirable for this linear fit. Examining Fig. 3, it can be easily seen that the other four models (Hunt94, LLAB, CIECAM97s, and ZLAB) show some chroma compression, which seems to be confounding the linear fit. From a statistical point of view, a higher-order fit would seem to be called for; however, there is no physical or perceptual justification for this. It is possible that these more complex appearance models are accounting for some perceptual phenomena that are not represented in the Munsell data. Hunt94, LLAB, CIECAM97s, and ZLAB were all derived from the LUT-CHI color-appearance dataset, not the Munsell data. The behavior of the chroma fit is not surprising, given the chroma compression used in these models.

Hue Linearity

There are two important metrics regarding hue prediction: hue linearity and hue spacing. Hue linearity implies that constant Munsell Hue input results in constant model hue output. Hue spacing implies that each plane of constant Munsell Hue input, on average, results in hue output that is evenly spaced from the next hue leaf. Figure 4(a) shows a sample model with perfect hue linearity, but poor hue spacing. Figure 4(b) shows a sample model with perfect hue spacing, but poor hue linearity. It is apparent why both metrics are needed for good overall hue prediction.

Figure 5(a)–(h) shows the hue linearity performance for all models. Note that some adjustment of predicted hue was required. This adjustment was performed as follows: assume that Munsell Hues are assigned numerical values from 0–360°. These hues are the abscissa of Fig. 5. The ordinate values, predicted hue, were adjusted in the following manner: whenever low Munsell Hues (<180°) resulted in high model hue (>180°), 180° was subtracted from the model hue. Likewise, when high Munsell hues (>180°) resulting in low predicted hue (<180°), 180° was added to these hues. (For reference, on this scale the primary Munsell hues are: 5R, 18°; 5Y, 90°; 5G, 162°; 5B, 234°; 5P, 306°.) This fix was required, because many lines of constant Munsell Hue input result in a curved locus of predicted hue, which cross the hue = 0° line. With this adjustment, hue loci for colors of a constant Munsell hue may be negative or exceed 360°. For example, using the set of all Munsell colors with hue of 351° as input, predicted hue may range from 345–365°, instead of 354–5°, which might confound the analy-

* The selection of this chroma circle to use as average chroma for the normalization is somewhat arbitrary. It is true that alternative choices might have changed the performance of the models. However, no investigation was made into "best" normalization point, because it would have been difficult to select a point and still treat all models impartially.

TABLE V. Hue linearity results, shown as percent unexplained variance of the principal component analysis. Lower values indicate that the data at that hue are better fit by a line, and, hence, more nearly linear. Bolded points are less than 0.5%, italicized points are between 0.5–1.0%. These values are relative, and not easily transformed to color coordinates. See Fig. 6 for examples to aid in visually understanding the relative size of these numbers.

Hue name	CIELAB	ZLAB	RLAB	LLAB	Hunt94	Nay95	CIECAM97s	IPT
10RP	3.33	5.81	2.56	5.48	2.50	*0.75*	2.64	2.30
2.5R	1.36	2.75	1.07	2.73	1.29	**0.33**	1.40	*0.99*
5R	*0.81*	1.79	*0.61*	1.87	1.07	**0.18**	1.17	*0.59*
7.5R	*0.51*	1.13	**0.42**	1.24	1.08	**0.13**	1.18	**0.36**
10R	*0.55*	1.02	**0.45**	1.17	1.84	**0.37**	2.01	**0.32**
2.5YR	**0.36**	*0.77*	**0.28**	*0.89*	1.93	*0.99*	2.10	**0.22**
5YR	**0.27**	*0.62*	**0.24**	*0.73*	2.06	1.13	2.22	**0.22**
7.5YR	**0.25**	*0.58*	**0.24**	*0.68*	2.27	1.33	2.39	**0.32**
10YR	**0.32**	*0.71*	**0.33**	*0.86*	2.54	*0.64*	2.64	**0.16**
2.5Y	**0.28**	*0.62*	**0.31**	*0.75*	1.83	**0.47**	1.86	**0.14**
5Y	**0.20**	**0.49**	**0.21**	*0.58*	1.21	**0.22**	1.16	**0.10**
7.5Y	**0.14**	**0.26**	**0.13**	**0.33**	**0.40**	**0.14**	**0.38**	**0.06**
10Y	**0.07**	**0.16**	**0.06**	**0.19**	**0.19**	**0.15**	**0.18**	**0.09**
2.5GY	**0.09**	**0.12**	**0.05**	**0.16**	**0.16**	**0.12**	**0.16**	**0.17**
5GY	**0.10**	**0.21**	**0.08**	**0.23**	*0.71*	**0.17**	*0.72*	**0.33**
7.5GY	**0.10**	**0.23**	**0.07**	**0.23**	*0.82*	**0.07**	*0.83*	**0.19**
10GY	**0.09**	**0.22**	**0.08**	**0.21**	*0.68*	**0.05**	*0.69*	**0.14**
2.5G	**0.30**	**0.62**	**0.27**	*0.56*	*0.88*	**0.08**	*0.87*	**0.39**
5G	**0.21**	**0.41**	**0.20**	**0.38**	**0.37**	**0.06**	**0.37**	**0.26**
7.5G	**0.23**	**0.46**	**0.23**	**0.42**	**0.34**	**0.09**	**0.34**	**0.33**
10G	**0.27**	*0.54*	**0.24**	*0.51*	*0.51*	**0.15**	*0.52*	**0.39**
2.5BG	1.10	1.99	1.04	1.86	1.52	**0.16**	1.52	1.57
5BG	1.74	2.98	1.62	2.91	2.33	**0.18**	2.38	1.94
7.5BG	2.03	3.17	1.84	3.25	2.58	**0.18**	2.68	1.66
10BG	2.12	3.06	2.11	3.31	2.10	**0.13**	2.21	1.24
2.5B	2.70	3.72	2.82	4.23	2.09	**0.08**	2.21	1.09
5B	2.43	3.05	2.63	3.67	1.57	**0.16**	1.66	*0.59*
7.5B	2.87	3.66	3.35	4.49	1.55	**0.37**	1.62	**0.43**
10B	2.03	2.66	2.32	3.29	1.34	*0.97*	1.42	**0.21**
2.5PB	1.26	1.68	1.31	2.10	*0.92*	2.13	*0.99*	**0.19**
5PB	*0.94*	1.30	*0.87*	1.57	*0.91*	3.55	*0.97*	**0.48**
7.5PB	*0.93*	1.54	*0.94*	1.61	*0.89*	*0.83*	*0.88*	**0.10**
10PB	*0.85*	1.59	*0.93*	1.53	1.03	**0.16**	*0.96*	**0.28**
2.5P	*0.52*	1.09	*0.60*	1.00	*0.60*	1.02	*0.53*	*0.54*
5P	**0.26**	*0.60*	**0.31**	*0.53*	**0.27**	1.19	**0.24**	*0.51*
7.5P	**0.04**	**0.10**	**0.05**	**0.09**	**0.28**	**0.09**	**0.27**	**0.12**
10P	**0.31**	*0.68*	**0.36**	*0.59*	*0.85*	**0.04**	*0.83*	*0.57*
2.5RP	**0.50**	1.12	*0.53*	*0.98*	*0.72*	**0.04**	*0.72*	*0.59*
5RP	1.34	2.73	1.24	2.43	1.31	**0.28**	1.35	1.23
7.5RP	1.65	3.31	1.39	3.03	1.44	**0.37**	1.52	1.29
average	0.89	1.49	0.86	1.57	1.22	0.49	1.27	0.57

sis. Note that this is a perfectly reasonable adjustment, given the nature of hue as an interval scale, without a meaningful zero.

A linear fit was made to the adjusted model hue data. There was statistically significant lack of fit found for all models. When a lack of fit is found, it is not typically appropriate to continue analyzing the linear fit. However, given the nature of the data, it is clear that the models should produce hue linear with Munsell Hue. Therefore, the statistics of the linear fit are shown in spite of the lack of fit. Table IV shows coefficients, *p*-values, and the lower and upper 95% confidence limits for the slope. Performance for most models is quite good, and slope is generally very close to unity, but only ZLAB can be said to have a slope of one. While these results are encouraging, the lack of fit is an important problem that is not easily dismissed.

Another quantification of hue linearity is shown in Table V. Using principal component analysis (PCA), a measure of hue linearity is derived as a function of hue. The nature of PCA is to rotate the data into a space where the variance is minimized. When applied to two-dimensional data, this is essentially fitting the line, which explains the most variance

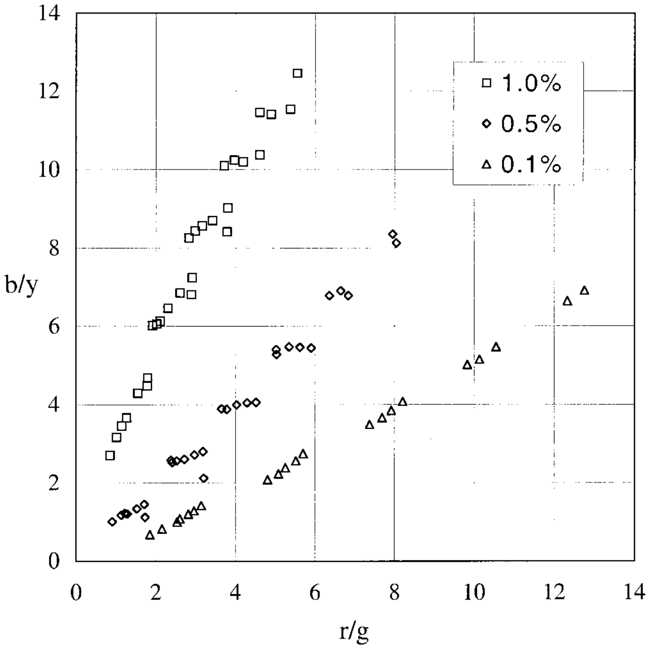


FIG. 6. Model predictions exemplifying various levels of unexplained variance. Relating to Table VI, simulated data are shown, which have 1.0, 0.5, and 0.1% variance unexplained by the first component of the principal component analysis. It can be seen that 1.0% or even 0.5% cannot be considered exceptional results for this metric.

of the data. The metric used here is how much variance is *not* explained by the fit. (That is, the variance left over from the first component.) These results are in Table V for all models and all hues. Readers are cautioned that the seemingly small unexplained variances are not insignificant. Figure 6 shows the spread of simulated output for three sets of data. Variances for the three hue lines are 1.0, 0.5, and 0.1%. While none of the fits appear particularly bad, it might be a mistake to believe that 1.0% or even 0.5% unexplained variance was exceptional. Table V shows two broad areas of hue space, where most models tend to perform well: reds, through yellows, and yellow-greens; and in the purple-blues. Overall, Nay95 and IPT were best, followed by CIELAB, RLAB, and the others.

The hue linearity results are not inconsistent with published results on the substantially improved linearity of the IPT model.[13] The colors used in this analysis were only those in the 1929 *Munsell Book of Color*, and are, hence, somewhat limited in their maximum chromas. The IPT model is able to make good hue predictions to much greater chromas and it is, therefore, not unexpected that IPT hue linearity performance is no better than the other models for this data set.

Hue Spacing

The 40 Munsell Hues fill the complete hue circle that most models define in terms of 360°. Therefore, good hue spacing performance means that predicted hue lines should be 9° apart. The statistical procedure used is as follows. The hypothesis to be tested is that two populations (hue leaves) differ by 9°. A *p*-value was generated comparing each set of colors at a given hue with the next set of colors at the next hue. Results for the various models are shown in Table VI. Since the goal is to have hue spacing equal to 9°, we desire the *p*-values to be high, indicating that the difference between the populations was not significantly different from 9°. No models universally predict well-spaced hue leaves. Bolded values are those above a significance of 95% ($\alpha = 0.05$). Some models have many borderline points; in some cases, a slight change in the selection of α would substantially alter the number of significant hues.

It is interesting to note that the models follow the same general trends with respect to the areas of color space where hue spacing is good. Most models have good spacing through the red-yellow region and also in the blue-green region. As with the previous metrics, it is possible that these trends are uncovering some underlying features in the Munsell data. This is especially true because the behavior of the models fit to other data (Hunt, ZLAB, LLAB, CIECAM97s) is very similar to those fit to Munsell data.

CONCLUSIONS

It is difficult to draw conclusions that point to a specific model as "best." The performance of most models varies depending on the metric chosen. This in itself is a useful, if not a particularly surprising conclusion. The appearance modeler is free to select the model best performing on the appearance scale of interest. Metrics that show hue dependency bring out the different performance of models with respect to hue. It is, therefore, possible to select a model that performs best in the area of color space in which the modeler is interested. This would be useful for industrial applications, where a specific color center is of a great importance (e.g., a logo or product label). In general, however, the separate nature of the Munsell appearance scales makes the unification of these metrics impossible.

Another intriguing question is whether these results can be used to make another correction to the Munsell Chromaticity coordinates. There are several occurrences of multiple models making predictions counter to the Munsell data. It is possible that these models, especially those not derived from the Munsell data, are uncovering systematic trends in the data that do not reflect perceptual phenomena. These trends may require subtle corrections not accounted for in the original scaling experiments. The verification of this adjustment would require another experiment of the scope of the previous one, which would be a large and complex undertaking.

ACKNOWLEDGMENTS

This work was done under the support of the Munsell Color Science Laboratory.

TABLE VI. Hue spacing results, shown as p-values for significance of average hue difference. Higher p-values indicate greater probability that the average hue difference between successive hue leaves is equal to 9°. Bolded values are those above the typical 95% point, or $\alpha = 0.05$.

Hue name	CIELAB	ZLAB	RLAB	LLAB	Hunt94	Nay95	CIECAM97s	IPT
10RP	**0.212**	**0.215**	**0.078**	**0.154**	0.013	0.001	0.001	0.013
2.5R	**0.140**	**0.109**	0.027	**0.092**	0.000	0.002	0.003	0.013
5R	**0.176**	**0.099**	0.046	**0.119**	0.014	0.020	0.013	0.013
7.5R	**0.397**	**0.405**	**0.382**	**0.476**	0.041	**0.405**	**0.072**	0.013
10R	**0.440**	**0.345**	**0.371**	**0.488**	**0.058**	**0.284**	**0.090**	0.017
2.5YR	**0.397**	**0.348**	**0.425**	**0.421**	**0.359**	**0.371**	**0.425**	**0.059**
5YR	**0.138**	0.043	**0.171**	**0.151**	**0.209**	**0.433**	**0.239**	0.018
7.5YR	**0.326**	**0.115**	**0.468**	**0.367**	**0.456**	**0.298**	**0.425**	0.029
10YR	**0.090**	0.024	**0.288**	**0.134**	**0.375**	**0.394**	**0.375**	0.037
2.5Y	0.035	0.007	**0.203**	**0.062**	**0.312**	**0.109**	**0.330**	0.011
5Y	0.000	0.000	0.000	0.000	0.002	0.003	0.013	0.000
7.5Y	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10Y	0.001	0.000	0.003	0.000	0.030	0.013	0.007	0.003
2.5GY	0.013	0.013	0.002	0.002	0.014	0.000	0.005	0.003
5GY	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7.5GY	0.013	0.001	0.005	0.006	0.000	0.000	0.001	0.000
10GY	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2.5G	0.031	**0.301**	0.022	0.027	**0.341**	0.012	**0.429**	0.000
5G	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.038
7.5G	0.000	0.003	0.000	0.000	0.000	0.001	0.000	**0.087**
10G	0.022	0.045	0.009	0.011	0.008	0.002	0.010	**0.284**
2.5BG	**0.337**	**0.308**	**0.448**	**0.433**	**0.352**	**0.195**	**0.397**	**0.305**
5BG	**0.460**	**0.413**	**0.345**	**0.363**	**0.084**	0.014	**0.113**	**0.161**
7.5BG	**0.452**	**0.444**	**0.417**	**0.456**	0.049	0.000	**0.071**	0.042
10BG	**0.456**	**0.305**	**0.337**	**0.375**	0.015	0.000	0.026	0.003
2.5B	**0.169**	**0.305**	**0.274**	**0.212**	0.049	0.001	0.079	0.002
5B	**0.067**	**0.151**	**0.115**	**0.076**	**0.106**	0.003	**0.166**	0.003
7.5B	**0.095**	**0.179**	**0.136**	**0.092**	**0.119**	0.037	**0.179**	0.003
10B	0.001	0.001	0.001	0.000	**0.308**	**0.233**	**0.215**	0.002
2.5PB	0.004	0.004	0.002	0.001	**0.106**	**0.433**	**0.068**	**0.058**
5PB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7.5PB	**0.055**	0.018	0.005	0.023	0.000	0.000	0.000	0.000
10PB	**0.248**	**0.472**	**0.480**	**0.363**	0.000	0.000	0.000	0.000
2.5P	0.014	**0.088**	**0.068**	0.029	0.000	0.000	0.000	0.000
5P	**0.079**	0.002	0.003	0.041	0.000	0.000	0.000	0.000
7.5P	0.018	**0.345**	**0.121**	0.026	0.001	0.001	0.002	0.000
10P	**0.149**	0.015	**0.100**	**0.166**	0.001	0.000	0.001	0.000
2.5RP	0.011	0.002	0.017	0.017	0.021	0.007	0.018	0.000
5RP	**0.326**	**0.452**	**0.198**	**0.258**	0.022	0.004	0.031	**0.308**
7.5RP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

1. Newhall SM, Nickerson D, Judd DB. Final report of the O.S.A. subcommittee on spacing of the Munsell colors. J Opt Soc Am 1943; 33:385–418.
2. The Munsell book of color. Baltimore: Munsell Color Company, 1929.
3. Newhall SM. The ratio method in the review of the Munsell colors. Am J Psych 1939;52:394–405.
4. MacAdam DL. Maximum visual efficiency of colored materials. J Opt Soc Am 1935;25:361–367.
5. Judd DB, Nickerson D. One set of Munsell re-renotations. National Bureau of Standards Report 192693 1967.
6. Colorimetry. CIE Publ No 15.2. Vienna: CIE; 1986.
7. Hunt RWG. The reproduction of colour. 5th Ed. London: Fountain; 1995.
8. Nayatani Y, Sobagaki H, Hashimoto K, Yano T. Lightness dependency of chroma scales of a nonlinear color-appearance model and its latest formulation. Color Res Appl 1995;20:156–167.
9. Fairchild MD. Refinement of the RLAB color space. Color Res Appl 1996;21:338–346.
10. Luo MR, Lo MC, Kuo WG. The LLAB(ℓ:c) colour model. Color Res Appl 1996;21;412–429.
11. The CIE 1997 interim colour appearance model (simple version). CIECAM97s. CIE Publ No 131. Vienna: CIE; 1998.
12. Fairchild MD. Color appearance models. Reading, Mass: Addison Wesley; 1997.
13. Ebner F, Fairchild MD. Development and testing of a color space (IPT) with improved hue uniformity. IS&T Color Imag Conf; 1998.
14. Luo MR, Clarke AA, Rhodes PA, Schappo A, Scrivner SAR, Tait CJ. Quantifying colour appearance. Part I. LUTCHI colour appearance data. Color Res Appl 1991;16:166–180.
15. Luo MR, Clarke AA, Rhodes PA, Schappo A, Scrivner SAR, Tait CJ. Quantifying colour appearance. Part II. Testing colour models performance using LUTCHI color appearance data. Color Res Appl 1991; 16:181–197.