Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

Metamer mismatching in practice versus theory

XIANDOU ZHANG,[1] BRIAN FUNT,[2,*] AND HAMIDREZA MIRZAEI[2]

[1]School of Media & Design, Hangzhou Dianzi University, Hangzhou 310018, China
[2]School of Computing Science, Simon Fraser University, Vancouver V5A 1S6, Canada
*Corresponding author: funt@sfu.ca

Received 8 October 2015; revised 4 January 2016; accepted 10 January 2016; posted 11 January 2016 (Doc. ID 251604); published 0 MONTH 0000

Metamer mismatching (the phenomena that two objects matching in color under one illuminant may not match under a different illuminant) potentially has important consequences for color perception. Logvinenko *et al.* [PLOS ONE 10, e0135029 (2015)] show that in theory the extent of metamer mismatching can be very significant. This paper examines metamer mismatching in practice by computing the volumes of the empirical metamer mismatch bodies and comparing them to the volumes of the theoretical mismatch bodies. A set of more than 25 million unique reflectance spectra is assembled using datasets from several sources. For a given color signal (e.g., CIE XYZ) recorded under a given first illuminant, its empirical metamer mismatch body for a change to a second illuminant is computed as follows: the reflectances having the same color signal when lit by the first illuminant (i.e., reflect metameric light) are computationally relit by the second illuminant, and the convex hull of the resulting color signals then defines the empirical metamer mismatch body. The volume of these bodies is shown to vary systematically with Munsell value and chroma. The empirical mismatch bodies are compared to the theoretical mismatch bodies computed using the algorithm of Logvinenko *et al.* [IEEE Trans. Image Process. 23, 34 (2014)]. There are three key findings: (1) the empirical bodies are found to be substantially smaller than the theoretical ones; (2) the sizes of both the empirical and theoretical bodies show a systematic variation with Munsell value and chroma; and (3) applied to the problem of color-signal prediction, the centroid of the empirical metamer mismatch body is shown to be a better predictor of what a given color signal might become under a specified illuminant than state-of-the-art methods.

OCIS codes: (330.1690) Color; (330.1715) Color, rendering and metamerism.

http://dx.doi.org/10.1364/JOSAA.99.099999

1. INTRODUCTION

Metamer mismatching [1] refers to the fact that two objects reflecting metameric light under one illumination may reflect nonmetameric light under a second, so two objects having the same color under one illuminant may have different colors under a second. Metamer mismatching has important consequences for human vision and computer vision since the light illuminating an object is frequently changing, for example as it moves from direct sun to shadow, or when the lights are turned on in a room, or the image is taken at a different time of day, or the object is viewed under fluorescent light at one moment and tungsten light at another.

Foster *et al.* [2] investigated the frequency with which nonidentical reflectances form metameric pairs under various daylight illuminants and found it to be rare. However, the relative frequency with which two objects reflecting metameric light under one illuminant then reflected nonmetameric light under a second illuminant was much higher. Based on 50 spectral-reflectance images of natural scenes under various phases of daylight ranging from correlated color temperatures of 4,000 to 25,000 K, they found that the frequency of occurrence of a metameric pair under one illuminant becoming distinguishable under a second illuminant was 10^{-2} to 10^{-1}. In a subsequent study, Feng and Foster [3] employed the conditional entropy of colors to predict the frequency of metamerism in natural scenes and again found it to be relatively low. Morovic and Haneishi [4] calculated the probabilities of metamer mismatching in 40 multispectral images with the illuminants changed from D65 to 173 different spectral power distributions and found a similar low frequency of metameric pairs. Prasad and Wenhe [5] consider the issue of metamer mismatching between three digital camera models and the human observer.

In contrast to these studies of the frequency of metamer mismatching in a typical scene, our focus here is not on the frequency of metamer mismatching but rather on the potential amount of metamer mismatching when it occurs. Given only the color signal produced in response to light reflected from an object of unknown reflectance under a given illuminant, we address the issue of what precisely can be said about what

the color signal from that same object is likely to be under a different illuminant. Logvinenko *et al.* [6] addressed this issue in terms of the degree of metamer mismatching (i.e., the volume of the metamer mismatch volumes/bodies) that can arise in theory. Here, we focus on the degree of metamer mismatching that arises in practice.

As Logvinenko *et al.* [6] argue, metamer mismatching imposes limits on color constancy since even when the full spectra of the two illuminants are known there is an inherent ambiguity in terms what a given color signal (i.e., camera sRGB or CIE XYZ coordinates) under a first illuminant will become under a second illuminant. In the color constancy and computer vision fields, it is generally assumed that the color of an object is an intrinsic property of the object, and hence the focus is on discounting the effects of the illuminant in order to recover the intrinsic color of the object. The intrinsic color is frequently expressed as the color signal that would be obtained from the object under some standard, "canonical" illuminant. However, Logvinenko [7] proves that color cannot be an intrinsic property of an object. His argument is straightforward: If two objects, A and B, are metameric matches (i.e., reflect light that generates an identical color signal) under the first illuminant but do not match under the second illuminant, then which of the two objects is to be considered the carrier of the "intrinsic" color? Clearly, a single color signal that becomes two different color signals cannot possibly map to some unique "intrinsic color" coordinate.

Metamer mismatching means that a color signal under a first light can become any color signal from an infinite convex set of color signals under a second light. This convex set is often called the metamer mismatch volume, or sometimes, the metamer mismatch body. In the present context the latter terminology is preferred because we wish to explore the volumes of metamer mismatch volumes/bodies and the multiple meanings of "volume" in a phrase such as "the volume of the metamer mismatch volume" can become very confusing. We will refer instead to "the volume of the metamer mismatch body" with the body referring to the (three-dimensional) convex set of color signals and the volume being the volume of that convex set.

To establish the extent of metamer mismatching in practice, we examine empirically the metamer mismatch bodies arising under several typical illumination changes for a large set of reflectance spectra obtained from multispectral images and other datasets of reflectances. A preliminary study [8] showed how the empirical mismatch bodies varied systematically with Munsell chroma and value. The present study expands the set of reflectances and illumination conditions used for testing and also compares the empirical metamer mismatch bodies to the theoretical metamer mismatch bodies calculated using the method of Logvinenko *et al.* [9].

The theoretical metamer mismatch body is based on the premise that the reflectances generating color signals on the boundary of the object color solid are special two-transition reflectances. The reflected values of such two-transition reflectances are either zero or one and make at most two transitions from zero to one or vice versa across the visible spectrum. Clearly, two-transition reflectances with either zero or one values seldom appear in practice, but there is no obvious, nonarbitrary way (e.g., an arbitrary degree of "smoothness") to constrain the set of reflectances further. The tests reported here show that the average volumes of the empirical and theoretical metamer mismatch bodies are clearly related, with the empirical bodies being substantially smaller than the theoretical ones.

Given a color signal arising from an object under a first illuminant, all that can be said definitely about its color signal under a second illuminant is that it could be any one of the color signals within its metamer mismatch body. Despite this lack of certainty, it is frequently the case (e.g., when white balancing an image) that we need to predict what the object's color signal is most likely to be under a second illuminant. Of course, any prediction can only be a guess since any of the color signals within the metamer mismatch body is a plausible answer. However, when forced to choose, what is a good choice to make? We explore this issue by making predictions based on several different measures (e.g., mean, median, centroid) of the metamer mismatch body and compare the mean prediction error to that obtained using the CAT02 [10] chromatic adaptation transform that underlies the CIECAM02 [11] color appearance model and to Mirzaei and Funt's [12] Gaussian metamer method of color signal prediction.

2. REFLECTANCE AND ILLUMINANT SPECTRA

In order to analyze the effects of metamer mismatching in practice, we construct a large dataset of reflectance spectra along with a sampling of illuminant spectra. The reflectance data are divided into disjoint training and test sets. Even though there is no machine learning involved, we use the term "training set" since we will be predicting results for the test data based on a prior set of reflectance data.

A. Dataset of Training Reflectances

A large dataset of spectral reflectances was assembled by gathering spectra from various sources in order to create a representative dataset of the spectral reflectances of natural and man-made objects that are likely to occur in practice. All the spectral reflectances are sampled from 400 to 700 nm at a 10 nm sampling interval.

The dataset was assembled from six main sources. The first group includes 11 multispectral images consisting of rocks, trees, leaves, grass, earth and urban scenes, and medieval and early modern illustrated works [2]. The second group includes 32 multispectral images [13] containing scenes of faces, hair, paints, food, drinks, and some other natural and man-made items. The third group includes 13 multispectral images [14] containing scenes of people, houses, hands, fruits, flowers and other natural and man-made items. The fourth group includes nine hyperspectral images containing scenes of textile, wood, leaves, painting, paper, and skin [15]. The fifth group includes 21 multispectral images mainly composed of different man-made items [16]. These five groups of images were all acquired with multispectral imaging systems. The sixth group includes spectral reflectances of man-made, natural, and industrial objects, which were measured using a spectrophotometer [17]. In total this leads to a set of 35,420,169 reflectance spectra.

Since many of these reflectance spectra are from multispectral images, it is likely that there will be many extremely similar or duplicate spectra in the datasets. To eliminate these similar/duplicate spectra, the numerical precision of the spectral data is first reduced to integer values 0–50 (i.e., multiplied by 50 and rounded), and then any spectra that are identical at that level of precision are removed. The spectra retained are kept at their full, initial precision. Although a little *ad hoc*, computationally this method is much faster than computing a distance metric (e.g., angular difference) between the approximately 10^{15} pairs of spectra. The final dataset contains 25,303,486 distinct reflectance spectra.

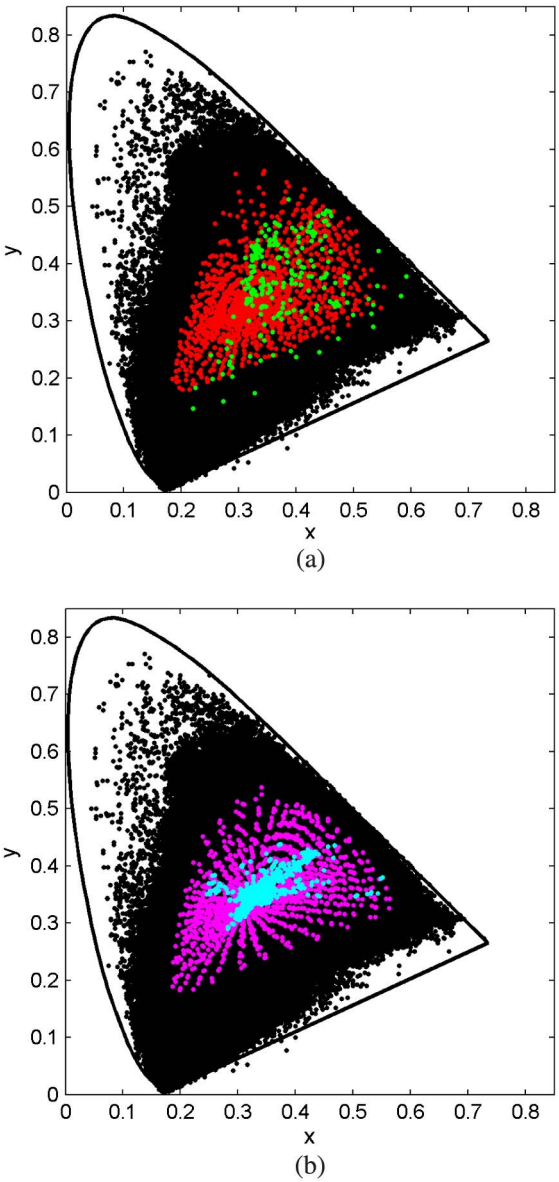


Fig. 1. Chromaticities of the reflectances in the various datasets under D65 plotted in the *xy*-chromaticity diagram. (a) Black dots indicate the samples from the full training dataset; red (gray in greyscale reproduction) dots are the Munsell papers; and green (white) dots indicate the Finland "Natural Colors" reflectances. (b) Black dots as in (a), bright purple (gray) dots indicate the NCS papers, and cyan (white) dots indicate the JPL reflectances.

B. Dataset of Test Reflectances

For testing, a second, smaller set of reflectance spectra is created by combining the 1600 reflectances of the Munsell glossy edition [18] papers, the 1950 reflectances of the Natural Color System (NCS) [19] samples, along with the 218 reflectances from the "Natural Colors" subset of the University of Eastern Finland's (UEF) spectral database [20] and 1301 reflectances of natural objects in the ASTER Spectral Library from the Jet Propulsion Laboratory (JPL) [21]. The total test set contains 5,069 reflectances.

C. Chromaticities of the Reflectances

As an indicator of how complete the set of reflectances is we computed the CIE1931 2-deg observer XYZ values under CIE D65 (daylight) of all the spectral reflectances in the training and test sets and plotted them in *xy*-chromaticity space [i.e., $x = X/(X+Y+Z)$, $y = Y/(X+Y+Z)$] as shown in Fig. 1. The plot shows that the full training set (black dots) covers a very significant portion of the *xy*-chromaticity diagram.

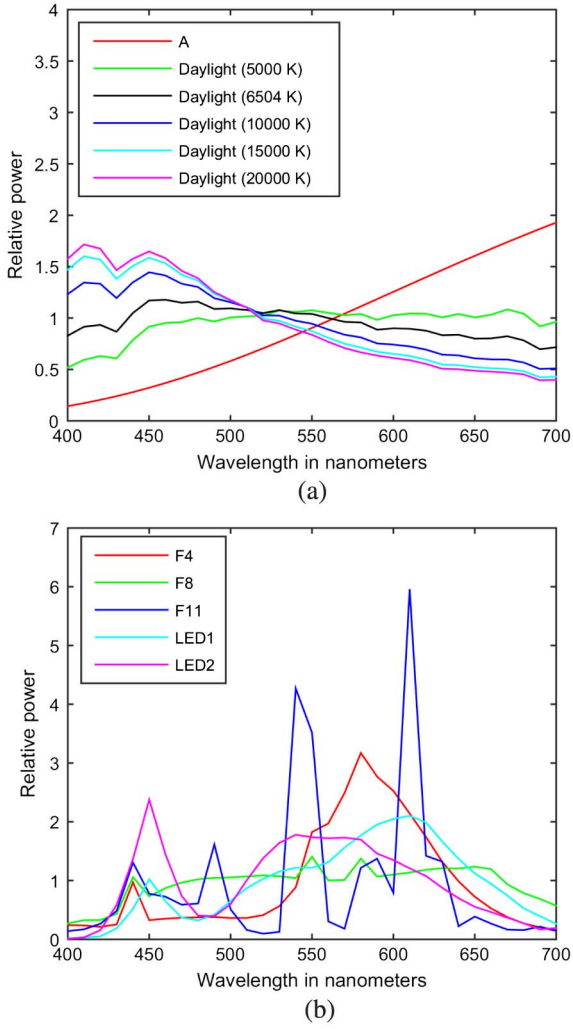


Fig. 2. Relative spectral power distributions of the 11 illuminants used for testing. (a) Illuminants D50, D65, D100, D150, and D200; (b) illuminants F4, F8, F11, LED1, and LED2.

D. Illuminant Spectra

Eleven illuminants, namely the CIE standard illuminants A, D50 (5000 K), D65 (6504 K), D100 (10000 K), D150 (15000 K), D200 (20000 K), F4, F8, and F11, along with two cellular/mobile phone LEDs, are used in evaluating the metamer mismatch bodies and color-signal prediction results. They were chosen as a representative test set since A is a typical tungsten light bulb; D50, D65, D100, D150, and D200 are

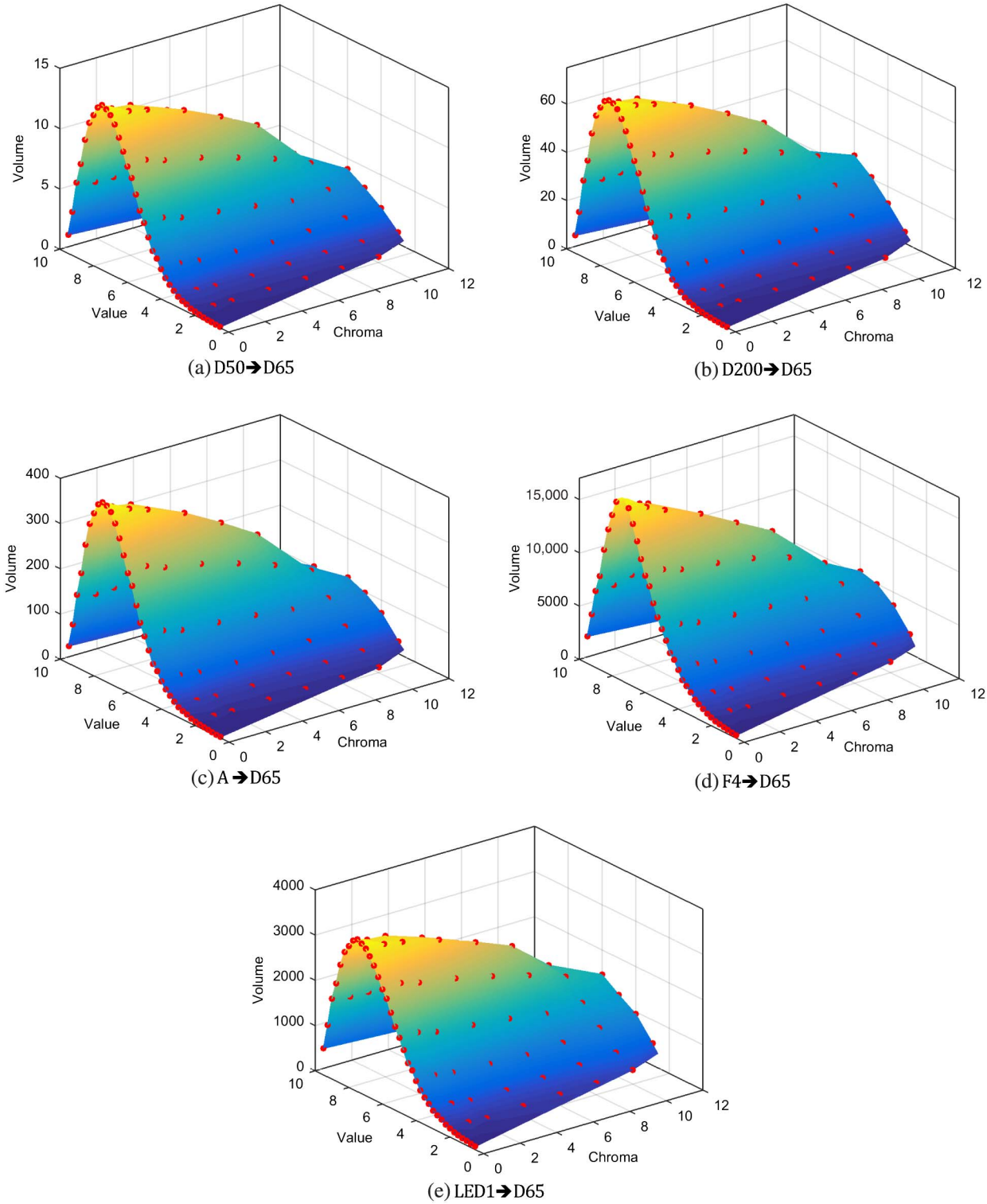


Fig. 3. Average volumes of theoretical metamer mismatch bodies obtained for all Munsell hues plotted as a function of Munsell chroma and value for the illuminant conditions D50, D200, A, F4, and LED1, respectively, changing to D65. Red dots indicate the actual data points. The surface is interpolated through the data points to aid in visualization. The plot colors are those provided by Matlab's "parula" colormap and are provided simply to aid in visualization. They indicate relative magnitude.

typical daylights with different correlated color temperatures; and F4, F8, and F11 are typical fluorescents with varying degrees of spikiness in their spectra. The two LEDs are typical light sources widely used in cellular/mobile phones. The relative spectral power distributions of these illuminants are shown in Fig. 2. In the calculations described below all the illuminants are first normalized so that CIE Y is 100 for the ideal reflector.

3. METAMER MISMATCH BODY VOLUMES

Since the range of possible color signals a given color signal under the first illuminant can become under the second illuminant is only limited by the metamer mismatch body, an interesting question is, How does the volume of the metameric mismatch body vary with the initial color signal? To address this question, we computed both the theoretical and empirical metameric mismatch bodies for each of the 1600 reflectances from the Munsell color atlas for a change from each of the 10 other illuminants (Fig. 2) to D65.

A. Theoretical Metamer Mismatch Body Volumes

Using the method and code from Logvinenko *et al.* [9] we calculated the volumes of the metamer mismatch bodies for the color signals obtained from each of the 1600 Munsell reflectances for the illuminant condition of an illuminant change from D50 to D65 (denoted D50→D65), and similarly each of the conditions D100→D65, D150→D65, D200→D65, F4→D65, F8→D65, F11→D65, A→D65, LED1→D65, and LED2→D65. Figure 3 shows how the theoretical volume varies with the value and chroma of the Munsell samples for the five illuminant conditions D50→D65, D200→D65, A→D65, F4→D65, and LED1→D65. Each red dot is a data point representing the average volume of the metamer mismatch bodies obtained for all hues of the samples having a given Munsell value and chroma.

Comparing the different panels of Fig. 3, it is clear that the overall shape of the plots is similar across all the illuminant conditions. Each plot clearly peaks for the achromatic (i.e., chroma zero) Munsell paper having value 7.5 and then decreases with increasing chroma. The achromatic sample with value 7.5 actually is the neutral gray with approximately constant reflectance of roughly 50%, as shown in Fig. 4. The results are consistent with those of Logvinenko *et al.* [9], showing that the theoretical metamer mismatch body is generally larger for color signals near the center of the object color solid (i.e., where the color signal of ideal 50% reflectance resides) and zero for color signals on the boundary of the object color solid.

The figures also show that the average metamer mismatch body volumes decrease smoothly from neutral gray to the highest chroma samples forming the boundary of the Munsell atlas. Although the plot shapes are qualitatively similar, quantitatively the size of the metamer mismatch bodies depends strongly on the illumination condition. As is evident from Table 1, the lights of similar chromaticity can lead to mismatch bodies of very different sizes, with the size more dependent on the type of light than its chromaticity.

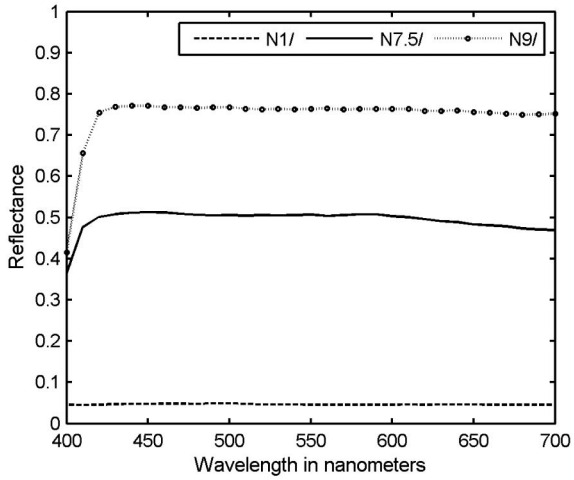


Fig. 4. Spectral reflectances of the neutral gray Munsell papers N 1/, N 7.5/, and N 9/of value 1, 7.5, and 9, respectively. F4:1 F4:2

B. Empirical Metamer Mismatch Body Volumes

The size of the theoretical metamer mismatch bodies shows a very distinct dependence on chroma and value but is based on the limiting case of two-transition reflectance functions. Can we expect similar trends in practice? To investigate this question, we calculated metamer mismatch bodies empirically using the large training set of reflectances described in Section 2.

Although the training set contains 25,303,486 distinct reflectances, it is still limited, and for many color signals there are not enough exact metameric matches to compute a metamer mismatch body reliably. Hence, we relaxed the definition of a metameric match slightly and consider any color signal within a small distance T to be a metameric match. In other words, two CIE XYZ color signals, (X_c, Y_c, Z_c) and (X_i, Y_i, Z_i), will be considered metameric matches whenever

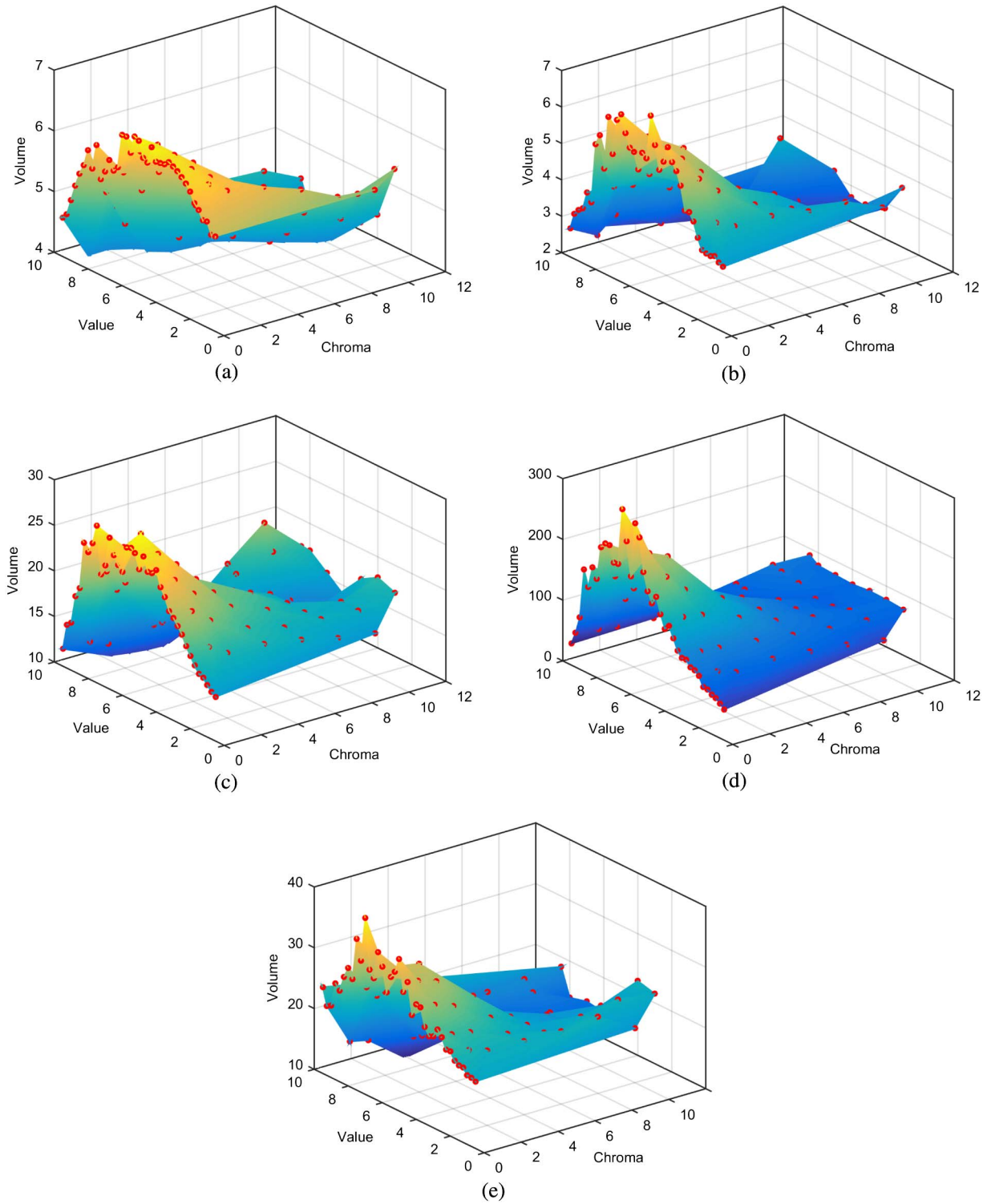
$$\sqrt{(X_i - X_c)^2 + (Y_i - Y_c)^2 + (Z_i - Z_c)^2} < T. \quad \textbf{(1)}$$

Using this definition of metameric matching, given a color signal (X_c, Y_c, Z_c), we find all the reflectances in the training

Table 1. Comparison of the Mean Volumes of the Theoretical Metamer Mismatch Bodies for the 1600 Munsell Samples for a Change from Each of the Different Illuminants to D65[a]

Illuminant Condition	First Illuminant's CIE *xy*	Distance to D65 CIE *xy* = (0.31, 0.33)	Mean Theoretical Volume	
A→D65	(0.45, 0.41)	0.16	143	T1:2
F4→D65	(0.46, 0.42)	0.18	6594	T1:3
LED1→D65	(0.44, 0.41)	0.15	1494	T1:4
D50→D65	(0.35, 0.36)	0.04	5.2	T1:5
F8→D65	(0.36, 0.37)	0.06	63	T1:6
LED2→D65	(0.34, 0.37)	0.05	1999	T1:7
D100→D65	(0.28, 0.29)	0.05	5.6	T1:8
D150→D65	(0.26, 0.27)	0.08	18	T1:9
D200→D65	(0.25, 0.26)	0.09	27	T1:10
F11→D65	(0.40, 0.39)	0.11	9474	T1:11

[a]Lights of similar CIE *xy*-chromaticity are grouped together.



F5:1 **Fig. 5.** Volumes (averaged across all Munsell hues) of the empirical metamer mismatch bodies as a function of Munsell chroma and value for the
F5:2 illuminant conditions (a) D50→D65, (b) D200→D65, (c) A→D65, (d) F4→D65, and (e) LED1→D65. Red dots indicate the actual data points
F5:3 with the surface interpolated through the data points to aid in visualization. The plot colors are as in Fig. 3.

set generating metameric color signals under the first illuminant. Using this set of reflectances, the *empirical metameric mismatch body* is then determined as the convex hull of the color signals generated by these reflectances under the second illuminant.

Threshold T is chosen so that there are enough approximately metameric samples to compute the convex hull of the metamer mismatch body reliably. In particular, T is chosen so that at least 60 approximately-metameric samples are found for 90% of the Munsell samples. The trade-off is that a small

T means more accurate metamers and potentially a more accurate metamer mismatch body, but too restrictive a T leads to too few samples, and hence an inaccurate estimate of the metamer mismatch body. $T = 1$ is used in the calculations reported in this section.

Figure 5 plots how the volume of the empirical metameric mismatch body varies with the chroma and value of the Munsell samples for each of the illuminant conditions D50→D65, D200→D65, A→D65, F4→D65, and LED1→D65. In comparing the empirical volumes depicted in Fig. 5 to the theoretical volumes from Fig. 3 it is clear that the volumes of the empirical metamer mismatch bodies are generally much smaller but show the same general pattern—peaking for the achromatic Munsell paper having value 7.5 and generally decreasing with increasing chroma.

Table 2 compares—for each of 10 illumination conditions—the mean empirical volume to the mean theoretical volume. The mean in each case is taken over all samples in the test set. The table also lists the mean number of approximately metameric ($T = 1$) samples found. For some test samples, too few metamers were found in the training set to compute the empirical metamer mismatch body reliably. In particular, if fewer than 60 metamers were found, then the test sample was excluded from further consideration for the given illumination condition. The numbers excluded in this way are listed in Table 2. The maximum fraction excluded is approximately 22%.

Based on the data from Table 2, Fig. 6 plots the cube root of the empirical volume as a function of the cube root of the theoretical volume, where a linear fit has R-squared 0.90. Although the metamer mismatch volumes tend to be more ellipsoidal than spherical, the cube root provides an approximate measure of a metamer mismatch body's "diameter" since volume varies as diameter cubed. The largest color difference between any two samples in a metamer mismatch body can be expected to relate more closely to the body's diameter than to its volume. The relatively shallow slope (0.15) of the line indicates the "diameter" of the empirical bodies is 15% of that of the corresponding theoretical bodies.

Table 2. Comparison of the Mean Empirical Volumes to the Mean Theoretical Volumes for the 10 Illumination Conditions[a]

	Illumination Condition	Mean Theoretical Volume	Mean Empirical Volume	Mean Number Metamers	Number Samples Excluded
T2:2	A→D65	140	18	45576	112
T2:3	F4→D65	6522	105	65157	66
T2:4	LED1→D65	1459	22	46714	120
T2:5	D50→D65	5	5	28403	235
T2:6	F8→D65	60	7	31101	209
T2:7	LED2→D65	1913	20	31912	198
T2:8	D100→D65	5	3	18978	326
T2:9	D150→D65	16	4	16632	339
T2:10	D200→D65	24	4	15742	349
T2:11	F11→D65	9121	68	34500	167

[a]The table also lists the mean number of (approximate) metamers found within the threshold distance $T = 1$. For some of the 1600 samples in the Munsell set not enough such metamers from the training set could be found to estimate the empirical metamer mismatch body accurately. The right-most column lists the number of Munsell samples excluded based on fewer than 60 metamers being found. Both the mean and theoretical volumes are based on the same subsets of Munsell samples. Lights of similar CIE *xy*-chromaticity are grouped together as in Table 1.

4. COLOR SIGNAL PREDICTION METHODS AND RESULTS

As mentioned above, given a color signal under one illuminant, all that can be definitively determined about what the color signal will become under a second illuminant is that it will lie within the theoretical metamer mismatch body. Of course, if the reflectance that led to the given color signal is known, then the new color signal can be simply calculated. However, in human vision and color imaging the reflectance is not available, and any prediction must be made based on the color signal alone. We describe a new method of making such a prediction based on the properties of the empirical metamer mismatch body and compare it to existing methods of color-signal prediction.

A. Metamer-Based Prediction Method

Mirzaei and Funt [12] proposed a method of color-signal prediction based on relighting a "wraparound Gaussian metamer." Given a color signal under the first illuminant, the idea is to find a Gaussian-like (the precise details are irrelevant for the present discussion) reflectance function producing that same color signal under the first illuminant and then to calculate what that reflectance's color signal would be under the second illuminant. They report excellent results using this Gaussian metamer (GM) method.

The GM method would appear to be limited in that the form of the metameric reflectance is fixed as something Gaussian-like. In comparison, the empirical metamer mismatch body is based on relighting the many reflectances from the training set, producing color signals that are approximately metameric to the given color signal under the first illuminant. The training set also contains only real reflectances, in other

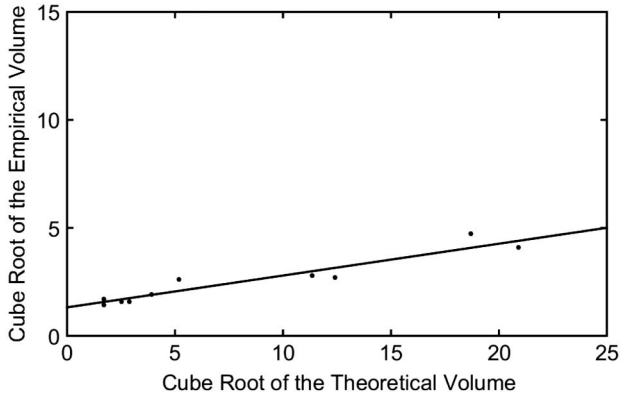


Fig. 6. Comparison across 10 different illumination conditions of the mean of the cube roots of the volumes (i.e., mean of the body "diameters") of the empirical metamer mismatch bodies as a function of mean of the cube roots of the volumes of the theoretical metamer mismatch bodies for the Munsell samples. The linear fit shown has slope of 0.15 with R-squared of 0.90.

words, ones measured in practice rather than an idealized Gaussian-like reflectance function. We hypothesize that basing the color signal prediction on the properties of the empirical metamer mismatch body rather than a single idealized reflectance will lead to more accurate predictions on average.

The empirical metamer mismatch body represents the range of possible color signals that might arise under the second illuminant so the centroid of the body's convex hull is one method of predicting the color signal under the second illuminant. Other choices we investigate are the mean of the possible color signals under the second illuminant, and similarly, their median.

In the previous section, the threshold $T = 1$ was used to ensure there would be enough metameric samples in the training set to estimate the metamer mismatch body accurately. For color signal prediction, we expand the set of reflectances and reduce the threshold to $T = 0.3$. We found the lower threshold resulted in better performance. The smaller threshold means that the samples classed as metamers are closer to being true metamers, and this leads to better predictions. In any case, the results do not depend very strongly on the choice of threshold.

We also expanded the set of reflectances by adding scaled versions of the reflectances in the original dataset because inspection of the dataset revealed that the majority of the reflectances $r(\lambda)$ were "dark," having $\max(r(\lambda)) < 0.5$ for $400\ \text{nm} \le \lambda \le 700\ \text{nm}$. To increase the total number of spectra and to better represent "brighter" reflectances, for each reflectance in the original dataset we added reflectances $r'(\lambda) = 2r(\lambda)$ so long as $\max(r'(\lambda)) \le 1.0$ for $400\ \text{nm} \le \lambda \le 700\ \text{nm}$. The expanded dataset contains 41,941,743 distinct reflectance spectra. Better predictions are obtained with this expanded dataset than the original dataset. In the previous section, we refrained from using the expanded dataset in our analysis of the volumes of the metamer mismatch bodies since there is no guarantee that such scaled reflectances will be found in practice, and we did not want them to affect the volume estimates. In the case of color signal prediction, however, what matters is the accuracy of the resulting predictions.

We also found that outliers were affecting the prediction results. In particular, some of the reflectances in the dataset that were obtained from the multispectral images contained very large spikes at individual wavelengths. Such spikes are likely caused by sensor noise and not from actual scene reflectances. These noise samples lead to outlier CIE XYZ values under the second illuminant that are far from the majority of the samples defining the metamer mismatch body. For color signal prediction, such outliers were removed using the median absolute deviation method [22]. Note that the average empirical volumes reported in the previous section include these spiky spectra since we did not want to prejudge what is and is not a naturally occurring reflectance. In contrast to the situation of color signal prediction where outliers matter because a prediction is made on the basis of a single metamer mismatch body, outliers will have little effect on the final average volumes since the averaging is over thousands of samples.

B. Prediction Results

The spectral reflectances of the Munsell, NCS, UEF Natural, and JPL datasets described above are used for testing. These datasets are distinct from the training set. Predictions are made for a change from illuminants A, D50, D100, D150, D200, F4, F8, F11, and two LEDs to D65 as the "canonical" illuminant. The centroid, mean, and median methods were all tested. The results for all the three methods are comparable, but the centroid method generally outperforms the others, so only the results for it are reported here.

For comparison, the GM method and the von-Kries-based CAT02 chromatic adaptation transform are tested as well. CAT02 is the chromatic adaptation step underlying the CIECAM02 color appearance model. CAT02 includes a spectral sharpening transform [23]. Chong *et al.* [24] propose a tensor-based method of choosing the basis for the diagonal transform. The prediction error is measured using the CIEDE2000 [25] color difference measure.

The results for the full test dataset of reflectances are listed in Table 3. The results indicate that the prediction accuracy of the GM method is higher than that of the CAT02 method, which is consistent with the conclusion of Mirzaei and Funt [12]. It is also clear that the proposed centroid method outperforms both the CAT02 and GM predictions in almost all cases. However, evaluating performance based on the mean and

Table 3. Color Signal Prediction Error of the Three Methods Each Applied to the Combined Set of Test Reflectances and Reported in CIEDE2000 (Mean, Median, 95th Percentile, Standard Deviation) for the 10 Illuminant Conditions[a]

	Centroid				GM				CAT02			
Illuminant Condition	Mean	Median	95th	Stdev	Mean	Median	95th	Stdev	Mean	Median	95th	Stdev
A→D65	1.07	0.76	2.95	0.98	1.40	1.03	3.82	1.17	1.83	1.50	4.48	1.31
F4→D65	1.66	1.22	4.51	1.46	1.77	1.33	4.92	1.52	3.69	2.76	10.32	3.24
LED1→D65	0.99	0.72	2.57	0.96	1.28	0.98	3.82	1.09	1.62	1.25	4.48	1.30
D50→D65	0.31	0.24	0.81	0.27	0.42	0.30	3.82	0.36	0.50	0.44	4.48	0.32
F8→D65	0.48	0.35	1.23	0.51	0.68	0.46	2.02	0.61	0.72	0.64	1.62	0.45
LED2→D65	0.76	0.58	1.87	0.73	1.04	0.89	3.82	0.72	2.08	1.81	4.48	1.48
D100→D65	0.36	0.29	0.88	0.27	0.45	0.31	3.82	0.40	0.54	0.49	4.48	0.34
D150→D65	0.52	0.42	1.31	0.40	0.70	0.47	3.82	0.63	0.85	0.76	4.48	0.54
D200→D65	0.60	0.48	1.50	0.47	0.81	0.54	3.82	0.74	0.99	0.88	4.48	0.63
F11→D65	1.44	1.02	4.10	1.37	1.59	1.12	4.28	1.42	1.57	1.16	4.66	1.50

[a]Lights of similar CIE *xy*-chromaticity are grouped together as in Table 1.

median statistics needs to be done with a note of caution. As Hordley and Finlayson [26] point out, these error distributions are generally not normally distributed because they are bounded by zero on the left. As a further way to evaluate the relative performance, the prediction errors for the three methods are histogrammed in Fig. 7. The distribution again shows that the centroid method outperforms the other two methods.

In predicting what a color signal will become under a second illuminant, we can expect that the larger its metamer mismatch body, the greater the prediction error is likely to be on average—a larger body represents a wider range of possible answers, and any prediction method is forced to choose only one. Figure 8 shows that the prediction error does increase as expected with increasing size of the empirical metamer mismatch body. Plots (not shown) of the error as a function of the size of the theoretical metamer mismatch bodies are qualitatively similar.

5. DISCUSSION

Based on a set of over 25 million spectral reflectances of real objects, estimates of the size of the potential metamer mismatch bodies were computed for the color signals generated from 5,069 test reflectances under 10 different illumination conditions. The average volumes of these empirically-determined bodies were compared to the average volumes of the corresponding theoretically-determined bodies and found to be roughly proportional but significantly smaller. The size of the bodies is important because metamer mismatching imposes a limit on the accuracy with which it is possible to predict the effect a change in illumination will have on a given color signal. The theoretical metamer mismatch body provides an upper limit on the size of a given metamer mismatch body, and the empirical body provides a measure of the lower limit. It is a lower limit since adding a reflectance spectrum to the training

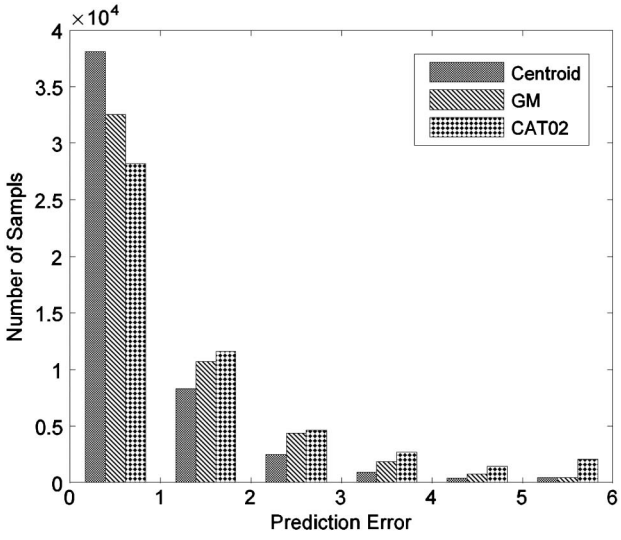


Fig. 7. Histogram of the CIEDE2000 prediction errors for the centroid, GM, and CAT02 methods across the combined set of test reflectances and all 10 illuminant pairs. The height of each bar indicates the number of samples falling within the respective interval, [0,1), [1,2), [2,3), [3,4), [4,5), or [5,∞).

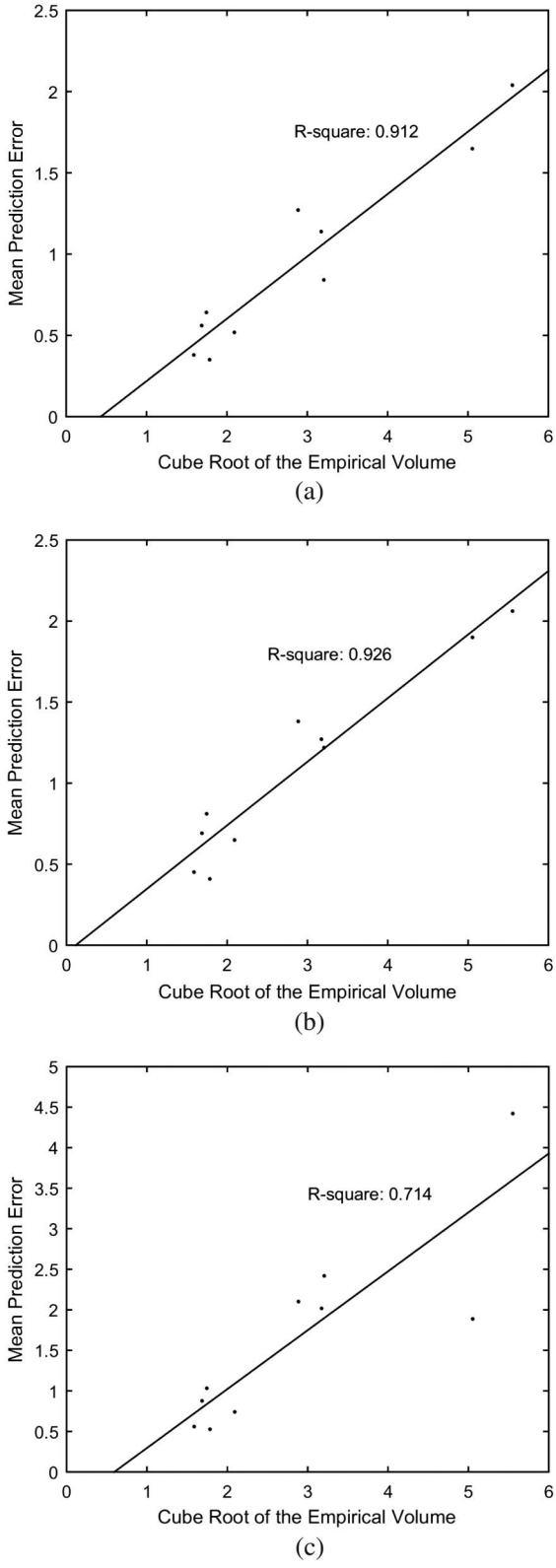


Fig. 8. Mean prediction error in CIEDE2000 units as a function of the cube root of the volume of the empirical metamer mismatch body (i.e., body "diameter") for the three prediction methods: (a) centroid method; (b) GM method; (c) CAT02 method.

set will either lead to a color signal inside the current metamer mismatch body and have no effect, or else lie outside it and therefore increase its size. The empirical bodies also represent a lower limit since the reflectance data is based on a 10 nm sampling interval using samples having a roughly 10 nm bandwidth [2]. In effect, the measured spectra are smoothed versions of the real spectra, and any such smoothing will potentially reduce the calculated amount of metamer mismatching. Since it is real spectral power distributions that enter the eye, the empirical metamer mismatch bodies reported here are likely to underestimate the true amount of metamer mismatching to a certain extent.

As a general rule, the volumes of the metamer mismatch bodies (both empirical and theoretical) were found to decrease with increasing distance of the color signal from mid-gray, as can be seen in Fig. 5. Concomitantly, the average error in predicting how a given color signal will change with a change in illumination was also found to decrease with increasing distance for the color signal from mid-gray. In terms of predicting what a given color signal may change to under a new illuminant, the centroid of the empirical metamer mismatch body performs better overall than the GM method, which in turn outperforms CAT02.

Funding. Natural Sciences and Engineering Research Council of Canada (NSERC)National Natural Science Foundation of China (NSFC) (61205168); National Science and Technology Support Program of China (2012BAH91F03).

Acknowledgment. The authors would also like to thank the anonymous reviewers for their exceptionally pertinent and constructive critical reviews.

REFERENCES

1. G. Wyszecki and W. S. Stiles, *Color science: concepts and methods, quantitative data and formulae* (Academic, 1982).
2. D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," J. Opt. Soc. Am. A **23**, 2359–2372 (2006).
3. G. Y. Feng and D. H. Foster, "Predicting frequency of metamerism in natural scenes by entropy of colors," J. Opt. Soc. Am. A **29**, A200–A208 (2012).
4. P. Morovic and H. Haneishi, "Quantitative analysis of metamerism for multispectral image capture," in *Proceedings of 9th International Symposium on Multispectral Color Science* (Academic, 2007), pp. 88–96.
5. D. K. Prasad and L. Wenhe, "Metrics and statistics of frequency of occurrence of metamerism in consumer cameras for natural scenes," J. Opt. Soc. Am. A **32**, 1390–1402 (2015).
6. A. D. Logvinenko, B. Funt, H. Mirzaei, and R. Tokunaga, "Rethinking color constancy," PLOS ONE **10**, e0135029 (2015).
7. A. D. Logvinenko, "Object-color manifold," Int. J. Comput. Vis. **101**, 143–160 (2013).
8. X. Zhang, B. Funt, and H. Mirzaei, "Metamer mismatching and its consequences for predicting how colors are affected by the illuminant," in *Proceedings of IEEE International Conference on Computer Vision Workshops* (IEEE, 2015).
9. A. D. Logvinenko, B. Funt, and C. Godau, "Metamer mismatching," IEEE Trans. Image Process. **23**, 34–43 (2014).
10. M. D. Fairchild, *Color Appearance Models* (Academic, 2013).
11. "A color appearance model for color management systems: CIECAM02," CIE Publication No. 159 (CIE Central Bureau, 2004).
12. H. Mirzaei and B. Funt, "Object-color-signal prediction using wrap-around Gaussian metamers," J. Opt. Soc. Am. A **31**, 1680–1687 (2014).
13. F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: post-capture control of resolution, dynamic range and spectrum," http://www.cs.columbia.edu/CAVE/databases/multispectral/.
14. Joensuu Spectral Image Database, "Spectral color research group," University of Eastern Finland, http://www.uef.fi/fi/spectral/spectral-database.
15. S. Moan, S. George, M. Pedersen, J. Blahova, and J. Hardeberg, "A database for spectral image quality," Proc. SPIE **9396**, 93960P (2015).
16. S. Hordley, G. Finlayson, and P. Morovic, "A multi-spectral image database and an application to image rendering across illumination," in *Proceedings of Third International Conference on Image and Graphics* (Academic, 2004). http://www2.cmp.uea.ac.uk/Research/compvis/MultiSpectralDB.htm.
17. C. Li, M. R. Luo, M. R. Pointer, and P. Green, "Comparison of real color gamuts using a new reflectance database," Color Res Appl. **39**, 442-451 (2014).
18. *Munsell Book of Color—Glossy Edition* (X-Rite Corporation).
19. A. Hard and L. Sivik, "NCS–natural color system: a Swedish standard for color notation," Color Res. Appl. **6**, 129–138 (1981).
20. J. Parkkinen, T. Jaaskelainen, and M. Kuittinen, "Spectral representation of color images," in *Proceedings of IEEE 9th International Conference on Pattern Recognition* (IEEE, 1988), pp. 14–17, http://www2.uef.fi/fi/spectral/natural-colors.
21. A. M. Baldridge, S. J. Hook, C. I. Grove, and G. Rivera, "The ASTER spectral library," Version 2.0, Remote Sensing Environ. **113**, 711–715 (2009).
22. C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median," J. Exp. Soc. Psychol. **49**, 764–766 (2013).
23. G. Finlayson, M. Drew, and B. Funt, "Spectral sharpening: sensor transformations for improved color constancy," J. Opt. Soc. Am. A **11**, 1553–1563 (1994).
24. H. Chong, S. Gortler, and T. Zickler, "The von Kries hypothesis and a basis for color constancy," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 2007), pp. 1–8.
25. "Improvement to industrial color-difference evaluation," CIE Publication No. 142 (CIE Central Bureau, 2001).
26. S. D. Hordley and G. D. Finlayson, "Reevaluation of color constancy algorithm performance," J. Opt. Soc. Am. A **23**, 1008–1020 (2006).

Queries

1. The funding information for this article has been generated using the information you provided to OSA at the time of article submission. Please check it carefully. If any information needs to be corrected or added, please provide the full name of the funding organization/institution as provided in the FundRef Registry (http://www.crossref.org/fundref/fundref_registry.html).

2. AU: Please provide complete details for Ref. [18].