Comparing Measures of Gamut Area

Michael P Royer[1]

[1]Pacific Northwest National Laboratory
620 SW 5th Avenue, Suite 810
Portland, OR 97204
michael.royer@pnnl.gov

This is an archival copy of an article published in *LEUKOS*. Please cite as:

Royer MP. 2018. Comparing Measures of Gamut Area. *LEUKOS*, 15:1, 29-53, DOI: 10.1080/15502724.2018.1500485.

PNNL-SA-132222 ▪ November 2018

Abstract

This article examines how the color sample set, color space, and other calculation elements influence the quantification of gamut area. The IES TM-30-18 Gamut Index (R_g) serves as a baseline, with comparisons made to several other measures documented in scientific literature and 12 new measures formulated for this analysis using various components of existing measures. The results demonstrate that changes in the color sample set, color space, and calculation procedure can all lead to substantial differences in light source performance characterizations.

It is impossible to determine the relative "accuracy" of any given measure outright, because gamut area is not directly correlated with any subjective quality of an illuminated environment. However, the utility of different approaches was considered based on the merits of individual components of the gamut area calculation and based on the ability of a measure to provide useful information within a complete system for evaluating color rendition. For gamut area measures, it is important to have a reasonably uniform distribution of color samples (or averaged coordinates) across hue angle—avoiding exclusive use of high-chroma samples—with sufficient quantity to ensure robustness but enough difference to avoid incidents of the hue-angle order of the samples varying between the test and reference conditions. It is also important to use a modern, uniform color space that is suitable for the quantification of color appearance and color difference.

1 Introduction

Over the past half century, gamut area has been the main alternative or counterpart to average color fidelity, which historically has been the primary method for quantifying color rendition. Average color fidelity measures are based on the average magnitude of difference in color appearance for a set of spectral reflectance functions ("color samples") as rendered by a test light source and reference illuminant. In contrast, gamut area measures are based on the polygonal area enclosed by a set of coordinates in a chromaticity diagram or color space (and also typically compared to a reference illuminant).

Both average color fidelity and gamut area are frequently *believed* to represent subjective experiences. Average color fidelity is often thought of as a characterization of how normal or natural a scene may appear. Gamut area measures have traditionally been intended to address color saturation or vividness, which subsequently has tied them to the concepts of color discrimination and color preference. It has been suggested that this dichotomy makes average color fidelity and gamut area important complements [Houser and others 2013; Rea and Freyssinier-Nova 2008; Rea and Freyssinier 2010; 2013], and in fact the two types of measures cannot be simultaneously maximized. Today, these associations between objective characterizations of color differences and subjective evaluations of illuminated environments are being challenged as new methods for characterizing color rendition are developed and new experimental evidence emerges [Esposito and Houser 2018; Royer and others 2017a; Royer and others 2017b; Royer and others 2016; Wei and others 2018; Zhang and others 2017].

Despite many efforts to introduce measures of gamut area—which are reviewed in Section 2—there has been little published literature exploring the functional elements of gamut area calculations. That is, there has been no formal discussion of how color samples, color spaces, averaging methods, or calculation procedures affect the quantification, although there has been some comparisons of existing measures having varying focus and depth [Houser and others 2013; Khanh and others 2016c; Smet and others 2011; Windisch and others 2017]. Further, unlike average color fidelity, there has been little discussion of the inherent limitations of measures of gamut area. This article focuses on comparing past and present measures of gamut area, as well as other closely related measures, such as gamut volume and average chroma shift. It covers the benefits and limitations of gamut area measures, when used alone or in combination with a measure of average color fidelity. Discussion of the association between gamut area and subjective evaluations is also included. A similar analysis of color fidelity measures was recently published [Royer 2017].

The Gamut Index (R_g) of the American National Standards Institute (ANSI) and Illuminating Engineering Society (IES) TM-30 method [IES 2015; 2018] is the only measure of gamut area formalized by a lighting organization responsible for standards, and as such it serves as a baseline for this analysis. The recently revised 2018 version of TM-30 is used, although there is no material difference for R_g in the 2015 and 2018 versions. The 2018 revision unifies the calculation framework of TM-30 with that of CIE 224 [CIE 2017]. IES R_g is compared to a variety of previously proposed measures of gamut area, as well as new formulations created specifically for this analysis that combine various elements of existing measures, such as the color samples, color space, or reference illuminant scheme. By mixing and matching, it is possible to investigate the key elements of gamut area measures and how they influence the resulting characterization. Undoubtedly, there are innumerable ways to construct a gamut area measure, and not all possible ways can be covered in one article.

2 Gamut Area Measures

At least 12 measures related to gamut area have been proposed and documented in scientific literature over the past 45 years, as documented in **Table 1**. The additional measures created specifically for this analysis are listed in **Table 2**.

Table 1. Previously-documented measures related to gamut area, gamut volume, or chroma shift. Citations provided in text.

Year	Measure and Abbreviation	Color Samples	Color Space/ Chromaticity Diagram	Reference Illuminant
2015, 2018	Gamut Index (R_g)	99/16 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
2017	Gamut Volume Index (GVI)	14	CIE LAB	None
2016	Relative Gamut Area Index (RGAI)	8 TCS [CIE 13.3-1995]	CAM02-UCS	Relative Planckian/CIE D Series
2016	Relative Gamut Area Index (G_a)	8 TCS [CIE 13.3-1995]	CIE 1964 U*V*W*	Relative Planckian/CIE D Series
2016	ΔC*	Varies	CIE LAB	Relative Planckian/CIE D Series
2010	Gamut Area Scale (Q_g)	15 VS	CIE LAB	Relative Planckian/CIE D Series
2009	Color Saturation Index (CSI)	1269 Munsell	None/MacAdam	Relative Planckian/CIE D Series
2008	Gamut Area Index (GAI)	8 TCS [CIE 13.3-1995]	CIE 1976 UCS	Fixed CIE Illuminant E
1993, 2007	Feelings of Contrast Index (FCI)	4	CIE LAB or CIE CAM02	Fixed CIE D65
1997	Cone Surface Area (CSA)	8 TCS [CIE 13.3-1995], *w*'	CIE 1976 UCS	None
1984, 1993	Color Rendering Capacity (CRC)	Theoretical All Colors	CIE LUV	None
1972	Color Discrimination Index (CDI)	8 TCS [CIE 13.3-1995]	CIE 1960 UCS	Fixed CIE Illuminant C

2.1 Gamut Area Measures Based on the Color Samples of the CIE Test Color Method

One key family of existing gamut area measures is the five that rely on the eight test color samples (TCS) used to calculate the Commission Internationale de l'Eclairage (CIE) General Color Rendering Index R_a (colloquially, CRI) [CIE 1995]. They differ from each other by employing different color spaces and reference illuminants.

The first of the TCS-based measures, the Color Discrimination Index (CDI), was proposed by Thornton in the early 1970s [1972; 1973]. It uses the CIE U*V*W* color space and CIE Illuminant C (a daylight simulator) is the reference illuminant for all test light sources, regardless of correlated color temperature (CCT). (Note:

Table 2. Newly derived measures of gamut area or gamut volume.

Type	Abbreviation	Color Samples	Color Space/ Chromaticity Diagram	Reference Illuminant
Gamut Area	R_g4	4 of FCI	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Area	R_g8	8 TCS [CIE 13.13-1995]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Area	R_g14	14 of GVI	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Area	R_g15	15 of CQS	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Area	G_{LAB}	8 TCS [CIE 13.13-1995]	CIE LAB	Relative Planckian/CIE D Series
Gamut Area	GAI_{rel}	8 TCS [CIE 13.13-1995]	CIE 1976 UCS	Relative Planckian/CIE D Series
Gamut Area	R_g99A	99 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Area	R_g99B	99 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Area	R_g99C	99 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Area	R_g4880	4880/16 TM-30 Reference Set	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Volume	R_gV	99 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Gamut Volume	R_g14V	14 of GVI	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Chroma Change	ΔC_{99A}	99 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Chroma Change	ΔC_{99B}	99 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series
Chroma Change	ΔC_{16}	99/16 CES [IES TM-30-18]	CAM02-UCS	Relative Planckian/Mixed/CIE D Series

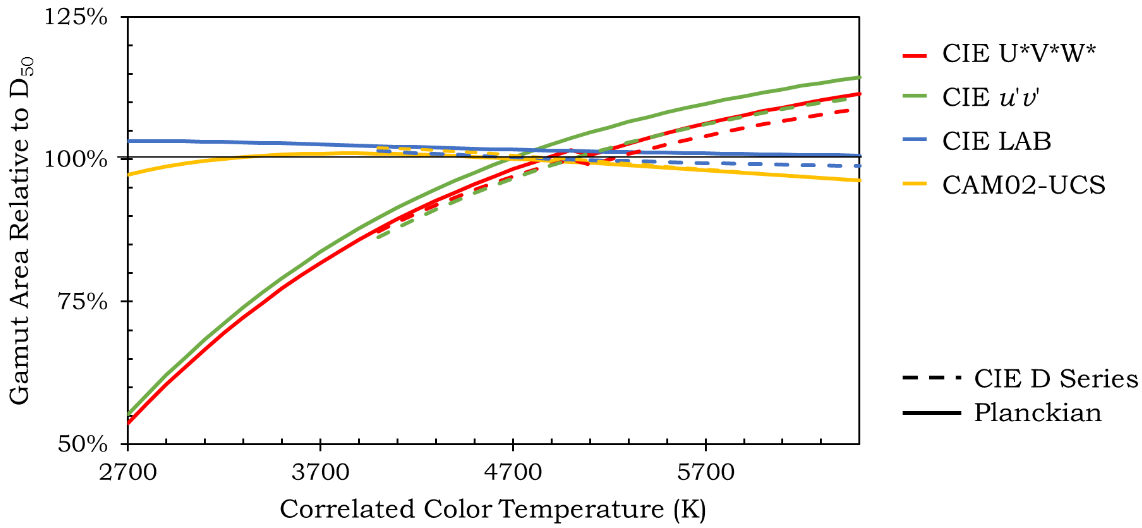

Figure 1. Comparison of the gamut area of Planckian radiator or a CIE D Series illuminant at the specified CCT, relative to the area of CIE D50, for four different gamut area measures based on the 8 TCS and reference illuminant scheme of CIE 13.3-1995.

Color spaces and standard illuminants are specified in [CIE 2004].) CDI received some attention in the research community [Boyce 1977], but perhaps due to a relative lack of variety of CDI values for then-available light sources, it did not become a mainstay of the broader lighting community.

The concept was revisited in 2008, when Rea and Freyssinier-Nova [2008] proposed the Gamut Area Index (GAI). GAI is very similar to CDI, but it is calculated in the CIE 1976 *u*'*v*' chromaticity diagram and uses CIE Illuminant E (the equal energy illuminant) as the reference for all test sources. Importantly, Rea and Freyssinier-Nova envisioned GAI as a complement to CIE R_a, suggesting the pair was capable of distinguishing preferred light sources, whereas Thornton identified CDI as a distinct measure characterizing color discrimination.

As expressed by Houser and colleagues [2013], CDI and GAI are differentiated by scale differences due to variation in the color space and reference illuminant, but they are perfectly correlated with one another. Given the use of a fixed reference illuminant and the specified color space, both measures have an inherent correlation with CCT, as shown in **Figure 1**. Low-CCT sources (e.g., 2700 K, 3000 K) have substantially lower CDI/GAI scores than sources with a CCT closer to the respective reference illuminant.

In 2016, Teunissen and colleagues proposed the Relative Gamut Area Index (G_a), which is the same as CDI other than using the relative-reference scheme of CIE R_a—a combination of Planckian radiation or a D Series illuminant depending on the CCT of the test source. This eliminates the CCT-dependence, while retaining the use of CIE U*V*W* and 8 TCS.

Another measure called Relative Gamut Area Index, but denoted RGAI, was also considered by Teunissen and colleagues [2016]. RGAI uses CAM02-UCS [Luo and others 2006] instead of CIE U*V*W*, but is otherwise equivalent to G_a. This concept has been modified for this article as the measure R_g8, which uses the IES TM-30-18 and CIE 224 reference scheme [CIE 2017; IES 2018; Royer 2016] instead of the CIE R_a reference scheme. This allows for a comparison to IES R_g that isolates the effect of the color samples.

One other measure based on the eight TCS took a different approach. Cone Surface Area (CSA) [Fotios and Levermore 1998], is determined by calculating the area of a cone, with the coordinates of the eight TCS in the 1976 *u*'*v*' chromaticity diagram comprising the base and the chromaticity of the light source (*w*') forming the point. This makes CSA a three dimensional calculation, although it is not a volume. No reference illuminant is used. CSA has been discussed in past reviews [Guo and Houser 2004; Houser and others 2013], and is not considered further here.

2.2 Other Gamut Area Measures

While gamut area measures based on the eight TCS used to calculate CIE R_a have proliferated, other gamut area measures have also been introduced. Perhaps the first was Color Rendering Capacity (CRC), which was introduced in 1984 [Xu 1984]. It is based on the theoretical volume of a color solid, originally in the CIE 1960 uvY color space. In 1993, it was updated to use the CIE LUV color space and normalized so that CIE Illuminant E had a value of 1.0 [Xu 1993]. CRC does not rely on color samples, but instead of the range of possible tristimulus values. It is not included in later comparisons.

The Feelings of Contrast Index (FCI) was first proposed in 1994 as a measure of visual clarity [Hashimoto and Nayatani 1994], and later updated and promoted as a measure of color preference [Hashimoto and others 2007; Windisch and others 2017]. In its latest form, FCI is determined from the areas of two triangles formed by coordinates of four test color samples in CIE CAM02 or CIE LAB, the ratio of the two areas is also scaled with an exponent, 1.5. Hashimoto and others [2007] found the two color spaces to produce equivalent results for a set of 20 SPDs. With the triangle-area methodology, FCI attempts to include all three dimensions of color space, rather than the projected area of a set of color samples onto a plane. The reference illuminant for FCI is fixed as CIE D_{65}. An alternative measure created for this analysis, R_g4, uses the four color samples of FCI but follows the methodology of IES TM-30-18 R_g (excluding hue-angle binning).

In 2010, Ohno and Davis published an article documenting the Color Quality Scale (CQS), which includes a gamut area measure (Q_g). All of the measures included in CQS rely on a set of 15 color samples, denoted VS, which are of higher chroma than the samples used to calculate CIE R_a, though they are still from the Munsell system and thus have a similar aggregate spectral sensitivity profile. The samples are approximately evenly spaced by hue angle in CIE LAB, which is used for CQS calculations. Like the rest of CQS, Q_g relies on a relative reference scheme that is the same as that used by CIE R_a. As with FCI, the color samples of CQS can be used to calculate a gamut area measure that otherwise follows the procedures of IES TM-30-18 R_g (excluding hue-angle binning); this measure is reported here as R_g15.

IES TM-30-15 was developed beginning in 2014, and revised in 2017 as IES TM-30-18. Given the prevalence of gamut area measures in scientific literature, including a measure of gamut area in IES TM-30 was a priority during its initial development. The rationale behind the procedure for calculating the IES Gamut Index, R_g, is covered by David and colleagues [2015] and Royer and colleagues [2017]. The most unique feature of R_g is that it relies on 16 average coordinates pairs instead of specific color samples. The coordinates are found via a process that begins by dividing the *a*'-*b*' plane of CAM02-UCS into hue-angle bins, each of which subtends 22.5° of the plane. Each hue-angle bin contains a portion of the 99 total color evaluation samples (CES) used in the method, varying somewhat based on the reference illuminant. The 16 coordinate pairs are the averages of the *a*' and *b*' coordinates for the color samples determined to be in each hue-angle bin. Other alternatives, such as calculating the area or volume enclosed by all 99 CES were considered at the time, and are compared to the specified method in this article.

2.3 Gamut Area Alternatives

The concept of gamut area has seen considerable attention, but at least two other related concepts have been proposed: average change in chroma and gamut volume. These both rely on modern object color spaces, which were not formally available when gamut area was first conceptualized. Average change in chroma (ΔC) can be calculated for any set of color samples (or any single color sample) as long as a color space intended to characterize object colors is used—that is, a color space that has a chroma correlate, such as CIE LAB or CIE CAM02. Only a couple of measures based on average change in chroma have been explicitly documented. Khanh and colleagues [2016; 2016a; 2016b] explored ΔC measures based on various color sample sets, including samples specific to what was being illuminated. They found these measures useful for describing perceptual ratings. Royer and Wei [2017] described how average change in chroma could be calculated within the IES TM-30-15 system (ΔC_{CES}), and compared the measure to IES TM-30-15 R_g using a set of approximately 400 SPDs. While the measures were reasonably well correlated, there were substantial differences in some cases, especially for sources that induce a relatively greater amount of hue shift.

One unique approach to characterizing chroma shifts is the Color Saturation Index (CSI) [Zukauskas and others 2009]. Within a larger framework, CSI defines the percentages of color samples for which chroma exceeds a three-step MacAdam ellipse, without regard to the actual difference in chroma. Thus, it is not expected to be closely correlated with gamut area over a large set of SPDs, and is not considered further in this analysis.

Another concept directly related to gamut area is gamut volume. Whereas gamut area focuses on changes in the hue-chroma plane, gamut volume includes changes in the lightness dimension of color space. One measure of gamut volume has been documented, the Gamut Volume Index (GVI) [Liu and others 2017]. It is uses CIE LAB color space and 14 color samples selected so that the measure was correlated with preference data from previous subjective evaluations. A gamut volume measure based on the 99 CES of IES TM-30-18 was calculated for this article, denoted R_gV. It utilizes CAM02-UCS and the same quick-hull algorithm used to determine GVI. Unlike GVI, it is based on the relative reference scheme of IES TM-30-18/CIE 224. R_g14V is a more direct alternative to GVI that uses the same color sample set but utilizes CAM02-UCS and a relative reference scheme.

3 Results: Comparing Measures

To get a complete picture of the differences between color rendition measures, a large set of SPDs with varied attributes—including average color fidelity, gamut area, and gamut shape—is necessary. Including a wide variety of SPDs helps to identify systematic patterns, which are not always present in smaller sets of SPDs focused on commercially-available products. This is further described and demonstrated by Royer [2017]. A set of 15,806 SPDs featuring 319 commercially-available, experimental, modelled, and theoretical light sources from the TM-30-18 Calculator Tool Example Library, 80 additional experimental LED sources [Royer and others 2016, Royer and others 2017], 825 real and theoretical SPDs collected by CIE R1-62, and 14,582 randomly-generated theoretical SPDs is used throughout this analysis. The set used in this article includes 10,000 new theoretical SPDs, comprised of five randomly-generated Gaussian components with a full-width-half-maximum (FWHM) between 2 and 120 nm, such that the resulting combination had a nominal CCT between 2700 K and 6500 K and D_{uv} between -0.018 and 0.006. All SPDs and calculated values are available as in a supplemental file that is linked to this article.

For the figures in this article, these SPDs are classified as commercial (n = 212), real (n = 806), or theoretical (n = 14,788). For clarity, many of the figures in this article show the gamut area range of 50 to 150. This includes approximately 14,500 of the SPDs, depending on the specific measures being compared.

3.1 Varying Color Sample Set

One of the attributes that gets a lot of attention in color rendition measures is the set of color samples used for evaluation. Typically, color sample sets are intended to mimic the color (spectral reflectance) of objects in architectural environments, either directly or in the values of the resulting measures. However, there is no explicit way to rate the "accuracy" of a color sample set—the best color sample set is the one most similar to a given set of real objects. For general use in rating the performance of light sources, a sufficiently large set with neutral spectral sensitivity, such as that used in IES TM-30 and CIE 224, is a logical choice [Royer 2017; Smet and others 2015]. For reference, **Figure 2** shows the CAM02-UCS (a', b') coordinates for five different color sample sets illuminated by CIE D_{50}: The 99 CES and 16 hue-angle-bin averages of IES TM-30, the 4 samples used to calculate FCI, the 8 samples used to calculate CIE R_a, the 14 samples used to calculate GVI, and the 15

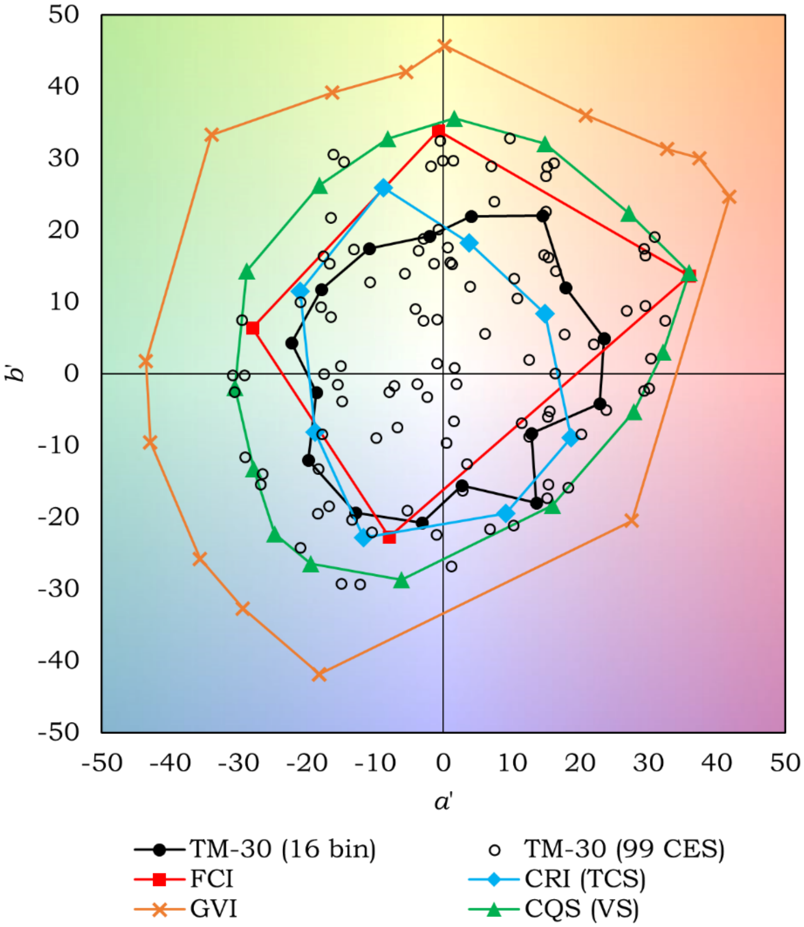

Figure 2. Comparison of (a', b') CAM02-UCS coordinates for five color sample sets. The illuminant is CIE D_{50}.

samples used to calculate CQS Q_g. Note that some of these color sample sets are not being used here as intended; in particular, the color sample set for GVI was not chosen with gamut area calculations in mind.

Figure 3 compares IES R_g to four alternatives: R_g4, R_g8, R_g14 and R_g15. Aside from the different color samples, all of the calculation procedures are the same for these measures, except that—by necessity—the alternatives do not use hue-angle binning to average coordinates. The differences between the measures are readily apparent. In all cases, the middle 95% of differences (the 2.5 to 97.5 percentile) exceed 10 points, with R_g4 being substantially more different from IES R_g than the others. Linear regression analysis may offer a slightly different impression, although it ignores some of the systematic differences that contribute to scale differences. In all cases the coefficient of determination (r^2) is greater than 0.85. The SPD-specific random variation (the residuals for a linear trend line fit to each comparison) shown in the charts in **Figure 3** tends to decrease as the quantity of samples in the alternative set increases. For R_g4 compared to IES TM-30-18 R_g, the mean squared error (MSE) for all 15,806 SPDs is 83.0 points. For R_g8, R_g14, and R_g15, the MSEs are approximately 11.0, 14.5, and 4.4 points, respectively. If only real (commercial and experimental) SPDs are considered—which excludes extreme values—the corresponding MSEs are 18.3, 10.1, 4.9 and 0.9 points, respectively. It is logical that measures derived from fewer samples have higher MSEs than measures derived from larger color sample sets, as a shift for a given color sample has more influence if it is part of a smaller color sample set. Further, small color sample sets cannot accurately detect the patterns of color shift in the three dimensions of color space. This has been explored by others [David 2013; David and others 2018; Smet and others 2015].

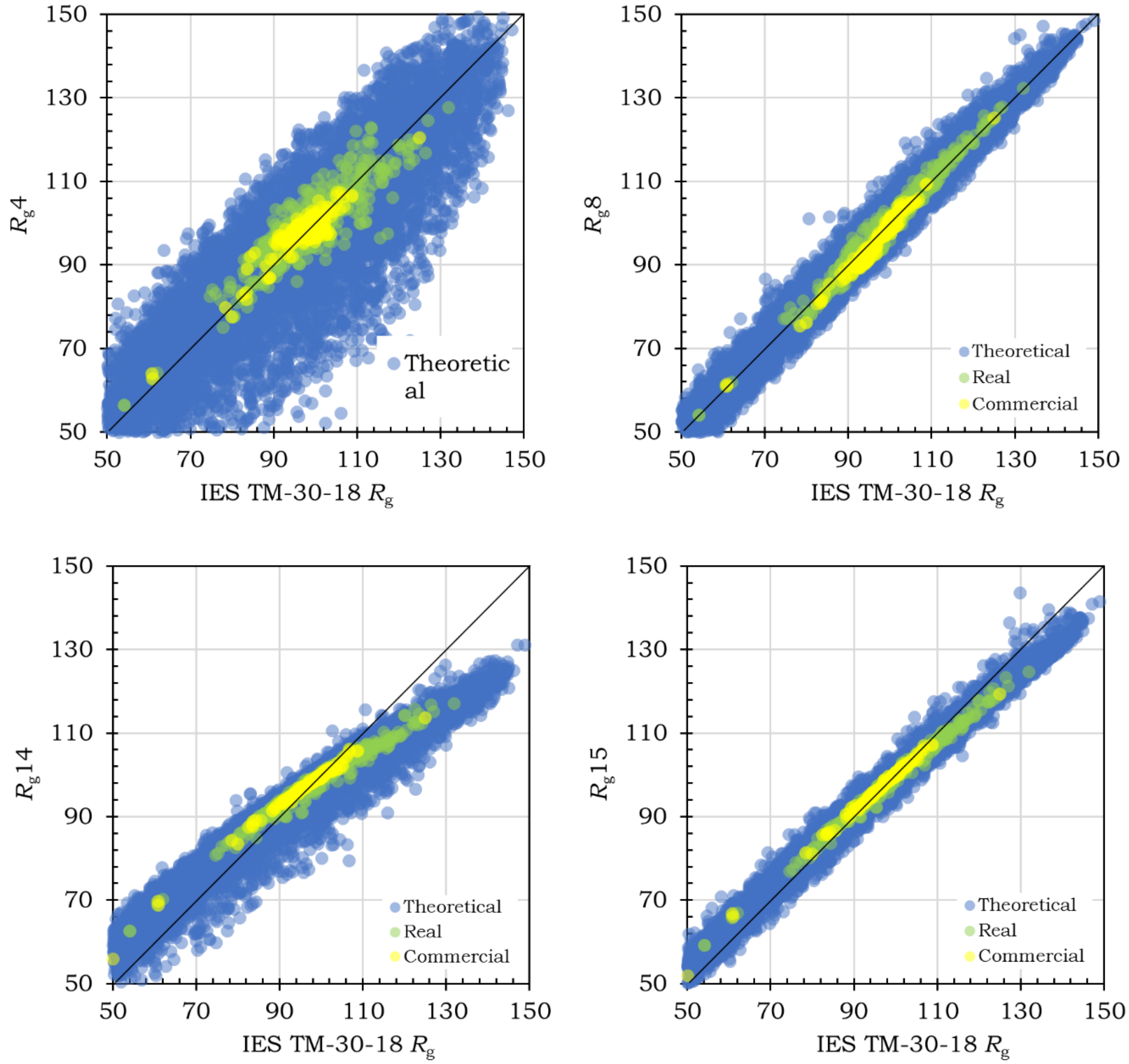

Figure 3. Comparison of four alternative gamut area measures using different color sample sets to IES TM-30-18 R_g. See Table 2 for a description of each alternative measure.

Another consideration is that the color sample sets with very high chroma tend to lead to lower values than IES R_g when gamut area values are very high, resulting in systematic variation. This is particularly noticeable for R_g14 versus IES R_g, and to a lesser extent for R_g15 versus IES R_g. Chroma cannot be increased indefinitely, so high-chroma color sample sets may result in a compression in gamut area values as the limits of the color volume are approached. This phenomenon is not likely representative of typical objects in architectural interiors. The 99 CES used in IES TM-30 include samples that reach similar chroma levels to the 15 VS used for R_g15. However, because they are averaged with color samples having lower chroma values, compression does not appear to be an issue; if it were, "decompression" would be seen in the comparison of R_g8 and IES R_g.

The systematic variation is also related to gamut shape [Royer and others 2017a], which may occur because of different sample-set spectral-sensitivity profiles [David and others 2015; Royer 2017; Royer and Wei 2017] and/or due to the chroma level of the color samples. This is illustrated in **Figure 4**, which compares the difference between IES R_g and the four alternative gamut area measures to Local Chroma Shift in hue-angle

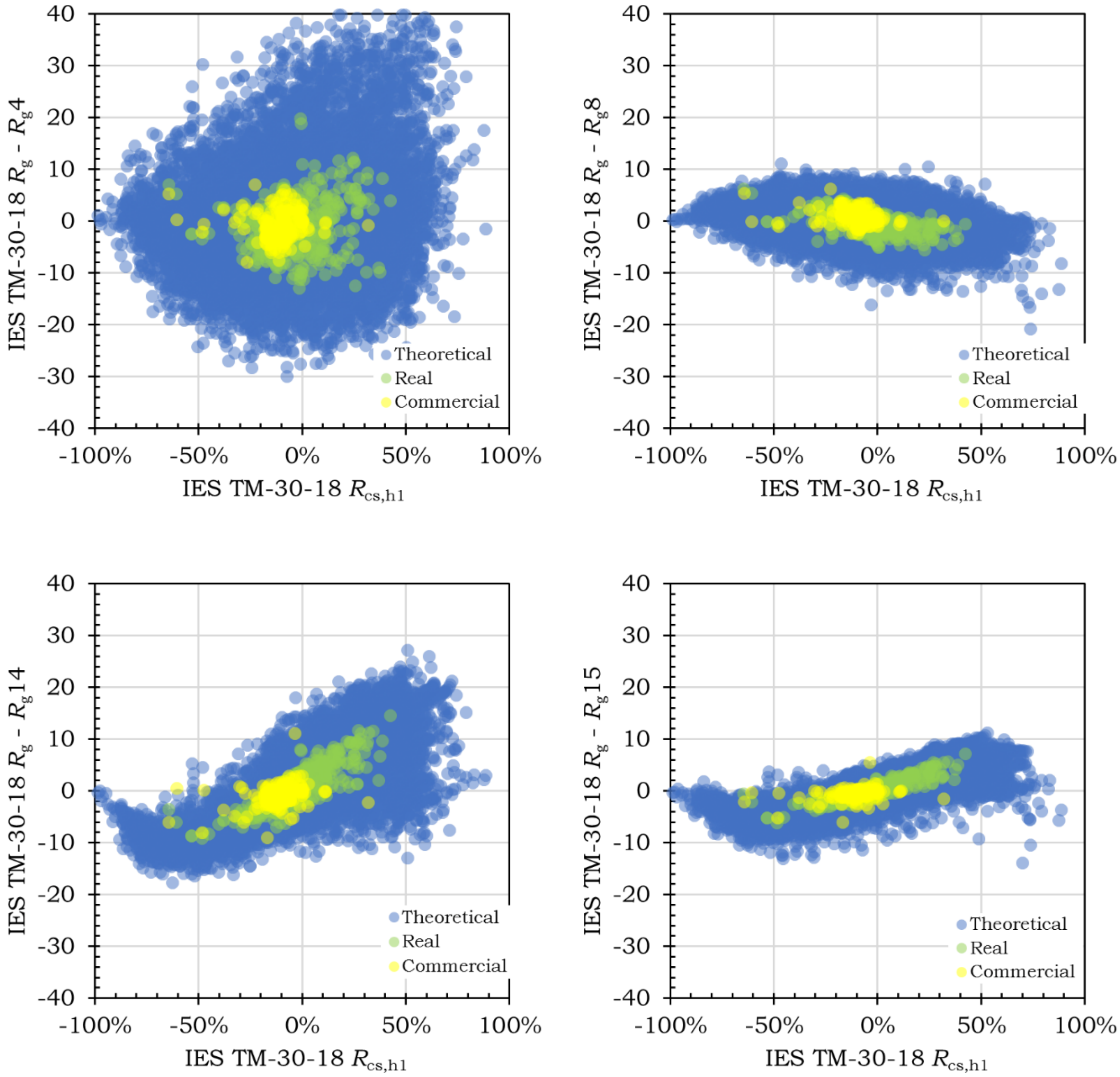

Figure 4. The difference between IES TM-30-18 R_g and alternative gamut area measures versus IES TM-30-18 $R_{cs,h1}$ (red chroma shift).

bin 1 (IES $R_{cs,h1}$). IES $R_{cs,h1}$ was chosen for this comparison to align with the subsequent analysis on color spaces, where non-uniformity in the red region is a documented issue. Other hue-angle bins can be examined in the supplemental data associated with this article. As shown, each color sample set has a unique relationship to the baseline:

- For R_g4 compared to IES R_g, any relationship with IES $R_{cs,h1}$ is obscured by the large amount of random, SPD-specific error.
- The difference between IES R_g and R_g8 is slightly more likely to be negative as IES $R_{cs,h1}$ increases—that is, the 8 TCS tend to lead to slightly higher gamut area values than the 99 CES as IES $R_{cs,h1}$ increases. This may occur because the chroma of TCS1 (red) is relatively lower than the other TCS and the coordinates used to calculate IES R_g (**Figure 2**).

- In contrast, R_g14 values tend to be lower than IES R_g values as IES $R_{cs,h1}$ increases, which may be linked to the uneven distribution of the 14 samples in hue space—there is only one sample in the positive a', negative b' quadrant (**Figure 2**)—and/or the high chroma levels of the color samples.
- R_g15 follows a similar trend as R_g14, but the slope of the linear relationship is much smaller.

An important factor in these relationships is that large increases in chroma (and thus gamut area) typically occur when red (and green) chroma is increased [Royer and others 2017a]. Therefore, the difference between IES R_g and R_g14 or R_g15 in relationship to IES $R_{cs,h1}$ may be directly related to the compression issue. No trends are present when examining IES $R_{cs,h5}$ (nominally yellow chroma shift) instead of IES $R_{cs,h1}$.

3.2 Varying Color Space

The influence of color spaces (or color appearance models) on gamut area values can be examined by comparing G_a (CIE 1964 U*V*W*), a relative-reference version of GAI (GAI_{rel}, CIE 1976 $u'v'$), newly-defined G_{LAB} (CIE LAB), and RGAI (CAM02-UCS), as shown in **Figure 5**. It is not possible to make this comparison

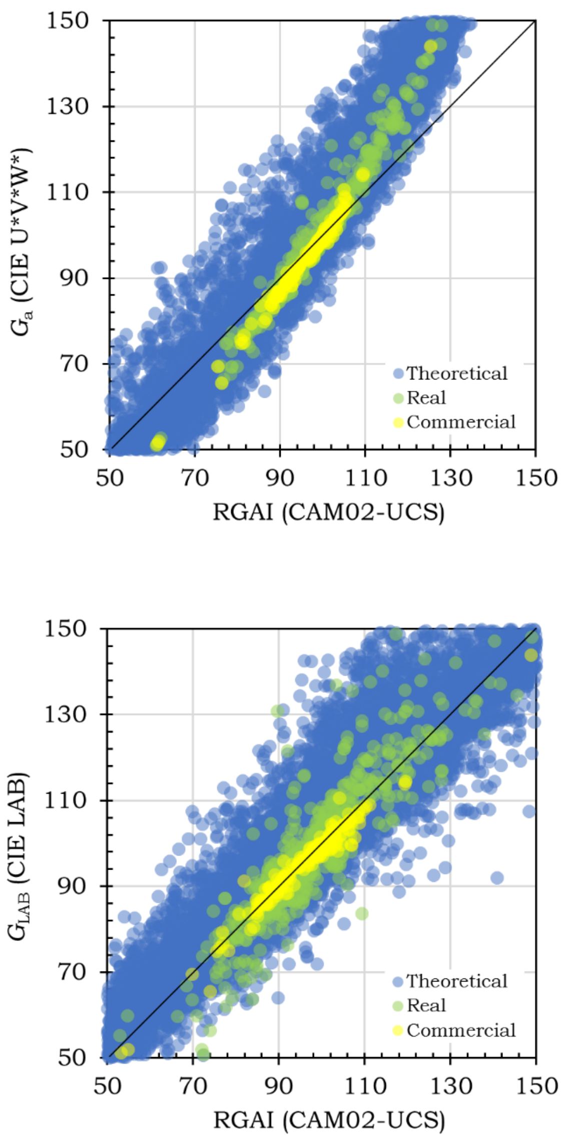

Figure 5. Comparison of gamut area measures with varied color space. CAM02-UCS is the baseline for each comparison. See Table 2 for a description of each alternative measure.

using the 99 CES with IES R_g as the baseline, because CIE 1964 U*V*W* and CIE 1976 *u'v'* were developed for the purpose of color difference calculations and do not include a hue correlate, which is required for the IES TM-30 hue-angle binning procedure. CIE LAB was also originally developed for the purpose of uniform color difference calculations, although it is frequently extended and used as a color appearance model [Fairchild 2013]. CAM02-UCS is an extension of the CIE CAM02 color appearance model that also allows for uniform color difference calculations. Given the distinction between color appearance models and uniform color spaces, it should be considered suspect to use uniform color spaces for calculating gamut area (or other values tied to hue and chroma perception), although it has been done in the past and continues to be proposed.

All of the measures shown in **Figure 5** use the same relative reference scheme and color samples, which are those of CIE R_a. Importantly, GAI_{rel} does not use a chromatic adaptation transformation (CAT), G_a uses a simple von Kries CAT as employed in CIE R_a, G_{LAB} uses the "wrong" von Kries CAT native to CIE LAB, and CAM02-UCS uses CIE CAT02; thus, the comparison is not only of the color space/color appearance model, but includes influences from the typical CATs used with each option. The use of different CATs with different color spaces/color appearance models was not investigated.

As with varying the color samples, the differences in gamut area quantifications with different color spaces are quite substantial when evaluating a large number of SPDs, with the range of the middle 95% of differences exceeding 40 points in all cases. This contrasts the conclusion of Teunissen and others [2016], who found that the difference between CIE U*V*W* and CAM02-UCS was not meaningful (using only a small number of SPDs). As with all comparisons in this document, the differences tend to be smaller—but often still meaningful—if only currently commercially-available light sources are considered, principally because currently available light sources offer little variation in gamut shape [Royer and others 2017a]. For this reason, it is insufficient to rely only on commercially-available light sources when evaluating the performance of color rendition measures.

The differences in gamut area characterizations due to color space are systematic, with a strong dependence on gamut shape. As shown in **Figure 6**, the difference between RGAI (CAM02-UCS) and the other three measures is strongly correlated with IES $R_{cs,h1}$. The different chromatic adaptation transformations may contribute to the random variation, but are not likely contributors to this systematic variation. The systematic difference caused by the color space should be considered flaws in G_a, GAI_{rel}, and G_{LAB}—which rely on *uniform color spaces*, not a color appearance model—because CAM02-UCS has been shown to be a more accurate representation of human color perception than older color spaces [Jost and others 2018; Luo and others 2006; Xu and others 2016]—this applies to both color appearance and color difference, including in the specific context of color rendition. **Figure 5** also illustrates some non-linearity for G_a and GAI_{rel} versus RGAI, which reflects the non-uniformity in those color spaces coupled with the typical gamut shape for very large increases in gamut area (red-green major axis ellipse). In short, using outdated color spaces can contribute to substantial errors in calculation of a visually-accurate measure of gamut area. This finding also applies to average color fidelity and all other measures of color rendition.

3.3 Alternative Calculation Methodologies

3.3.1 TM-30 Gamut Area without Hue-Angle Binning

Because IES TM-30-18 R_g is used here as the baseline for comparison, it is important to consider how the unique hue-angle bin averaging method influences its performance. To do this, IES TM-30-18 R_g can be compared to measures that eliminate hue-angle bin averaging and instead calculate the area enclosed by the (*a*', *b*') coordinates of all 99 CES. Rather than a (relatively) circular shape that is typical of smaller color sample sets used to calculate relative gamut area measures, the polygon formed by the 99 CES has a spikey shape. It should be noted that the 99 CES were chosen from the color volume, not from the hue-chroma (*a*', *b*') plane. Because individual color samples shift somewhat independently, the hue-angle order of the samples can vary between SPDs. This presents at least three options for calculating a relative gamut area measure based directly on all 99 CES.

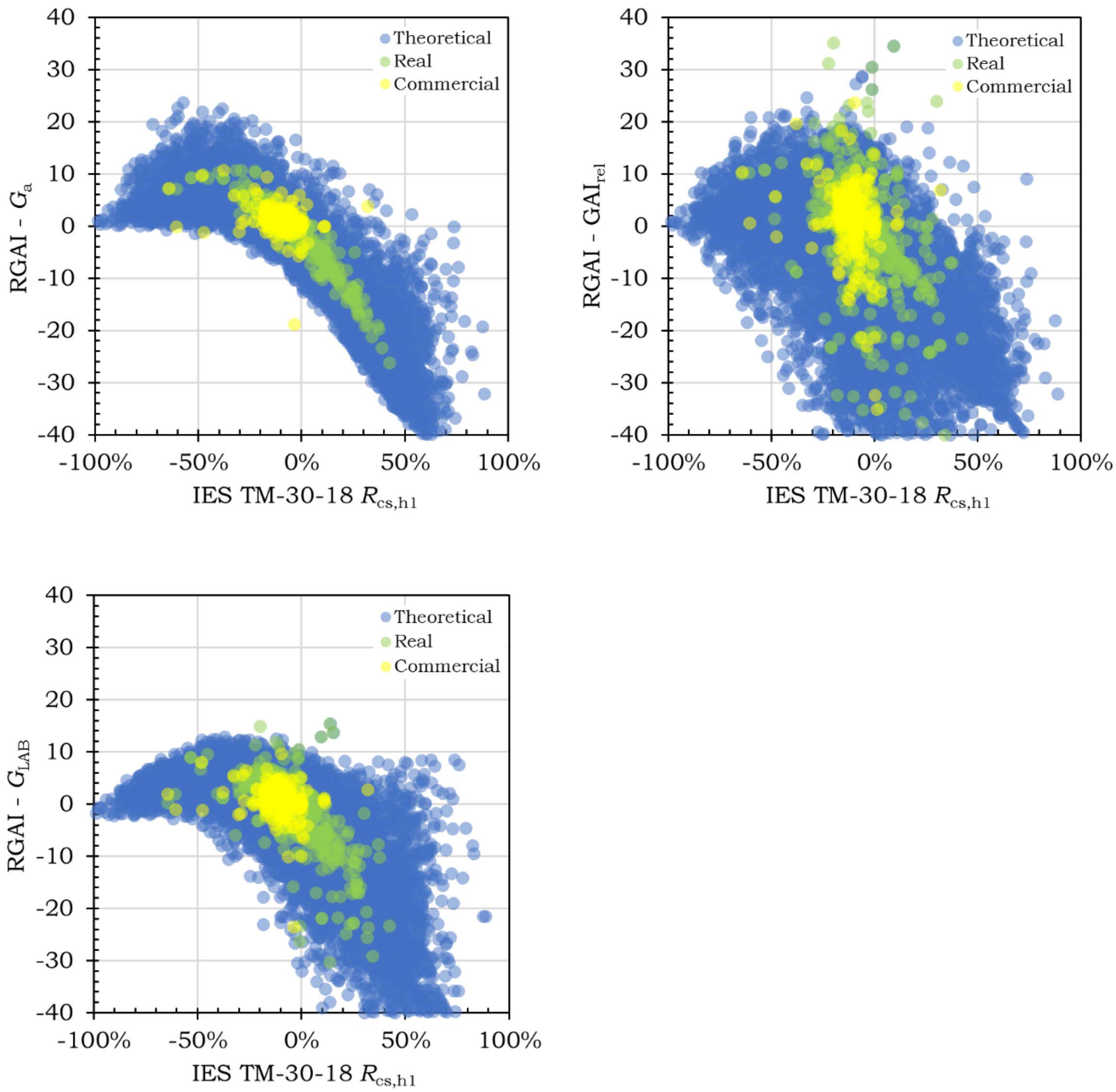

Figure 6. The difference between gamut area measures using varying color spaces versus IES $R_{cs,h1}$ (red chroma shift).

For this article, the first approach was to maintain the numerical order of the samples defined in IES TM-30-15, which was determined based on the hue angle of each sample under the 5000 K reference illuminant (a mix of 5000 K Planckian radiation and CIE D_{50}); this presents the possibility of a discontinuous area, where individual segments of the polygon may intersect. This approach is identified as R_g99A. The second approach was to re-order the samples based on their hue angle under the specified, CCT-dependent reference illuminant, using the same order for the test source; with this approach, the order of the samples can vary with the test light source. This approach is identified as R_g99B. The third option evaluated was to re-order the samples, based on hue angle, independently for the test light source and reference illuminant; in this case, the order of the samples, and thus the polygons, may not be the same for the test and reference conditions. This approach is identified as R_g99C.

Figure 7 shows R_g99A, R_g99B, and R_g99C versus IES TM-30-18 R_g. In all cases, there is substantial variation between the measures, both for commercially-available light sources and for theoretical SPDs. Given the stronger correlation of IES TM-30-18 R_g than any of the three variants of R_g99 with the other color sample sets

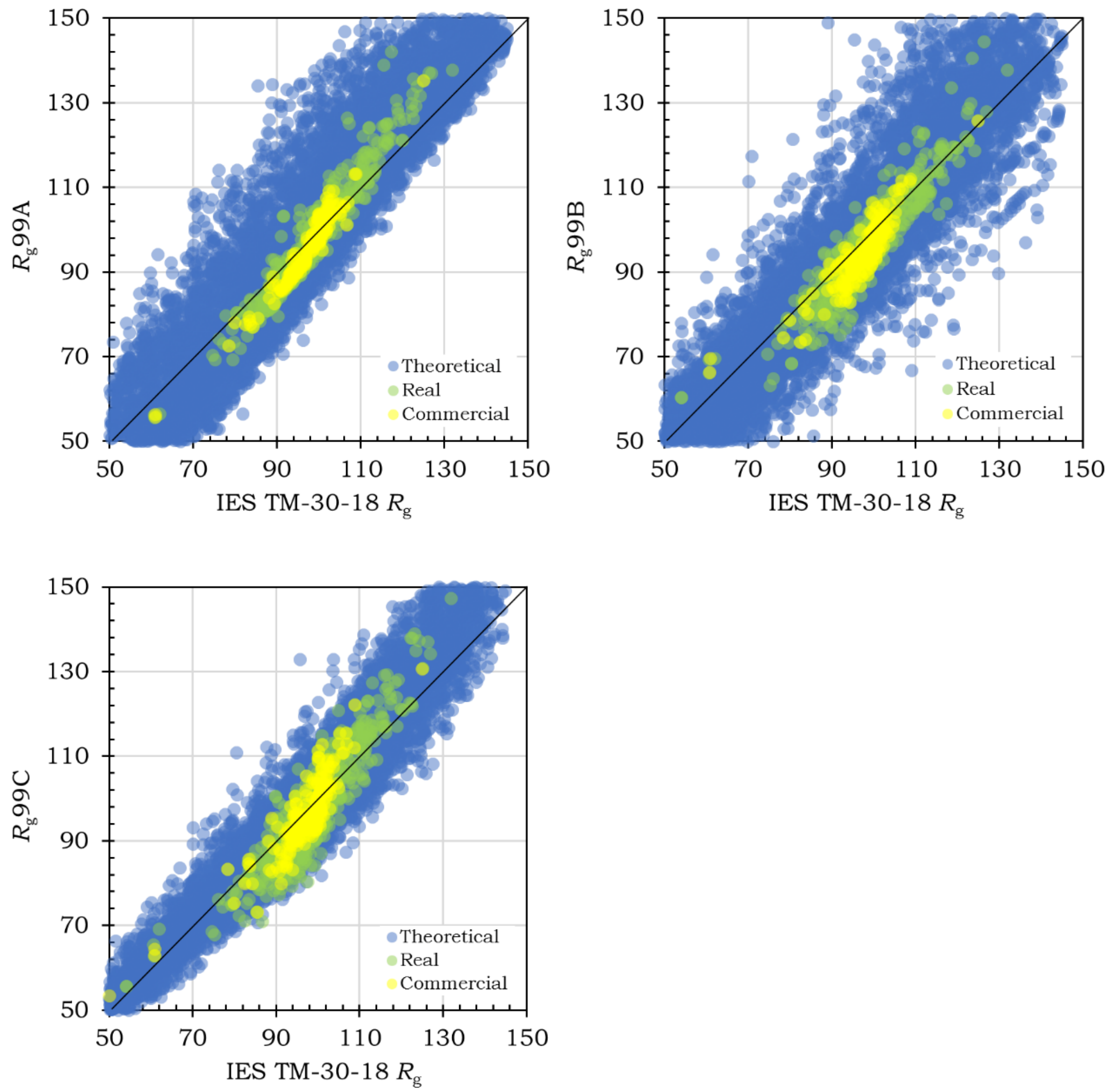

Figure 7. Three alternatives for computing gamut area based directly on all 99 CES versus IES TM-30-18 R_g.

described in Section 3.1 (not shown), it appears that the aforementioned issues with all three alternatives for using all 99 CES are problematic in practice. This is also evidenced by similarly large differences when comparing R_g99A, R_g99B, and R_g99C to one another, and to measures of average change in chroma (also not shown). In short, large sets of color samples forming irregular polygons are not equivalent to regularly-spaced color samples for calculating gamut area, and there is no evidence to suggest irregular polygon approach is appropriate.

Another option for avoiding hue-angle binning in calculating a gamut area with the 99 CES is to use an algorithm to find a convex area, but this may suffer from the compression issue discussed previously as well as issue of different samples forming the perimeter. It was not explored further for this analysis.

3.3.2 TM-30 Gamut Area with Larger (Reference) Color Sample Set

The 99 CES used in IES TM-30 were chosen to closely mimic the quantification results from a larger set of 4,880 samples chosen to evenly sample the color volume with (approximately) neutral spectral sensitivity in aggregate, which is referred to as the reference set [David and others 2015]. It is possible to calculate any of

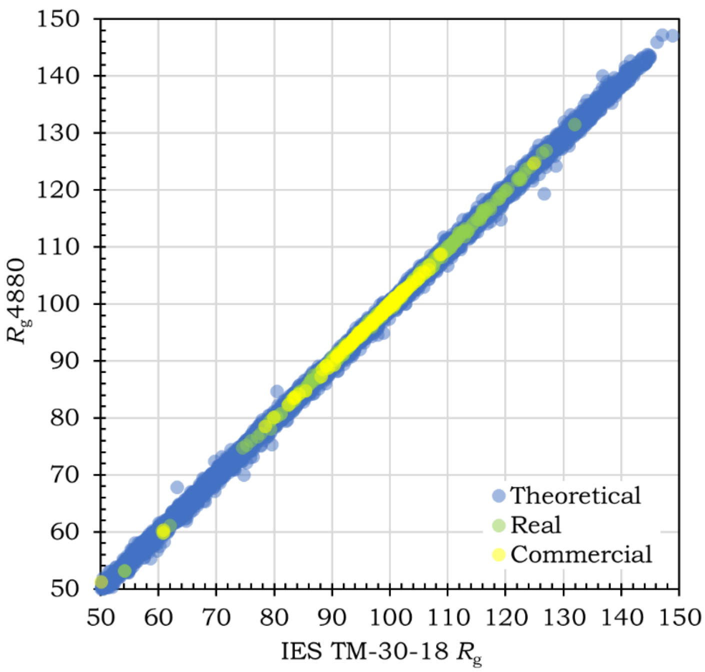

Figure 8. Gamut area calculated using the 4,880 sample reference set in lieu of the 99 CES versus the standard IES TM-30-18 R_g calculation.

the measures included in IES TM-30 using the 4,880-sample reference set instead of the 99 CES. **Figure 8** compares R_g4880, which uses the full reference set, to IES TM-30-18 R_g, which uses the 99 CES, showing that the difference is small and consistent throughout the range of IES TM-30-18 R_g values. For the 15,806 SPDs considered in this analysis, 95% of the differences (IES TM-30-18 R_g – R_g4880) are between -1.4 and 2.2 points. Thus, IES TM-30-18 R_g is a reasonably strong correlate of the reference gamut area. This also demonstrates that, when carefully selected, somewhat smaller color sample sets can provide a reasonable approximation for much larger color sample sets.

3.3.3 Gamut Area Versus Gamut Volume

The emergence of larger sets of color samples filling the color volume instead of just the hue-chroma plane has fomented the idea of calculating gamut volume instead of gamut area. **Figure 9a** compares R_gV (a gamut volume metric based on the 99 CES) and IES TM-30-18 R_g, showing fairly strong correlation between the two with a typical range in R_gV of approximately ±5 points at any given value of IES TM-30-18 R_g. At least some of the difference may be attributed to the fact that the convex hulls [Barber and others 1996] of the CES under the test and reference illuminants may include a different number of facets—a similar situation to what is encountered when trying to calculate the gamut area directly from all 99 CES.

Not all corresponding volume and gamut measures will produce similar results. For example, R_g14V (a gamut volume measure calculated from the 14 color samples used to calculate GVI) and R_g14 do not have a linear relationship (**Figure 9b**), which is a function of the color samples used in the calculation, including their distribution in the color volume. The problem of facets is exacerbated with only 14 samples; SPDs from the examined set with a volume including 22 facets for the reference condition had as few as 10 and as many as 24 facets under the test condition. This may be due to the fact that the 14 samples are arranged in a ring, which does not relate to the theoretical spherical volume of color space. R_gV and R_g14 are more closely related, despite the use of different color sample sets, because the greater number of samples used for R_gV reduces the level of facet mismatch.

Of course, the 14 color samples of GVI were not selected for use in a relative measure. However, when they are used in the reference-free GVI calculation, they induce a CCT-dependence that is remarkably similar to the

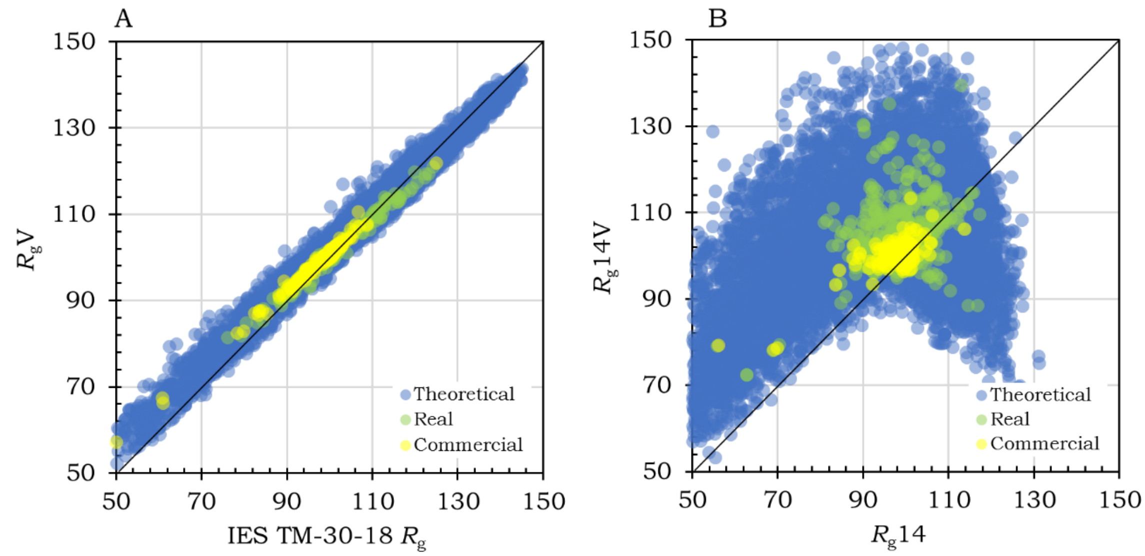

Figure 9. A: Gamut volume versus gamut area using the 99 CES of IES TM-30-18. B: Gamut volume versus gamut area using the 14 color samples of the recently proposed Gamut Volume Index (GVI).

CCT-dependence induced by the non-uniformity of the CIE U*V*W* color space (**Figure 10a**). Note that GVI values were calculated with a corrected formula that includes division by 10,000 [Liu 2017]. The CCT-dependence is not supported by experimental evidence when chromatic adaptation and color rendition characteristics are carefully controlled [Royer and others 2017b; Zhang and others 2017]. As a consequence of its CCT-dependence and other factors, GVI shows a relatively weak relationship with other measures of gamut volume or gamut area, such as IES TM-30-18 R_g (**Figure 10b**) or R_gV (**Figure 10c**). Although GVI was proposed as a measure of color preference and fit to existing color preference data, it does not reward increases in red chroma (**Figure 10d**) and cannot entirely account for gamut shape, which calls into question its utility, as both are important factors driving evaluations of color preference [Esposito 2016; Esposito and Houser 2018; Ohno and others 2015; Royer and others 2017a; Royer and others 2017b; Royer and others 2016; Zhang and others 2017].

3.3.4 Gamut Area Versus Chroma Shift

Gamut area measures are often thought of as a measure of the average change in objects' saturation or chroma induced by a test light source relative to the reference illuminant. Modern object color spaces have direct correlates for chroma and saturation, but these post-date the initial conceptualization of gamut area. Nonetheless, it is now possible to calculate change in chroma directly. **Figure 11a** documents the difference between ΔC_{99A} and IES TM-30-18 R_g for the large SPD set, where ΔC_{99A} is the average of the absolute chroma difference between test and reference for all 99 CES, with each individual chroma difference scaled by the CIE 224/IES TM-30 scaling factor (6.73) and added to 100. **Figure 11b** documents the difference between ΔC_{99B} and IES TM-30-18 R_g for the large SPD set, where ΔC_{99B} is the average of the relative chroma difference between test and reference for each of the 99 CES. **Figure 11c** documents the difference between ΔC_{16} and IES TM-30-18 R_g for the large SPD set, where ΔC_{16} is the average Local Chroma Shift for all 16 hue-angle bins.

A clear difference is visible, regardless of the specific ΔC formulation, and it occurs because hue shifts influence gamut area but not average chroma shift. Thus, it should be kept in mind that gamut area is not entirely a measure of saturation or chroma levels, especially as overall color shifts compared to the reference illuminant become large.

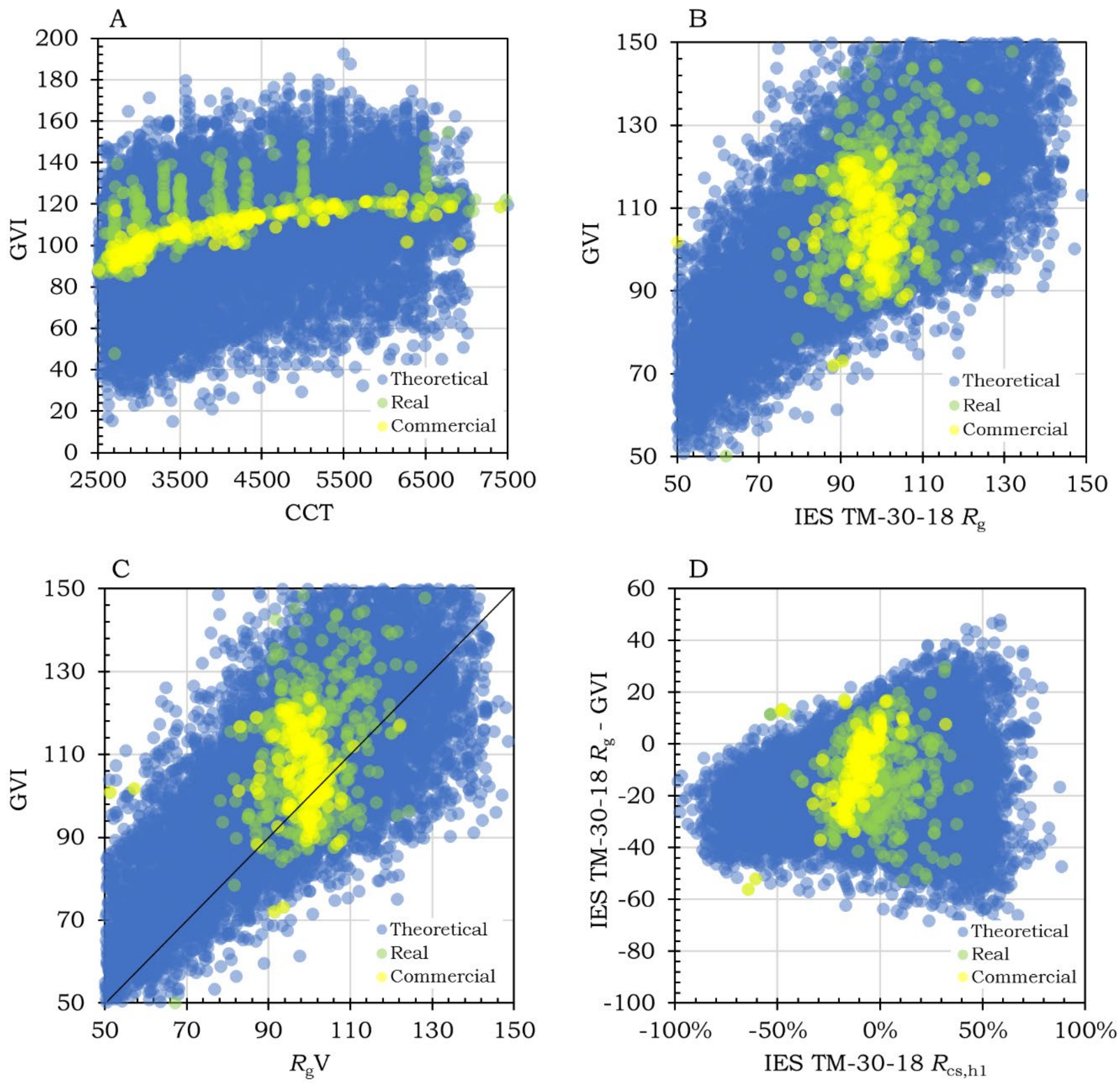

Figure 10. A: The Gamut Volume Index (GVI) is dependent on CCT. B: GVI is not well correlated with IES TM-30-18 R_g. C: GVI does not favor increases in red chroma compared to IES TM-30-18 R_g. D: The difference between IES R_g and GVI is no related to red chroma, casting doubt on the utility of GVI as a measure of color preference.

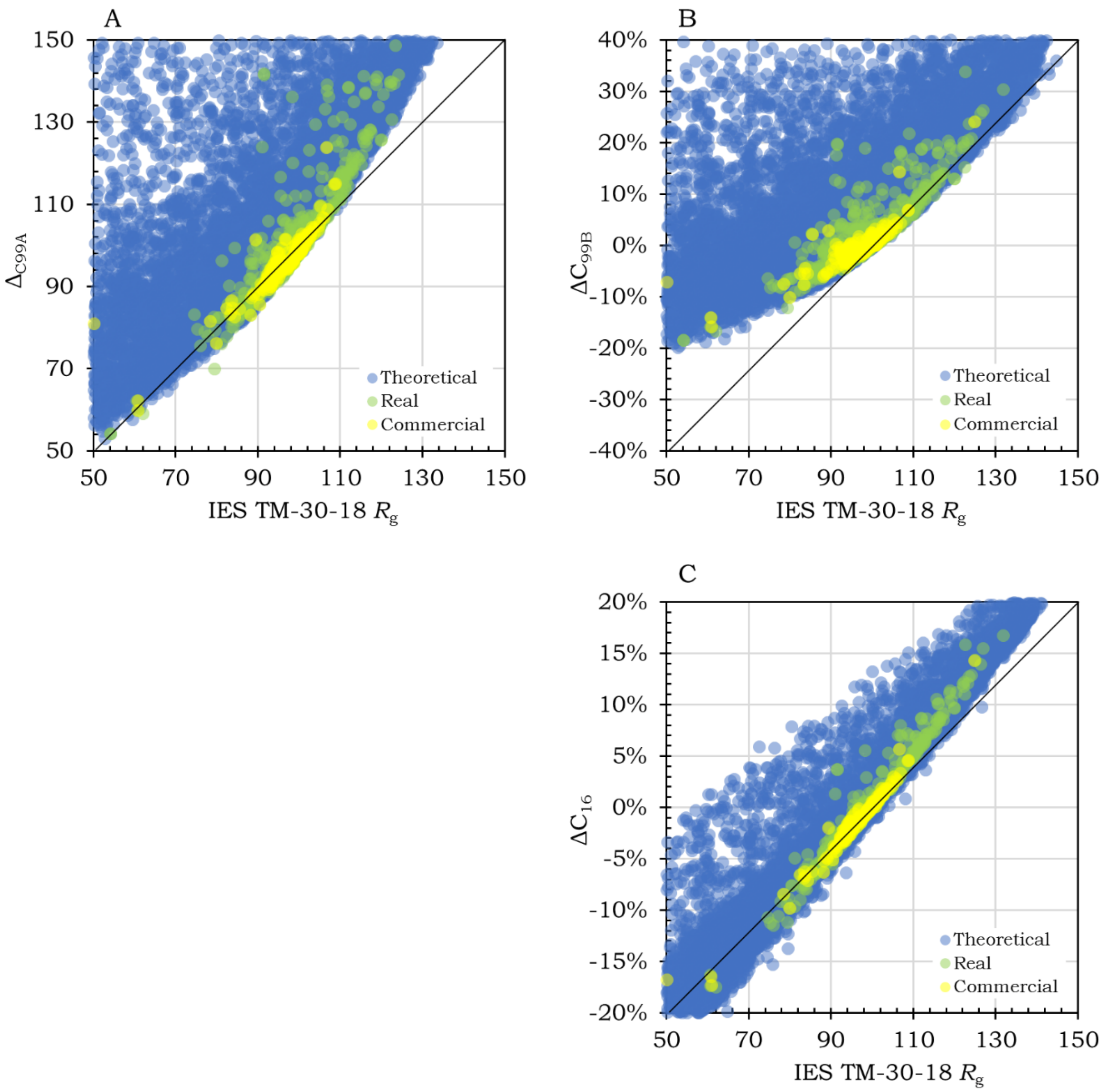

Figure 11. Three measures of average change in chroma based on the IES TM-30 system compared to IES TM-30-18 R_g.

4 Discussion

4.1 Combined Effects

This analysis has demonstrated that changing the color sample set or color space can have a substantial effect on how the gamut area of a light source is characterized. This occurs absent of other differences in calculation methods—such as volume versus area—that may also lead to different characterizations. The combined effects of these differences are also important. For example, **Figure 12a** demonstrates the relationship between IES TM-30-18 R_g and G_a, which use different color samples, different color spaces, and a slightly different reference illuminant scheme. G_a values have a range exceeding 30 points at any given value of IES TM-30-18 R_g within the range of typical architectural light sources, and the range of the middle 95% of differences (IES TM-30-18 $R_g - G_a$) for all SPDs in the set examined was -31.4 to 11.9 points. Even commercially-available

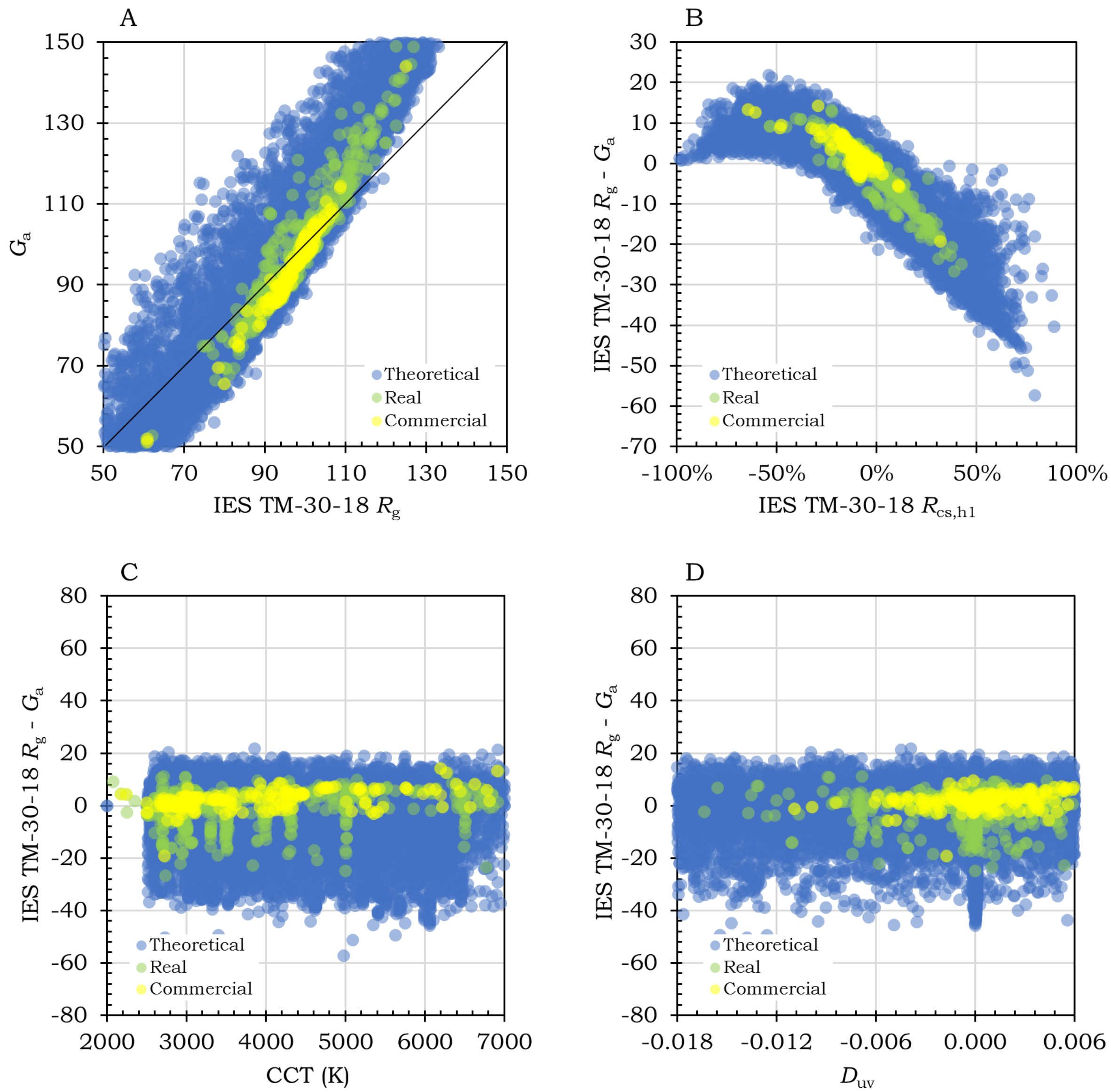

Figure 12. A: Comparison of G_a and IES TM-30-18 R_g. B: The difference between IES TM-30-18 R_g and G_a versus IES $R_{cs,h1}$.

sources can have differences ranging from -19.1 points (a particular RGB LED) to 14.3 points (plasma). Thus, IES TM-30-18 R_g and G_a provide substantially different characterizations.

As with their counterpart measures for average color fidelity (IES R_f and CIE R_a), the primary contribution to this difference comes from the fact that G_a relies on a non-uniform color space, although the color sample set also plays a role. Unlike how CIE R_a unduly penalizes increases in red chroma, G_a values tend to be greater than IES TM-30-18 R_g values for such sources (**Figure 12b**). *When each is used alone*, this tends to make G_a more strongly correlated with subjective evaluations of color preference, saturation (vividness), and normalness (naturalness), although neither is the best fit to subjective evaluation data when an appropriate variety of light sources is considered. Additionally, G_a is still an average value, which means that two sources with equal G_a values can be perceived very differently, and because it essentially functions as a weighted measure of gamut area—where all increases in chroma are not treated equally—it is less useful in applications where red chroma enhancement is not important. Although some have advocated for gamut measures based on the eight TCS of CIE R_a because of their apparent ease of implementation, they carry with them significant limitations that unnecessarily complicate product development and specification, as further discussed in the next section.

Appendix A documents the net relationship between four additional existing measures of gamut area (FCI, GAI, Q_g, and RGAI) and IES TM-30-18 R_g, further illustrating the points made previously based on individual calculation elements. When an appropriately diverse set of SPDs is considered, CCT does not contribute to differences between relative-reference measures of gamut area (for example, **Figure 12c**), nor is gamut area correlated with CCT, as suggested by Khanh and colleagues [2016c]. However, CCT heavily influences gamut area quantifications with GAI (lower CCT determined to have lower gamut area) and somewhat influences gamut area quantifications with FCI (lower CCT determined to have higher gamut area). As mentioned earlier, CCT has been shown to have no influence on subjective evaluations of color rendition when chromatic adaptation is carefully controlled [Royer and others 2017b; Zhang and others 2017]. Likewise, D_{uv} and gamut area are independent quantities, and do not contribute to differences between G_a and IES TM-30-18 R_g (**Figure 12d**).

4.2 Complementary Measures?

In many cases, gamut area measures have been presented as complements to average color fidelity measures. For example, G_a was proposed as a complement to CIE R_a, Q_g as a counterpart to Q_f (and Q_a) in the CQS system, and IES TM-30-18 R_g as a component of the broader IES TM-30 method. These three systems all share the tenet of being derived from a common calculation framework, with a unified set of color samples, model of human color perception, and reference illuminant scheme. Each system has unique characteristics, however, which arise from the underlying calculation elements. For example, the distortion in the CIE U*V*W* color space that influences both CIE R_a and G_a values means that the two provide less independent information than IES TM-30-18 R_f and R_g, as illustrated in **Figure 13**. This reduces the overall utility of the system—even if average measures alone are already insufficient for characterizing subjective evaluations of color quality. For example, specification criteria for such measures cannot be independent.

4.3 Gamut Area Versus Perception: Is Gamut Area Useful?

Gamut area measures originated under the premise that they were useful indicators of color preference or color discrimination. Experimental evidence on these suppositions is mixed, which may be a result of the inherent limitations and/or features of gamut area (or gamut volume) measures. Important considerations include:

1. Global average measures—that is, measures that average all hues together—inherently discard important information. Because certain hues (for example, red) may influence subjective evaluations more than other hues, it is impossible for any gamut area measure to completely capture perceived color quality attributes [Royer and others 2016]. This remains true even if a gamut area measure is paired with an average color fidelity measure. However, experiments that present stimuli with limited variation in gamut shape—a frequent occurrence—may find gamut area to be a better predictor of subjective evaluations, which is similar to what happens with color fidelity [Royer 2017].

2. Increases in gamut area require hue shifts. Because it does not appear to be possible to uniformly increase chroma for all hues, increasing gamut area requires increasing chroma for some hues more than others, which in turn means intermediate hues shift toward those where chroma is being enhanced. For example, meaningfully increasing red chroma typically results in oranges and purples shifting towards red. While increasing chroma may increase the difference between colors, thus aiding color discrimination, the accompanying hue shifts may counteract any benefit, even juxtaposing the hue order of color samples versus a reference condition [Esposito and Houser 2017; Royer and others 2012].

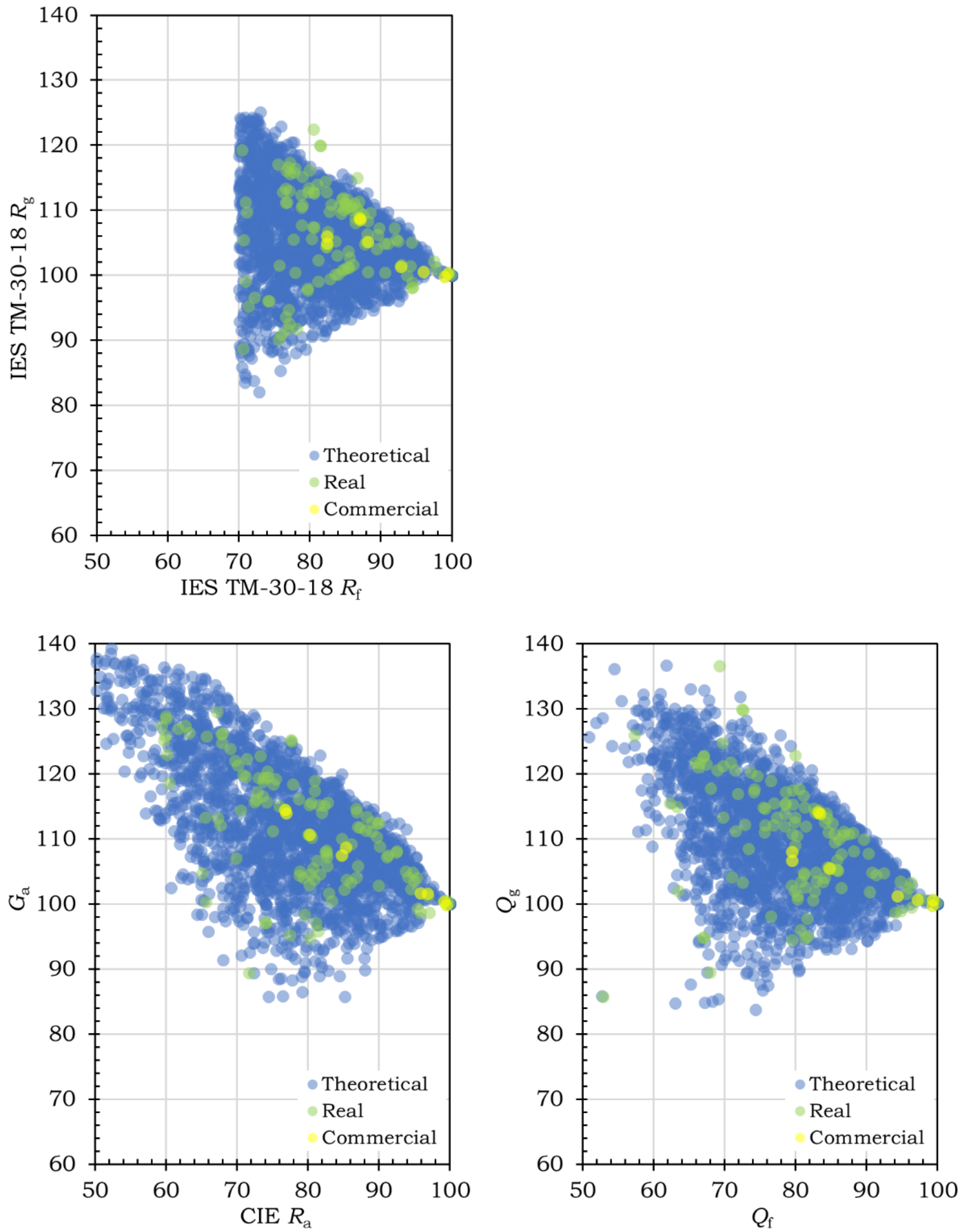

Figure 13. Gamut area versus average color fidelity for SPDs in the large set having IES $R_g \geq 70$ and IES $R_{cs,h1} \geq 0\%$. For these 2,307 SPDs, IES R_g and IES R_f are independent, but the alternative pairs (G_a and R_a; Q_g and Q_f) are not.

3. Though not exactly a limitation, it is important to remember that higher gamut area values are not necessarily better for color preference, color discrimination, or any other aspect of color quality [Wei and Houser 2017]. Increasing gamut area requires reductions in color fidelity, and at extreme levels increases in chroma can make scenes appear cartoonish. If developing models of subjective evaluations is an important experimental goal, it is important to include stimuli that capture this non-linear effect, otherwise the model may simply reward any amount of increase in gamut area.

4. No current methods for characterizing light source color rendition account for illuminance, instead always comparing the test source to a reference illuminant with the same Y tristimulus value. This stipulation is necessary for commerce, where light sources are characterized independent of their use. However, the Hunt effect [Fairchild 2013; Hunt 1952] identifies that colorfulness is dependent on luminance. Preliminary results suggest that preference for increases in chroma are tied to illuminance [Kawashima and others 2017; Wei and others 2018], which obscures the meaning of relative gamut area measures with regard to subjective evaluations of color preference or vividness. If the Y tristimulus value is allowed to vary—which can be done for measures based on a color appearance model, such as CAM02-UCS—relative gamut area values would be affected.

These considerations reveal that gamut area (or volume), when used alone, is not a particularly useful or informative quantity. It is a quantification that is not predictive of subjective evaluations, such as saturation or vividness, or task performance. This makes trying validate gamut area measures in experimental settings a dubious strategy, because correlation will likely depend on other factors that may or may not be controlled. In contrast, red chroma shift (IES $R_{\mathrm{cs,h1}}$) alone has been shown to be extremely well correlated with ratings of saturation or vividness in polychromatic environments [Esposito 2016; Royer and others 2017b; Royer and others 2016], likely due to the psychological importance of red [Elliot and Maier 2014]. Nonetheless, several studies have identified IES TM-30-18 R_{g} as a component of multi-measure models that show strong correlation with subjective evaluations of illuminated environments [Esposito 2016; Esposito and Houser 2018; Royer and others 2017b; Royer and others 2016; Zhang and others 2017]. This indicates that a coordinated system of measures—including average color fidelity, gamut area, and measures of gamut shape—is necessary to convey color rendition.

5 Conclusions

Calculations of gamut area are influenced by the color sample set, model of color perception, and other calculation elements. The magnitude of the differences can be substantial. Color sample sets that utilize too few samples or samples that are not approximately evenly spaced in hue angle appear to produce results that diverge from other sets in a way that is unrelated to other aspects of color rendition, such as gamut shape. Likewise color sample sets that include only high-chroma color samples may result in compression of gamut area values for sources that substantially increase gamut area. Gamut area measures that use too many samples can present methodological challenges (related to mismatched test and reference polygons) that have no logical resolution. A reasonable compromise is to use a moderate number of samples (approximately 16) with moderate chroma, either directly or derived using a grouping method—both of these solutions provide similar results. In contrast, differences in the color space lead to systematic differences based on gamut shape. Using a modern color appearance model vetted with color difference data (e.g., CAM02-UCS) is advisable.

Gamut area measures are related to the average change in chroma, but are also influenced by hue shifts. Thus, gamut area measures and average chroma shift measures are not well correlated. In some cases, gamut area is more closely related to gamut volume, but this relationship is dependent on the color sample set. No form of a gamut area measure is capable of characterizing object color appearance for different hues—a similar limitation of gamut volume or average chroma shift measures. As a result, gamut area alone has not been linked to any perceptual attribute when a large quantity and diverse range of SPDs are considered. This makes the meaning of any of the differences between measures identified in this analysis difficult to determine. Still, some gamut area measures, such as IES R_{g}, are more useful than others when included in multi-measure models of subjective evaluations of color quality, because they can provide information that is more independent of other measures in the system.

For lighting specifiers, manufacturers, and other end-users less concerned with the underlying complexities of gamut area calculations, there are two important takeaways. First, given the inherent limitations of hue-averaged measures, gamut area should be considered, at best, a tertiary measure of color rendition. Gamut shape and color fidelity will have more bearing on light source quality in most situations, and simply increasing gamut area may have no benefit. Second, if gamut area is considered, it is important to use an up-to-date measure. Simply pairing CRI (CIE R_a) with its complementary gamut area measure (G_a) is unlikely to address any of CRI's fundamental problems.

References

Barber CB, Dobkin DP, Huhdanpaa H. 1996. The quickhull algorithm for convex hulls. ACM Trans. Math. Softw. 22(4):469-483.

Boyce PR. 1977. Investigations of the subjective balance between illuminance and lamp colour properties. Lighting Research and Technology 9(1):11-24.

CIE. 1995. 13.3: Method of measuring and specifying colour rendering properties of light sources, 3rd Ed. Vienna (Austria): Commission Internationale de l'Eclairage. p. 16.

CIE. 2004. 15:2004 Colorimetry, Third Edition. Vienna, Austria.: Commision Internationale de L'Eclairage.

CIE. 2017. 224:2017 CIE 2017 Colour Fidelity Index for accurate scientific use. Vienna, Austra: Commision Internationale de L'Eclairage.

David A. 2013. Color Fidelity of Light Sources Evaluated over Large Sets of Reflectance Samples. Leukos 10(2):59-75.

David A, Esposito T, Houser K, Royer M, Smet K, Whitehead L. 2018. A Vector Field Color Rendition Model for Characterizing Color Shifts and Metameric Mismatch. Leukos Submitted for Publication.

David A, Fini PT, Houser KW, Ohno Y, Royer MP, Smet KA, Wei M, Whitehead L. 2015. Development of the IES method for evaluating the color rendition of light sources. Opt Express 23(12):15888-906.

Elliot AJ, Maier MA. 2014. Color psychology: effects of perceiving color on psychological functioning in humans. Annu Rev Psychol 65:95-120.

Esposito T. 2016. Modeling color rendition and color discrimination with average fidelity, average gamut, and gamut shape [Doctoral]. [University Park, PA]: Penn State University.

Esposito T, Houser K. 2017. A new measure of colour discrimination for LEDs and other light sources. Lighting Research & Technology Online Before Print. DOI: 10.1177/1477153517729200.

Esposito T, Houser K. 2018. Models of colour quality over a wide range of spectral power distributions. Lighting Research & Technology Online Before Print. DOI: 10.1177/1477153518765953.

Fairchild MD. 2013. Color Appearance Models. Chichester, United Kingdom: Wiley. p. 472.

Fotios SA, Levermore GJ. 1998. Chromatic effect on apparent brightness in interior spaces I: Introduction and colour gamut models. Lighting Research and Technology 30(3):97-102.

Guo X, Houser KW. 2004. A review of colour rendering indices and their application to commercial light sources. Lighting Research and Technology 36(3):183-199.

Hashimoto K, Nayatani Y. 1994. Visual clarity and feeling of contrast. Color Research & Application 19(3):171-185.

Hashimoto K, Yano T, Shimizu M, Nayatani Y. 2007. New method for specifying color-rendering properties of light sources based on feeling of contrast. Color Research and Application 32(5):361-371.

Houser KW, Wei M, David A, Krames MR, Shen XS. 2013. Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition. Opt Express 21(8):10393-411.

Hunt R. 1952. Light and dark adaptation and the perception of color. J Opt Soc Am 42:190-199.

IES. 2015. IES-TM-30-15 Method for Evaluating Light Source Color Rendition. New York, NY: The Illuminating Engineering Society of North America. p. 26.

IES. 2018. IES-TM-30-18 Method for Evaluating Light Source Color Rendition. New York, NY: The Illuminating Engineering Society of North America. p. 26.

Jost S, Cauwerts C, Avouac P. 2018. CIE 2017 color fidelity index Rf: a better index to predict perceived color difference? Journal of the Optical Society of America A 35(4):B202-B213.

Kawashima Y, Ohno Y, Oh S. Vision Experiment on Verification of Hunt Effect for Lighting. 11th Biennial Joint CIE/USNC and CNC/CIE Technical Conference; 2017; Gaithersburg, Maryland.

Khanh T, Bodrogi P. 2016. Colour preference, naturalness, vividness and colour quality metrics, Part 3: Experiments with makeup products and analysis of the complete warm white dataset. Lighting Research & Technology Online Before Print. DOI: 10.1177/1477153516669558(0):1477153516669558.

Khanh T, Bodrogi P, Vinh Q, Stojanovic D. 2016a. Colour preference, naturalness, vividness and colour quality metrics, Part 1: Experiments in a room. Lighting Research & Technology. Online Before Print: DOI: 10.1177/1477153516643359(0):1477153516643359.

Khanh T, Bodrogi P, Vinh Q, Stojanovic D. 2016b. Colour preference, naturalness, vividness and colour quality metrics, Part 2: Experiments in a viewing booth and analysis of the combined dataset. Lighting Research & Technology. Online Before Print: DOI: 10.1177/1477153516643570(0):1477153516643570.

Khanh TQ, Vinh QT, Bodrogi P. 2016c. A numerical analysis of recent colour rendition metrics. Lighting Research and Technology 49(5):561-573.

Liu Q. 2017. Gamut Volume Index. Personal Communication. August 3, 2017.

Liu Q, Huang Z, Xiao K, Pointer MR, Westland S, Luo MR. 2017. Gamut Volume Index: a color preference metric based on meta-analysis and optimized colour samples. Optics Express 25(14):16378-16391.

Luo MR, Cui G, Li C. 2006. Uniform colour spaces based on CIECAM02 colour appearance model. Color Research & Application 31(4):320-330.

Ohno Y, Fein G, Miller C. Vision Experiment on chroma saturation for color quality preference. 28th CIE Session; Jun 28 - Jul 4 2015; Manchester, UK: Vienna (Austria): Commision Internationale de l'Eclairage. p. 2124.

Rea MS, Freyssinier-Nova JP. 2008. Color rendering: A tale of two metrics. Color Research and Application 33(3):192-202.

Rea MS, Freyssinier JP. 2010. Color Rendering: Beyond Pride and Prejudice. Color Research and Application 35(6):401-409.

Rea MS, Freyssinier JP. 2013. White lighting. Color Research & Application 38(2):82-92.

Royer M. 2017. Comparing Measures of Average Color Fidelity. Leukos 14(2):69-85.

Royer M, Houser K, David A. 2017a. Chroma Shift and Gamut Shape: Going Beyond Average Color Fidelity and Gamut Area. Leukos Online Before Print. DOI: 10.1080/15502724.2017.1372203.

Royer M, Wei M. 2017. The Role of Presented Objects in Deriving Color Preference Criteria from Psychophysical Studies. LEUKOS 13(3):143-157.

Royer M, Wilkerson A, Wei M. 2017b. Human Perceptions of Color Rendition at Different Chromaticities. Lighting Research & Technology Online before print. DOI: 10.1177/1477153517725974.

Royer MP. 2016. What is the Reference? An Examination of Alternatives to the Reference Sources Used in IES TM-30-15. Leukos 13(2):71-89.

Royer MP, Houser KW, Wilkerson AM. 2012. Color discrimination capability under highly structured spectra. Color Research & Application 37(6):441-449.

Royer MP, Wilkerson A, Wei M, Houser K, Davis R. 2016. Human perceptions of colour rendition vary with average fidelity, average gamut, and gamut shape. Lighting Research & Technology 49(8):966-991.

Smet K, David A, Whitehead L. 2015. Why Color Space Uniformity and Sample Set Spectral Uniformity Are Essential for Color Rendering Measures. Leukos 12(1-2):39-50.

Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. 2011. Correlation between color quality metric predictions and visual appreciation of light sources. Opt Express 19(9):8151-66.

Teunissen C, van der Heijden F, Poort SHM, de Beer E. 2016. Characterising user preference for white LED light sources with CIE colour rendering index combined with a relative gamut area index. Lighting Research & Technology 49(4):461-480.

Thornton WA. 1972. Color-discrimination index. J Opt Soc Am 62(2):191-4.

Thornton WA. 1973. Fluorescent Lamps with High Color-Discrimination Capability. Journal of the Illuminating Engineering Society 3(1):61-64.

Wei M, Bao W, Huang H. 2018. Consideration of Light Level in Specifying Light Source Color Rendition. Leukos Online Before Print. DOI: 10.1080/15502724.2018.1448992.

Wei M, Houser KW. 2017. Systematic Changes in Gamut Size Affect Color Preference. Leukos 13(1):23-32.

Windisch R, Heidel G, Binder U, Bergenek K. 2017. Impact of spectral features of common LED lighting systems on TM-30 color indices. Optics Express 25(3):1824-1830.

Xu H. 1984. Color Rendering Capacity of Illumination. Journal of the Illuminating Engineering Society 13(2):270-276.

Xu H. 1993. Colour rendering capacity and luminous efficiency of a spectrum. International Journal of Lighting Research and Technology 25(3):131-132.

Xu W, Wei M, Smet K, Lin Y. 2016. The prediction of perceived colour differences by colour fidelity metrics. Lighting Research and Technology Online before print. DOI: 10.1177/1477153516653650.

Zhang F, Xu H, Feng H. 2017. Toward a unified model for predicting color quality of light sources. Applied Optics 56(29):8186-8195.

Zukauskas A, Vaicekauskas R, Ivanauskas F, Vaitkevicius H, Vitta P, Shur MS. 2009. Statistical approach to color quality of solid-state lamps. IEEE Journal of Selected Topics in Quantum Electronics 15(4):1189-1198.

Appendix A: Additional Direct Comparisons

Figures A1 through A4 provide summary comparisons between GAI, FCI, Q_g, or RGAI versus the baseline of IES R_g. In each figure, the upper left chart is a direct comparison, whereas the other three charts explore potential underlying causes of differences based on treatment of gamut shape (exemplified with red chroma shift, $R_{cs,h1}$), CCT, and D_{uv}. The differences in underlying calculation frameworks can lead to very large differences (between GAI or FCI and IES R_g) or moderate differences (between Q_g or RGAI and IES R_g), depending on the specific elements of each calculation.

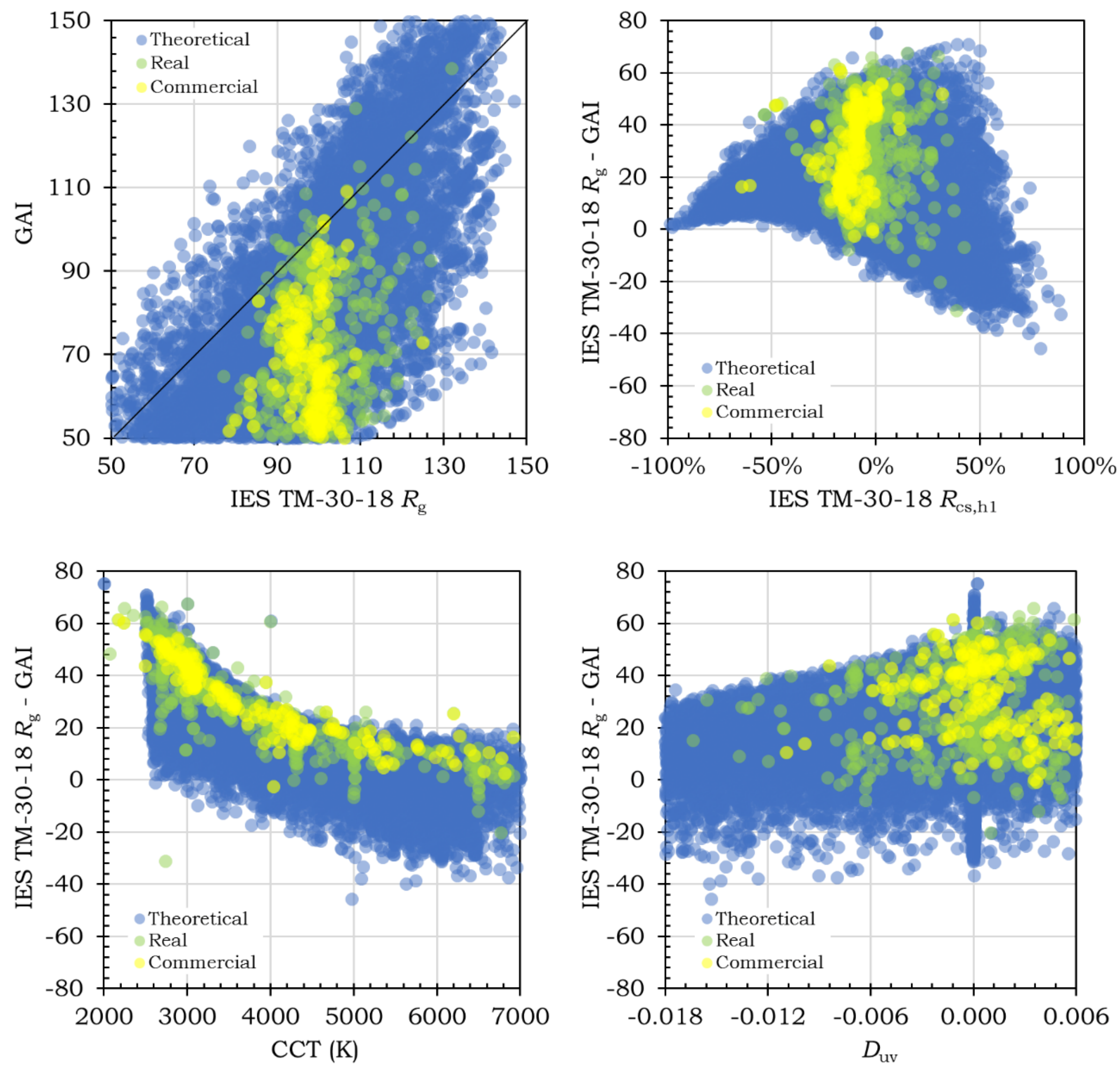

Figure A1. GAI versus IES R_g. The use of a small number of color samples, fixed reference illuminant, and non-uniform color spaces leads to differences in how GAI treats different CCTs and gamut shapes.

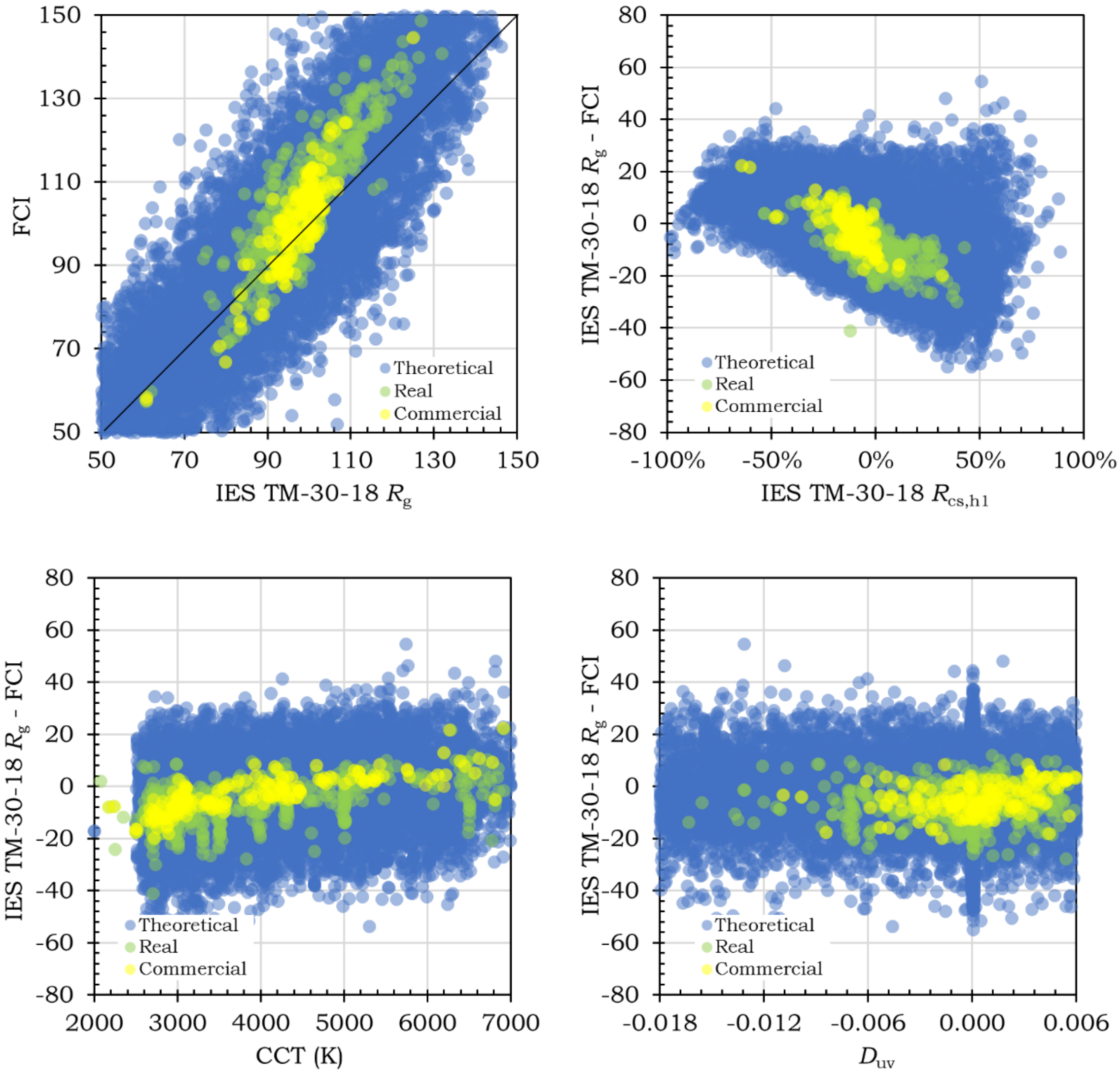

Figure A2. FCI versus IES R_g. Here, FCI is computed in CAM02-UCS. The principle difference is due to the color samples: because FCI only uses 4 color samples, there is a large amount of variation that results from specific SPD features. The used of a fixed reference illuminant is less problematic in a uniform color appearance model, but there is still some correlation with CCT.

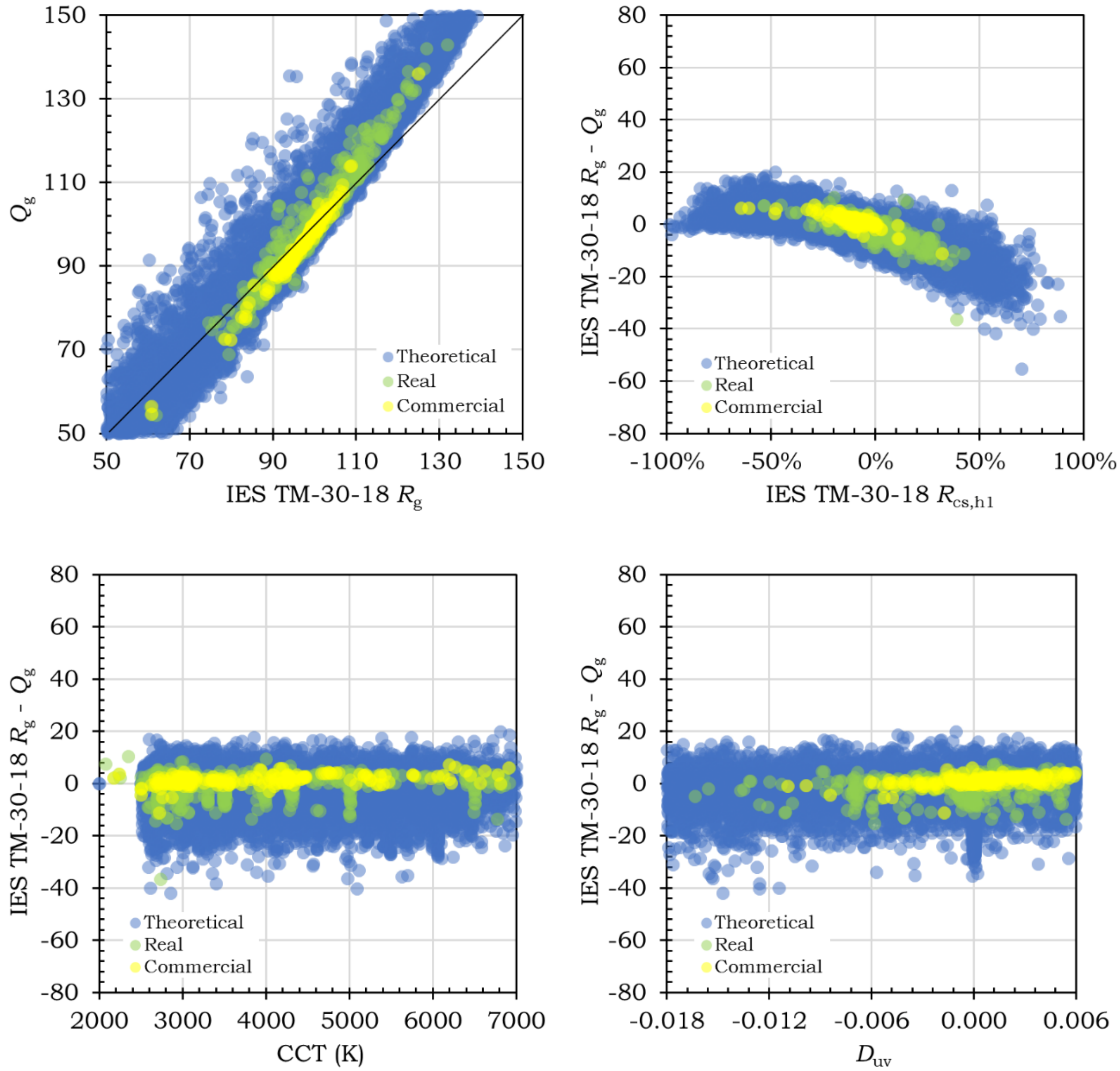

Figure A3. CQS Q_g versus IES R_g. The use of a relative reference scheme means that Q_g treats CCT and D_{uv} similarly to IES R_g. However, the combination of the color samples and color space (CIE LAB) lead to different treatments of gamut shape: Q_g increases more than IES R_g as red chroma is increased.

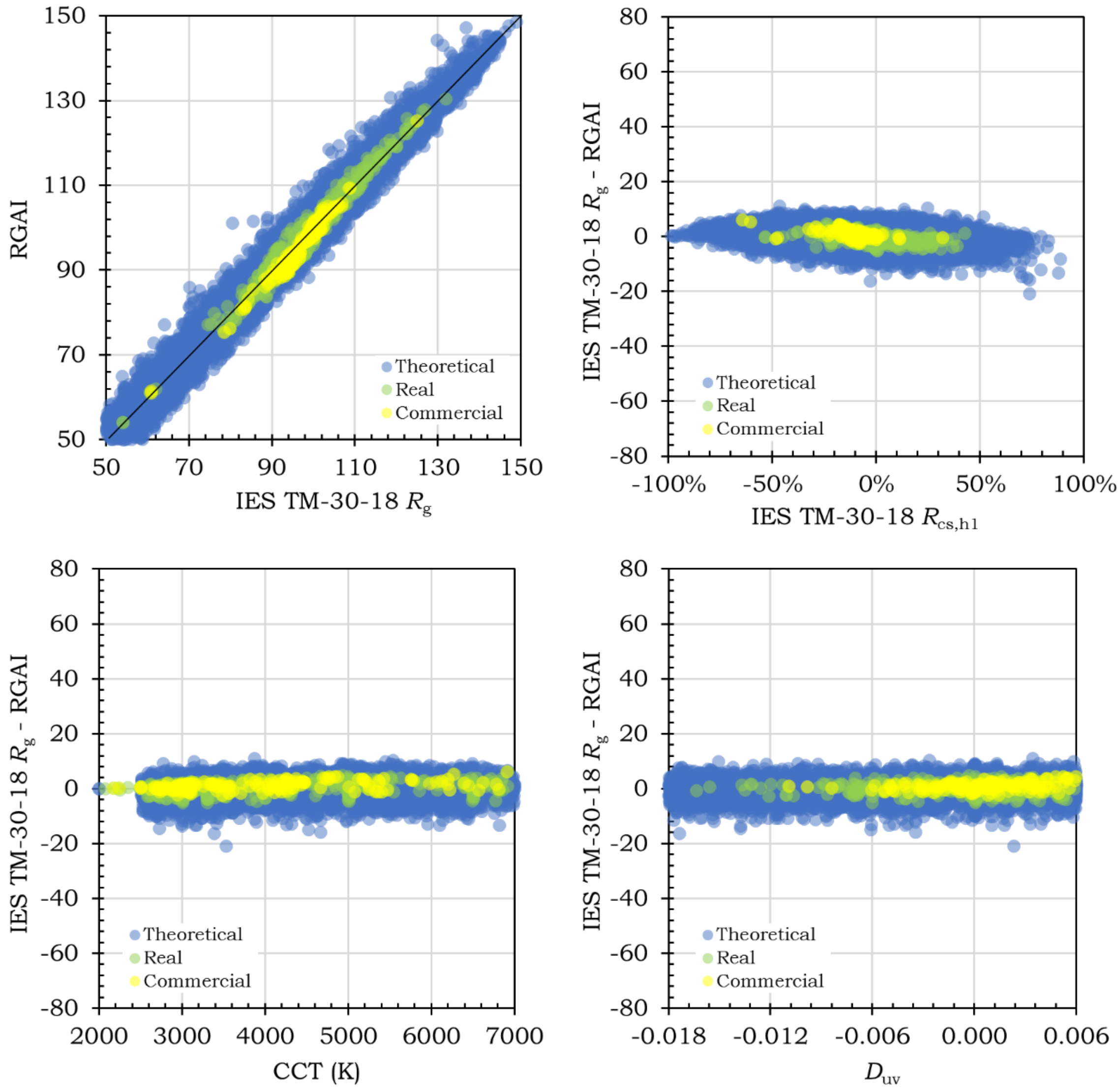

Figure A4. RGAI versus IES R_g. Of the measures evaluated, RGAI performs most similarly to IES R_g. These two measures vary only in their color samples. The differences are SPD-specific—rather than systematic—and are still notably large, especially for theoretical light sources.