What is the Reference? An examination of alternatives to the reference sources used in IES TM-30-15

Michael P Royer, Ph.D.[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. 2016. What is the Reference? An examination of alternatives to the reference sources used in IES TM-30-15. LEUKOS 13(2):71-89. DOI: 10.1080/15502724.2016.1255146.

Abstract

This paper documents the role of the reference illuminant in the IES TM-30-15 method for evaluating color rendition. TM-30-15 relies on a relative reference scheme; that is, the reference illuminant and test source always have the same correlated color temperature (CCT). The reference illuminant is a Planckian radiator, model of daylight, or combination of those two, depending on the exact CCT of the test source. Three alternative reference schemes were considered: 1) either using all Planckian radiators or all daylight models, while maintaining the CCT match of the test and reference; 2) using only one of ten possible illuminants (Planckian, daylight, or equal energy), regardless of the CCT of the test source; 3) using an off-Planckian reference illuminant (i.e., a source with a negative D_{uv}).

No reference scheme is inherently superior to another, with differences in metric values largely a result of small differences in gamut shape of the reference alternatives. While using any of the alternative schemes is more reasonable in the TM-30-15 evaluation framework than it was with the CIE Test Color Method, the differences still ultimately manifest only as changes in interpretation of the results. Reference illuminants are employed in TM-30 to provide a familiar point of comparison, not to establish an ideal source.

PNNL-SA- 119373

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov

1 Introduction

The International Commission on Illumination (CIE) defines color rendering as the "effect of an illuminant on the color appearance of objects by conscious or subconscious comparison with their color appearance under a reference illuminant." Implicit in this definition is the fact that communicating the color rendering characteristics of a light source requires comparison with another light source, arguably one with familiar characteristics.

Almost all color rendering metrics that have been adopted or proposed over the past 50 years are based on comparing a test source to a reference illuminant—for a review see [Houser and others 2013]. This includes the CIE Test Color Method [CIE 1995], of which R_a (commonly known as CRI) is a principal component. A limited number of proposed metrics, ranging from the Preferred Color Index [Sanders 1959] to the Memory Colour Rendition Index [Smet and others 2010], have eschewed a traditional reference illuminant, although they have still employed a target for comparison. In most cases, the reference has been *relative*, where it is tied to the correlated color temperature (CCT) of the test source. Planckian radiation and CIE D Series illuminants are typical reference illuminants used in this scheme. Alternatively, a single *fixed* reference can be used for all test sources, regardless of chromaticity. For example, the Gamut Area Index (GAI) [Rea and Freyssinier-Nova 2008; Rea and Freyssinier 2010], uses the equal energy illuminant (CIE Standard Illuminant E).

IES TM-30-15 (TM-30) [David and others 2015; IES 2015] uses a relative reference scheme similar to that of the CIE Test Color Method. Some people have questioned whether a different reference scheme should have been used. The goal of this article is to explore, through discussion and numerical analysis, how the TM-30 output values—and interpretation of those values—would be different if an alternative reference scheme was used.

1.1 The IES TM-30-15 Reference Scheme

TM-30 is a relative-reference-based color rendition evaluation framework. For sources with a CCT of 4500 K or less, the reference is a Planckian radiator. At 5500 K or above, the reference illuminant is a CIE D Series illuminant, which is a mathematical model of daylight. Between 4500 K and 5500 K, the reference illuminant is a proportional blend of Planckian radiation and the D Series illuminant, each at the specified CCT. For example, at 4750 K, the reference illuminant is 75% Planckian radiation (at 4750 K) and 25% CIE D4750. The blended reference differs from the reference scheme used in the CIE Test Color Method, and was implemented to avoid a sharp transition at 5000 K, which could result in oddities for color-tuning products as well as manipulation to achieve a different score. Like the reference illuminants themselves, the blending range is, essentially, arbitrary.

Importantly, assigning any source as the reference may be misconstrued as implying that it renders colors in an ideal way—or "the way the colors are supposed to look." This is especially true with single-number measures, such as CIE R_a, where the reference source achieves the highest rating, and maximizing the value becomes a logical goal. When an evaluation framework with multiple measures, such as TM-30, is used, the potential to misconstrue the reference source as ideal is reduced. This is principally because not all of the included measures can be maximized simultaneously. Rather than just identifying the average magnitude of the difference between the test source and reference source, which is the singular purpose of average fidelity measures, TM-30's combination of measures allows one to identify if the test source increases or decreases saturation, on average, as well as how hue and saturation changes for specific colors.

While the TM-30 reference scheme is somewhat arbitrary, in that any source(s) could have been assigned as the reference and thus given Fidelity Index (R_f) and Gamut Index (R_g) values of 100, the choice was based on several practical considerations. Primarily, it builds on the familiar reference system of the CIE Test Color Method, maintaining a level of continuity to help ease adoption of the new evaluation framework. In addition to continuity, the choice of the IES Color Metrics Task Group to use Planckian radiation and the CIE D Series illuminants was made in part because they are mathematically defined international standards that can be calculated across a range of CCTs. Using different reference illuminants (for example, a set of illuminants with negative D_{uv} values) would have required inventing them first and then seeking recognition by a standards body. This would have been a difficult task, and the illuminants would not have been as familiar to observers as daylight and Planckian radiation. Inventing such illuminants may have been more important if there was a universally preferred chromaticity and if TM-30 included a universal color preference metric, but the premise of TM-30 is the exact opposite: it provides an evaluation framework, but does not identify a universally preferred type of color rendering or attempt to rank light sources.

While the specific, initial motivation for the choice of references used for the CIE Test Color Method is unknown, one possibility is that they were chosen in part because they are naturally occurring. More practically, the basic intended use of CIE Test Color Method was to compare fluorescent lamps to incandescent lamps, which emit light that is very similar to Planckian radiation. The fact that two different types of reference illuminants are used is mostly inconsequential, as shown in **Figure 1**, since the sun is a Planckian radiator; the only difference between Planckian radiation and a CIE D Series illuminant at the same CCT is due to filtering of light by the Earth's atmosphere. While the need to explicitly compare test sources to incandescent-like, continuous spectrum sources has been reduced by the decreased use of such products for architectural lighting, they remain a relatively familiar and consistent source. Besides that, daylight—though itself variable—is something that cannot be changed.

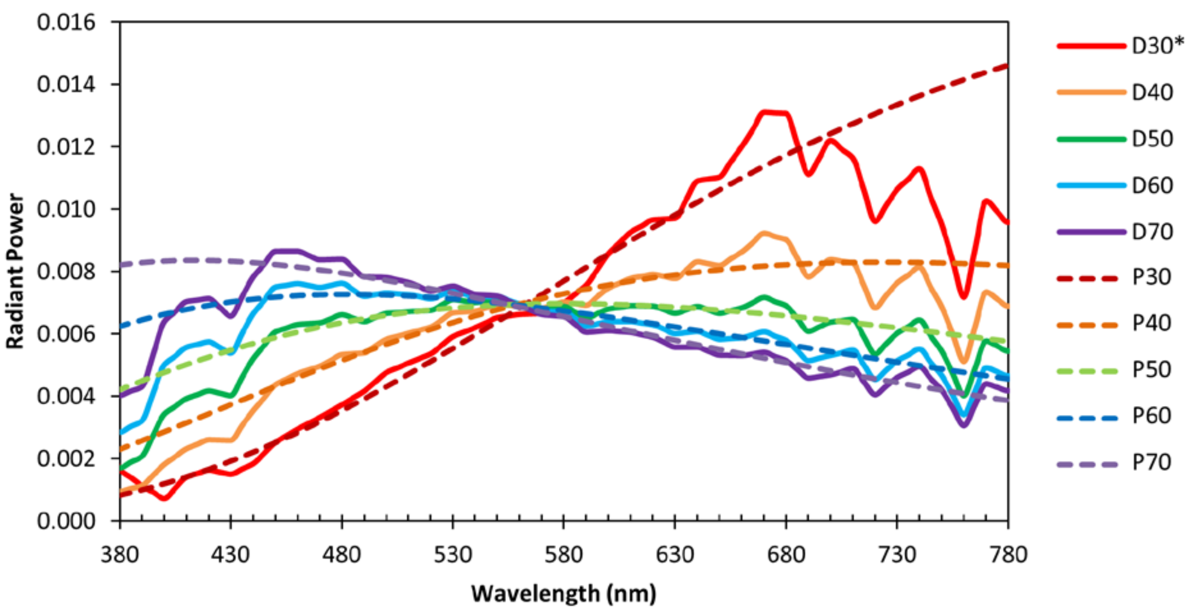


Figure 1. Comparison of Planckian radiation (P) and CIE D Series illuminants (D) at five correlated color temperatures (CCTs), with each source normalized for equal luminous flux. The abbreviations indicate the CCT; for example, P30 is Planckian radiation at 3000 K. At any given CCT, the difference between Planckian radiation and a CIE D Series illuminant is due to the filtering of light by the Earth's atmosphere. *D30 is not officially a CIE Standard Illuminant.

The idea of matching the CCT of the test and reference source was likely another important consideration when the relative reference scheme was first adopted. Conceptually, this distinction allows a specifier to choose the overall tone of an environment first, before subsequently evaluating color rendition. It also avoids introducing systematic differences into the values, which is discussed in later sections. Importantly, it was necessary because chromatic adaptation formulas were less refined than those that are currently used.

This article explores three possible alternatives to the TM-30 reference scheme: 1) using all Planckian radiation or all D Series illuminants as relative reference illuminants, 2) using one of several fixed reference illuminants, and 3) using an off-Planckian reference.

2 Methodology

A series of calculations was completed, each adjusting the reference in the documented TM-30 procedure to produce an alternative set of metric values. The reference was the only element of the TM-30 calculation procedure that was altered. This means that the scaling factor was not adjusted. For comparison purposes, the same modifications were made to CIE R_a, which allows for a better understanding of how the color space and chromatic adaptation models inherent to the two schemes [Smet and others 2015b; Xu and others 2016] are related to the results. A summary of the variations included in this calculation is provided in **Table 1**. The identifier given for each calculation is used in the subsequent discussion and figures, with the base measure (e.g., R_f) augmented with the alternative shown in parentheses [e.g., R_f(PXX)].

One notable absence from the set of calculations is a set of relative references with chromaticity not on the Planckian locus. This is because no such set of sources has been developed, likely because research into perception of off-Planckian sources is still nascent [Dikel and others 2013; Ohno and Fein 2014; Rea and Freyssinier 2013; Smet and others 2015a; Smet and others 2014; Wei and Houser 2016]. As such, this concept will only be considered through the use of a single, off-Planckian source, neodymium incandescent. Note that this reference source also increases average gamut area relative to a standard Planckian radiator at the same CCT.

For each of the reference alternatives, a complete set of TM-30 calculations was performed. That is, all metric values, from R_f and R_g to all 32 fidelity and chroma shift values for the 16 hue angle binswere calculated. The most interesting results are presented herein.

2.1 Scaling Factors and Data Interpretation

Typically applied in fidelity calculations but not in gamut area or chroma shift calculations, scaling factors adjust the range of scores by multiplying the raw average color difference. As a multiplier, they do not affect

Table 1. List of alternative reference scenarios considered.

Identifier	Type of Reference	On Planckian Locus?	Description
PXX	Relative	Yes	Planckian Radiation
DXX[1]	Relative	Yes[2]	CIE D Series Illuminant
P27	Fixed	Yes	2700 K Planckian Radiation
P30	Fixed	Yes	3000 K Planckian Radiation
P35	Fixed	Yes	3500 K Planckian Radiation
P40	Fixed	Yes	4000 K Planckian Radiation
P50	Fixed	Yes	5000 K Planckian Radiation
P65	Fixed	Yes	6500 K Planckian Radiation
D40	Fixed	Yes[2]	CIE D Series 4000 K
D50	Fixed	Yes[2]	CIE D Series 5000 K
D65	Fixed	Yes[2]	CIE D Series 6500 K
NI	Fixed	No	Neodymium Incandescent (2870 K)
P28	Fixed	Yes	2870 K Planckian Radiation

1. The CIE D Series is only defined at or above 4000 K. For this analysis, lower CCT sources were calculated using the same equation.
2. The CIE D Series is slightly above the Planckian locus.

the rank order of sources, and their main purpose is to help the range of typical values be more easily interpreted.

While adjusting scaling factors is helpful for addressing differences in the properties of standardized color samples for different methods, a change could also be made to help align scores from an alternative reference to be similar to the range of existing methods. In the past, a scaling factor has been determined based on the average fidelity of the CIE F Series illuminants [David and others 2015; Davis and Ohno 2010], although the scale could be based on the average value of any set of sources. For this analysis, the scaling factor was *not* adjusted for each alternative reference scheme, although it likely would be necessary to adjust if the alternative were to be implemented in practice. This approach helps to preserve transparency in the differences between alternatives, but also makes the mean difference between alternatives more pronounced. For many of the alternative schemes (PXX, DXX, D65. D50, P40, P50, P65, and EE) the scaling factor would be less than 2.6% different from the standard. In contrast, the scaling factor using P27 or neodymium incandescent as fixed references would need to be 22% and 37% less than the standard, respectively, to maintain equivalent ranges of scores compared to TM-30 R_f.

Although some mean differences are included in this article, the most important data to consider is the pattern in differences that result from other characteristics of the light sources, such as CCT, D_{uv}, or gamut shape. The analysis focuses on examining correlations as well as the range of differences across all included sources versus standard TM-30 calculations. Where ranges of differences are used, the central 95% is reported instead of the full range. A few sources in the dataset, such as 2000 K Planckian radiation, were often outliers; they help to show important patterns in scatterplots, but may obscure the true level of difference relevant to architectural lighting practice. The 95% criterion was arbitrarily chosen.

2.2 Sources

The set of test sources used was common to all calculations. It includes the 318 sources in the TM-30 Calculator Tool Library [IES 2015], which are a subset of those documented in [Houser and others 2013]. It also includes the 26 sources from [Royer and others 2016], for a total of 344 sources. These sources are not representative of the installed base or proportional to the quantities of various source types that are on the market today. They are not an exhaustive compilation of all types of sources, although they are a fairly comprehensive representation of what is currently available. They include commercial sources, modelled sources, and experimental sources, as well as some theoretical sources, such as the equal energy illuminant. **Figure 2** provides a breakdown of the sources, based on several characteristics. Understanding the underlying properties of the sources is critical for interpreting the results.

Importantly, the chromaticity of the sources (including CCT and D_{uv}) was not considered when building the dataset, and there are observable trends in the prevalence of different chromaticities. For example, the dataset is weighted toward products with CCTs of 3000 K, and 3500 K. Additionally, these nominal CCTs include many experimental SPDs that were created specifically to test color rendition perceptions; many of them exaggerate gamut or have chromaticity that is off the Planckian locus. Simply stated, this is not a randomly generated set of SPDs, nor intended as a random set of commonly available sources. Because of biases in the data set and different effects of the reference schemes at different chromaticities, mean values must be carefully interpreted.

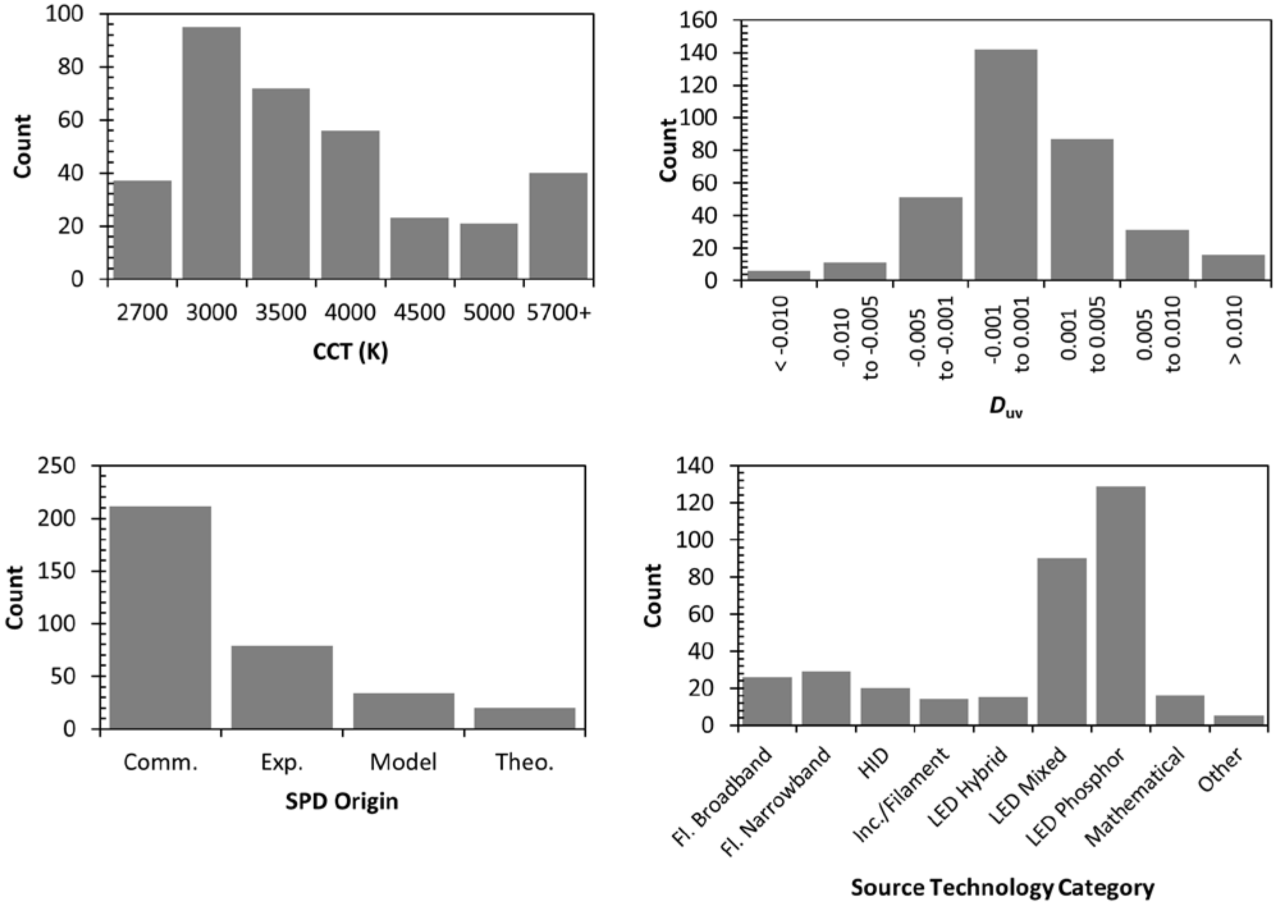


Figure 2. **Characterization of the 344 sources comprising the dataset used in this evaluation.**

3 Results

3.1 Evaluation of Different Relative Reference Illuminants

Perhaps the most straightforward alternatives to the TM-30 reference scheme would be to use only Planckian radiation (PXX) or CIE D Series (DXX) illuminants, regardless of CCT, and maintain the stipulation that the reference is always at the same CCT as the test source. Either would be simpler than the existing TM-30 and CIE Test Color Method reference schemes. Note that the CIE D Series is only officially defined for CCTs greater than or equal to 4000 K, but D Series illuminants at lower CCTs were calculated to meet the needs of this study.

For the DXX scenario (applying to sources with a CCT of less than 5500 K) the central 95% of differences were between -4.8 and 4.0 points for average fidelity (8.8 point range) and between -1.4 and 1.4 points for average gamut area (2.8 point range). Similarly, for the PXX scenario (applying to sources with a CCT of greater than 4500 K) the central 95% of differences were between -1.7 and 1.0 points for average fidelity (2.7 point range) and between -2.4 and 0.0 points for average gamut area (2.4 point range). It is important to note the range of CCTs to which these data apply, as well as the number of sources included, which was much smaller for the PXX scenario. Also, sources between 4500 K and 5500 K could have two possible alternate values, because in that range the TM-30 scheme uses a blended reference. Those sources were included in both calculations.

Results for both scenarios are shown in **Figure 3**. Note that the differences between each alternative and standard TM-30 calculations did not follow a normal distribution, owing to the tendency of sources in the dataset to have certain performance attributes (e.g., increases in yellow-green saturation) [Royer and others 2016]. In other words, the differences are not random, but due to specific features of the changing reference illuminants and the features of the light sources in the dataset.

Also shown in **Figure 3**, changes in fidelity for individual hue bins were variable, with some having a range of difference for the middle 95% as high as 15 points, and others as low as 5 points. The range in chroma shift values was as low as 3% and as high as 7%, depending on the hue bin. **Figure 4** shows standard TM-30 hue angle bin chroma shift and fidelity results for D40 and P65, which are compared to P40 and D65 (the standard references), respectively. This figure, which is critical to the analysis presented in this article, highlights the differences—small, yet notable—across the hue bins between the alternative references, which are not captured in an average gamut area value. These specific hue and chroma characterizations can be collectively referred to as gamut shape [Royer and others 2016]. For example, D Series illuminants around 4000 K tend to reduce chroma in the yellow region compared to Planckian radiation; in the alternative where D Series is the reference at low CCTs (DXX), sources would therefore tend to be rated as having greater yellow chroma shift. This is visible in **Figure 3**. Sources rated as having high fidelity in the yellow hue region (approximately hue angle bins 4 and 5) may be rated as having worse fidelity in that region with the alternative reference, but could also be rated as having better fidelity, if the shift made the source more similar to the new reference. This is true of other hue regions as well.

A deeper look reveals some interesting patterns in the data. **Figure 5** shows the difference between standard (TM-30 R_f and R_g) and alternate calculations for average fidelity, R_f(DXX) or R_f(PXX), and average gamut area, R_g(DXX) or R_g(PXX), versus the CCT of each source. Again, the charts combine the two alternative scenarios, because each only applies to half of the sources. Note that D_{uv} was also explored as an explanatory factor, but the range of chromaticity across D_{uv} in the dataset is much smaller than the range of chromaticity across CCT; with minimal chromaticity difference, no obvious patterns were detected.

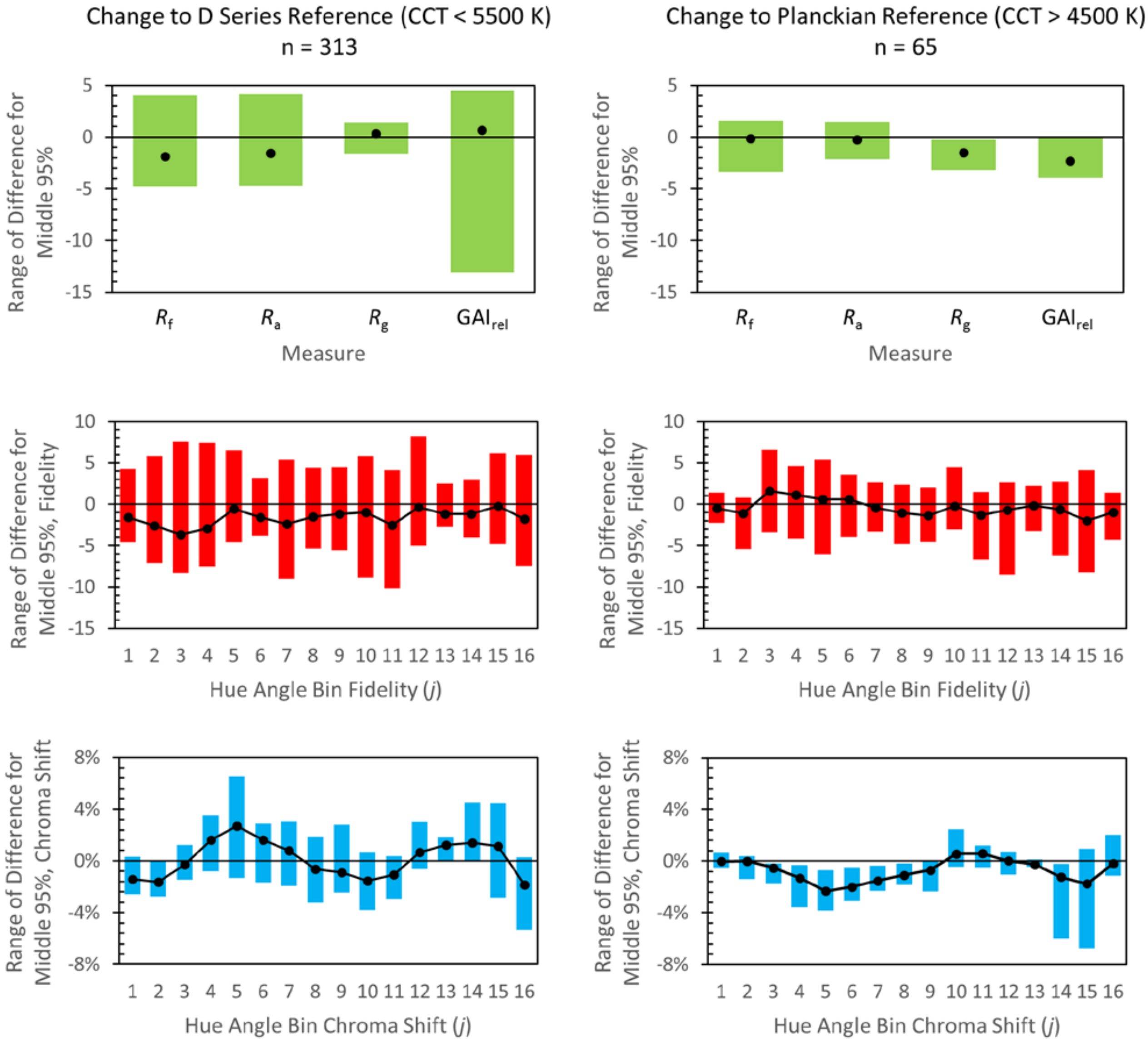


Figure 3. Difference in score (middle 95%) based on each of the TM-30 measures, along with CIE R_a and GAI_{rel}, for sources where the reference was changed from a D Series illuminant to a Planckian radiator (right) and for sources where the reference was changed from a Planckian radiator to a D Series illuminant (left). The black dots/lines indicate the mean values.

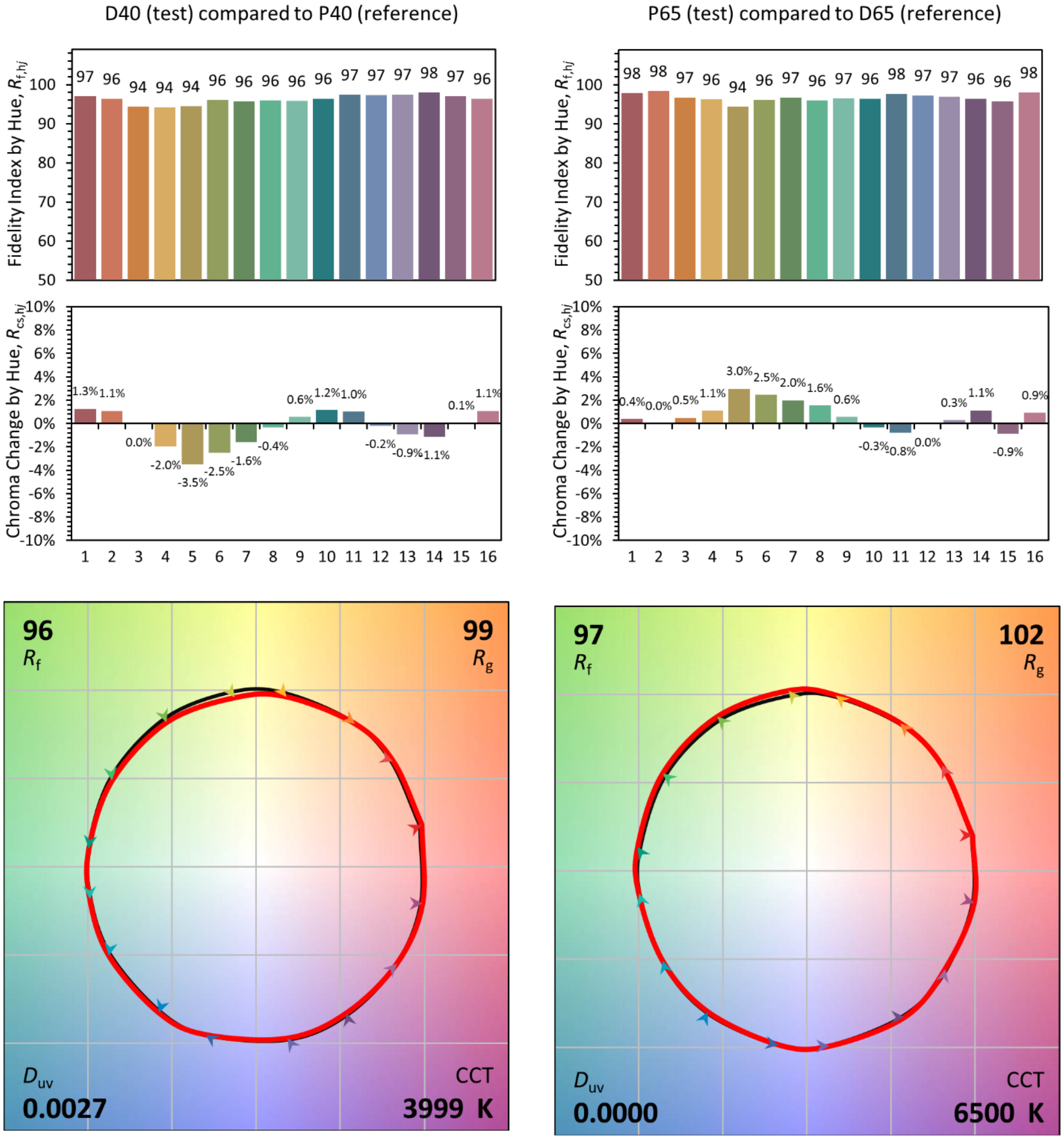
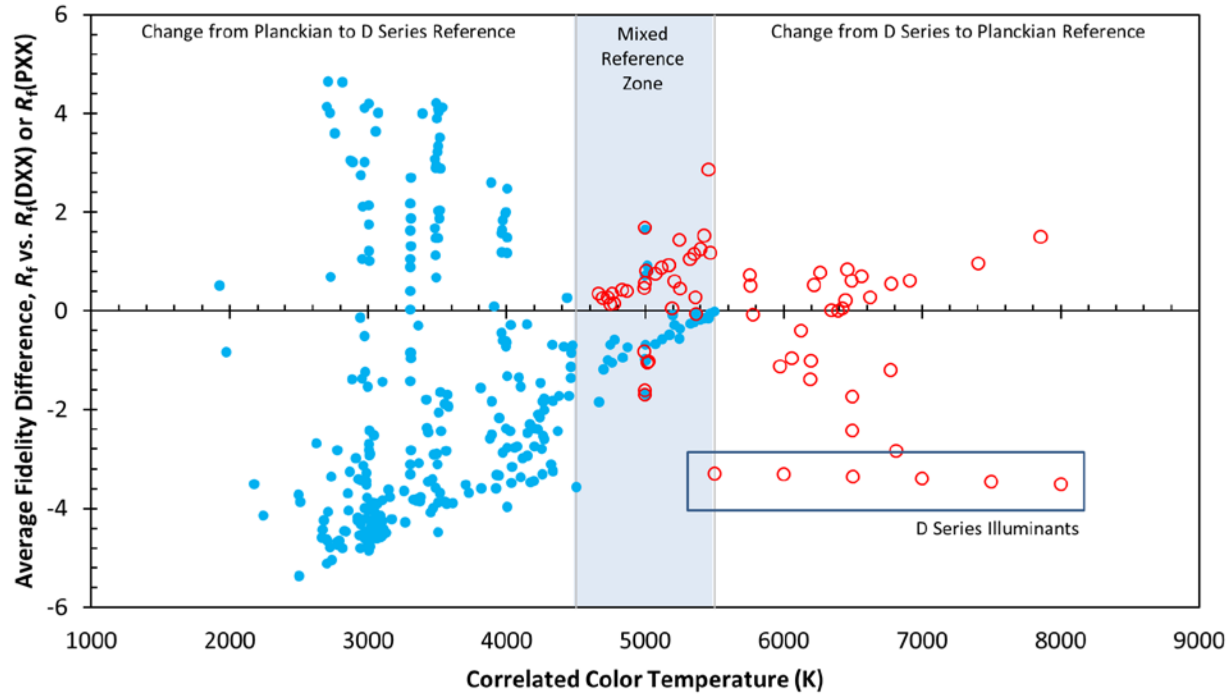


Figure 4. Color rendering over hue for two alternative reference illuminants compared to the standard TM-30 reference at that CCT. While average gamut is relatively constant between D Series illuminants and Planckian radiators, there is a slightly different gamut shape, which contributes to difference when alternate reference illuminants are used.



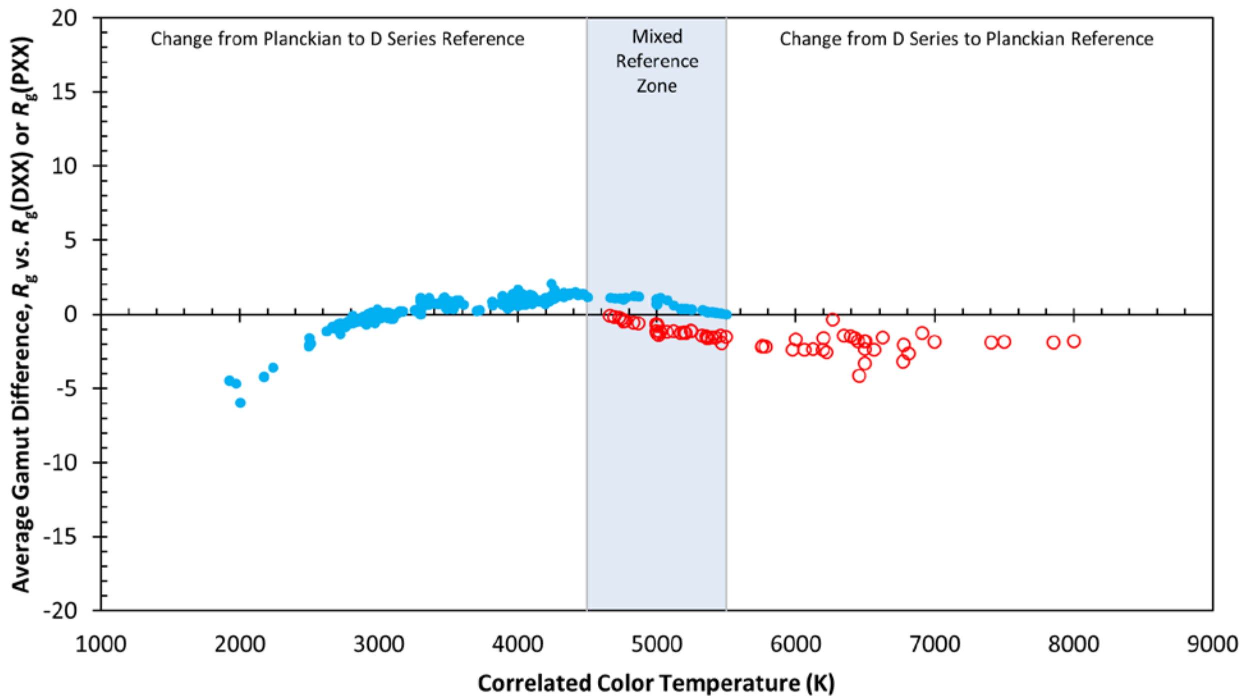


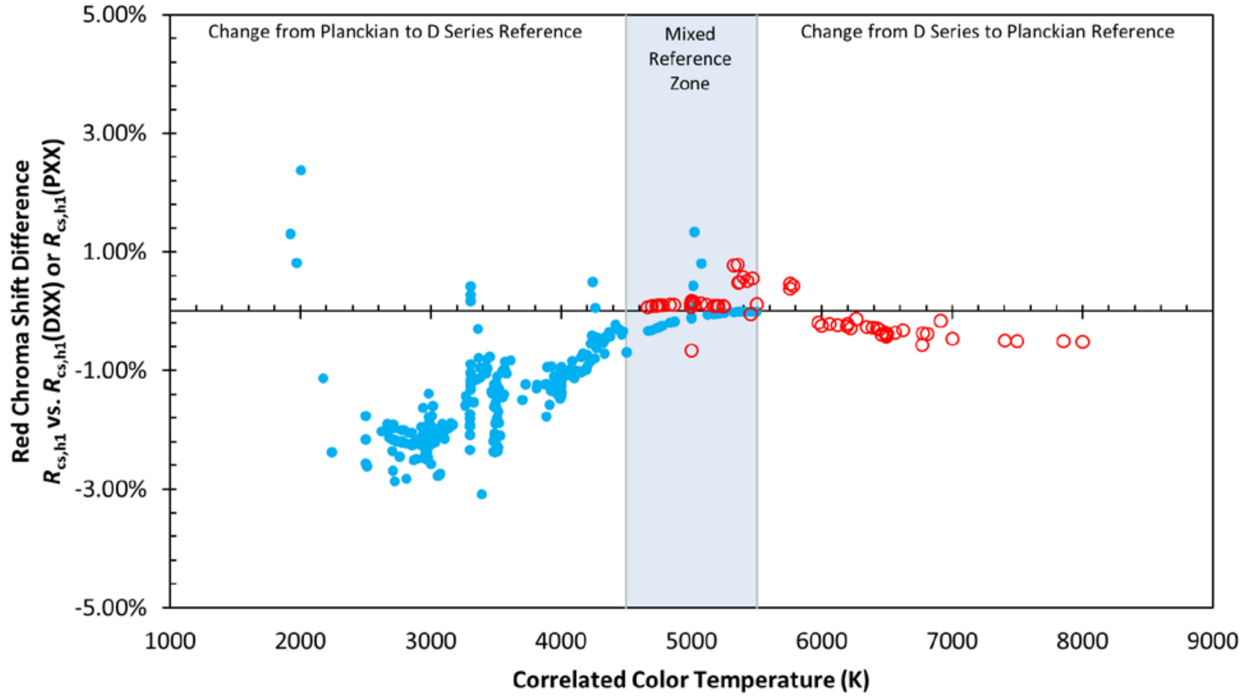
Figure 5. Difference between standard (TM-30 R_f and R_g) and alternate calculations for average fidelity (top) and average gamut area (bottom) versus the CCT of each source in the dataset.

Depending which gamut shape the test sources is more similar to overall—that of the alternative reference or the standard reference—the alternative average fidelity value may be higher or lower than the standard calculation. For example, a 4000 K source that increases red chroma is likely to have a higher average fidelity score with a D40 reference than a P40 reference, since D40 saturates reds compared to P40 (Figure 4). On the contrary, average gamut area is fairly similar for the two types of reference illuminants, with substantial differences only occurring below about 2500 K, where (unofficial) D Series illuminants have greater average gamut areas compared to Planckian radiators. Thus, the difference between R_g and R_g(DXX) or R_g(PXX) are relatively small.

Figure 6 shows the differences for standard versus alternative (PXX, DXX) chroma shift values for hue angle bins 1 and 5, based on the CCT of the test source. These two hue angle bins—nominally red and yellow, respectively—were selected to illustrate the possible outcomes among the 16 hue angle bins included in TM-30. As CCT is reduced, D Series illuminants have increasingly more red saturation than Planckian radiators, which results in lower red chroma shift values than standard calculations for almost every source in the dataset used for this analysis. In contrast, they have less yellow saturation, as documented in **Figure 4**. This is why the differences in chroma shift values tend to be homogenous for all test sources, which is unlike from the pattern seen for average fidelity (Figure 5, top), where the difference in value may be positive or negative depending on the specific test source. These results can be extended to other hue angle bins (not shown), with patterns reflecting the differences in gamut shape between Planckian and D Series illuminants at any given CCT. The complement the data in Figure 6, **Figure 7** shows fidelity differences for hue angle bins 1 and 5. The differences for hue angle bin fidelity values tend to mimic those for average fidelity.

From this set of data, it can be understood how gamut shape—the specific hue and chroma shifts—contributes to the CCT-dependent differences for average fidelity. **Figure 8** shows this in greater detail for those sources below 5000 K with the reference changed to a D Series illuminant (DXX)—this subset is shown because there is a larger sample size with a greater variety of performance. The difference in average fidelity [R_f(DXX) versus R_f] is related to $R_{cs,h1}$, the TM-30 measure for red saturation and a proxy for gamut shape. Forty-nine of 73 sources with $R_{cs,h1}$ values greater than or equal to 0% had R_f(DXX) values that were higher than standard R_f values—these were overwhelmingly color-mixed LED sources. Likewise, of the 79 total sources that had higher R_f(DXX) values, 49 had an $R_{cs,h1}$ value greater than or equal to zero. Note that $R_{cs,h1}$ is an incomplete characterization of gamut shape, which is likely why the correlation between average fidelity difference and $R_{cs,h1}$ is not higher.

Finally, it is important to note that very similar trends emerge if the same procedure is followed in deriving average fidelity alternatives to CIE R_a. However, a complementary gamut area measure, GAI_{rel}, did not perform similarly to TM-30 R_g under alternative reference schemes. GAI_{rel} is defined here as the area of the eight samples used to calculate CIE R_a in the 1976 $u'v'$ chromaticity diagram, compared to the area of the same samples with a reference at the same CCT. In other words, it is similar to the Gamut Area Index described by Rea and Freysinnier-Nova [2008, 2010], but with a relative reference instead of a fixed reference. Values for GAI_{rel} had a middle 95% range of -10.2 to 4.5 points for the DXX scenario, much larger than the range for TM-30 R_g. This results from the underlying mechanics of the calculations, including the color space and sample set. This is discussed further in the next section, and the issue of color space uniformity has also been explored by Smet and colleagues [Smet and others 2015b].



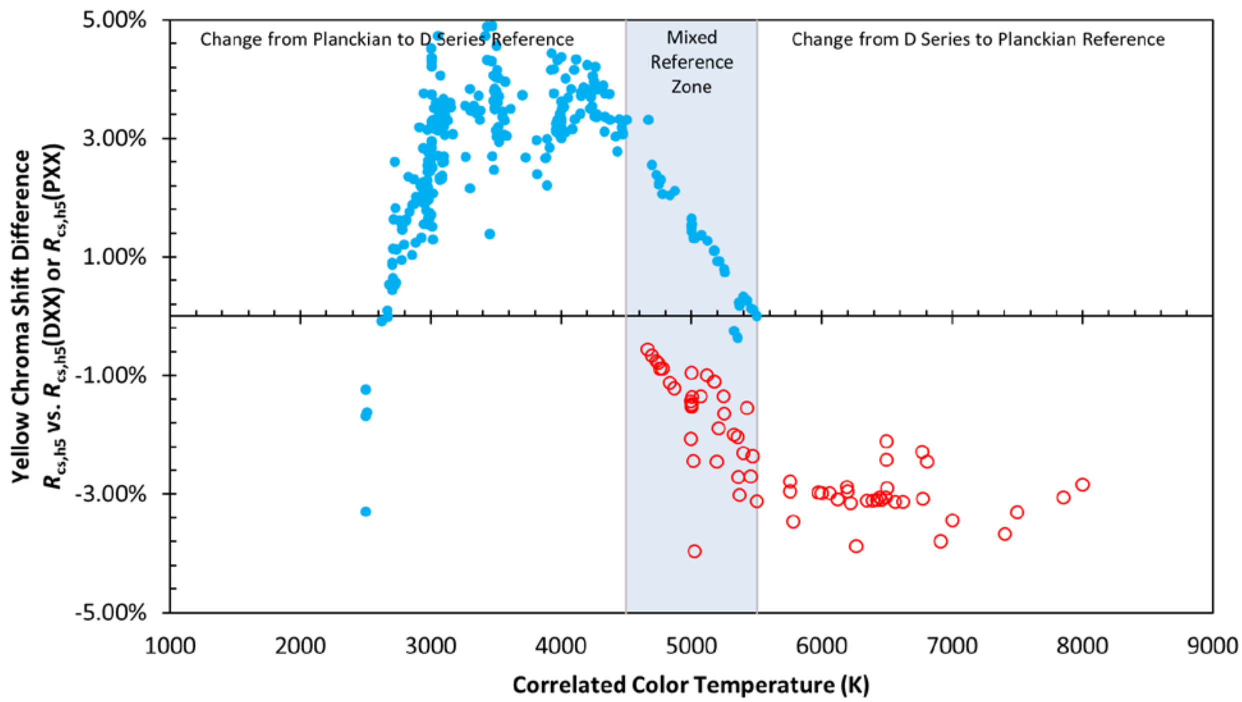


Figure 6. **Difference between standard (TM-30 $R_{cs,h1}$ and $R_{cs,h5}$) and alternate calculations for chroma shift in hue angle bins one (top) and five (bottom) versus the CCT of each source in the dataset.**

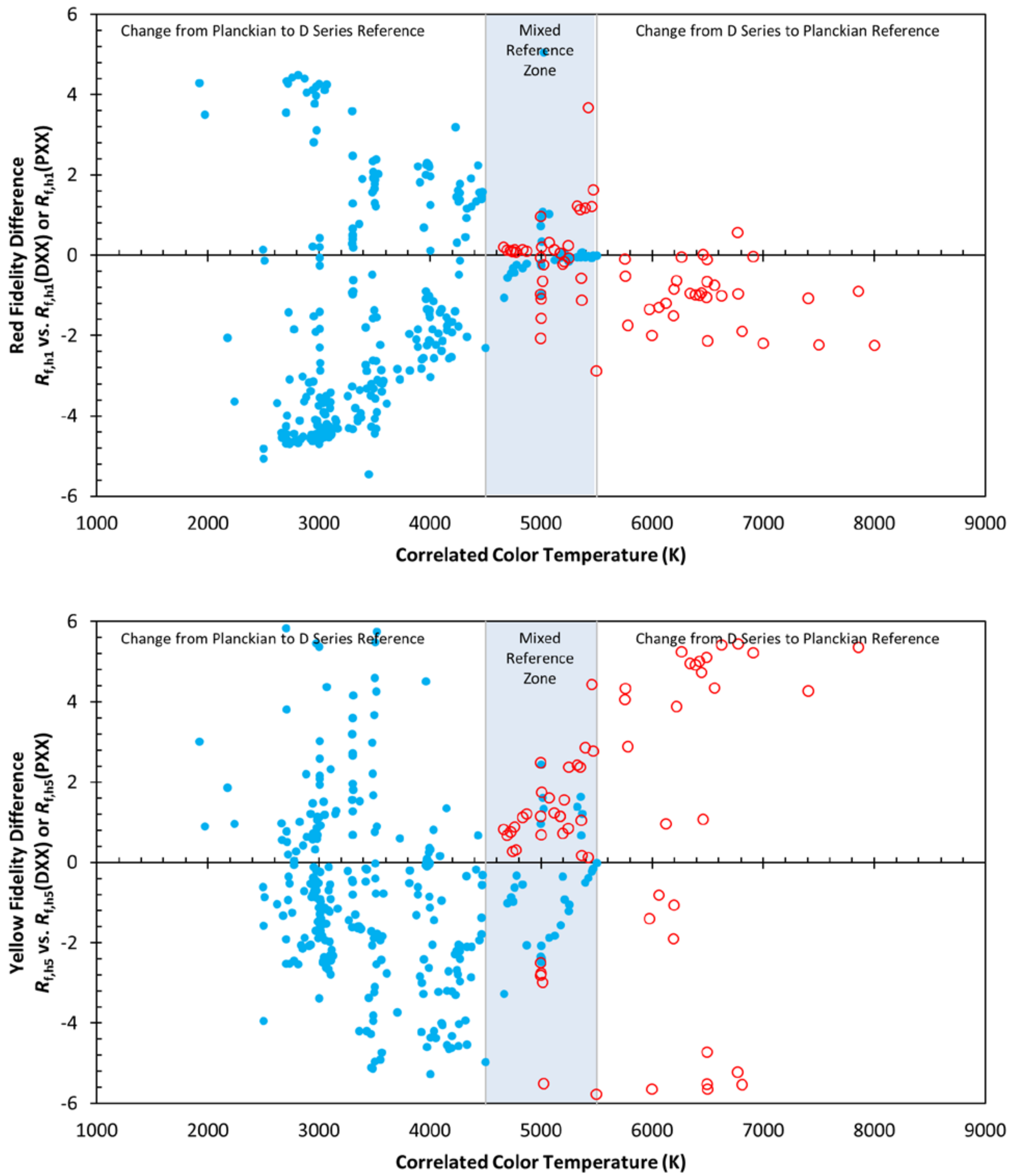


Figure 7. Difference between standard (TM-30 $R_{f,h1}$ and $R_{f,h5}$) and alternate calculations for fidelity in hue angle bins one (top) and five (bottom) versus the CCT of each source in the dataset. Differences for hue angle bin fidelity values tend to be more varied than for hue angle bin chroma shift values because they average multiple dimensions of color differently.

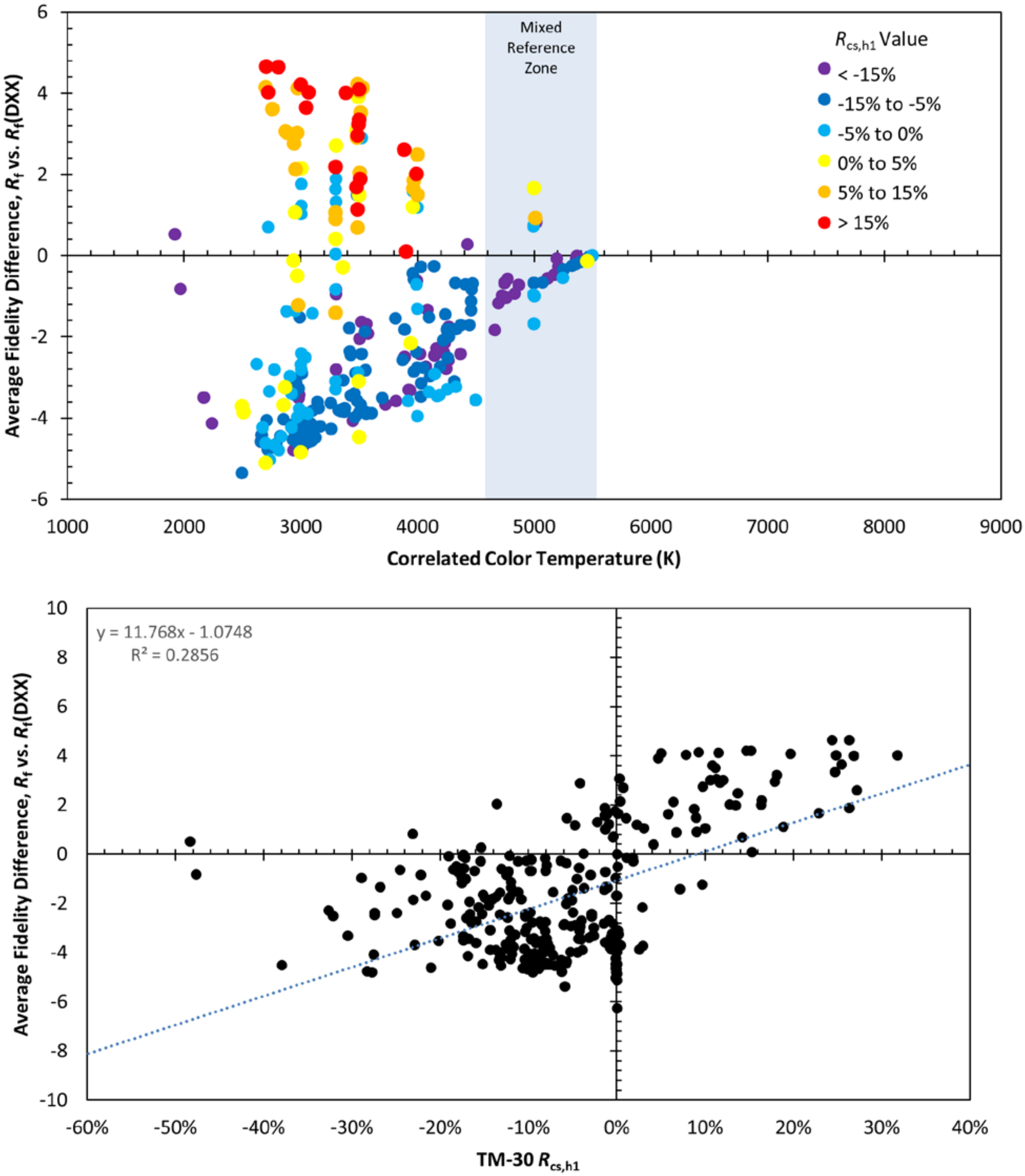


Figure 8. The difference in average fidelity value resulting from changing the reference from a Planckian radiator (or mixed) to a D Series illuminant is dependent on the gamut shape of the test source. In general, sources with greater red saturation have a higher average fidelity score with a D Series illuminant reference than a Planckian radiation reference.

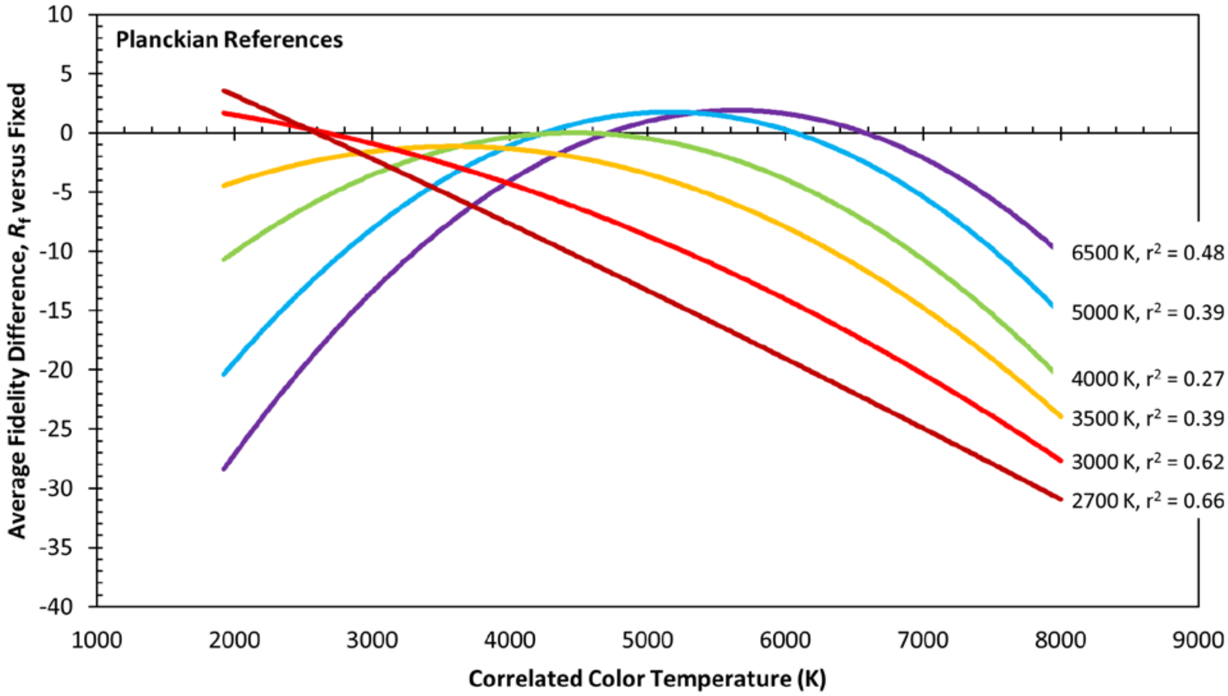
3.2 Evaluation of Different Fixed Reference Illuminants

A single reference illuminant, which does not vary with CCT, has been used for previously-proposed color rendition measures, and may be considered an alternate method to use for TM-30. This analysis does not focus on which fixed reference could be best; rather, it focuses on understanding the consequences of various selections.

Using a fixed reference results in increasing disparity between standard R_f values and the alternative average fidelity values as the difference in CCT between the test and reference increases, with the fixed-reference scheme generally resulting in lower values. This is demonstrated in **Figure 9**, where second-order polynomial best-fit lines for each of the alternatives considered are plotted, and in **Figure 10**, which compares the full dataset for two of the fixed-reference alternatives (3000 K and 6500 K Planckian radiation). Although not shown, the same trends in differences are also identifiable in the fidelity values for specific hue angle bins. The trend is more or less pronounced for different hues—likely due to gamut shape relationships, as previously discussed—but always present to some degree. The relationship between fidelity value and CCT that results from using a fixed reference is an important reason why a relative reference scheme was chosen for TM-30; the task group members felt that a relative scheme allowed easier comparisons that are more in-line with lighting design practice. That is, one does not have to mentally adjust the scale of "good" or "bad" when given a CCT.

When using the TM-30 framework, the amount of difference in average fidelity values for any given fixed reference versus the standard reference scheme may appear large, but it is small compared to a similar analysis based on CIE R_a. The range of the middle 95% of difference values for TM-30 R_f-derived values versus the alternatives was between 48% and 61% of the same range for CIE R_a-derived values, depending on the exact reference. That is, the difference was generally about twice as much for the CIE Test Color Method as it was for the TM-30 framework. This is due to non-uniformity of the CIE U*V*W* color space and the inaccuracies of the von Kries chromatic adaptation transformation used in the CIE Test Color Method. In the CIE Test Color Method, any of the fixed reference alternatives resulted in numerous average fidelity values that were more than 20 points lower than standard CIE R_a calculations. The disparity was typically smaller for fixed reference illuminants with higher CCTs than those with lower CCTs, likely due to the especially poor uniformity in the red region of the U*V*W* color space. This non-uniformity manifests itself in the large scale differences for R_9 compared to R_a; that is, an R_9 value of 0 is generally considered a rough equivalent to an R_a value of 80.

With regard to average gamut area, under the TM-30 framework the results are similar to those described in Section 3.1: average gamut area values were fairly stable, regardless of the reference source being used. Of the 11 fixed reference conditions considered, the greatest range for the middle 95% of difference values was 4.2 points (using the neodymium reference). This can be attributed to the fact that all the possible reference sources considered have approximately the same average gamut area in the CAM02-UCS, as shown in **Figure 11**. This is in stark contrast to the gamut area of the reference sources in the CIE 1976 $u'v'$ chromaticity diagram, also shown in **Figure 11**. Because of the non-uniformity of the CIE 1976 chromaticity diagram, the gamut area of lower-CCT sources is smaller, regardless of source type. Accordingly, values for fixed-reference alternatives to GAI_{rel} were drastically different depending on the CCT of the fixed reference source. If the CCT of the reference was lower, the values of the test sources were greatly inflated, and if the CCT of the reference was higher, the values of the test sources were greatly deflated. This is one of the major limitations of the previously proposed GAI metric that uses a fixed reference (equal energy illuminant): it systematically reduces average gamut values for sources lower than about 4000 K.



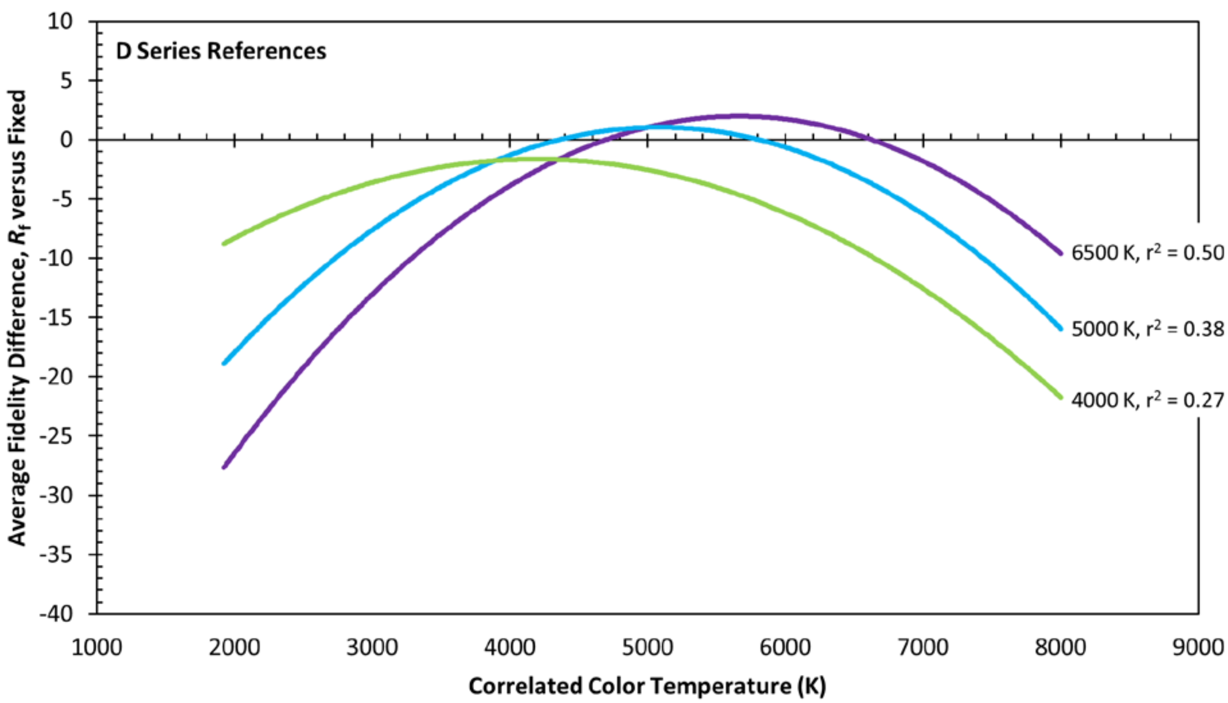


Figure 9. Best fit lines for the difference in average fidelity score for nine fixed reference illuminants versus standard TM-30 calculations. Differences for all 344 SPDs in the dataset (not shown), were used to determine the second-order polynomial best fit lines. The r^2 values ranged from 0.26 to 0.66.

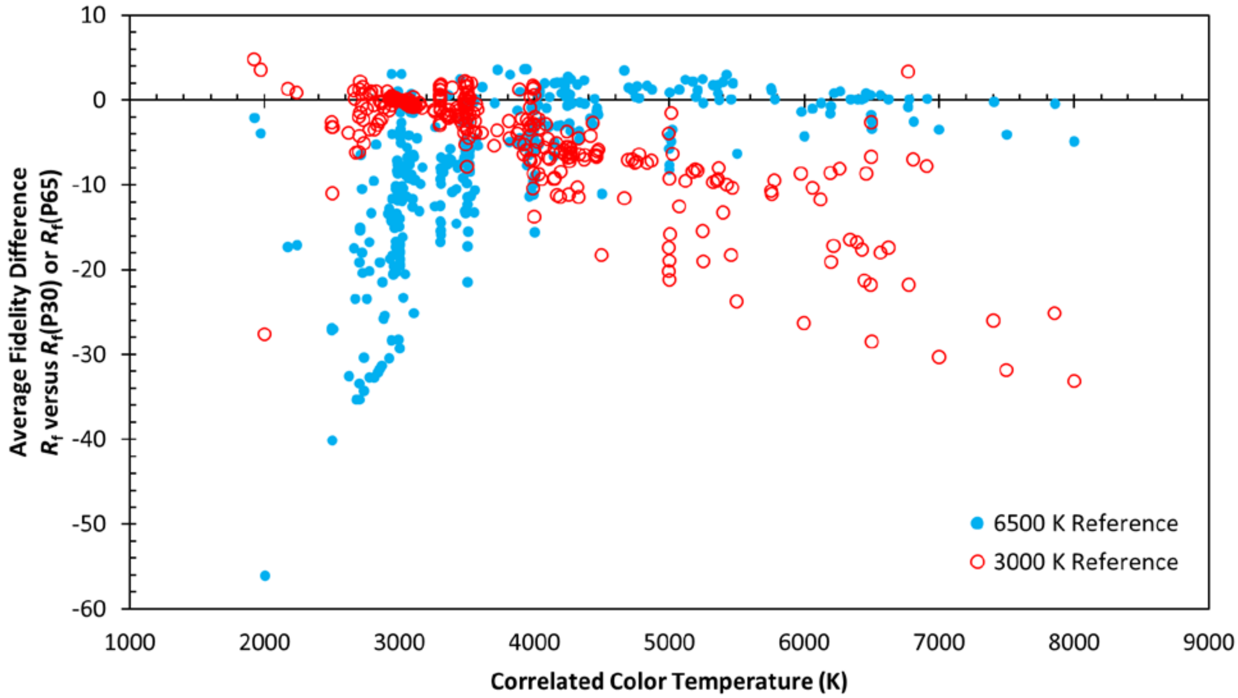


Figure 10. Difference in average fidelity for each of the 344 SPDs using a fixed reference of 3000 K Planckian radiation (P30) and 6500 K Planckian radiation (P65) versus the standard TM-30 calculation procedure. As the difference between the CCT of the fixed reference and the CCT of the test source increases, the range of difference in average fidelity for the two calculations also increases.

Importantly, these differences are unrelated to scaling factors, which would not address these concerns and are generally not used for average gamut area measures.

Despite the constant values for R_g-based average gamut area values regardless of the fixed reference, there are notable differences for chroma shift in different hue angle bins, as shown in **Figures 12 and 13**. Each displays the pattern of change across the 16 hue angle bins for the fixed reference (3000 K or 5000 K Planckian radiation) versus the standard TM-30 calculations. Two key patterns emerge. First, both sets show that the CCT of the fixed reference is a key point, with differences in many of the hue angle bins tending to change from positive to negative change at that value. This is particularly pronounced for hue angle bins 1, 2, 5, 6, 15, and 16. Second, there is a clear trend for hue angle bins 1 through 6 and 15 and 16, whereas there is less or no relationship between CCT and the difference in chroma shift for bins 7 through 14. Looking more closely at bins 1 through 6, it can be observed that as red saturation increases, yellow saturation decreases. These trends are probably not an artifact of the color space, but a reflection of the energy distribution at different CCTs (Figure 1), combined with the realities of human vision and chromatic adaptation.

Another pattern is shown in **Figure 14**, which shows the change in chroma shift for hue angle bin 1 (red) for nine of the evaluated fixed reference alternatives. Somewhat intuitively, red chroma shift values (for example, $R_{cs,h1}$) tend to be lower when the CCT of the fixed reference is lower and higher when the CCT of the fixed reference is higher. In contrast (not shown), yellow chroma shift values are higher when the CCT of the fixed reference is lower. Green and blue chroma shift values are relatively more stable, with little change regardless of the CCT of the reference illuminant.

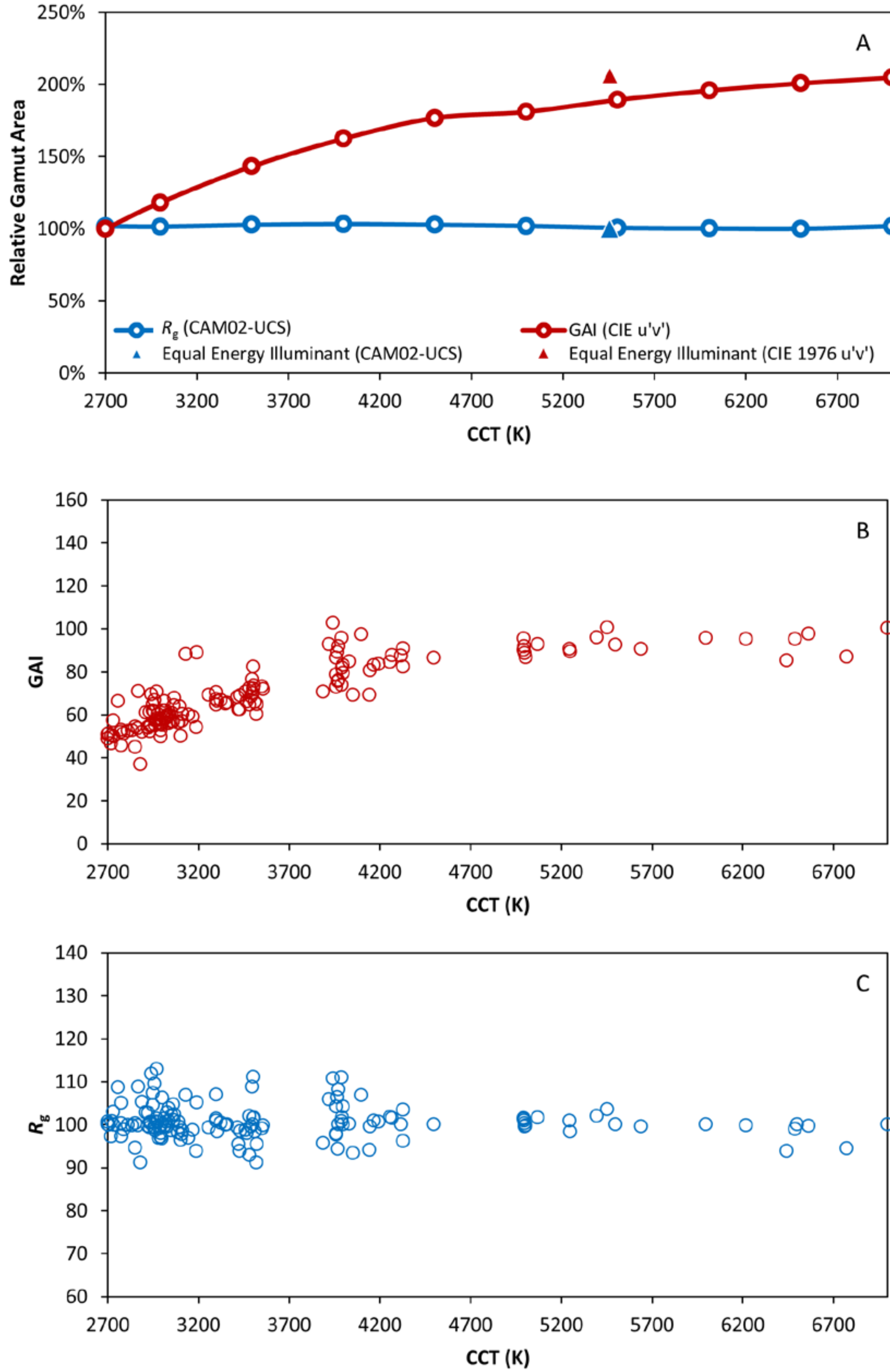


Figure 11. **A:** Relative area of Planckian (<5000 K) and D Series (≥ 5000 K) reference illuminants, as well as the Equal Energy Illuminant, in two color spaces (CIE u'v' or CAM02-UCS) using their respective calculation methods (GAI or R_g). The average gamut area of the broad-emitting reference illuminants stays constant in the R_g/CAM02-UCS scheme, but not in the GAI/CIE u'v' scheme. **B/C:** Commercial sources follow the same pattern for gamut area versus CCT, which systematically lowers scores for sources with lower CCTs in the GAI/CIE u'v' scheme.

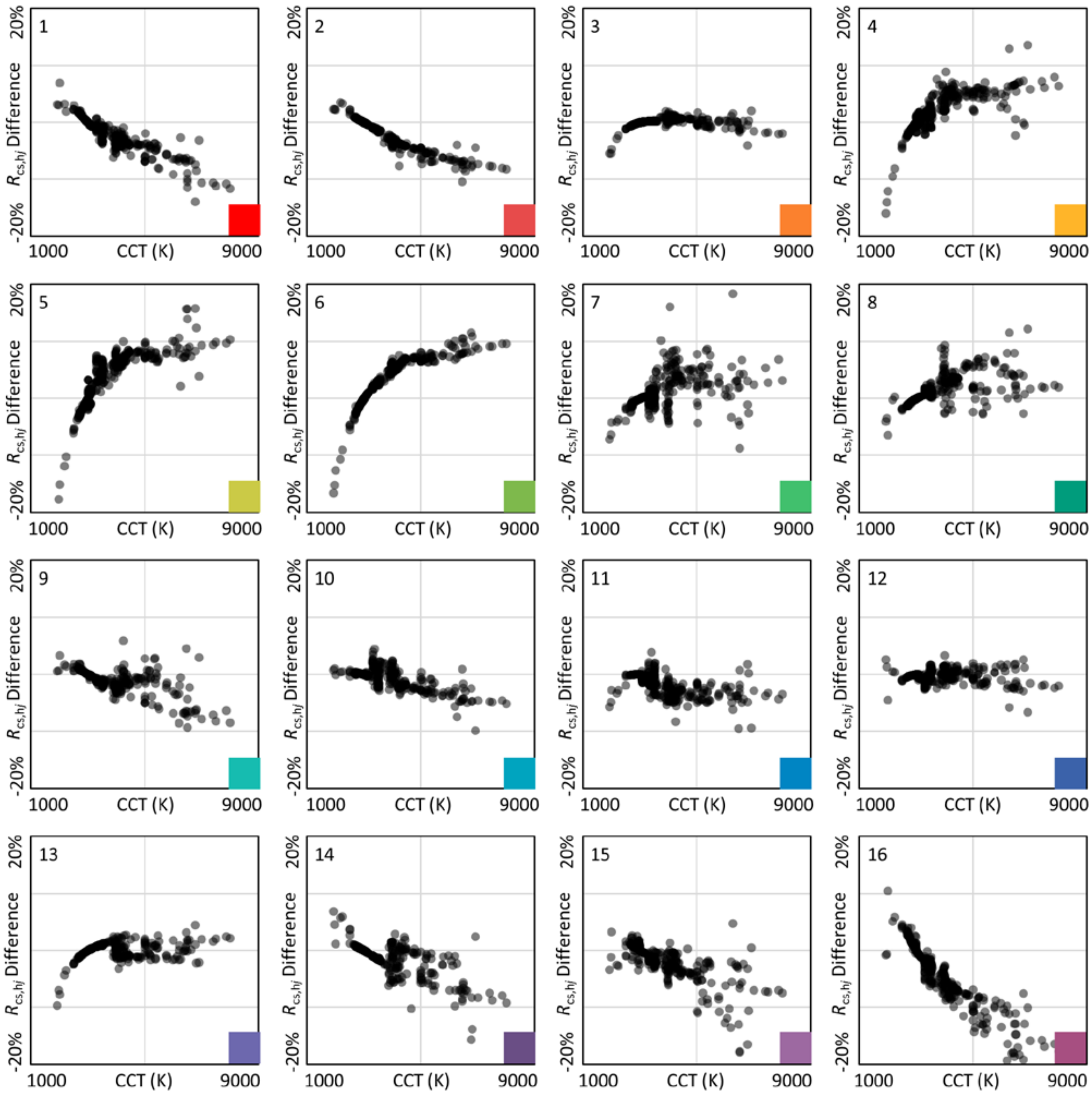


Figure 12. Differences between standard TM-30 hue angle bin chroma shift values and equivalent values using a fixed, 3000 K Planckian radiation reference. The number in the upper left for each plot is the hue angle bin designation. The points are plotted with transparency to show density. This figure is intended to show how the variation in chroma shift changes from bin to bin, with a consistent transformation going around the hue circle. It is intended to be viewed in conjunction with Figure 13, which shows the same data for a fixed, 5000 K Planckian radiation reference.

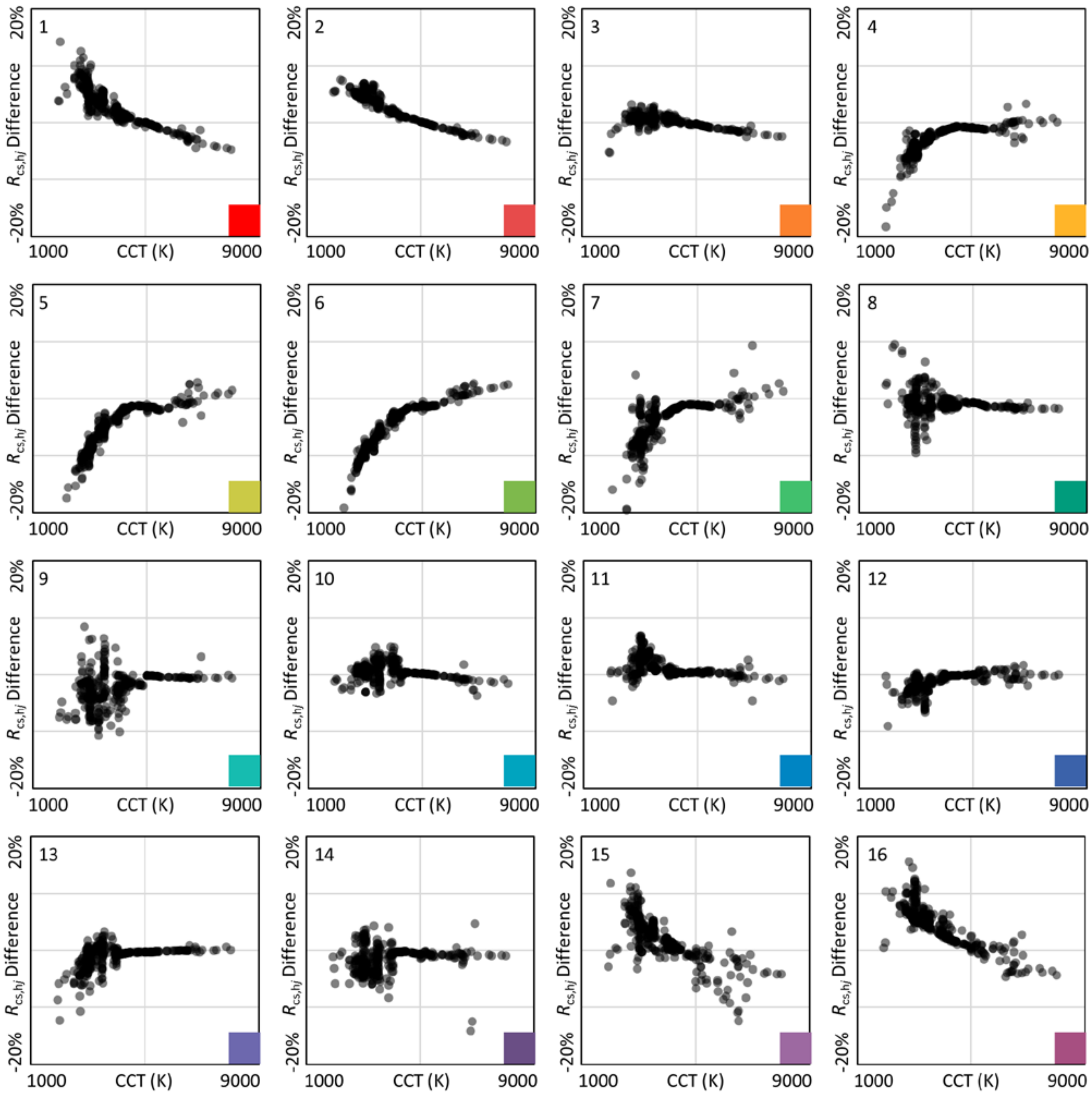


Figure 13. Differences between standard TM-30 hue angle bin chroma shift values and equivalent values using a fixed, 5000 K. The number in the upper left for each plot is the hue angle bin designation. The points are plotted with transparency to show density. This figure is intended to show how the variation in chroma shift changes from bin to bin, with a consistent transformation going around the hue circle. It is intended to be viewed in conjunction with Figure 12, which shows the same data for a fixed, 3000 K Planckian radiation reference.

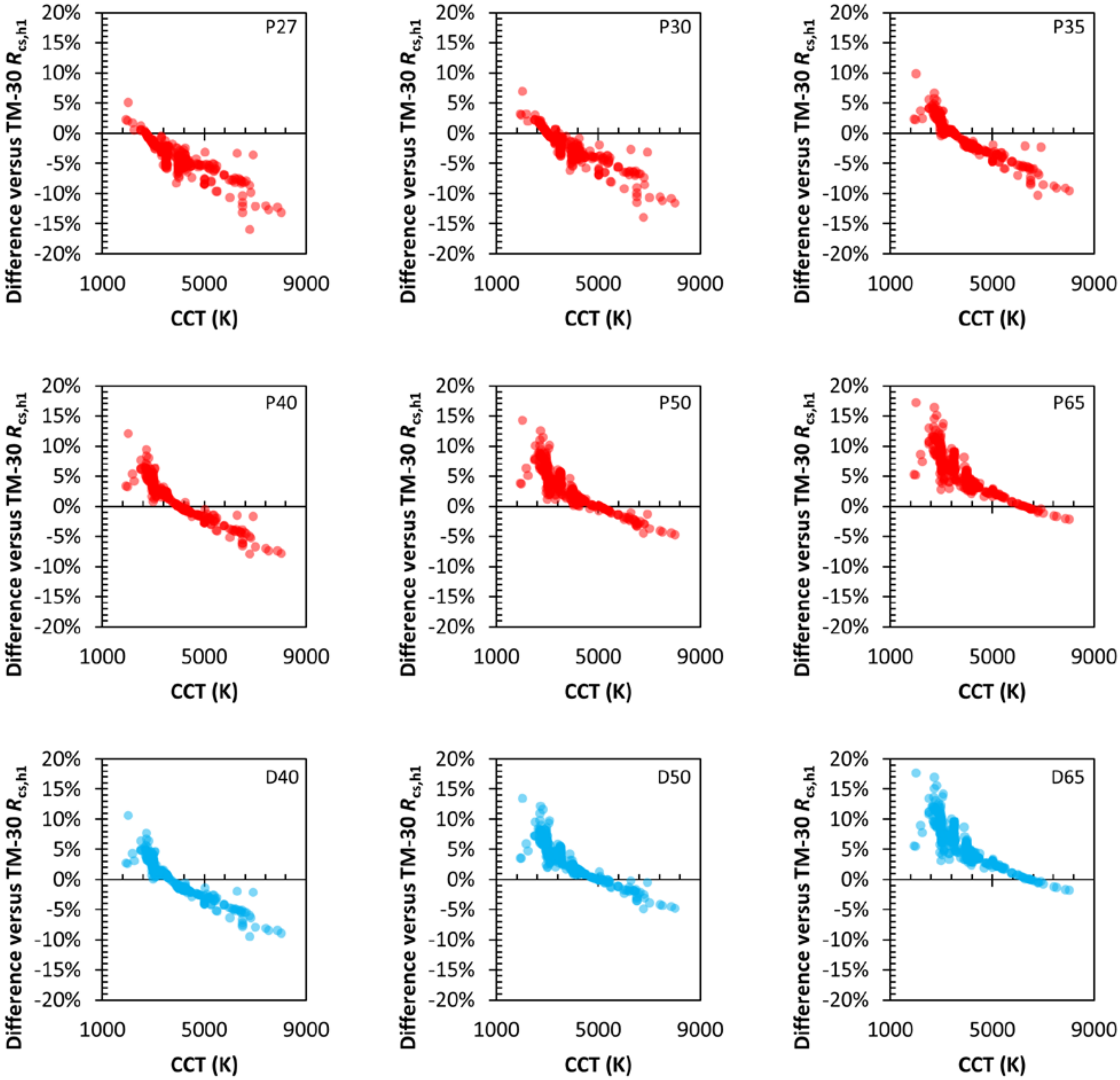


Figure 14. Difference between standard TM-30 and fixed-reference alternative calculations for hue angle bin 1 chroma shift. The points are plotted with transparency to show density. The relative pattern of change is the same from reference to reference: test sources with a lower CCT than the fixed reference have a higher $R_{cs,h1}$ value than the standard calculation, and test sources with a higher CCT than the fixed reference have a lower $R_{cs,h1}$ value.

3.3 Conditional Evaluation of Off-Planckian Reference

Because there is no series of mathematically-derived, off-Planckian reference sources, it is not easy to evaluate an alternative reference scheme that uses relative, off-Planckian reference sources. For this analysis, neodymium incandescent was chosen as a fixed reference, with the results compared to both the standard TM-30 calculations as well as a scheme with a fixed, on-Planckian reference with the same CCT as the chosen neodymium incandescent lamp (2870 K). Comparing the standard TM-30 results to the neodymium-reference results (NI) includes two changes—moving to a fixed reference and to an off-Planckian reference—which makes it difficult to understand the exact effect of simply moving the reference to below the Planckian. Also, it is critical to note that neodymium incandescent lamps increase the average gamut area compared to a standard incandescent lamp, and in particular saturated reds and greens. Much of the effects described in this section are a function of this behavior, rather than the off-Planckian chromaticity, although off-Planckian sources often increase average gamut area [Wei and Houser 2016].

Figure 15 shows a standard plot of R_f versus R_g for a subset of the 344 products, identified by source technology type. The figure includes 252 total products, including:

- 129 phosphor LEDs, relying on either a blue LED pump or a violet LED pump.
- 64 color mixed LEDs, combining three or more discrete LEDs.
- 15 hybrid LEDs, combining a phosphor LED with at least one additional discrete LED.
- 10 standard halogen/incandescent lamps.
- 4 halogen/incandescent lamps with additional filtering to alter the spectrum, such as a neodymium coating.
- 7-series linear fluorescent lamps (CIE R_a in the 70s).
- 14 8-series linear fluorescent lamps (CIE R_a in the 80s).
- 2 9-series linear fluorescent lamps (CIE R_a in the 90s).
- 10 metal halide lamps, including ceramic and quartz metal halide.

In short, as average fidelity is reduced, the range of average gamut area values increases. This is a basic tenet of the TM-30 two-axis system of average fidelity and average gamut area. **Figure 16** shows the same information, but with an alternative reference, fixed Planckian radiation at 2870 K (P28). This serves as an intermediate step toward understanding the effect of a fixed, off-Planckian reference. Comparing the two plots, it can be seen that most sources (79%) have a lower fidelity value, but little change in average gamut area. The mean difference in average fidelity was -5.9 points, and the middle 95% range was -27.5 to 2.5. The mean difference could be addressed by changing the scaling factor, but the large range could not. The mean difference in average gamut area was -3.3 points, with a middle 95% range of -5.3 to -2.1. As previously identified, the change in fidelity value is strongly related to CCT, and somewhat related to gamut shape, as is the case any time a fixed reference is used.

Figure 17 is again the same type of chart, but using the fixed, off-Planckian reference (neodymium incandescent). Again, neodymium incandescent is somewhat typical of sources with a negative D_{uv}, in that it increases saturation [Wei and Houser 2016]. Accordingly, compared to the values represented in **Figure 15**, 88% of average fidelity values and 100% of the average gamut area values are lower when using the neodymium reference instead of the standard TM-30 reference. The 12% of sources that see an increase in fidelity have an average gamut area and gamut shape that are more similar to neodymium incandescent, which strongly saturates reds and greens. In comparison to the previously described fixed-reference condition, P2870 (Figure 16), the numbers are the same: 88% of average fidelity values and 100% of average gamut area values are lower. This illustrates the substantial effect that moving to a saturating, off-Planckian reference would have on familiar characterizations.

The interpretation of **Figures 15 to 17** is constant, but because the reference changes, the meaning changes. This is, essentially, an exaggeration of the previous cases studied in this report. Although only one (saturating) off-Planckian source was examined herein, the trends would be the same for other similar off-Planckian sources at different CCTs. Generally, an off-Planckian reference series would mean that existing sources would be rated as desaturating, with lower fidelity values—assuming the reference series follows trends for negative D_{uv} sources by increasing saturation. Such a change would require a recalibration of expectations and revision of existing specifications. One challenge is that most people are not as familiar with the color rendering of neodymium-like sources (i.e., negative D_{uv} and saturating colors) as they are with standard incandescent sources. Further, daylight would then be rated as having modest fidelity and being desaturating, which is likely to seem counterintuitive to many. One possibility for defining a set of off-Planckian reference sources would be to more thoroughly document and model the SPDs of daylight at various times of day, locations, and atmospheric conditions.

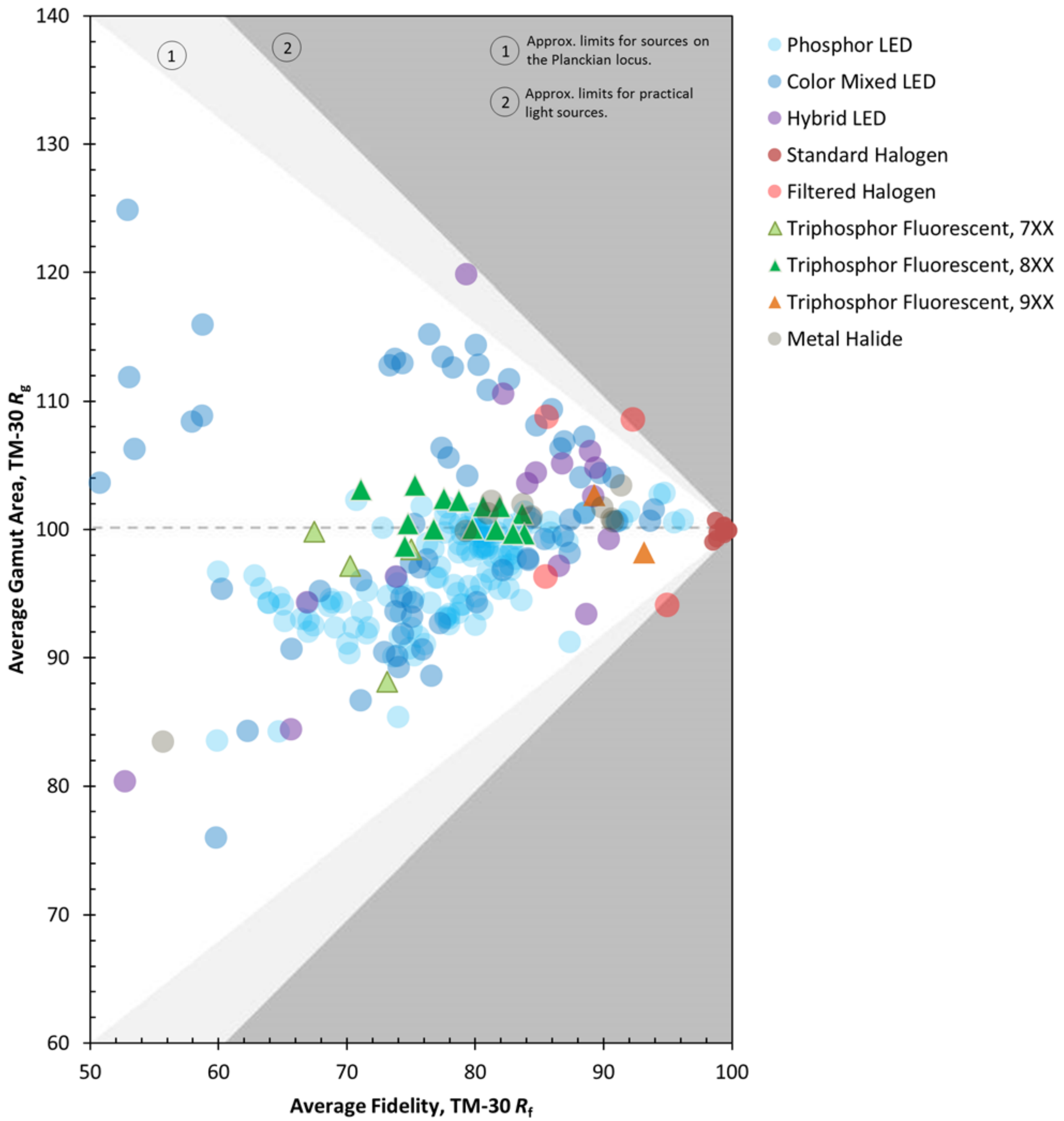


Figure 15. Standard plot of TM-30 R_g versus TM-30 R_f for a sample of 252 sources, identified by type.

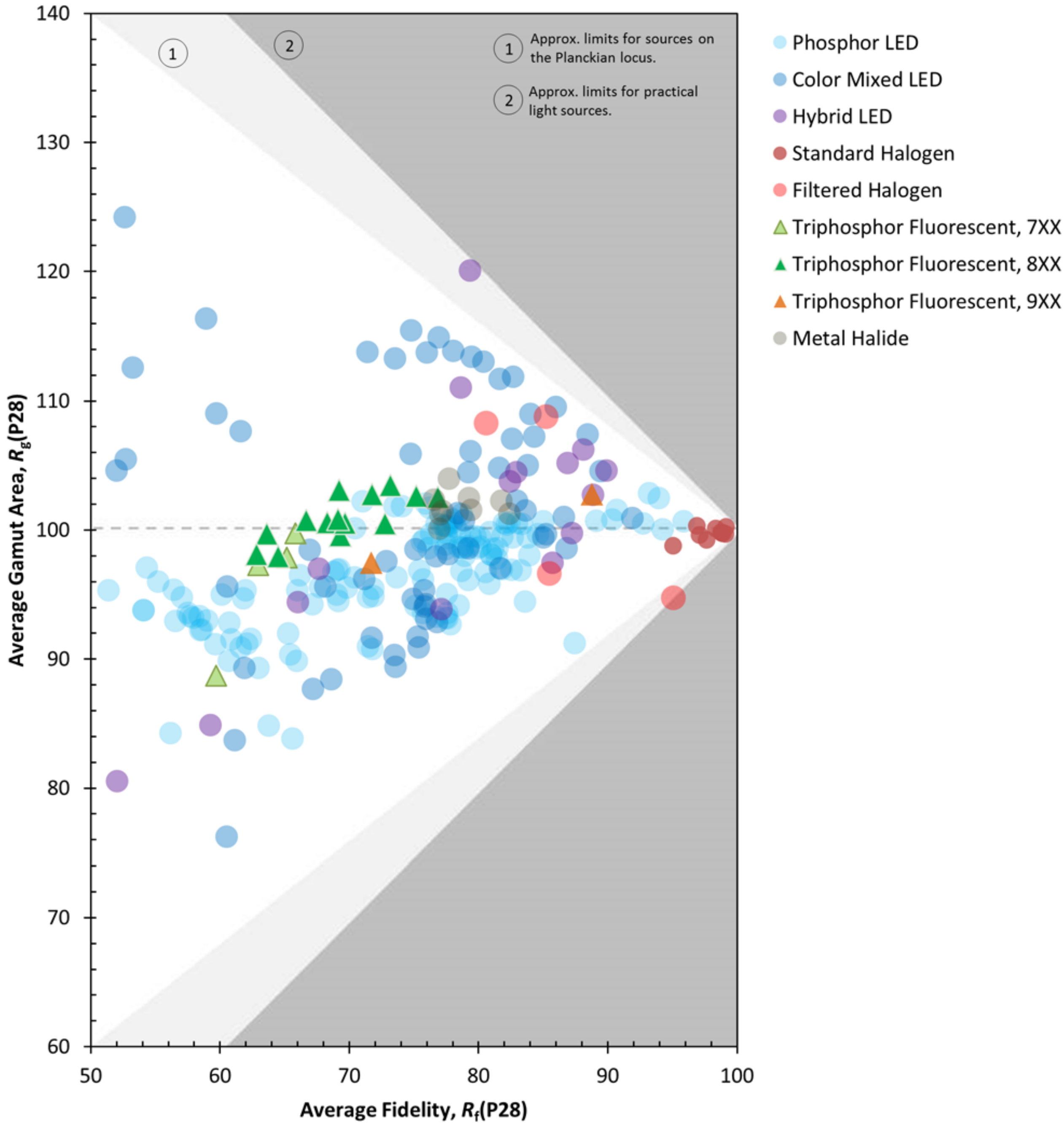


Figure 16. Altered plot of average gamut versus average fidelity, using a fixed reference of 2870 K Planckian radiation (P28), for the same sample of 252 sources shown in Figure 15. While average gamut remains relatively constant between Figures 15 and 16, average fidelity is generally reduced, depending on the difference in CCT between the test and reference source, as well as the gamut shape.

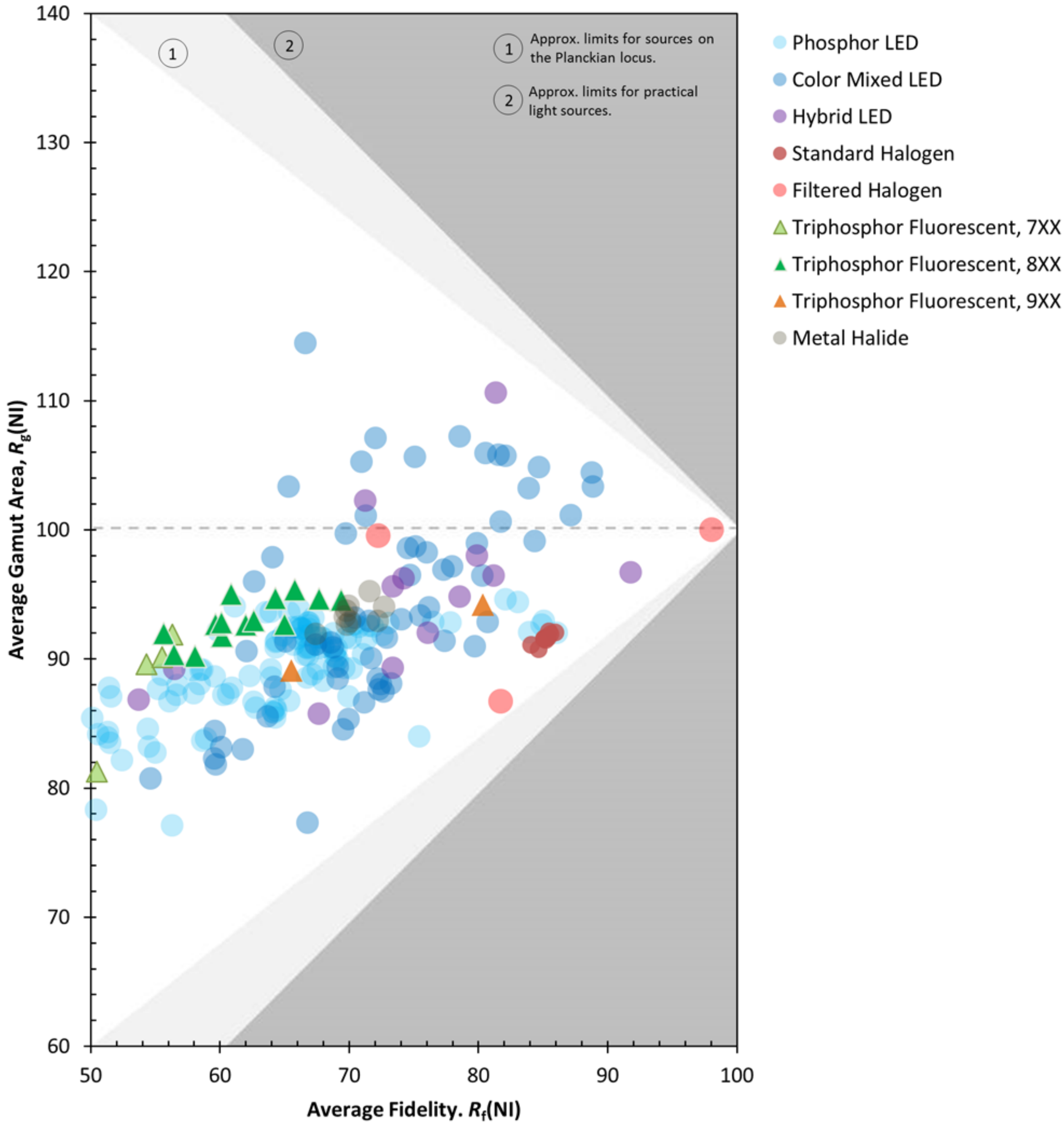


Figure 17. Altered plot of average gamut versus average fidelity, using a fixed reference of 2870 K neodymium incandescent (off-Planckian, saturating), for the same sample of 252 sources shown in Figures 15 and 16. With a reference illuminant that is saturating, like off-Planckian sources tend to be, the rated average gamut values for all other sources decrease, with almost all being classified as desaturating.

4 Discussion

Planckian radiation and CIE D Series illuminants do not necessarily provide ideal color rendering for architectural interiors, but they are familiar light sources which allow lighting specifiers, manufacturers, and others to formulate an idea of how a test source will make a space appear. Because TM-30 provides a much-improved framework for understanding how a test source is different than the reference—along with relying on multiple average measures that cannot be simultaneously maximized—the mistake of assuming "a higher number is better" is less likely to be made. Still, the reference illuminants used in a color rendering evaluation framework are an important consideration that affects interpretation of the results, and their role must be understood in order to best apply the results to lighting practice. Examining alternatives is one way to gain this knowledge.

This analysis considered three alternatives to the standard TM-30 reference scheme, each of which had varying levels of effect on the output values of the TM-30 evaluation framework. First, using all Planckian radiation or all D Series illuminants in a relative reference scheme had relatively small effects on average fidelity (R_f) values for the 344 light sources in the dataset, and almost no effect on average gamut area values (R_g). There are somewhat greater differences for fidelity and chroma shift of individual hue bins. These results stem from the small differences in color rendition between Planckian radiation and D Series illuminants at a given CCT (**Figure 4**). It is unclear if one of these alternatives would better match human perception. It is probably indeterminable given the small differences and the lack of direct correlation between average measures of color rendition and perceptual attributes, such as normalness, saturation, or preference [Royer and others 2016]. Use of either all Planckian or all D Series reference illuminants in a relative scheme is plausible, though any potential advantages in simplicity might be outweighed by changes to convention.

Second, changing to a fixed reference scheme, regardless of the chromaticity of the chosen reference, has a much larger effect than the first scenario on all output measures except average gamut, which again remains relatively constant in the TM-30 framework. Regardless of the improved chromatic adaptation transformation included in TM-30, there are predicted differences in color appearance of the samples under different reference illuminants (as intended). This should not be a surprise, otherwise there would be no validity to using CCT to establish the tone of an environment. The end result is that using a fixed reference in the TM-30 framework would bias fidelity and chroma shift calculations to favor sources with a chromaticity similar to the reference. This characteristic does not match studies of human perception, which show that perceived attributes of color rendering, such as saturation, are independent of chromaticity [Ohno and others 2015].

Finally, the option of using an off-Planckian reference with color-saturating characteristics has perhaps the most definitive effect on fidelity and gamut calculations. Light sources rated as having high average fidelity in standard TM-30 or CIE Test Color Method calculations are instead characterized as lower fidelity and smaller average gamut area (desaturating) when using a high gamut area (saturating) reference source. This drastic change in interpretation is an impediment to using an off-Planckian, saturating reference source in an objective evaluation framework, such as TM-30. However, such a reference may be more useful for a dedicated measure of subjective color preference, where the goal would be to render colors similarly to a reference that is known to be preferred. Critically, preference must be tied to a specific application [Lin and others 2015], so any such measure would not be universally applicable.

Reference illuminants are important to the evaluation of color rendition because they provide a common and familiar point of comparison. A fourth alternative option not previously discussed in this article is to facilitate calculations based on a custom reference. In fact, this has been done in specific cases where color rendering is a critical aspect of the lighting design, such as with the Sistine Chapel [Schanda and others 2016]. However, for the purposes of commerce, it is important that measures are based on standardized procedures—which, in this case, includes the reference illuminant—so that reported values are comparable to one another. While it is reasonable to expect that users may rely on the TM-30 framework to calculate custom metrics based on the comparison of a test source to a non-standard reference, it is important that TM-30 itself remain a referenceable standard.

Lastly, it is important to note that the TM-30 framework tends to be more stable than the CIE Test Color Method, thanks to the improved uniformity of CAM02-UCS, and its underlying chromatic adaptation transformation. Thus, the specific reference scheme chosen has less effect on use of the measures. CAM02-UCS has been shown to be a better model of human vision [Jost-Boissard and others 2014; Luo and others 2015; Sandor and Schanda 2006], and better for characterizing human perception of color difference when used in fidelity measures [Xu and others 2016]. Regardless of the reference scheme used, CAM02-UCS is a clear improvement.

5 Conclusions

The results of this numerical experiment show that a variety of outcomes could be achieved by applying a different reference scheme to IES TM-30-15. Using all Planckian radiation or all CIE D Series Illuminants tended to have relatively less effect on values; the differences compared to the standard reference scheme were typically less than a few points, and were generally related to the slight differences in gamut shape between the different reference illuminants. Using a fixed reference, such as CIE D65, for all test sources dictates that sources with CCTs different from the reference will tend to have relatively lower values, with some variation based on individual gamut shapes. Finally, off-Planckian and saturating reference illuminants would result in a majority of existing sources being rated as desaturating, with somewhat lower fidelity.

Ultimately, reference sources are important for communicating color rendering, but no possible reference is inherently superior. Given their status as naturally-occurring (non-electric) light sources with which most people are familiar and which are already in established use, Planckian Radiation and Daylight seem as reasonable as any possible choice. Still, as color rendering metrics are revised in the future, the reference scheme should be scrutinized, especially if new standard illuminants are defined.

References

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.

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.

Davis W, Ohno Y. 2010. Color quality scale. Optical Engineering 49(3):033602.

Dikel EE, Burns GJ, Veitch JA, Mancini S, Newsham GR. 2013. Preferred Chromaticity of Color-Tunable LED Lighting. Leukos 10(2):101-115.

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.

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

Jost-Boissard S, Avouac P, Fontoynont M. 2014. Assessing the colour quality of LED sources: Naturalness, attractiveness, colourfulness and colour difference. Lighting Research and Technology 47(7):769-794.

Lin Y, Wei M, Smet K, Tsukitani A, Bodrogi P, Khanh T. 2015. Colour preference varies with lighting application. Lighting Research and Technology.

Luo MR, Gu HT, Liu XY, Wang BY. Testing colour rendering indices using visual data under different LED sources. In: 1(1) CP, editor. Proceedings of the 28th CIE Session; 2015; Manchester, UK.

Ohno Y, Fein G, Miller C. Vision Experiment on chroma saturation for color quality preference. 28th CIE Session; 2015; Manchester, UK: Commision Internationale de l'Eclairage.

Ohno Y, Fein M. Vision experiment on acceptable and preferred white light chromaticity for lighting. CIE 2014 "Lighting Quality and Energy Efficiency"; 2014; Kuala Lumpur, Malaysia.

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, 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 and Technology In Press.

Sanders C. 1959. Assessment of color rendition under an illluminant using color tolerances for natural objects. Journal of the Illuminating Engineering Society(54):640-646.

Sandor N, Schanda J. 2006. Visual colour rendering based on colour difference evaluations. Lighting Research & Technology 38(3):225-239.

Schanda J, Csuti P, Szabo F. 2016. A New Concept of Color Fidelity for Museum Lighting: Based on an Experiment in the Sistine Chapel. Leukos 12(1-2):71-77.

Smet KA, Deconinck G, Hanselaer P. 2015a. Chromaticity of unique white in illumination mode. Opt Express 23(10):12488-95.

Smet KA, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. 2010. Memory colours and colour quality evaluation of conventional and solid-state lamps. Opt Express 18(25):26229-44.

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

Smet KAG, Geert D, Peter H. 2014. Chromaticity of unique white in object mode. Opt Express 22(21):25830-41.

Wei MC, Houser KW. 2016. What Is the Cause of Apparent Preference for Sources with Chromaticity below the Blackbody Locus? Leukos 12(1-2):95-99.

Xu W, Wei M, Smet K, Lin Y. 2016. The prediction of perceived colour differences by colour fidelity metrics. Lighting Research and Technology.