Meta-Standards for Color Rendering Metrics and Implications for Sample Spectral Sets

Kevin Smet[1] and Lorne Whitehead[2];[1]KAHO Sint-Lieven, Gent, Belgium; [2]University of British Columbia, Vancouver, Canada

Abstract

There is compelling, largely agreed-upon evidence that the CIE Color Rendering Index (CRI) is not a sufficiently accurate measure of the average color rendering difference between two different spectral irradiance distributions. This has become particularly evident with the consideration of white light sources employing several narrow band light emitters. In this presentation an approach is suggested for reducing the error, based on consideration of certain meta-standards for evaluation of proposed CRI calculation methods. Here we focus mainly on one such meta-standard –the idea that the spectral sensitivity of the CRI should be reasonably uniform and largely independent of the details of the selected sample set. A key problem with the current CRI calculation method, and some suggested improvements, has been their lack of this important characteristic. By employing this guiding principle, we have devised an improved sample set that substantially eliminates the problem of non-uniform spectral sensitivity.

Introduction

This paper concerns an important component of the CIE Color Rendering Index (CRI) – the calculation of the average color rendering difference between two spectral irradiance distributions. The calculation method used in the current CRI has known inaccuracies. One is that the color space employed is out of date. Others have made excellent suggestions for solutions to this problem; these need not be discussed further here.

We focus on the problem of the selection of the set of sample spectral reflectance distributions that are employed within the CRI calculation. The set currently used is shown to have serious problems that cause inaccuracy and we discuss the development of a new set that is free of this problem.

In order to understand this development, a reader should be familiar with the background concepts of chromaticity, metamerism, uniform color spaces and color rendering. It is impractical to review them here, but the information is available in a wide variety of readily available reference texts [1].

Errors in evaluating color rendering difference

It is well known that evaluation of the average color rendering difference between two spectral irradiance distributions uses two largely independent mathematical components:

- a set of defined colored reflectance samples with known spectral reflectance distributions – we will call this here the "sample set" for short; and
- a color difference calculation and averaging algorithm that calculates the color difference (between the two illumination states) for each of the samples, and then calculates an average value of these differences – here we will call this the "calculation engine" for short.

In the CRI, the current versions of each of these components introduce problematic errors [2-5]. The most readily addressed part of the problem involves the calculation engine. The current CRI uses an outdated color calculation system. Fortunately, there is little difficulty in using a more modern calculation engine. Based on a proposal from the group of Luo, Schanda et al. [6], one may update the engine to use modern color space and r.m.s. averaging instead of arithmetic averaging. This improvement seems not to be contentious, although certain minor outstanding matters may still need to be addressed.

The other component – the sample set – is more problematic and is the primary subject of this paper. Various parties have claimed that the previous sets of 8 and 14 samples used in the CRI are inadequate and suggested a different small set of real samples. Alternative proposed sample sets have been selected so that their chromaticity coordinates form a regular pattern in color space [7]. However, uniformity in color space does not in general transfer into a set of spectra with uniformly distributed spectral features. This was probably known by the experts who designed the CRI, but at the time, in the pre-computer era, there was no alternative. Fortunately, also, most of the light sources at the time had spectra with relatively broad features for which the resultant error was tolerable.

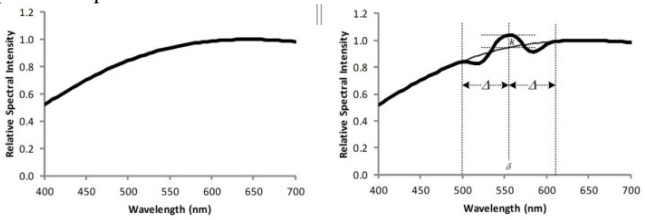
To study these ideas more carefully, it is helpful to consider the idea of the responsiveness of the CRI score to localized spectral features. Here we will call this idea the "spectral sensitivity" of the CRI calculation.

Spectral sensitivity of CRI

As an initial example, consider two illuminant spectral distributions, as follows: Spectrum (a) is a 4500K Planckian and Spectrum (b) is the same function, but multiplied by $(1+k(\,p(\lambda-\delta)))$ where k is a perturbation coefficient, δ is a spectral offset value that can be adjusted and p is a smooth, symmetrical perturbation function that integrates to 0, is zero outside the "perturbation zone" within $\pm\Delta$, and has a maximum magnitude of 1.0 within the perturbation zone. Figure 1 compares the two spectra – they are the same everywhere except within the perturbation zone, the location of which can be shifted continuously along the visible spectrum. Within the perturbation zone the two spectra differ, but the total amount of light is unchanged – it is simply more intense near the center and less intense near the edges.

Now consider what happens when these two spectral distributions illuminate test samples, as is done notionally in the color rendering difference calculation. As one might expect, the average color rendering difference is proportional to k; in other words the ratio of average color difference to k has a constant value, which we will call the sensitivity of the color rendering

difference. However the constant of proportionality will generally be a function of the central wavelength, δ, and will depend on the particular perturbation function used.



Spectrum (a) *Spectrum (b)*

Figure 1. *Comparison of perturbed (b) to the original (a), in this example a 4500K Planckian*

It is helpful to consider how this sensitivity to perturbation varies as the central wavelength is shifted across the visible spectrum. We would expect this functional relationship to have a smooth character that represents, in some complex manner, the underlying spectral response characteristics of human vision, while also depending on the particular selected perturbation function $p(\lambda-\delta)$). Thus we would generally expect the spectral sensitivity function to be fairly smoothly distributed across the visible spectrum, and that it would not be critically dependent on specific details of the particular reflectance samples used.

It is helpful to consider how this idea constrains the selection of the reflectance sample set. Specifically, let us consider how individual test samples contribute to the overall spectral sensitivity. The reflectance values of a test sample, for wavelengths lying outside the "perturbation zone", are irrelevant because at those wavelengths the two illuminants are identical. Further, if the reflectance sample has uniform reflectance within the perturbation zone, there will be little difference between the two illumination states because the perturbation function averages to zero. It is even true that a linear gradient in the test sample reflectance will yield little color difference, because of the symmetric shape of the perturbation function. Therefore, what is required to generate a significant color difference for the sample, between the two states of illumination, is curvature in the plot of reflectance vs. wavelength, within the perturbation zone. By curvature, we mean the magnitude of the second derivative of the function $R(\lambda)$.

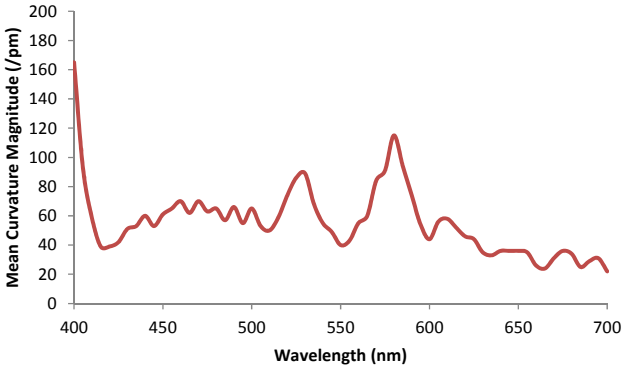


Figure 2. *The average magnitude of curvature for the 8 basic CRI reflectance samples.*

To help make this clearer, consider the 8 samples that are used in the basic CRI calculation. Figure 2 shows the average magnitude of curvature of these 8 spectra, (which have first been smoothed with a 20 nm rolling average, to reduce the influence of noise in the data). The resultant arithmetic average is very non-uniform, with the shape determined by the arbitrary spectral details of the individual reflectance samples. If we were to replace each of the 8 samples with a different, but metameric sample, the sensitivity curve in Figure 2 would be quite different – it therefore is not representative of human vision.

The resultant error in the CRI is large for LEDs and also for the narrow-band spectral perturbations that lamp manufactures may soon wish to consider as quantum dot phosphors become available. For this reason, this is now a major problem.

We have concluded that it is impossible to find a small number of real reflectance samples that will yield a suitably smooth spectral sensitivity response. However, there is good reason to hope that using a large number of samples could reduce this problem, because features from individual spectra should tend to average out statistically.

Finding a large data set that has sufficiently uniform results

From this perspective, a very important resource has recently been available to the scientific community – the Leeds 100,000 set of measured spectral reflectance distributions of real samples of all kinds of natural and artificial surfaces [8]. Before this set was available, there was insufficient data to enable the work described here, and only fairly recently has this work become computationally practical.

One possible use of this new set could be simply to select a manageably sized subset of spectra from it. From simple statistical considerations, we have estimated that about 1000 spectra could be enough to yield a sufficiently uniform spectral sensitivity. However, there are two significant problems in this regard:

First, it would not be enough to simply carry out a random sampling of the Leeds set, because that data was not selected with the goal of uniformly representing the vast space of possible reflectance distributions. For this reason, in discussion with a number of color experts, it has generally been agreed that a more reasonable weighting would be to select spectra whose color coordinates are fairly uniformly distributed in the color volume within $L^*a^*b^*$ space.

We tried such a sampling and discovered a second, more subtle problem. Such a selection did not substantially reduce the problem of non-uniformity of mean curvature depicted in Figure 2. At first this seemed surprising, because for independent entities one generally expects the magnitude of the average of random data to decrease with the inverse square root of the number of samples. In this case this would mean that non-uniformities shown in Figures 2 should be reduced by a factor of about (8/1000)^.5, which is 0.09. But as shown in Figure 3, clearly that is not the case for the basic Leeds samples, depicted in the curve labeled "1000 Non-hybrid Leeds samples". We believe the reason for this result is that the Leeds samples are not mathematically independent, which is not surprising since there is no reason they should be, because most colored surfaces are composed of various ratios of a reasonably small set of common dyes. These dyes have numerous spectral features in fixed locations that will contribute to curvature at the same wavelengths in the spectrum, so averaging of large numbers of them would not result in very much statistical averaging.

An alternative approach that has been considered is to use mathematically generated reflectance spectra made by a random algorithm that treats all spectral locations equivalently – a so-called "Monte Carlo" approach [9]. This would solve the spectral non-uniformity problem, but unfortunately it would introduce another problem: We could not be sure that such generated spectra would contain sufficiently "natural" shapes. It would be better, if possible, for the spectra to containing only spectral features that are found in real samples, and for those features to have approximately the same frequency distribution as found in real samples.

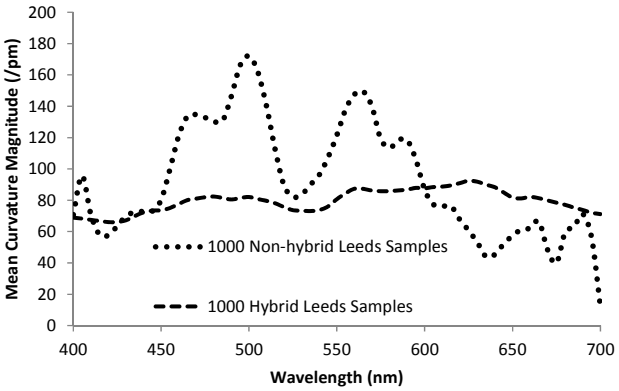


Figure 3. *The average magnitude of curvature is not improved by using 1000 Leeds samples.*

Considering this, we have found a solution that adapts the Leeds spectra in a hybrid, composite manner, as follows:

- The first step was to select, from the Leeds 100,000 set, 10,000 spectra with approximately uniform density in *L*a*b** space. The purpose of this is to reduce selection bias in sampling spectral features.
- These spectra were then truncated to the range 450 nm to 700 nm, because some artefacts were noted in some of the spectra outside this range.
- Using these 10,000 selected spectra, we carried out the following steps in each of 1,250 independent cycles:
  - First, a spectrum was randomly selected from the 10,000 spectra subset, as a "seed".
  - Next, a spectrum was selected for which the starting slope and starting reflectance best matched the end slope and reflectance of the seed, and was "stitched" onto its end. This provided a smooth, fairly continuous, natural looking connection.
  - This match/stitch process was repeated three times, producing a continuous spectrum long enough for the sampling operation described next:
  - Selected spectral subsets were copied, to populate in 5nm intervals, the range 380 to 780 nm (81 values) in an output file. This was done 80 times, and on each successive sampling, the sample interval was shifted by 5nm intervals. This ensured that every wavelength was treated the same as every other. In this way 80 full spectra were created for each of the 1,250 stitched spectra. This yielded 100,000 final hybrid stitched spectra.
- Each of these spectra was slightly smoothed by carrying out a 20nm rolling average, which did not change spectral shapes, but removed any discontinuity in curvature at the connection points. The resulting spectra look very "natural" and indeed, all of the features are "real".
- This final set of 100,000 spectra was then used for the selection of the sets of hybrid spectra, as follows:
- For each of the 100,000 spectra, the local density was estimated by counting the number of a random subset of 100 of these spectra that were located within a DE distance of 20 units in *L*a*b* ** space. Rather than using a discrete boundary, a Gaussian weighting was used, to give smoother and non-zero values for all points.
- A selection probability was calculated for each point, in order to achieve a uniform selection density in *L*a*b* ** space while ensuring reasonable diversity in multiple selections. By experimentation we found it best to set the selection probability to the minimum of 0.05 and .0038 divided by the number of 20nm neighbours. This created a good uniform distribution in *L*a*b* ** space with no unusual discontinuities and good coverage of the color volume.
- 16 independent random selections were carried out to select 16 different subsets of 1,000 spectra. Averaging the 16 sets allowed accurate determination of the corresponding spectral sensitivity functions, and we selected as our final set the 1000 spectrum set that had results closest to the mean of the 16.

As shown in Figure 3, the curvature distribution is much smoother, as expected, for the resultant set of 1000 Hybrid Leeds Samples. Importantly, and not surprisingly, the stitched hybrid spectra have the same "look and feel" as the original Leeds spectra from which they are formed, as compared in Figure 4.

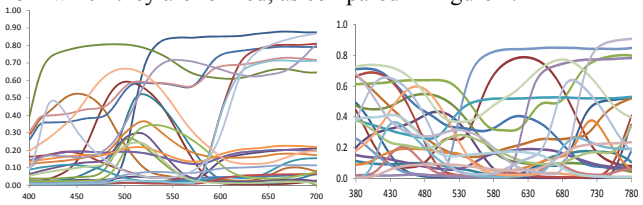


Figure 4. *Comparing 24 randomly selected spectra from the original Leeds set and from the hybrid stitched Leeds spectra generated from the original Leeds set.*

Although there is no practical difficulty in using 1000 spectra, nevertheless some feel this number is too big. Therefore, having developed a good spectral set, it is reasonable to try to find a way to reduce the number of samples without adding unacceptable error, using the Hybrid 1000 sample as a guide.

Designing a smaller spectral set that maintains reasonable uniformity

One way of obtaining spectral uniformity would be to have a fairly smooth spectral feature shift, from one sample to the next, through the spectrum. Figure 5 shows the general form that we have studied in detail. We settled on this by first optimizing with a larger number of free parameters using a general polynomial model, and the general trend seemed to be toward a shape of the form shown. For this reason we decided to optimize, with greater precision, using this model which has only a few adjustable parameters. We varied the forward slope, the rear slope, the width of the non-curved fraction of the slope, the minimum and maximum reflectance and the spacing of samples across the

spectrum. Our goal was to find a minimal simple sample set that emulates the 1000 samples set with sufficient accuracy.

In so doing, there was also a question of whether we should include in the sample spectra that are complementary in the sense that they would equal the difference between unity and the value of the primary spectrum shown in Figure 5. This would allow bright, saturated purple samples, for example. Generally, the graphs of such functions would have an opposite appearance in that they would have a trough of low reflectance in a surrounding of high reflectance, rather than a peak of high reflectance in a surrounding of low reflectance. Although intuitively this seemed like a good idea, we found that in fact it did not make any significant improvement to the quality of the fit to the larger set. This is not particularly surprising, given the nature of the mathematical calculations involved. Overall, since using such a complementary set would double the number of samples but provide no practical benefit, we decided not to do so, in recognition of the desire for simplicity.

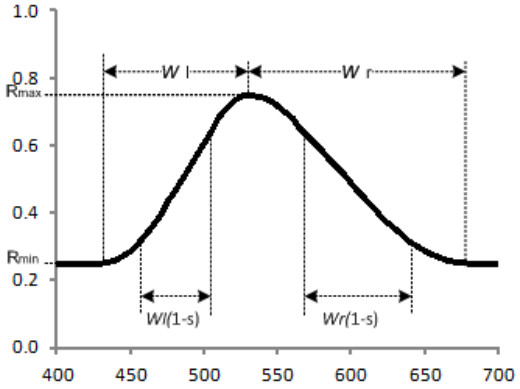


Figure 5. *The test mathematical generator spectrum. The general form of the function is an asymmetrical triangle with slope width wl on the left and wr on the right, reflectance ranging from Rmin to Rmax , and with quadratic rounding applied at the corners leaving a fraction of the slope, 1-s, uncurved.*

We did however find a different factor that we concluded needed to be taken into account. In studying the Hybrid 1000 set, we found that the average reflectance of the 1000 samples increases gradually with wavelength throughout the spectrum. Presumably this is due to the fact that the color volume in $L^*a^*b^*$ space has a somewhat irregular shape. We found that if we adjust the mathematically generated sample set to correspondingly increase across the spectrum, the quality of the fit improved, from very good to excellent.

In these studies, we used a computer optimization algorithm to find the best match between the results from the proposed smaller test set and those for the Hybrid Leeds 1000 sample set. In doing so, we considered the match in spectral sensitivity (which we found to be the most stringent criterion), the match in the R_a calculations for 53 common reference lamps (a somewhat less stringent criterion), and human observation of color rendering difference observations (the least stringent criterion because of the relatively large spread of values in existing data). We found that quite good results were possible – essentially no information is lost in using this vastly smaller, but mathematically regular, spectral set.

Our resultant spectral set consists of only 17 spectra. We call this the "HL17" set to designate it as 17 spectra emulating the Hybrid Leeds 1000 set. The design parameters (as defined in Figure 5) have the following values: w_l=w_r=100nm, s=0.5, R_{min} = 0.01 and R_{max} = 0.5 + 0.001nm^{-1} * (λ_c-550nm). (This last term provides the gradual increase in reflectance with wavelength mentioned above.) The centers of the distributions increment, from one sample to the next, in 16 intervals of 25nm, starting with the first at 350nm and ending with the last at 750 nm. The end points of this spectral series were selected to include all, and only, spectra for which the luminous reflectance exceeded, by 0.001, that of a spectrally uniform reflector reflectance 0.01. This was determined to be a good cut-off, by checking that samples with lower values had negligible effect on the overall results. Of these parameters studied, the only two that were really critical were w_l and w_r; these were constrained quite tightly by the optimization process. Although the others were not as tightly constrained, they all needed reasonably careful adjustment to get the best overall fit. In particular, R_{min} was quite sensitive.

Figure 6 shows the plot of these 17 spectra. They can be labeled by an integer i which ranges from 1 to 17. Each sample function has a defined central wavelength λ_{ci} given by λ_{ci} = 550nm + 25nm * (i -9). (Thus λ_{ci} has a value of 350 nm for sample #1, 550 nm for the central sample #9 and 750 nm for the last sample #17, with the increment from one sample to the next being 25nm.) Each sample function has a peak reflectance M_i given by M_i = 0.5 + 0.001nm^{-1} * (λ_{ci} – 550nm) . (Thus M_i has a value of 0.3 for sample #1, 0.5 for the central sample #9 and 0.7 for the last sample #17.) The formula for the reflectance values depends on the value of ABS(λ – λ_{ci}), in four distinct ranges, as follows:

A) If ABS(λ – λ_{ci}) ≥ [125nm] :

$R(\lambda)$=[0.01]

B) If [125nm]>ABS(λ–λ_{ci})≥[75nm] :

$$R(\lambda)=[0.01]+(Mi-[0.01])*\left[\frac{1}{8750nm^2}\right]*\left([125nm]-ABS(\lambda-\lambda_{ci})\right)^2$$

C) If [75nm]>ABS(λ – λ_{ci})≥[25nm] :

$$R(\lambda)=[0.01]+(Mi-[0.01])*\left(\left[\frac{8}{7}\right]-\left[\frac{2}{175nm}\right]*ABS(\lambda-\lambda_{ci})\right)$$

D) If [25nm]>ABS(λ – λ_{ci}) :

$$R(\lambda)=[0.01]+(Mi-[0.01])*\left([1]-\left[\frac{1}{4375nm^2}\right]*(\lambda-\lambda_{ci})^2\right)$$

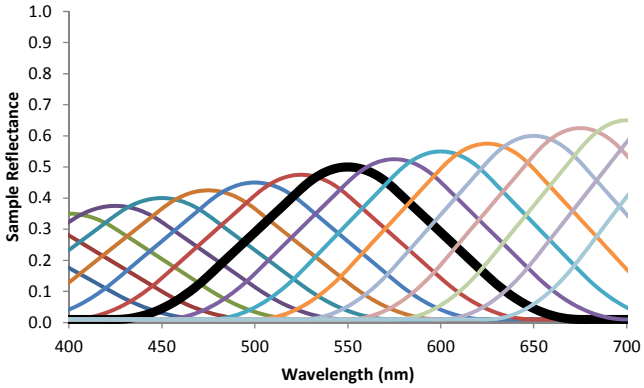


Figure 6. *The HL17 set of spectra developed to emulate the 1000 Hybrid Leeds spectra for the purposes of calculation of average color rendering difference between two illuminant spectral distributions. One of the spectra is shown in bold, as an example.*

Because of the mathematical form of these spectra and the values selected, the average magnitude of the curvature of these spectra is quite uniform, which as described previously is a valuable characteristic that would not be possible with any small set of naturally occurring spectra.

It is also interesting to consider the calculated *L*a*b** values for these spectra. They are shown in the three plots in Figure 7, for a D65 illuminant. To help evaluate the appropriateness of their locations, the plots include an overlay of the chromaticity points of the 1000 Hybrid Leeds set. The optimized HL17 spectra demark a path that seems reasonably representative of the larger spectra set.

As mentioned earlier, one of the primary criteria in designing the spectral set was to accurately match the CRI results for the 1000 Hybrid Leeds set. Additionally, our goal was to match the spectral sensitivity functions mentioned previously. In making that comparison, we used two different sizes of perturbation function, one with a value of Δ of 20nm, and another with a value of 40nm, and we perturbed three different spectra, a 3000K Planckian, a 4100K Planckian and a 6500K Daylight spectrum. The nature of the fitting exercise was similar in all cases, so for simplicity here Figure 8 shows the results for the average of all six combinations.

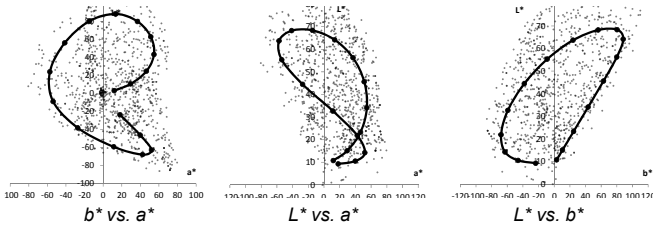


Figure 7. *Distribution of the HL17 spectral set under D65 illumination in L*a*b* space. Overlayed on these plots are the colors of the 1000 Hybrid Leeds spectral set.*

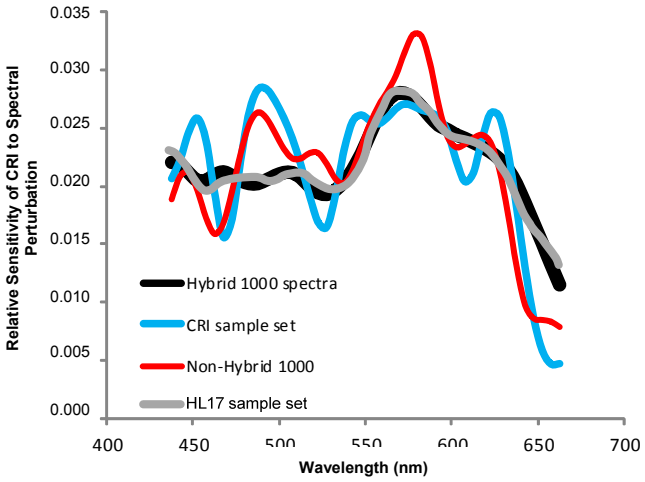


Figure 8. *Plot of the average sensitivity of the CRI to small spectral perturbations, as a function of the central wavelength of the perturbation. The HL17 set matches the sensitivity of the Hybrid 1000 set well. In contrast, the results for the CRI sample set and a non-hybrid 1000 sample set match poorly.*

The characteristic that may be of more immediate interest to the lighting industry is the results when the new spectral set is used to calculate *Ra* values for common light sources. The following Figure 9 compares the results of using the proposed new CRI calculation engine, for the different spectral sets described above.

Probably the most important conclusion from Figure 9 is that the fit of the new "HL17" spectral set to the Hybrid 1000 spectra is about six times better than is the case for the CRI sample set or the set of 1000 non-hybrid spectra selected directly from the large Leeds database. (The correlation R^2 value was only 0.94 for both the CRI sample set and the Non-hybrid Leeds 1000 set, and was 0.99 for the HL17 set. This shows that we have very significantly reduced the size of sample-induced error.

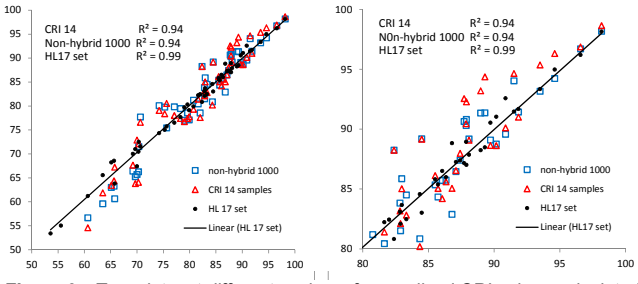


Figure 9. *Two plots, at different scales, of normalized CRI values calculated using the new calculation engine with the different spectral sets. The horizontal axis represents the value obtained using the Hybrid Leeds 1000 set. The vertical axis depicts the result for the CRI samples, the Non-hybrid 1000 set, and the proposed HL17 set, which shows a very good correlation with the Hybrid Leeds 1000 set.*

We also examined the spectra of specific light sources that score higher using the CRI sample set than using the 1000 Hybrid Leeds set or HL17 set. We found that some narrow band phosphor fluorescent lamps, with a major narrow spectral peak near 550 nm, score up to 5 CRI points higher with the historic CRI sample set. This is a wavelength region for which the CRI samples have an anomalously low spectral sensitivity, so the observed difference is not surprising. In all probability, had the CRI been updated sooner, such spectra would have been designed differently. More generally, we believe that a significant advantage of the HL17 data set is that it is impervious to such "spectral gaming".

Conclusions

We propose the HL17 set as an improved sample set for more accurate evaluations of the average color rendering difference between two spectral irradiance distributions.

However, even with this improved accuracy, we are not suggesting that the CRI should be used alone for evaluation of the general level of human preference of a given light source. In particular, this new work does not imply that Planckian and/or daylight spectra are necessarily optimal for human vision.

Nevertheless, the CRI has been an important tool for many years, and we feel that with a new sample set as described here, it can remain a useful metric in the future, as energy efficiency concerns grow and narrow band light emitters become common.

Acknowledgment

The authors gratefully acknowledge the important contributions and guidance of Prof. Ronnier Luo and Prof. Janos Schanda, who have supported this project through collaborative discussions.

References

[1] For example, Janos Schanda, Colorimetry: Understanding the CIE System (Wiley, 2007).

[2] C. Li and M.R. Luo, “Evaluation of light sources based on visual color rendering”, Proc. 16th Color Imaging Conference, 201–204 (2008).
[3] X. Guo and K. Houser, “A review of color rendering indices and their application to commercial light sources”, Lighting Res. Technol. 36, 183–199 (2004).
[4] A. Zukauskas, R.Vaicekauskas, F. Ivanauskas, H.Vaitkevicius, P. Vitta and M. Shur, “Statistical approach to color quality of solid state lamps”, IEEE J. Sel. Top. Quant. Electron. 15, 1189–1198 (2009).
[5] P. van der Burgt and J. van Kemenade, “About color rendition of light sources: The balance between simplicity and accuracy”, Color Res. Appl. 35:85–93 (2010).
[6] Ronnier et al private communication within TC 1-69 and co-presentation at this conference
[7] Y. Ohno and W. Davis, “Toward an improved color rendering metric,” Proc. SPIE Fifth International Conference on Solid State Lighting 5941, doi: 10.1117/12.615388 (2005).
[8] Ronnier private communication with the authors
[9] L. Whitehead and M. Mossman, “A Monte Carlo method for assessing color rendering quality with possible application to color rendering standards”, Color Res. Appl. doi: 10.1002/col.20649 (2010).