Update to Evaluation of Reference Materials for Standardization of Spectrophotometers

John Seymour

Keywords: spectrophotometer, calibration, inter-instrument agreement, standardization

Abstract

Two new sets of reference materials have been evaluated for their ability to standardize one spectrophotometer to another. Standardization is the act of applying corrections to spectral data from one spectrophotometer in order to attain better agreement with another.

This paper focuses on one aspect of the suitability of the sets of reference materials: whether the spectra of the reference materials are rich enough to provide a numerically stable calibration of the equations which correct one instrument to another. Additional sets of potential standardization materials are also analyzed in order to provide perspective on the results. These additional sets are likely unsuitable because of lack of physical durability.

The results demonstrate that 1) It is physically possible to create a set of reference materials that can provide numerically stable standardization, but 2) The two sets of reference materials, at best, provide marginally acceptable numerical stability.

Statement of problem to be solved

Reference materials are physical standards that can be used to improve the color measurement process in several ways. They are also known under the names color standards, reference standards, or verification standards. There are three ways that they may be used.

Use case 1: Reference materials can be used to verify that a color measurement device has not changed. The set of physical standards is measured initially, like when the instrument is purchased, and then again at regular intervals. If the

---

John the Math Guy, LLC

measured values do not change appreciably from the initial measured values, then it is assumed that the measurement device continues to measure faithfully.

This technique can be robust provided the set of reference materials is physically stable, that is, resistant to scratches and fading. Ceramic color tiles are popular as a reference material for this reason.

For this use case, there is not a strong need for a broad set of colors; it has been suggested [Berns and Reniff, 1997] that a single cyan tile may be adequate.

Use case 2: Reference materials can be used to quantify the agreement between two color measurement devices. This quantification may advise the user as to the suitability of mixing the two devices in an work flow.

The requirements of the reference materials for this second case are somewhat different than for the first case. The physical stability of the reference materials is less critical, especially if the two instruments are to be compared sided-by-side and at the same time. But the need for a wide range of colors (or more accurately, spectra) is essential to accurately assess the agreement of the instruments.

A less obvious requirement is that the reference materials should be similar in nature to the samples that will be measured. It has been found that differences in the illumination and acceptance angles between two different models can be a substantial source of disagreement. In addition, differences in aperture size may create disagreement in measurement due to the lateral diffusion of light within certain of the ceramic tiles. Light that scatters broadly may exit the tile in a place where the aperture of one instrument may allow its collection, but not the aperture of another [Seymour, 2014, ISO/TS 23031].

Thus, instrument agreement based on measurements of, for example, high-gloss ceramic reference materials may not be representative of disagreement that will be seen when two color measurement devices are measuring samples of print.

Use case 3: Reference materials can be used to standardize one spectrophotometer to another [Robertson 1986, Berns and Peterson 1988, Berns and Reniff 1997, Rich and Martin 1999, Van Aken 2000, 2003, 2006, Chung et al. 2002, Rich 2004, Nussbaum et al. 2011, Seymour 2013]. This is the practice of using a set of reference materials to determine the parameters for a transform which converts spectral reflectance values from one spectrophotometer so as to more closely agree with another spectrophotometer. The most common technique for standardization involves quantifying an offset term, an overall gain, and a wavelength shift.

Evaluation criteria for a standardization set

There are a number of requirements that must be met for a set of samples to be considered a good standardization set. The physical properties must be stable under the conditions it is expected to be used. If it must be shipped between plants, it should be able to recover from extremes of temperature that it is likely to encounter. If it is to be used for several years, it must not be subject to fading in that period of time.

The well-known set of Lucideon *Colour Standards Series II (CCSII)* (AKA Ceram, AKA BCRA) has been shown to provide this stability. Aside from some well understood issues with thermochromism (transient change in color due to temperature), this set of tiles has been proven useful for verifying that a color measurement device has not changed (use case #1).

A second criterion for a standardization set is that it must provide a rich enough set of spectra for the correction technique to be mathematically stable. If one wishes, for example, to correct for a spectral shift between two instruments, then the samples must include spectra which shows transitions, i.e. changes in reflectance with respect to wavelength. This was investigated previously [Seymour 2013].

Quoting from that paper: "standardization with a seemingly reasonable set of samples and a seemingly reasonable underlying mathematical model can be a worthless endeavor, and can often significantly worsen intra-model agreement." The well-known set of BCRA tiles was found to not be rich enough for this purpose. In another paper by Seymour [2014] it was shown that one of the causes for disagreement between instruments with different designs had to do with the geometry of illumination and detection. The differences in design even within the group of ostensibly 45:0 (or the equivalent 0:45) geometry was significant.

For samples with either extremely rough or highly polished surfaces, the difference between two instruments with similar geometry is minimal. Unfortunately, almost all the material to be measured in the print industry fall between these two extremes. The amount of difference between measurements depends on where the measured samples lie in the range from matte to glossy.

It is important to note that this difference depends almost entirely on the surface roughness and index of refraction of the material– one instrument will capture more of the specularly reflected light than the other. Therefore, the difference between two instruments will likely be an offset. On the up side, the magnitude of the offset is not significantly dependent on any of the bulk characteristics of the sample.

Thus, it is possible to accurately standardize one instrument to another, provided the standardization set has a similar roughness and index of refraction as the samples to be measured.

Here again, the standard set of BCRA tiles are not suitable for standardization. They are considerably glossier than any print.

Thus, we have a set of ceramic reference materials which is in common use for verification of spectrophotometers in the graphics arts, but which has two undesirable properties when it comes to standardization of one instrument to another. The ceramic tiles are considerably glossier than printed samples, and they are less than ideal in their ability to calibrate at all wavelengths.

Evaluation of two potential standardization sets

The ISO technical specification ISO/TS 23031 recommends the ChromaChecker™ Instrument Inspector target as a set of reference materials for comparison between different models of color measuring device. It further suggests that this set may be used to diagnose causes for disagreement. It stops just short of suggesting ChromaChecker™ as a set of reference materials for standardization of one color measurement device to another.

There are two physical reasons to prefer this set of reference materials over the BCRA tiles. First, the gloss of the ChromaChecker™ is within the range of print, whereas ceramic tiles are generally considerably glossier. Second, the patches are a thin film, so lateral diffusion is minimized, or is at least similar to what may be found in graphic arts samples. However, the spectral suitability of ChromaChecker™ Instrument Inspector target has yet (to the knowledge of this author) to be investigated.

Another candidate for a set of reference materials for standardization of color measurement devices is a set of reference materials which Lucideon has developed specifically for the print industry. The Lucideon Print Standards are available in either glossy or matte finish. The latter was investigated for this paper, since they are deemed to be closer to the finish of print. As with the ChromaChecker™ set, this author is unaware of any published assessment of the Lucideon Print Standards for the purpose of spectrophotometer standardization.

This paper fills that need by investigating the spectral suitability of these two potential standardizations sets and comparing this with other sets of samples. The assumption is made that these two sets of standards meet the requirements for physical durability and stability.

It is also assumed that the gloss and lateral diffusion characteristics of the two standardization sets are sufficiently close to those of printed material. This seems reasonable, but it is regardless an untested assumption.

Theoretical background

The list of potential reasons for two instruments to disagree is rather daunting. The following list was compiled by Seymour [2014] (with reference to Spooner [1991]): repeatability, black level, rejection of scattered light, white level, measurement geometry, nonlinearity, aperture size, wavelength alignment, bandwidth difference, and fluorescence.

Much of the work around standardization of instruments has been based on the assumption that white level, black level, wavelength alignment, and bandwidth difference are the major sources of disagreement. Standardization methods have been developed to quantify and correct the differences between to instruments based on this assumption. A set of samples are measured with both instruments, and the discrepancies are used to calibrate the resulting equations.

Differences in white and black level can be readily quantified from a white and a black sample. Wavelength alignment and bandwidth are a far more complicated matter. Seymour [2013] demonstrated that there is likely to be a significant nonlinearity between the measurements of two instruments. Regression techniques used to determine correction parameters may confuse wavelength shift and nonlinearity, since the two will have a similar numerical effect. This confusion can lead to a failure of the standardization process. If different physical causes are confused, standardization may worsen agreement between two instruments.

This confusion can be ameliorated if a standardization set has moderate collection of both positive and negative derivatives (reflectance as a function of wavelength) at each wavelength.

The mathematical models

Four different mathematical models are investigated in this paper, each with a certain combination of assumptions about the form of the correction needed. Variations on these models can be found in the literature [Robertson (1986), Berns and Reniff (1997), Rich and Martin (1999), Van Aken (2001), Chung et al. (2002), Rich (2004)].

While it is tempting to create a model that includes all possible corrections, the more individual parameters there are, the more samples are required to calibrate these parameters and the greater the requirements on the type of data required. The most appropriate mathematical model is dependent on the specific instruments

that are to be used. If the two instruments are of the same design, for example, adding terms for nonlinearity and black level may not be necessary, and in fact, may worsen the results due to misattribution. Therefore, it makes sense to look at multiple mathematical models.

The first four models tested in this paper assume that there could be a difference in black level, in white level, and a shift in wavelength. The four models here either include or exclude a correction term for bandwidth, and either include or exclude a correction for nonlinearity. A fifth model is added that excludes any wavelength dependence.

Model #1 – black level, white level, and wavelength shift

$$R_c(\lambda) = R_m(\lambda) + \beta_1(\lambda) + \beta_2(\lambda) R_m(\lambda) + \beta_3(\lambda) \frac{dR_m(\lambda)}{d\lambda} \tag{1}$$

Model #2 – include bandwidth

$$R_c(\lambda) = R_m(\lambda) + \beta_1(\lambda) + \beta_2(\lambda) R_m(\lambda) + \beta_3(\lambda) \frac{dR_m(\lambda)}{d\lambda} + \beta_4(\lambda) \frac{d^2 R_m(\lambda)}{d\lambda^2} \tag{2}$$

Model #3 – include nonlinearity

$$R_c(\lambda) = R_m(\lambda) + \beta_1(\lambda) + \beta_2(\lambda) R_m(\lambda) + \beta_3(\lambda) \frac{dR_m(\lambda)}{d\lambda} + \beta_5(\lambda)((1 - R_m(\lambda)) R_m(\lambda)) \tag{3}$$

The use of $\beta_5(\lambda)((1 - R_m(\lambda)) R_m(\lambda))$ for the nonlinearity term deserves a bit of explanation. Note that setting $\beta_5(\lambda)$ to something other than zero will not change the value of the righthand side when $R_m(\lambda)=0$ or $R_m(\lambda)=1$, since the product inside the parentheses is zero in either case. Because of this, the parameter for scaling, $\beta_1(\lambda)$, will not need to change appreciably if the nonlinearity term is removed. Thus $\beta_1(\lambda)$ in Equation 3 retains the meaning of a scaling parameter. This form can also be found in Ingleson and Brill, 2007.

Model #4 – include bandwidth

$$R_c(\lambda) = R_m(\lambda) + \beta_1(\lambda) + \beta_2(\lambda) R_m(\lambda) + \beta_3(\lambda) \frac{dR_m(\lambda)}{d\lambda} + \beta_4(\lambda) \frac{d^2 R_m(\lambda)}{d\lambda^2} \tag{4}$$
$$\beta_5(\lambda)((1 - R_m(\lambda)) R_m(\lambda))$$

Model #3 – include nonlinearity

$$R_c(\lambda) = R_m(\lambda) + \beta_1(\lambda) + \beta_2(\lambda) R_m(\lambda) + \beta_5(\lambda)((1 - R_m(\lambda)) R_m(\lambda)) \tag{5}$$

Definitions of variables

$R_m(\lambda)$ is the measured reflectance of a sample at wavelength λ, as measured on the instrument to be corrected.

$R_c(\lambda)$ is the corrected reflectance of the sample at wavelength λ.
$\beta_1(\lambda)$ is the calibrated offset parameter at wavelength λ.
$\beta_2(\lambda)$ is the calibrated gain parameter at wavelength λ.
$\beta_3(\lambda)$ is the calibrated wavelength shift parameter at wavelength λ.
$\beta_4(\lambda)$ is the calibrated bandpass correction parameter at wavelength λ.
$\beta_5(\lambda)$ is the calibrated nonlinearity correction parameter at wavelength λ.

The first and second derivatives are estimated from the measured spectrum according to the following.

$$\frac{d^2 R_m(\lambda)}{d\lambda^2} \approx \frac{R_m(\lambda+\Delta) - R_m(\lambda-\Delta)}{2\Delta} \tag{6}$$

$$\frac{d^2 R_m(\lambda)}{d\lambda^2} \approx \frac{R_m(\lambda+\Delta) - 2R_m(\lambda) + R_m(\lambda-\Delta)}{2\Delta} \tag{7}$$

The parameter Δ is the wavelength interval of the instrument to be standardized, typically 10 nm or 5 nm. Since the units are arbitrary (they could be nanometers or Angstroms or microns) we will set the units conveniently so that $\Delta=1$, which is to say, the units are 10 nm for an instrument that reports every 10 nm.

Note: There was an error in a previous paper by this author: Equation 5 in Seymour (2013), which gives the formula for the approximation of the second derivative, is incorrect. Equation 6 corrects this.

Note: These mathematical models all require measurements at a collection of adjacent wavelengths in order to determine the derivatives. As such, the models are not appropriate for standardization which involves one or more colorimeters.

Determination of the correction parameters

The explanation for the determination of correction parameters that follows assumes the mathematical model #1, which is the simplest of the four models. Derivation of the other three models follows along the same lines.

We start with the following values

$R_m(i,\lambda)$ is the reflectance value of sample i at wavelength λ on the instrument which is to be standardized.

$R_r(i,\lambda)$ is the reflectance value of sample i at wavelength λ on the reference instrument.

If we take model #1 as an example, we must determine the set of correction parameters $\beta_1(\lambda)$, $\beta_2(\lambda)$, and $\beta_3(\lambda)$ which minimize the error in the following set of equations, with one equation for each of the n samples in the standardization set.

$$R_r(1,\lambda)-R_m(1,\lambda)=\beta_1(\lambda)+\beta_2(\lambda)R_m(1,\lambda)+\beta_3(\lambda)\,\frac{dR_m(1,\lambda)}{d\lambda}$$

$$R_r(2,\lambda)-R_m(2,\lambda)=\beta_1(\lambda)+\beta_2(\lambda)R_m(2,\lambda)+\beta_3(\lambda)\,\frac{dR_m(2,\lambda)}{d\lambda}$$

(8)

...

$$R_r(n,\lambda)-R_m(n,\lambda)=\beta_1(\lambda)+\beta_2(\lambda)R_m(n,\lambda)+\beta_3(\lambda)\,\frac{dR_m(n,\lambda)}{d\lambda}$$

Note: There is one equation for each wavelength, and a set of correction parameters βi will be determined at each wavelength. This is the most general case, and is most appropriate for instruments that utilize a separate filter for each wavelength that is measured. In this case, it is expected that there not be a clear relationship between the wavelength shift and one wavelength and the next. If both instruments utilize a grating and an array of photodetectors, there may be a correlation between the wavelength shift at neighboring wavelengths and a simple relationship between them. In this case, it would be beneficial to broaden the regression to look at all points at once. This is beyond the scope of this paper.

This set of equations can be rewritten in matrix form.

$$\begin{bmatrix} R_r(1,\lambda)-R_m(\lambda) \\ R_r(2,\lambda)-R_m(\lambda) \\ \vdots \\ R_r(n,\lambda)-R_m(\lambda) \end{bmatrix} = \begin{bmatrix} 1 & R_m(1,\lambda) & \frac{dR_m(1,\lambda)}{d\lambda} \\ 1 & R_m(2,\lambda) & \frac{dR_m(2,\lambda)}{d\lambda} \\ \vdots & \vdots & \vdots \\ 1 & R_m(\text{n},\lambda) & \frac{dR_m(\text{n},\lambda)}{d\lambda} \end{bmatrix} \begin{bmatrix} \beta_1(\lambda) \\ \beta_2(\lambda) \\ \beta_3(\lambda) \end{bmatrix}$$

(9)

In a more compact form, we have

$$\vec{R}_r(\lambda)-\vec{R}_m(\lambda)=M(\lambda)\vec{\beta}(\lambda)$$

(10)

The least-squares solution for the vector of correction parameters $\beta(\lambda)$ is

$$\vec{\beta}(\lambda)=(M(\lambda)^T \bullet M(\lambda))^{-1}M(\lambda)^T \bullet (\vec{R}_r(\lambda)-\vec{R}_m(\lambda)) \qquad (11)$$

While the correction parameters βi can be determined at any point in time, there is no point in repeating the exercise if neither of the instruments have changed. In fact, doing so will only increase the variability of the measurement system. It is recommended that they be recalculated whenever one instrument or the other is recertified at the factory, or when a significant drift is noted by routine verification against a set of physical standards (use case 1).

For the purposes of this paper, the correction parameters are determined on a wavelength-by-wavelength basis, without consideration of the correction parameters at neighboring wavelengths. For some pairs of instruments, it may be desirable to determine the parameters en masse to take advantage of relationships between the correction parameters at different wavelengths. For example, if both instruments are grating instruments, then there is likely to be a simple equation that estimates the wavelength shift as a function of wavelength. Such an approach would not be appropriate if one or both of the instruments utilize discrete bandpass filters at each wavelength.

Spectral suitability of a standardization set

It has been stated before that the standardization set must be sufficiently rich so as to provide a reliable standardization. One way to test the reliability is test the standardization on a large number of instruments. A standardization would be done for the standardization set under test for each pairing of instruments. This standardization would be tested by looking at how well the method standardized each pairing on a different set of samples. This was the approach followed in Seymour [2013], with a total of three instruments.

Such a study is invaluable in that it is grounded in reality. Any conclusions derived from such an experiment are certainly indicative of performance of those specific instruments (model and serial number). The conclusions can potentially be generalized to different specific instruments of the same family (same model, but different serial numbers). Generalizing the conclusions to parings of instruments that were not included in the initial test is somewhat less reliable. The conclusions depend on the magnitudes of the various sources of disagreement between the instruments.

Another way to test the standardization sets would be to do simulations of a wide number of possible magnitudes of the various sources of disagreement between the instruments. While this could provide valuable theoretical information about what might happen, it is unclear how the conclusions might apply to specific instruments.

Does the simulation include realistic magnitudes of each of the possible sources of disagreement?

A third approach has been taken in this paper. This method is theoretical, but it comes to the heart of the issue of assessing the richness of a standardization set for a given mathematical model.

Consider the case where a hapless standardization set only included white, black, and gray samples, each of which were perfectly flat spectrally. In this degenerate case, the derivatives, as computed through Equation 6, and as used in Equation 8, would all be zero. As a result, the matrix $M(\lambda)^T \bullet M(\lambda)$ in Equation 11 would not be invertible, and it would not be possible to solve for $\vec{\beta}(\lambda)$. This is just a mathematical explanation for what we already know. If the standardization set is spectrally flat, it isn't possible to determine wavelength shift.

Take this example a step further. It is virtually impossible for a measured spectrum to have exactly the same reflectance at all wavelengths. So, if a standardization set includes actual measured spectra of white, black, and gray and those samples were virtually flat, spectrally, there would be a non-zero derivative and the matrix $M(\lambda)^T \bullet M(\lambda)$ would be invertible. This would appear to solve the issue mathematically, but yet from a practical standpoint, we know that the results would be disastrous.

The issue in this discussion is that we have used a binary characterization of $M(\lambda)^T \bullet M(\lambda)$; either the matrix is invertible or it is not. We need a way to quantify the numerically stability of the results. Numerically stability is a measure of how much a calculation is affected by small changes in the input data. A numerically unstable calculation will show big changes due to small amounts of noise in the data.

Numerical stability

There are two parts to the computation to be considered:

1) The determination of the vector of correction parameters $\beta(\lambda)$. The determination of $\beta(\lambda)$ is an infrequent event which is based on measurements of a standard reference set.
2) The use of those correction parameters to correct a measurement from one instrument to better agree with a measurement from another.

These will be addressed one at a time.

Error in determination of correction parameters

There are two considerations for assessing the error in determination of correction parameters. The first consideration is the error in the measurements and the second is the amplification of those measurements.

To recap the equation for the determination of $\beta(\lambda)$

$$\vec{\beta}(\lambda) = (M(\lambda)^T \bullet M(\lambda))^{-1} M(\lambda)^T \bullet (\vec{R}_r(\lambda) - \vec{R}_m(\lambda)) \tag{11}$$

where

$\vec{R}_r(\lambda)$ is the vector of reflectance values of the standardization samples at wavelength λ on the reference instrument. There is one element in this vector for every sample in the standardization set.

$\vec{R}_m(\lambda)$ is the vector of reflectance values of the standardization samples at wavelength λ on the instrument which is to be standardized. There is one element in this vector for every sample in the standardization set.

M is the matrix derived from the measurements of the standard reference set taken with the reference instrument. An example of this matrix (for Model #1) is shown below. Other models would include a different collection of columns.

$$M(\lambda) = \begin{bmatrix} 1 & R_m(1,\lambda) & \frac{dR_m(1,\lambda)}{d\lambda} \\ 1 & R_m(2,\lambda) & \frac{dR_m(2,\lambda)}{d\lambda} \\ \vdots & & \vdots \\ 1 & R_m(n,\lambda) & \frac{dR_m(n,\lambda)}{d\lambda} \end{bmatrix} \tag{12}$$

Error in measurements

One common specification for spectrophotometers is repeatability. This metric quantifies the degree that replicated measurements of the same sample agree when they are taken over a short period of time. A typical spec for this is 0.1% reflectance. Assume that both instruments will have this variation. The difference is computed between the two, so the standard deviation of the repeatability of the difference is 1.4 X 0.1% = 0.0014.

This is only one source of variation. The placement on the sample will necessarily be somewhat different between the two instruments. This paper will assume that the samples are uniform enough and the placement of the spectrophotometers on the samples is smaller than the basic repeatability, but still of minor significance. It

is thus assumed that each entry in the vector $(\vec{R}_r(\lambda) - \vec{R}_m(\lambda))$ has an error which is on the magnitude of 0.002 reflectance.

Sidebar - numerical instability of matrix inversion

The error magnification depends critically on that nature of the matrix. A brief introduction to this source of numerical instability is in order.

Matrix inversion may cause an unexpected amplification of small noise in measurements. The following simple example will give an appreciation for the concept of numerical instability. Suppose one wished to solve the following matrix equation for the values of x and y.

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{no solution} \qquad (13)$$

The equation clearly has no solution. To satisfy the equation, the sum of x and y must both be 3 and 4. Such an equation is not common in the real world where the entries in the matrix are likely measured quantities. One may instead see an equation like the following, which can actually be solved. Note that the equation is only very slightly different from the previous, with the 1 in the lower right corner replaced by a 1.01.

$$\begin{bmatrix} 3.00 \\ 4.00 \end{bmatrix} = \begin{bmatrix} 1.00 & 1.00 \\ 1.00 & 1.01 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad x = -97 \text{ and } y = 100 \qquad (14)$$

But the solution to this is numerically instable. A small change to the vector on the left may cause a comparatively large change in the solution. By changing the values of the vector on the left by only 0.01, the solution changes by about 2. The small perturbation was amplified by a factor of about 200.

$$\begin{bmatrix} 3.01 \\ 3.99 \end{bmatrix} = \begin{bmatrix} 1.00 & 1.00 \\ 1.00 & 1.01 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad x = -94.99 \text{ and } y = 98 \qquad (15)$$

Not all perturbations in the input vector will show this large of a change. In this equation, if the changes are in the same direction, the change in the solution is roughly the same size as the perturbation.

$$\begin{bmatrix} 3.01 \\ 4.01 \end{bmatrix} = \begin{bmatrix} 1.00 & 1.00 \\ 1.00 & 1.01 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad x = -96.99 \text{ and } y = 100 \qquad (16)$$

The matrix causes this numerical instability because it is very close to a matrix that is non-invertible. This similarity can be measured with the eigenvalues. The eigenvalues of the matrix are 2.005 and 0.0049875. It is no coincidence that we found an amplification of 200, since this is close to the reciprocal of the smallest eigenvalue.

Application to this problem

Ideally, a set of reference materials will be rich enough to provide a stable correction for the correction model that is chosen. This might not be the case, however, if a model is used which requires derivatives and the reference materials are limited to gently sloping reflectance spectra. This dearth of information is evinced by the matrix $M(\lambda)^T \cdot M(\lambda)$ being nearly non-invertible. Near non-invertibility means that the determination of the correction parameters will be prone to error.

The condition number is a traditional measure of how close a matrix is to being non-invertible. For a square matrix, this is the ratio of largest to smallest eigenvalue. However, this standard approach is not appropriate for this application, since the magnitude of the smallest eigenvalue is irrelevant.

A related measure is used in this paper to directly assess the propagation of error. The measure is based on the observation that an eigenvalue of a matrix is a measure of the extent that multiplying the corresponding eigenvector by the matrix magnifies the eigenvector. Therefore, the appropriate measure is the largest eigenvalue.

Since the matrix $(M(\lambda)^T \cdot M(\lambda))^{-1} M(\lambda)^T$ is in general not square, the singular value decomposition was computed and the largest singular value was identified as the noise magnification factor.

Statistical versus maximum

Repeatability is a statistical measure; it quantifies typical behavior for an instrument. The magnification factor described above is a worst-case measure; it quantifies how large the magnification could be for a particularly unlucky collection of measurement noise.

Outliers are to be expected from any measurement device. As such, allowances are made for the occasional aberrant measurement – it may be filtered out or otherwise disregarded. Ant harm that may be caused by an outlier is transitory, since it is only a single measurement. There is generally no benefit to be gained in analyzing worst case behavior.

Examination of the maximum errors in computing correction parameters is justified in this case because of the implications. While it may be unlikely that any given combination of measurement noise aligns substantially with the eigenvector of the largest eigenvalue, if this should happen, then all subsequent corrections may substantially ruin the agreement between two instruments.

$$\epsilon_{\beta 3} \approx 0.002 V_{max} \qquad (17)$$

$$\epsilon_{\beta 4} \approx 0.002 V_{max} \qquad (18)$$

where

$\epsilon_{\beta 3}$ is the magnified noise that is included in the estimate of $\beta_3(\lambda)$ that occurs during the process of determining the correction parameters,

$\epsilon_{\beta 4}$ is the magnified noise that is included in the estimate of $\beta_4(\lambda)$ that occurs during the process of determining the correction parameters, and

V_{max} is the largest eigenvalue of $(M(\lambda)^T \bullet M(\lambda))^{-1} M(\lambda)^T$.

Error in correction of a measurement

The second part, the correction of subsequent measurements is shown below for Model #2, which is the simplest model which includes a correction for bandwidth.

$$R_c(\lambda) = R_m(\lambda) + \beta_1(\lambda) + \beta_2(\lambda) R_m(\lambda) + \beta_3(\lambda) \frac{dR_m(\lambda)}{d\lambda} + \beta_4(\lambda) \frac{d^2 R_m(\lambda)}{d\lambda^2} \qquad (20)$$

The error due to noise in the bandwidth correction can be seen by adding noise terms to Equation 2.

$$R_c(\lambda) = R_m(\lambda) + \beta_1(\lambda) + \beta_2(\lambda) R_m(\lambda) + \beta_3(\lambda) \frac{dR_m(\lambda)}{d\lambda} + (\beta_4(\lambda) + \epsilon_{\beta 4})\left(\frac{d^2 R_m(\lambda)}{d\lambda^2} + \epsilon_{md}\right) \qquad (20)$$

where

$R_m(\lambda)$ is the measured reflectance of a sample at wavelength λ, as measured on the instrument to be corrected,

$R_c(\lambda)$ is the corrected reflectance of the sample at wavelength λ,

$\epsilon_{\beta 4}$ is the magnified noise that is included in the estimate of $\beta_4(\lambda)$ that occurs during the process of determining the correction parameters, and

ϵ_{md2} is the noise in the estimation of the second derivative from the measure sample.

The final term expands out to

$$(\beta_4(\lambda)+\epsilon_{\beta4})\left(\frac{d^2R_m(\lambda)}{d\lambda^2}+\epsilon_{md}\right)=\beta_4(\lambda)\frac{d^2R_m(\lambda)}{d\lambda^2}+\beta_4(\lambda)\epsilon_{md2}+\epsilon_{\beta4}\frac{d^2R_m(\lambda)}{d\lambda^2}\epsilon_{\beta4}\epsilon_{md2} \tag{21}$$

The error due to bandwidth correction is

$$Err_{bw}=\beta_4(\lambda)\epsilon_{md2}+\epsilon_{\beta4}\frac{d^2R_m(\lambda)}{d\lambda^2}\epsilon_{\beta4}\epsilon_{md2} \tag{22}$$

Similarly, the error due to wavelength shift is

$$Err_{ws}=\beta_3(\lambda)\epsilon_{md1}+\epsilon_{\beta3}\frac{d^2R_m(\lambda)}{d\lambda^2}\epsilon_{\beta3}\epsilon_{md1} \tag{23}$$

The magnitude of the noise in the estimates of the first and second derivatives can be determined from Equations 6 and 7, using the fact that the standard deviation of the noise of a weighted sum of random variables is equal to the sum in quadrature of the weights times the standard deviation of the random variables. The 0.002 in the following equations is the standard deviation of the noise for a single reflectance measurement at one wavelength.

$$\epsilon_{md1}\approx0.002\sqrt{\left(\tfrac{1}{2}\right)^2+\left(\tfrac{1}{2}\right)^2}\approx0.0014 \tag{24}$$

$$\epsilon_{md1}\approx0.002\sqrt{(1)^2+(-2)^2+(1)^2}\approx0.005 \tag{25}$$

xxxxxx

$$Err_{ws}\approx0.0014\beta_3(\lambda)+0.002V_{max}\frac{dR_m(\lambda)}{d\lambda}+0.0014\times0.002V_{max} \tag{26}$$

$$Err_{ws}\approx0.0014\beta_3(\lambda)+0.002V_{max}\frac{dR_m(\lambda)}{d\lambda}+2.8\times10^{-6}V_{max} \tag{27}$$

$$Err_{bw}\approx0.0014\beta_4(\lambda)+0.002V_{max}\frac{d^2R_m(\lambda)}{d\lambda^2}+0.005\times0.002V_{max} \tag{28}$$

$$Err_{bw}\approx0.0014\beta_4(\lambda)+0.002V_{max}\frac{d^2R_m(\lambda)}{d\lambda^2}+1\times10^{-5}V_{max} \tag{29}$$

Based on analysis of several hundred spectra, it is not unusual to see first derivatives of 0.20, and second derivatives of 0.10.

$$Err_{ws}\approx0.0014\beta_3(\lambda)+0.002\times0.002V_{max}+2.8\times10^{-6}V_{max} \tag{30}$$

$$Err_{ws}\approx0.0014\beta_3(\lambda)+4\times10^{-4}V_{max}+2.8\times10^{-6}V_{max} \tag{31}$$

$$Err_{ws}\approx0.0014\beta_3(\lambda)+4\times10^{-4}V_{max} \tag{32}$$

$$Err_{bw}\approx0.005\beta_4(\lambda)+0.002\times0.10V_{max}+1\times10^{-5}V_{max} \tag{33}$$

$$Err_{bw}\approx0.005\beta_4(\lambda)+2\times10^{-4}V_{max}+1\times10^{-5}V_{max} \tag{34}$$

$$Err_{bw}\approx0.005\beta_4(\lambda)+2\times10^{-4}V_{max} \tag{35}$$

The correction terms are $\beta_3(\lambda)\dfrac{dR_m(\lambda)}{d\lambda}$, and $\beta_4(\lambda)\dfrac{d^2R_m(\lambda)}{d\lambda^2}$, and the first and second derivatives are generally smaller than 0.20 and 0.10, respectively. The error term of $0.005\beta_4(\lambda)$ can be treated as insignificant in comparison. This brings us to a simple estimate for the magnitude of the error.

$$Err_{ws} \approx 4 \times 10^{-4}V_{max} \tag{36}$$

$$Err_{bw} \approx 2 \times 10^{-4}V_{max} \tag{37}$$

Potential standardization sets

As indicated above, there are two potential standardization sets which are being investigated for this study: The ChromaChecker and the Lucideon Print Standards.

ChromaChecker

ChromaChecker is a laminated card as shown in the image below. There is a total of 42 patches.
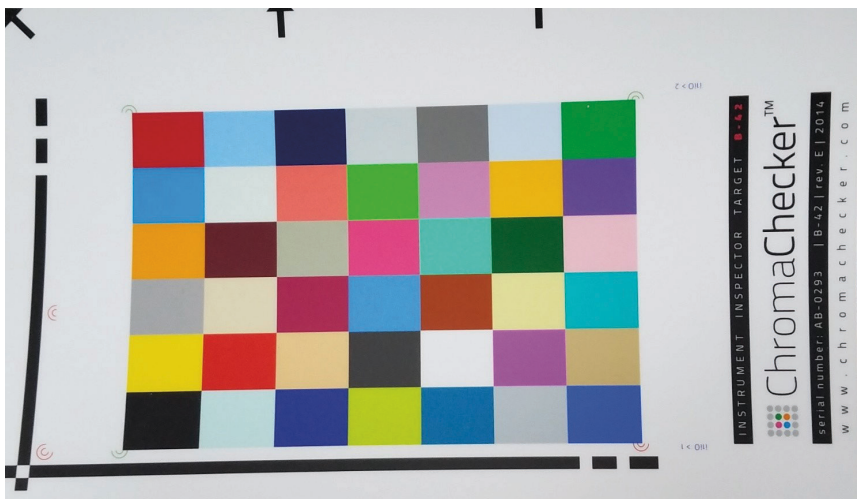
Figure 1: The Chromachecker

The spectra of the 42 patches is shown in the next graph. Two limitations are apparent from this plot. First, there is a lack of diversity in the lower wavelengths. At 380 nm, the highest reflectance is just over 20%, making it difficult to determine a gain parameter. For the most part, the spectra start low at 380 nm and rise in reflectance up to around 420 nm, meaning that nonlinearity will be hard to distinguish from wavelength shift. It is surmised that this set of patches were printed on paper with largely transparent inks. A different choice of may improve this.

Second, the spectra are largely flat above 650 nm. As a result, it will be hard to determine wavelength-related parameters.
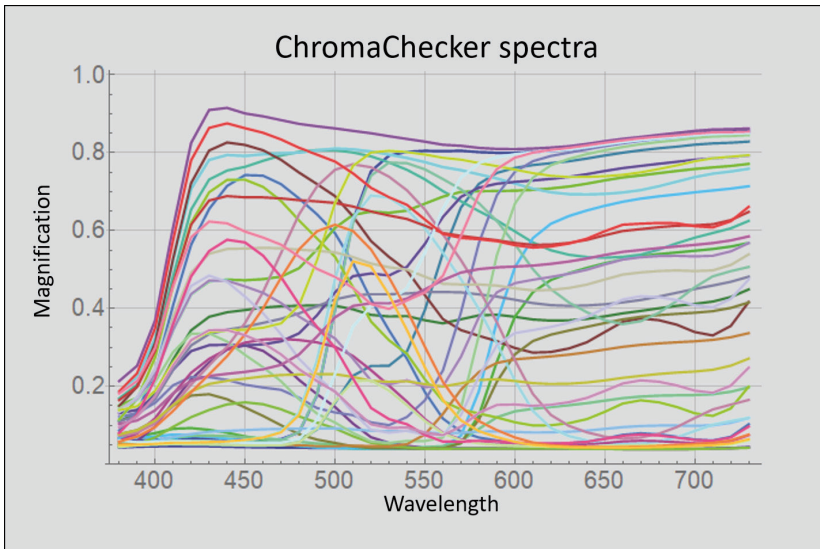
Figure 2: *Spectra of the 42 ChromaChecker patches*

The first derivatives of the spectra are shown in the next image. This graph underscores the lack of strong derivatives above 650 nm.
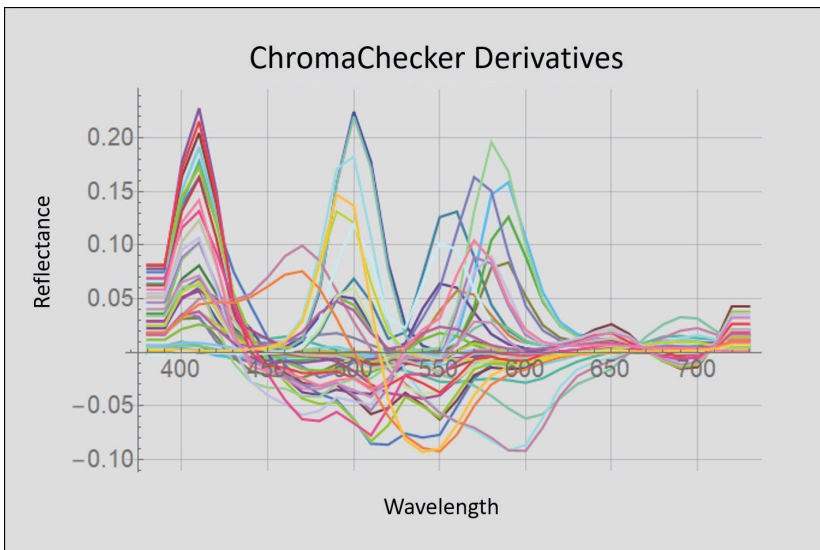

Figure 3: *First derivative spectra of the 42 ChromaChecker patches*

The next image shows the second derivative of the spectra. Ideally, one would like to have strong positive and negative second derivatives at all wavelengths. This is perhaps met in the middle of the spectrum, but there is a shortage elsewhere.
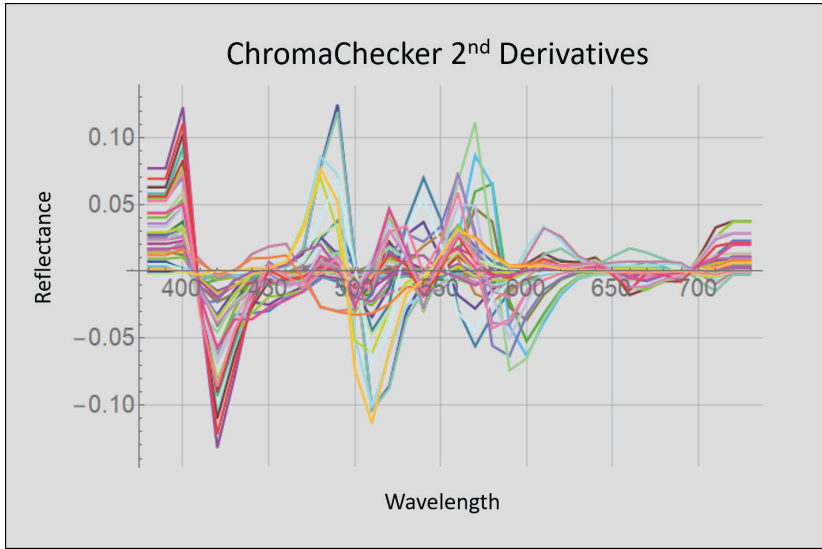
Figure 4: Second derivative spectra of the 42 ChromaChecker patches

Finally, we have a plot of the nonlinearity function. With the possible exception of 380 and 390 nm, there seems to be a decent diversity across the spectrum.
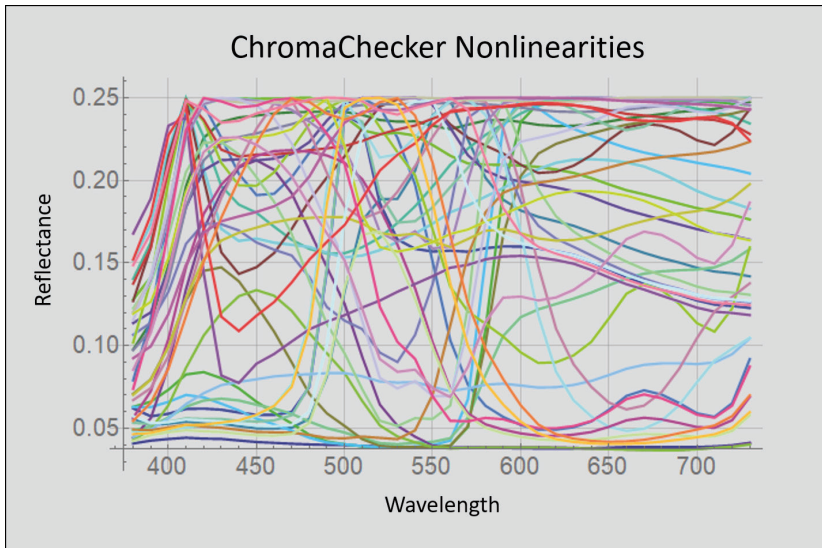
Figure 5: Nonlinearity function spectra of the 42 ChromaChecker patches

From this cursory analysis, it would appear that with the ChromaChecker, the red and blue ends of the spectrum may prove more difficult to accurately standardize one instrument to another. If the standardization algorithms are stable and the spectra are only to be used for conversion to CIELAB under one light source, this may not be a particular problem. The eye is less response at the ends of the spectrum.

Lucideon Print Standards

The Lucideon Print Standards are comprised of 14 ceramic tiles and are available with gloss, semi-gloss, and matte finish. They are mounted on a flat plate in a plastic case, as shown below. The plate is used to hold the spectrophotometer flat and stable while measurements are taken.

Figure 6: The Lucideon Print Standards, in their case

The next image is the spectra of the 14 Lucideon tiles. There are considerably fewer tiles, so the graph appears considerably less full. At the blue end, there is a considerably wider range of reflectance values available. It is hard (from this view) to tell if there is sufficient derivative information available.
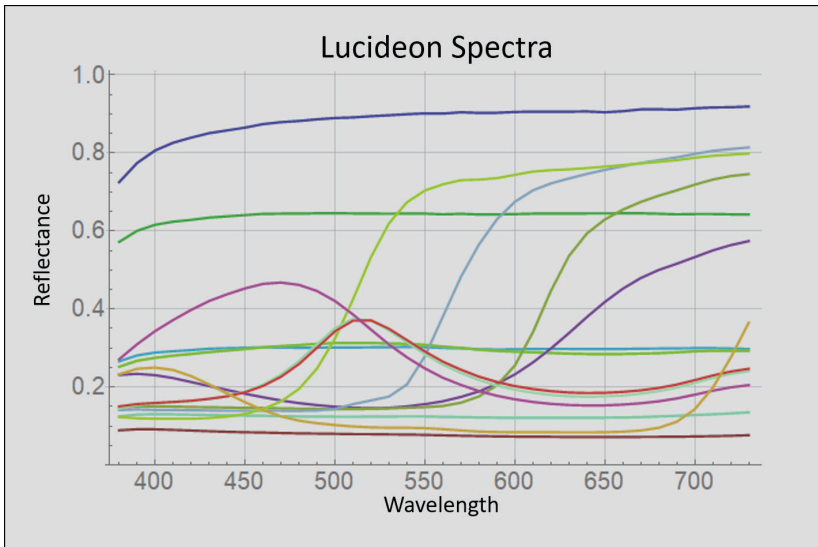
Figure 7: Spectra of the Lucideon Print Standards

Next, we have the first derivatives. There appears to be a more even coverage than the ChromaChecker set, but the overall magnitude is smaller – this graph is scaled up to only 0.10.
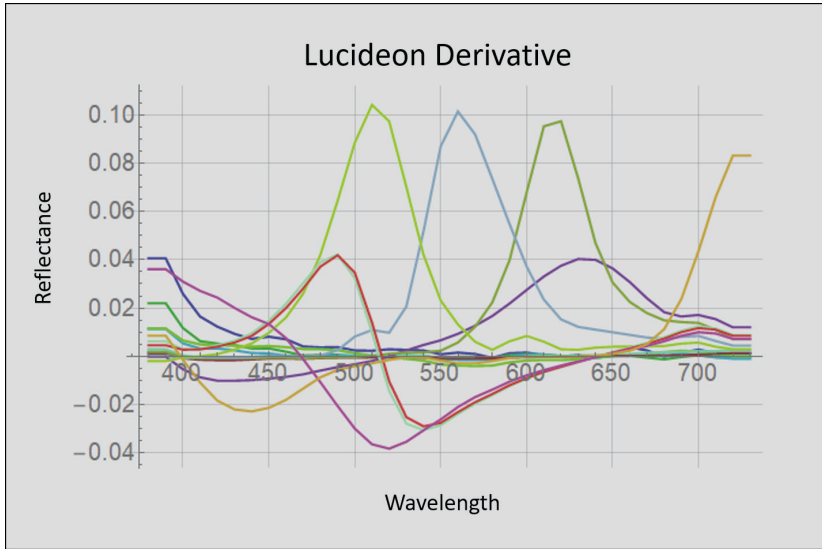
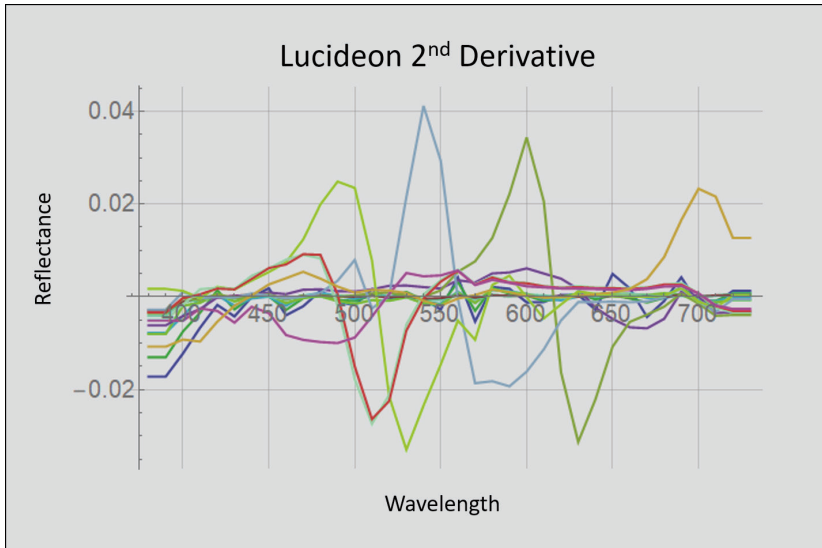*Figure 8: First derivative spectra of the Lucideon Print Standards*

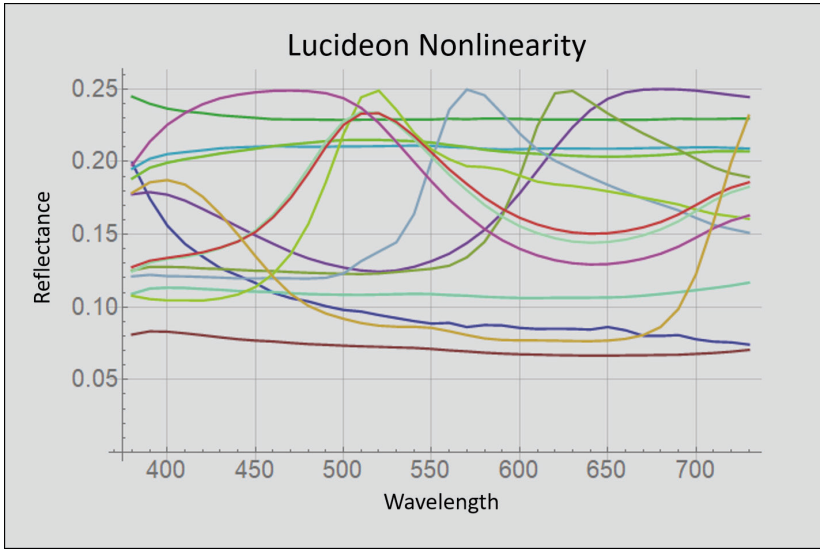*Figure 9: Second derivative spectra of the Lucideon Print Standards*

Figure 10: Nonlinearity spectra of the Lucideon Print Standards

For comparison, six other sets of standards were included in the analysis. None of these additional sets of six are being recommended as a standardization set, due to limitations in the aforementioned required properties. The additional six are BCRA ceramic tiles, Behr paint ramps, Munsell ColorChecker, Pantone primaries, and set of 155 Sherwin-Williams paint samples. The Behr paint ramps were used in a previous study [Seymour 2013].

Figure 11: The 24 colors in the Behr paint samples

Figure 12: Sherwin-Williams set of 155 paint samples

Error magnification in computing correction parameters

The plots below show the first singular values for the ChromaChecker set with each of the five mathematical models. The vertical axis is V_{max}, the largest eigenvalue of the matrix used to compute the correction parameters.
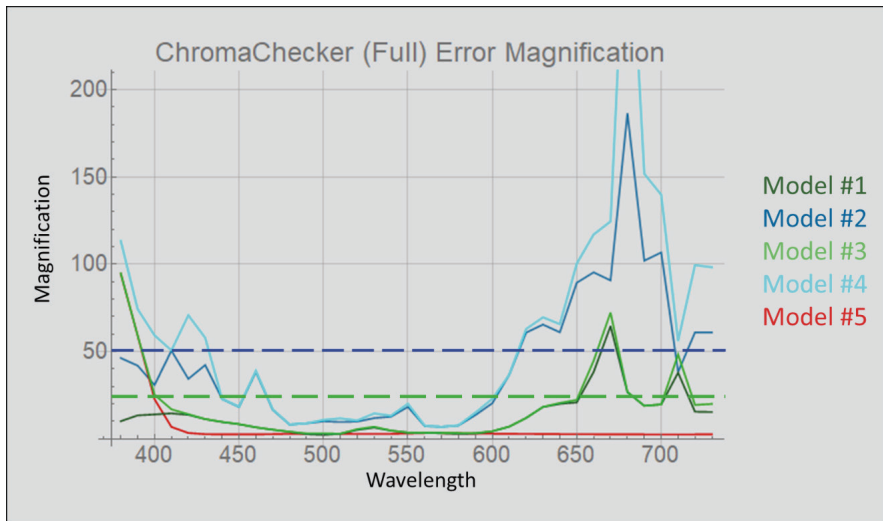

Figure 13: Error magnification for the 5 models on the ChromaChecker

Note first that the magnitude of the error magnification groups according to the number of derivatives required in the mathematical model. Models #2 and #4 (lighter and darker blue graphs) require first and second derivatives and have the highest errors. Models #1 and #3 (lighter and darker green graphs) require only first derivatives and have intermediate errors. Model #5 (red) does not require any derivatives, and shows the most stability.

The differences in stability caused by the addition of a nonlinear term (comparing light and dark blue, or comparing light and dark green) is insignificant in comparison. Basically, having a black, a white, and a gray sample is sufficient for stable determination of nonlinearity.

The most startling thing about this plot is the magnitude of the instability of the models involving the second derivative. In certain parts of the spectrum, any error in measurement could by magnified by a factor of up to 300 in computing the correction parameters. From Equation 37 (repeated below for convenience), we see that error due to an unlucky calibration could be on the order of 0.2, which is to say, 20% points in reflectance. Adding a potential 6% R noise to correct for 1% change in reflectance is completely unsatisfactory.

$$Err_{bw} \approx 2 \times 10^{-4} V_{max} \tag{37}$$

The dashed lines in Figure 13 are the cutoff where Equation 37 would predict errors of up to 1% R, which might be considered marginally acceptable for some non-demanding applications. The blue dashed line is the cutoff for the models which incorporate the second derivative. The green dashed line is the cutoff for the models which incorporate only the first derivative.

The graph below shows the same test applied to the Lucideon Print Standard. The magnitude of the errors for models #2 and #4 are on par with those of the ChromaChecker.
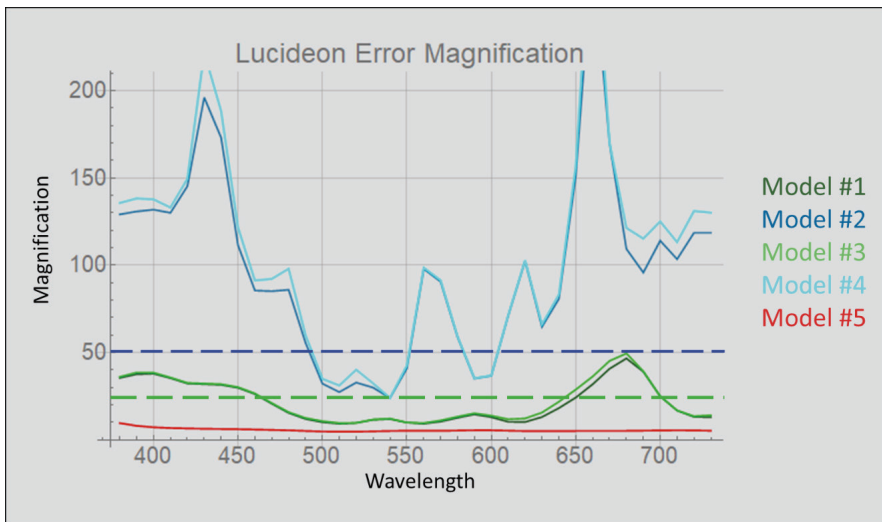

Figure 14: Error magnification for the 5 models on the Lucideon Print Standard

This same analysis was applied to the other sets. It was found that the Munsell ColorChecker (shown below) had the best performance for the models which use the second derivative. If we ignore the problem area at 400 nm and below, the maximum error magnification is marginally acceptable for both models with second derivatives (Models #2 and #4), and also for models with only first derivatives (Models #1 and #3).
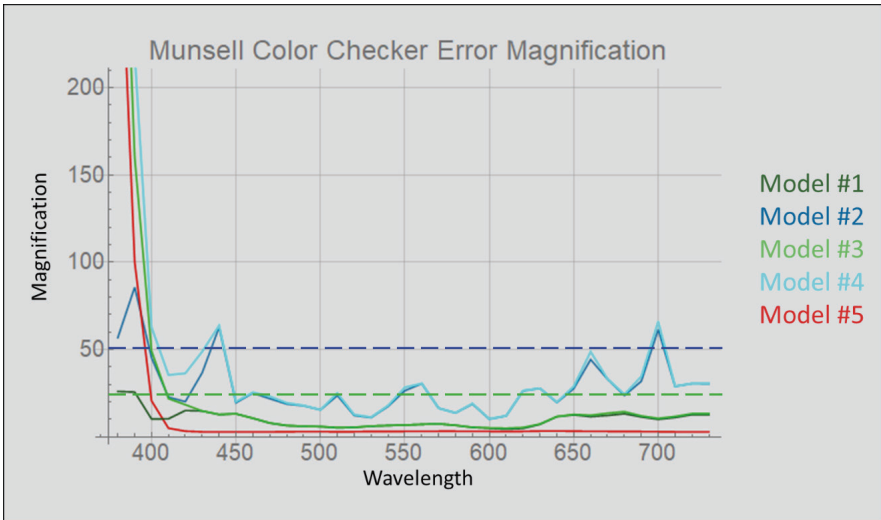

Figure 15: Error magnification for the 5 models on the Munsell ColorChecker

All of the sample sets have been shown to be completely unacceptable at many wavelengths for determining a correction between two spectrophotometers for bandpass.

Is an acceptable set possible?

Various combinations of sample sets were evaluated on the assumption that deficiencies in one set could be covered by strengths in another sample set. If the two key sets are combined (the Lucideon and the ChromaChecker), the result is an improvement over either. Much of the spectrum is marginally acceptable for the second derivative models, and all but a region around 670 nm is acceptable for the first derivative models.
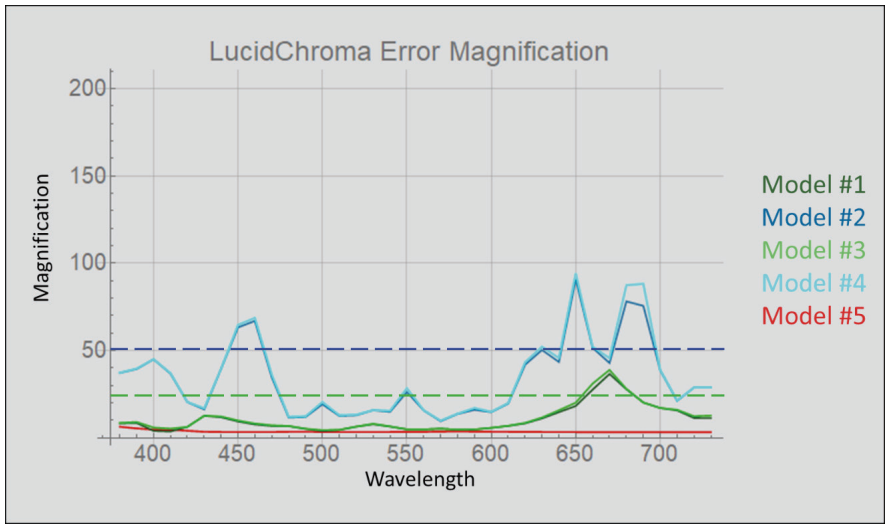
Figure 16: Error magnification for the combined Lucideon and ChromaChecker sets

A better combination was found by combining the Lucideon and the Munsell samples. As seen in Figure 17, the set is at least marginally acceptable at all wavelengths. For the methods which do not include the second derivative, it appears quite acceptable at all wavelengths. However, the Munsell ColorChecker is potentially not durable enough for this application.
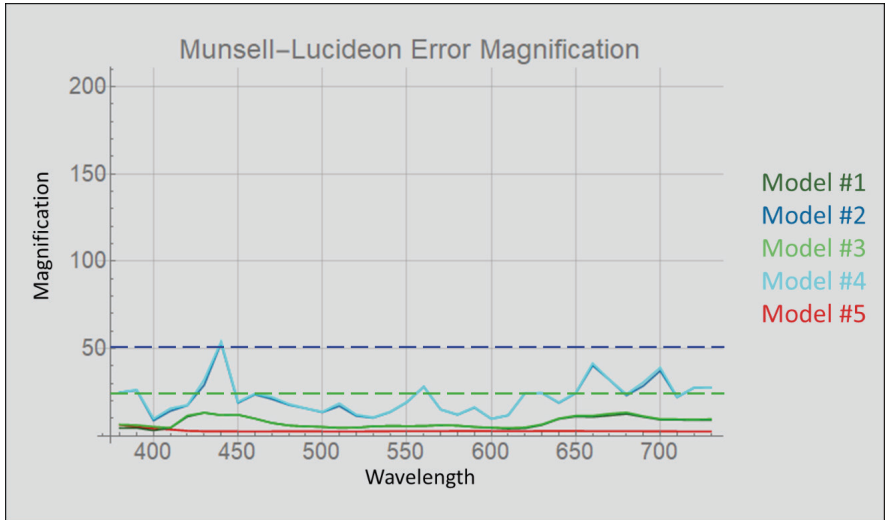

Figure 17: Error magnification for the combined Lucideon and Munsell sets

As an extreme test, all of the samples (301 in total) were combined into one potential standardization set. The error magnification values are shown in Figure 18. Note that this plot has been scaled differently that the previous, having a maximum of 20 rather than 200. This shows that the combination of all the samples would provide a standardization that is quite stable numerically.
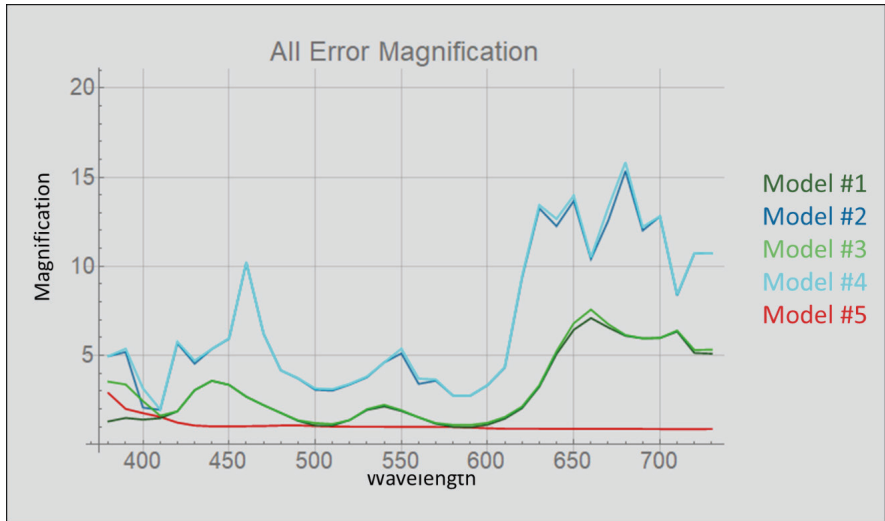

Figure 18: Error magnification for all the samples combined

These results show that, while the two existing sets of reference materials may only be marginally acceptable, a set which is numerically stable is physically possible.

Conclusions

Two sets of potential standardization materials were evaluated for numerical stability with five different mathematical models. These sets are assumed to be physically durable and stable enough for standardization of spectrophotometers in an industrial setting.

Certain of the mathematical models utilize the second derivative of the spectra with respect to wavelength to quantify (and hence correct) a difference in bandpass between two instruments. It was found that the two candidates standardization sets were numerically instable on these models for much of the spectrum. The end result is a potential exacerbation of the discrepancies between instruments.

Certain other mathematical models utilize the first derivative to account for differences in wavelength position. This calculation is found to be somewhat more stable, but the two candidate sets are only marginally numerically stable at many wavelengths.

Seven other sets of physical standards were evaluated. These sets are likely to be lightfast, but none of them are expected to have the durability required for industrial use. However, certain of them have spectral that have the requisite numerical stability. This demonstrates the potential for the development of a set of standardization materials which are durable, lightfast, and provide for numerically stable standardization of one spectrophotometer to another.

Acknowledgments

Bibliography

Berns, Roy S. and Kelvin H. Peterson. *Empirical Modeling of Systemic Spectrophotometric Errors*, Color Research and Application 13.4, pp. 243–256 (Cited on pages 172, 173). (1988)

R.S. Berns, and L. Reniff, *An Abridged Technique to Diagnose Spectrophotometric Errors*, Color Research and Application, 22(1), 51(10) (1997)

Chung, Sidney Y., K.M. Sin, and John H. Xin, *Comprehensive comparison between different mathematical models for inter-instrument agreement of reflectance spectrophotometers*, Proc. SPIE 4421, 789 (2002)

Ingleson, Alan and Michael H. Brill, *Methods of Selecting a Small Reflectance Set as a Transfer Standard for Correcting Spectrophotometers*, Color Research and Application,Volume 31, Number 1, February 2006

ISO/NP TS 23031, Graphic technology — Assessment and validation of the performance of spectrocolorimeters and spectrodensitometers, ISO TC 130, currently under development

Nussbaum, Peter, Aditya Sole, Jon Y. Hardeberg, *Consequences of Using a Number of Different Color Measurement Instruments in a Color Managed Printing Workflow*, TAGA 2009

Rich, D. C., and D. Martin, *Improved model for improving the inter-instrument agreement of spectrocolorimeters*, Analytica Chemica Acta, 380, (2-3), 263-276, (1999).

Robertson, A. R., *Diagnostic Performance Evaluation of Spectrophotometers*, presented at Advances in Standards and Methodology in Spectrophotometry, Oxford, England, 1986

Seymour, John, *Evaluation of Reference Materials for Standardization of Spectrophotometers*, TAGA 2013

Seymour, John, *Assessment of the Sources for Disagreement Between Two Spectrophotometers*, TAGA 2014

Spooner, David L., *Translucent blurring errors in small area reflectance spectrophotometer and densitometer measurements*, TAGA 1991

Van Aken, Harold, and Ronald Anderson, *Method for maintaining uniformity among color measuring instruments*, US patent 6,043,894, 2000

Van Aken, Harold, and Ronald Anderson, *Method for maintaining uniformity among color measuring instruments*, US patent 6,559,944, 2003

Van Aken, Harold, Shreyance Rai, Mark Lindsay, Richard Knapp, *System and method for transforming color measurement data*, US Patent 7,116,336, 2006

Appendix

ChromaChecker (Full) Error Magnification

Behr Paint Ramps Error Magnification

ChromaChecker (Abridged) Error Magnification

ChromaChecker (Full) Error Magnification

Munsell Color Checker Error Magnification

Munsell–Lucideon Error Magnification

Pantone Primaries Error Magnification

Sherwin Williams Error Magnification