



Inter-Instrument Agreement

Software to the Rescue?

To be useful, spectrophotometers must agree in their measurements. Software is now available to characterize a spectrophotometer parametrically and correct it to measure like a reference instrument. When should such software be used as a diagnostic tool and when can we trust it as a correction tool? Comparison of two manufacturers' hardware and software shows that correction is most trustworthy when the software and instruments are from the same manufacturer, and the measured test specimens are similar to the BCRA tiles used to perform the correction.

INTRODUCTION

The exchange of reflectance data among retail and branded apparel companies with their suppliers has increased substantially over the past five years. These electronic color programs are creating a "flatter color space" by connecting tiers of suppliers across continents via a common color language and protocol. Unlike text, digital music, and images, the reliability of reflectance data is often challenged and the variables associated with this process have been previously reported. The instrument, or spectrophotometer, in most cases provides the baseline upon which all other factors are built. In this paper, we will review sources of error in reflectance measurement, especially as it pertains to textiles. We will also discuss the process of spectrophotometer correlation (profile-based correction)---whereby a spectrophotometer is characterized parametrically and corrected to measure like a standard instrument by software and a series of reflectance standards. The content of the present paper is based on two earlier presentations to the Inter-Society Color Council¹ and to the American Association of Textile Chemists and Colorists².

PROFILE-BASED CORRECTION

To profile a color-measuring or color-producing device means to quantify its input-output relationships. In spectrophotometry, profiling is used to effect instrument correction---whereby a spectrophotometer is characterized parametrically and corrected to measure like a reference instrument by software and a series of reflectance standards. Such profile-based correction (PBC) has been discussed for a long time.^{3,4,5} It is time to look at the question of when instruments should be profiled for correction. Profile-based correction typically starts from a model equation such as the following:

$$R_{ci} = A + B R_{mi} + C R'_{mi} + D R''_{mi} + E R_{mi} (100 - R_{mi}). \quad (1)$$

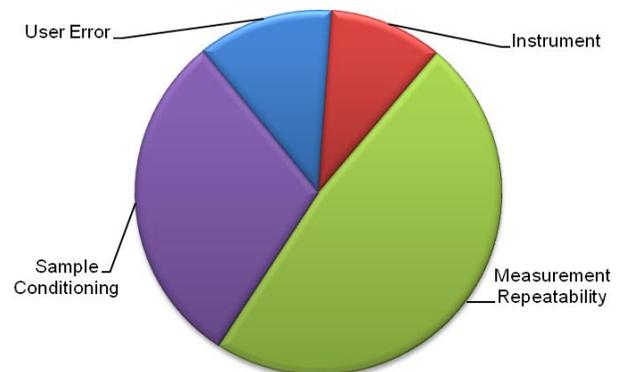
Here, R_{mi} is the i 'th measured reflectance, R_{ci} is the i 'th corrected reflectance, all variables implicitly depend on

wavelength, and ' and '' refer to the first and second derivatives of R_{mi} with respect to wavelength. The corrections in Eq. 1 are offset (A), gain change (B), wavelength-scale change (C), bandwidth change (D), and some nonlinearities (E). The particular nonlinearity in Eq. (1) is taken from Berns and Petersen, 2 but other expressions have been used as well. Equation (1) is used twice in profile-based correction: once with known specimens such as the BCRA tiles with parameters A-E (at each wavelength) in the solve state, and again with measured reflectances, now-known A-E, and R_{ci} in the solve state.

LIMITATIONS OF PBC METHOD

Profile-based correction has limitations. First and foremost, it can correct only consistent instrument error, not other errors such as user error, sample conditioning dependence, and random errors affecting measurement repeatability. These errors are apportioned approximately as in Figure 1, and are explained in Ref. [6]. Clearly, one cannot expect user errors or repeatability (random) errors to be corrected by software. The same is true of conditioning errors, if the software is made to depend on the instrument with no knowledge of the samples. Conditioning error refers to failure to control the temperature and humidity in the sample environment: For textiles, ASTM Standard Practice D1776-98 specifies a temperature of 21 +/- 1° C and 65 +/- 2% relative humidity (RH). Substantial deviations from these numbers create the conditioning errors in spectrophotometry.

Figure 1 Kinds of error in spectrophotometry



Given a consistent instrument error, PBC can sometimes come to the rescue. It can compensate instrument malfunction due to changes in wavelength scale, bandwidth, or gain. However, it

will not fix optical alignment or bad pixels. Also, it cannot correct, e.g., the numerical noise due to wavelength error (from mismatch between the sample and reference signals.) Also, PBC cannot fix the following typical symptoms of aging spectrophotometers: dirty/dusty optics; stained/hazy filters and lenses; degraded spheres and specular ports; light-source degradation; and imprecise sample mounting. Finally, a corrected spectrophotometer may lose traceability if its native data are unavailable, especially if the reference instrument is un-traceable. Despite the fact that all of color management is based on the philosophy implicit in PBC, the tolerances of industrial spectrophotometry are much tighter than in such worlds as ICC---too tight to permit digital vagaries.

Profiling a spectrophotometer may be useful in diagnosis, and such diagnosis entails a threefold decision: Do nothing (if the performance is good), service the instrument (if its performance is bad), or perform software correction using PBC. That decision should depend not only on the mean ΔE error, but also on the specimens for which the error occurs. Chromatic colors with $\Delta E > 0.5$ can still support PBC, but neutral colors that are off by as little as 0.1 ΔE in the presence of much lower error on the white tile indicate un-fixable errors (e.g. stray light, sample-to-reference crosstalk, gloss/sphere problems, sample-mounting/ aperture-size problems and translucency problems.) Because of the above risks, it is important not to perform PBC on instruments that are performing too well or too poorly. To inform customers of when PBC is appropriate, a test of the candidate algorithm (in conjunction with its implied reference instrument) is needed now that PBC has become more popular than ever.

FIELD TEST OF PBC METHOD

Given this need, a test was performed to determine the quality improvement to be expected from PBC, across several kinds of reflecting specimens, across spectrophotometers of the same model and manufacturer, and across spectrophotometers of different models and manufacturers. A total of 42 test specimens were examined: 22 textile and 20 non-textile. Eight units each of d/8 (sphere) spectrophotometers from two manufacturers (A and B) were tested. Three sets of measurements were performed: "as is" (without PBC), profiled using software from manufacturer A, and profiled using software from manufacturer B. Quantities ΔE_{cmc} and ΔE_{cielab} were evaluated for comparison with the arithmetic mean $L^*a^*b^*$ of instruments split by manufacturer, of all 16 instruments together, and of various specimen subsets (BCRA tiles, textiles, and all specimens). Also, quantities ΔE_{cmc} and ΔE_{cielab} were evaluated for comparison with specific instruments (A7 and B7) for various specimen subsets (BCRA tiles, textiles, and all specimens). We examined in each case how many samples "got better", "got worse", or did not change when profiling was brought to bear.

ΔE_{cielab} from Mean of Devices

As an example of the results obtained, Table 1 shows that both profiling programs were able to improve the inter-instrument agreement with the mean on BCRA tiles, but with consistency only for instruments of the same manufacturer. (Note: Here, OEM means each instrument was corrected by software from its own manufacturer.)

Table 1 ΔE_{cielab} comparison with mean of all 16 instruments, over BCRA tiles.

	As Is	Software		
		A	B	OEM
Instrument A	0.25	0.12	0.35	-
Instrument B	0.19	0.07	0.10	-
All Instruments	0.28	0.12	0.28	0.21

In contrast, it is clear from Table 2 that improvements in agreement with the mean are insignificant when all 16 instruments are corrected to their mean. As borne out by both Table 1 and Table 2, the values improve when the instruments are corrected to the mean of eight from the same manufacturer.

Table 2 ΔE_{cielab} comparison with mean of all 16 instruments, over all samples.

	As Is	Software		
		A	B	OEM
Instrument A	0.28	0.23	0.38	-
Instrument B	0.31	0.23	0.25	-
All Instruments	0.35	0.28	0.36	0.31

ΔE_{cmc} from a Real Instrument

The above analysis might fail to convince because day-to-day textile processes use a metric of ΔE_{cmc} rather than ΔE_{cielab} , and also that inter-instrument agreement is measured with respect to a particular spectrophotometer unit (not the master unit, but one in the supply chain) and not with respect to an average of units. Therefore, the following results were obtained by computing the ΔE_{cmc} error from fielded instruments A7 and B7, which are units from manufacturers A and B respectively. The results for BCRA tiles, analogous to Table 1, are shown in Table 3. Again it can be seen that there are mixed results when non-native software is used, but results are always good with native software. These results are the same as those compared with the mean instrument and using ΔE_{cielab} .

Table 3 ΔE_{cmc} comparison with instruments A7 and B7, over BCRA tiles.

	As Is	Software		
		A	B	OEM
Instrument A7	0.26	0.11	0.28	0.18
Instrument B7	0.31	0.14	0.28	0.19

The results for all samples, analogous to Table 2, are shown in Table 4. As in Table 2, the results were mixed, but did not achieve the improvement predicted with BCRA tiles. Use of non-native software again seemed to degrade the inter-instrument agreement.

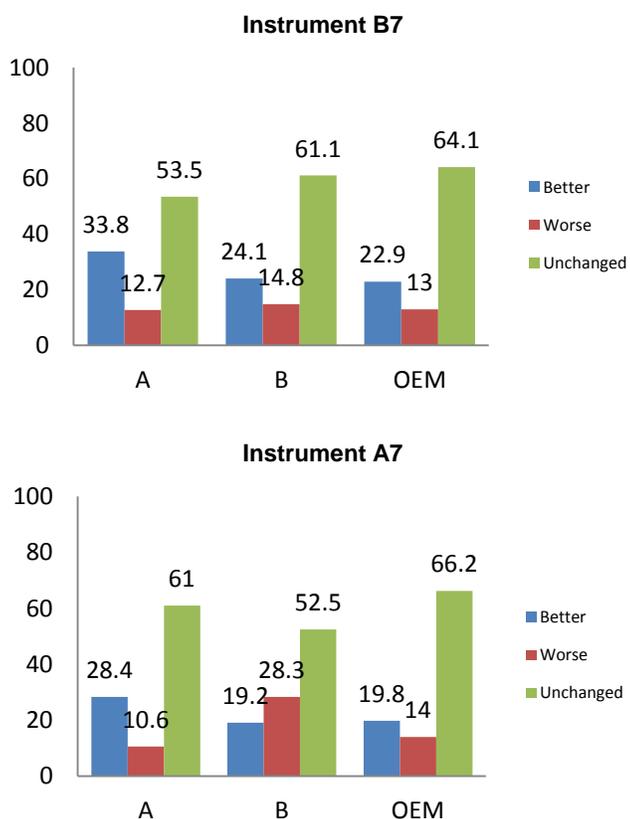
Table 4 ΔE_{cmc} comparison with instruments A7 and B7, over all samples.

	As Is	Software		
		A	B	OEM
Instrument A7	0.26	0.21	0.34	0.22
Instrument B7	0.32	0.24	0.30	0.27

How Many Samples Improved?

For the ΔE_{cmc} metric and specific comparison instrument used in Section 4.2, we then asked what percentage of samples did PBC improve by at least 0.05, how many were made worse by at least 0.05, and how many were unchanged (by more than 0.05 ΔE_{cmc} units). The results appear in Figure 2.

Figure 2 How the samples fared



There were more improved samples than worsened ones only when each instrument was profiled by software from its own manufacturer. In each case, most of the samples were not changed significantly. Our data analysis generally indicated that the textile samples did not improve significantly with PBC. The number of textile samples that improved was only slightly greater than the number that got worse after profiling. Not all the news was bad: In cases where the BCRA values improved by more than 0.15 CIELAB unit, textile performance improved significantly though to a lesser degree.

Diagnostic Tests

It was found that diagnostic tests on non-native instruments failed to identify or falsely identified problems with these instruments. Consequently, instrument diagnostic tests and PBC should generally be performed by the manufacturer's own software. As with PBC tests, one should also expect diagnostic performance on real-world samples to be worse than those obtained from the samples that were used for profiling (BCRA tiles).

CONCLUSION

In conclusion, it is good to keep in mind that improvements in inter-instrument agreement gained through PBC of a properly functioning spectrophotometer will be insignificant compared to the error induced by poor measurement, variation in sample conditioning, and operator error. Furthermore, although diagnostic profiling and PBC with your instrument manufacturer's profiling program is a critical part of a total measurement program, it is not a substitute for attention to detail in the measurement.

REFERENCES

1. C. McLellan, D. Randall, A. Ingleson, and M. H. Brill, "To profile or not to profile: the user's question," ISCC Annual Meeting, Ottawa, Canada, 14-15 May 2006.
2. K. R. Butts, M. H. Brill, and A. Ingleson, "Reflectance in perspective," AATCC Symposium, Atlanta, GA, 31 October to 2 November 2006.
3. R. S. Berns and K. H. Petersen, "Empirical modeling of systematic spectrophotometric errors," *Color Res. Appl.* 13 (1988), 243- 256.
4. D. C. Rich and D. Martin, "Improved model for improving the inter-instrument agreement of spectrophotometers," *Analytica Chimica Acta* 380 (1999), 263-276.
5. A. Ingleson and M. H. Brill, "Methods of selecting a small reflectance set as a transfer standard for correcting spectrophotometers," *Color Res. Appl.* 31 (2006), 13-17.
6. K. R. Butts, "Keys to reliable digital color communication," *AATCC Review* 4, No. 4 (2004) 15-19.