Considerations and Shortfalls of Instrument Evaluation Using ASTM E2214

David R. Wyble[1], Michael H. Brill[2], Danny C. Rich[3], Peter Fortini[4]
1.Munsell Color Science Laboratory, RIT; 2. Datacolor;
3. Color Research Laboratory, Sun Chemical Corporation; 4. Wyeth

ASTM has published the practice E2214-02, *Standard Practice for Specifying and Verifying the Performance of Color Measuring Instruments.* The goal of this practice is to standardize the procedures and terminology used for instrument evaluation, and also to improve communication between users and manufacturers and between customers and suppliers. The experimental procedures and mathematical techniques recommended in E2214 have been analyzed and will be published in a pair of upcoming *Color Research and Application* articles[1,2]. Most of the contents of those articles has been presented to ISCC previously. Briefly, they reported the results of a long term instrument evaluation of twelve commercial spectrophotometers. These instruments were analyzed for short-, medium-, and long-term repeatability, as well as inter-instrument comparison and accuracy. The primary focus of E2214 is on a multivariate analysis rather than the traditional univariate approach (eg: average ΔE). While some of the E2214 multivariate techniques appear mathematically sound, other do not. This paper will address some mathematical shortfalls, provide a few cautions, and propose solutions that are mathematically rigorous and should result in more useful analyses for the user community.

The recommended multivariate analyses are extensions of typical hypothesis testing to three-dimensional data. One fundamental requirement for many of these tests is that the underlying data be distributed multivariate-normally. Establishing normality can be accomplished using several procedures. Using the medium-term color difference results, we will report the results of applying some common methods of determining the normality of data. Medium-term results are hourly measurements, and should be representative of the drift user might experience during a single day of instrument use. Measurements are treated as 80 pooled colorimetric triplets. The normality evaluations will be made on XYZ and CIELAB values as well as their differences and univariate color difference ΔE^*_{ab}.

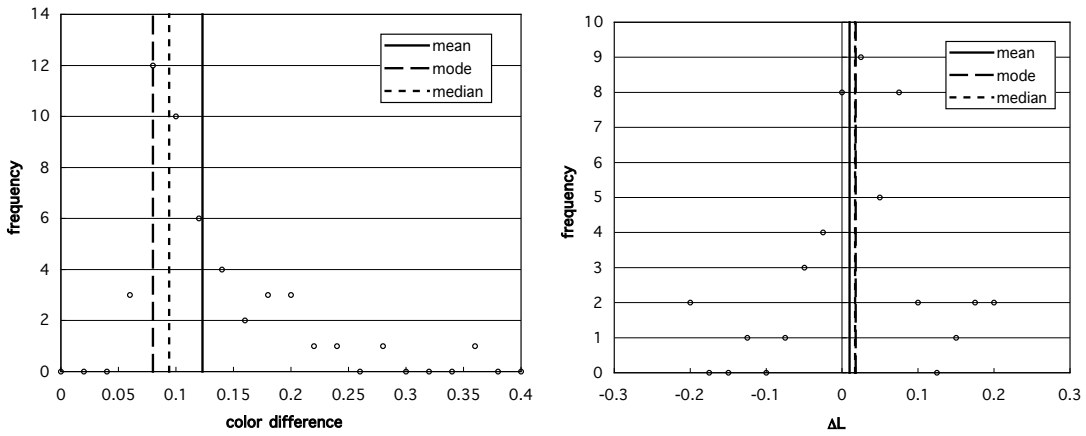

Figure 1. Qualitative evaluation of normality for color difference histograms of 80 medium-term measurements. Vertical lines show the mean, mode, and median of each distribution. The left chart shows DE^*_{ab}. which shows how the three metrics vary for non-normally distributed data. The right chart shows only the DL component the same measurements. This exhibits more normal behavior*

Following the discussion of normality, we will review proposed solutions directly addressing the mathematical techniques in E2214. The discussion starts by exposing some of the known problems with E2214. The multivariate approach has at its heart the covariance matrix of color difference values. This matrix defines an ellipsoid which, to a given statistical significance, can be considered to describe the expected behavior of an instrument. Color difference can either be examined along each respective dimension or as a distance in a 3D space. Differences equal to zero indicate an instrument exhibiting no drift over time. For either the 3D or single-dimension case, the statistical test answers the question "is this color difference statistically different from zero?" E2214 techniques based on the variances along the cardinal directions in CIELAB color space and their covariances result in tolerances that are extremely tight, generally rejecting all measurements as significantly different from zero. As in the Billmeyer and Alessi[3] study this may arise because the actual distributions have a kurtosis greater than expected from the trivariate Normal distribution.

The results presented here are twofold. First, an examination of the normality of colorimetric data should provide an understanding of the appropriateness of all statistical testing require normal data. Second, the description and techniques of E2214 instrument evaluation will be improved and therefore much more useful for the color measurement community.

1. DR Wyble and DC Rich, Evaluation of Methods for Verifying the Performance of Color-Measuring Instruments. Part I: Repeatability **32** in press, *Color Research and Application* (2007)
2. DR Wyble and DC Rich, Evaluation of Methods for Verifying the Performance of Color-Measuring Instruments. Part II: Inter-instrument Comparison **32** in press, *Color Research and Application* (2007)
3. FW Billmeyer Jr. and PJ Alessi, Assessment of Color-Measuring Instruments, *Color Research and Application* **6** 195-202 (1981)