

A White Paper

Made Available by

Committee for Graphic Arts Technologies Standards (CGATS)

Graphic technology — Improving the inter-instrument agreement of spectrocolorimeters

**Dr. Danny C. Rich
Color Physicist**

January 2004

DISCLAIMER

Publication of this Technical White Paper is at the request of members of the Committee for Graphic Arts Technologies (CGATS) Subcommittee 3 (Metrology), and with the permission of the author, Dr. Danny C. Rich.

Questions and comments regarding this White Paper should be addressed to the CGATS Secretariat, NPES The Association for Suppliers of Printing, Publishing and Converting Technologies, 1899 Preston White Drive, Reston, Virginia 20191-4367. Such questions and comments will be forwarded to the author.

This document is being made publicly available at no cost for informational and reference purposes only. Neither NPES The Association for Suppliers of Printing, Publishing and Converting Technologies, Sun Chemical, nor the Committee for Graphic Arts Technologies Standards (CGATS) makes any claim to the validity of the contents of this document.

This document does not relieve the reader of any burden imposed by law, and the reader should independently investigate and verify the performance, results, safety and efficacy of his or her own product or process, even if it follows any recommendations contained in this document. CGATS, its subcommittees, and NPES expressly disclaim liability for any damages or injury arising out of or related to the use of information contained in this document.

**Copyright ©2004 – NPES The Association for Suppliers of Printing, Publishing and Converting Technologies
All rights reserved.**

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission from either NPES or the author.

Contents

Abstract iv
 Introduction iv
 1 Scope 1
 2 Theory 1
 3 Experiments 3
 4 Results 4
 5 Discussion 5
 6 Conclusions 7

Figures

1 Spectral transmittance of a red specimen (solid curve) 8
 2 Extension of Figure 1, showing first and second derivatives 9
 3 Schematic illustration of an ideal linear reflectometer versus two real instrument curves 10
 4 Test sample set of 420 specimens displayed on CIELAB a*, b* plane 11
 5 Test sample set of 420 specimens displayed on CIELAB a*, L* plane 12
 6 Test sample set of 420 specimens displayed on CIELAB b*, L* plane 13
 7 Test samples of 718 ink-on-film specimens displayed on CIE a*, b* plane 14
 8 Test samples of 718 ink-on-film specimens displayed on CIE C*, L* plane 15
 9 Plot of model parameter B₃ versus wavelength showing the linear change in bandwidth for the ChromaSensor CS-5 correlated to a Spectraflash 500 16
 10 Histogram of the 420 color differences between the ChromaSensor CS-3 and the Spectraflash 500 before application of the correlation model 17
 11 Histogram of the 420 color differences between the ChromaSensor CS-3 and the Spectraflash 500 after application of the correlation model 18
 12 Histogram of the 718 color differences between the 0°:45° and the di:8° before application of the correlation model 19
 13 Histogram of the 718 color differences between the 0°:45° and the di:8° after application of the correlation model 20

Tables

1 Regressing coefficients for CS-5 regressed onto Spectraflash 500, specular component included 21
 2 Model fitting results 22
 3 Statistical analysis of the histograms shown in Figures 9 and 10 24
 4 Results of the average inter-instrument model in terms of CIELAB ΔE test instruments versus Spectraflash 500, with the specular component included for the 420 total samples 24

References 25

Biographical Information on the Author 27

Abstract

The first edition of ANSI CGATS.5 contained an appendix that discussed a method for improving the agreement between two instruments. The technology described was taken from a paper given at the first Oxford Conference, in which Alan Robertson gave a presentation in which he proposed using spectral differences to predict systematic errors in spectrophotometric measurements. Subsequent to that, Berns and Petersen published a paper in *Color Research & Application* that modified the Robertson method slightly by using numerical derivatives and derived a model for improving inter-instrument agreement. That model has found limited success in practical application. This technical report reviews the progress that has been made since the original release of CGATS.5-1993 and demonstrates that improvements in inter-instrument agreement are now possible at a level that equals or exceeds that of national standardizing laboratories. In addition, it will be shown that it may also be possible to bring together the readings taken with diffuse illumination and directional viewing with those collected with directional illumination and directional viewing measurement geometries.

Introduction

With the continually increasing number of offerings from instrument and software companies vying for the color measurement business of publication designers, ink makers and printers it is becoming more and more difficult to make color measurements that can be communicated from one part of the production chain to the next. What was needed is a method to adjust the readings from one manufacturer's instrument so that they are compatible with those from another manufacturer's instrument – so called inter-instrument agreement. This need resulted in the call for the development of a model of instrumental measurement correlation.

The model must be generally applicable across all modern instrument configurations and measurement geometries. It is also to be specifically applicable within groups of some of the higher volume instrument types. A fundamental model was required. The historically poor instrumental colorimetric reproducibility has forced the printers of colored images to rely on material standards for color-difference evaluation and as production targets throughout their production sites. Such material standards are difficult to produce, maintain and distribute. A system to reduce instrumental differences to acceptably small amounts would be very attractive to such printing companies.

Graphic technology — Improving the inter-instrument agreement of spectroradiometers

1 Scope

This Technical Report is intended to provide information indicating that current methodologies can result in improved inter-instrument agreement that can equal or exceed that of national standardizing laboratories. It will also show that it may be possible to correlate readings taken with diffuse illumination and directional viewing with those collected with directional illumination and directional viewing.

2 Theory

In selecting a model, the literature was searched and a set of two papers was identified that could contribute directly to this effort. The first was a paper presented by Alan Robertson[1] of the NRC in Canada. He presented a method for diagnosing instrumental errors at a joint meeting of the British UV Spectrophotometry Group and the American Council on Optical Radiation Measurements at Oxford, England in 1986. This method diagnosed instrument errors based on a method of spectral differences. He noted that the algebraic difference between two spectral reflectance factor measurements of the same material on two different instruments, one a reference instrument and one a test instrument, contained information about the photometric and wavelength scales. Figure 1 shows the divided differences between two real spectral curves differing in wavelength scale and bandwidth.

Robertson went on to show that the spectral differences between two instruments with identical photometric scales but differing only in the wavelength scale had the shape of the first derivative of the spectra. Similarly, he showed that two instruments differing only in spectral bandwidth exhibited a spectral difference with the shape of the second derivative of the spectra. He did not calculate derivatives from measurements on the NRC reference spectrophotometer but used only the divided differences between the test and reference instruments. Figure 2 shows the same spectral curve as above along with the first and second numerical derivatives. Note the similarity in the two sets of curves.

In a subsequent paper, Berns and Peterson[2] modified Robertson's method to use it as a model to improve inter-instrument agreement. Here they recommended calculating numerical derivatives on the test instrument in order to derive a model, which will indicate how to adjust the readings from the test instrument to obtain maximum agreement with the reference instrument. This resulted in a software model to "correct" instrumental errors. In building their model they assumed that all wavelengths are related and that errors are uniformly distributed and applied to each wavelength. Thus, they derived a least squares "average" zero error, photometric error, wavelength error and bandwidth error. They also tried to model very complicated photometric and wavelength scale errors. They could add terms for non-linearities in the scales and even for cyclic errors in the wavelength scale. All of the instruments in their paper were scanning instruments. Even so, they did not achieve the instrument correction results they had hoped. Still, the method showed a consistent improvement in the residual color differences between the test and reference instruments. A second effect, which they did not highlight in their paper, was that the agreement between the two test instruments also improved. This application was subsequently published by Reniff[3] and also by Berns and Reniff[4].

Berns and Peterson as well as Reniff found that with some instruments, calculating an "average" error did not significantly improve the results. Most spectroradiometers do not exhibit large, average errors in their scales.

Improving the inter-instrument agreement of spectroradiometers

Diode array instruments, in particular, have significant point-to-point errors due to the approximations used in the corrections of the aberrations in the spectrometer optics and the flattening of the spectral image. Similarly, the zero level and the photometric scale vary from wavelength to wavelength due to differences in the signal to noise ratio and to uncorrected differences in the general instrument function.

The only fundamental, intrinsic concept to the Berns and Peterson model is the use of numerical derivatives. Derivative spectroscopy has become a routine technique in modern analytical chemistry. All "intelligent" analytical spectrophotometers have derivative calculation available from the operator's keypad up to 5th or 7th order. Such derivatives are useful in locating and identifying peak positions and spectral line width differences, indicative of hidden lines. As shown in Figures 1 and 2, linear spectral scale changes are proportional to the first derivative and linear spectral bandwidth changes are proportional to the second derivative.

In quantitative spectroscopy, the accepted scale of diffuse reflectance is the perfect reflecting diffuser. This is a theoretical material with the unique property of reflecting 100% of the radiant power incident on it in such a way that the radiance (power per unit area per unit solid angle) has the same value for all directions in the hemisphere. No real material exhibits this property. All real materials (BaSO₄, Halon™, ceramic tile, opal glass, etc.) exhibit some degree of specular reflectance, some retro-reflectance and some variation in the radiance distribution as a function of angle. Differences in optics and electronics will affect the linearity and offset settings. These differences will normally be a function of wavelength, since all optical materials exhibit a wavelength dependence of their optical properties.

There may be nonlinear errors in the instrument response function as well. Those errors are generally small compared to the linear ones and are very difficult to quantify and model. Figure 3 illustrates the radiance to voltage signal transfer function for an ideal instrument and two typical real instruments at one wavelength. There may also be nonlinear and cyclic errors in the wavelength scale of the instrument. These are not modeled globally but are actually compensated by deriving a unique model at each 10-nm interval. This also compensates for instruments with variable bandwidths such as the Chroma Sensor CS-5®.

The general model given in equation 1 comes from the publication of Rich and Martin[6].

$$R_o(\lambda) = \beta_0 + \beta_1 \cdot R_t(\lambda) + \beta_2 \cdot \frac{dR_t(\lambda)}{d\lambda} + \beta_3 \cdot \frac{d^2 R_t(\lambda)}{d\lambda^2} \quad (1)$$

where the $R_o(\lambda)$ term is the reflectance factor measured on the reference instrument at each wavelength and the $R_t(\lambda)$ term is the reflectance factor measured on the test instrument at each wavelength. Following Robertson, the four coefficients can be related to the four types of linear systematic errors described above. The β_0 term represents the difference in zero offset or black level, the β_1 term represents the difference in linear scaling between the black and white points or scale calibration factor, the β_2 term represents the linear difference in the wavelength scale and the β_3 term represents the difference in the bandwidth. These parameters should be determined at each 10-nm wavelength position from 400 nm to 700 nm.

In reviewing this model it can be seen that there is no explicit term for a correction for differences in geometry. As the geometry of illumination and view is changed, the amount of light flux captured by the instrument relative to the flux from the perfect reflecting diffuser changes and with its change the resulting reflectance factor reported by the instrument changes. Thus, the effect of geometry is confounded with the effect of the reflectance factor scale adjustment applied to each instrument. For a nearly ideal distribution of flux this difference is negligibly small. For materials with textured or modulated surfaces this error becomes very large.

Instruments with integrating spheres have ports that are opened or backed by a photometric detector to eliminate or record the amount of surface reflectance relative to the body reflectance. The body reflectance will be very nearly

Lambertian but the surface reflectance will be a function of the roughness of that surface as well as the optics of the instrument. The surface reflectance will generally range between 3% and 5% depending primarily upon the chemistry of the ink or coating and not upon the color of the ink.

During the development of ANSI CGATS.5, several instrument vendors participated in an internal round-robin using SWOP[®] press sheets and a press target based on the Kodak Q60[®]. The participants were instructed to make the color measurements using a 45° : 0° instrument. One participant did not read or did not follow the directions and made the measurements using a di : 8° instrument. The instrument data were not used in the round-robin report. The measurements were taken separately and run through a Berns & Peterson correlation program. The results were astonishing. When measuring only the web-fed heat set ink on a single standard paper, the model converted the di : 8° measurements such that when compared with the measurements from a mixed group of 45° : 0° instruments, the differences in the measurements were so small that the color coordinates were indistinguishable.

3 Experiments

The four parameters described in the Theory section must be statistically estimated. Such estimates require a minimum of 12 to 20 material samples to provide a sufficient number degrees of freedom to evaluate the goodness of the fits. The surface character of the samples determines how optimum the fits will be. This is because the specular reflectance component is confounded with the photometric scale linearity parameter as described in the theory section. This is true for both diffuse and bi-directional instruments. Preliminary tests were performed using only BCRA tiles. The results indicated that when estimated using glossy samples, the model predicted the reflectance factor of the ceramic tiles to within the normal reference instrumental reproducibility. When the same model is applied to semi-gloss or matte specimens, the model predictions are often worse than the unmodeled measurements. In the paper by Rich, and Martin[6], an experiment was conducted to investigate the general model's effectiveness and to determine the limits of its applicability. The requirements for a good model were that the average ΔE^* should show a consistent improvement. It was also required that specimen sampling come from a Munsell spacing such that the specimens included 5 or 6 hues at 3 or 4 chroma levels and 4 or 5 lightness levels each. Sets of specimens were obtained to meet these requirements plus some additional materials that possessed special characteristics. The sample set consists of 72 colored samples each from the glossy edition of the Munsell Book of Color[®][7], the matte Color Curve[®][8] atlas, the semi-gloss NCS[®][9] atlas, and polyester swatches from the SCOTDIC[®][10] collection. There are eight hue pages with three value levels and three chroma levels for each of the four collections. Samples were taken from each set to be visually closest to the color of the glossy Munsell paper. In addition, gray scales were taken from each of the four sample collections consisting of 13 to 18 steps in value. Also included in the test were 12 plastic standards from FTS[11], 12 Ceramic Colour Standards (BCRA)[12] tiles, and 8 matte Spectralon[13] plaques. Figures 4-6 illustrate the sampling of color space possible with this set of 420 samples. The samples were mounted on white foam board to provide a stiff, uniform substrate for the sample positioning mechanism of the instruments. The board also allowed the operator to handle the samples without touching their surfaces. Each sample was labeled for identification and kept, when not in use, in a white plastic box with a tightly fitting lid. The instruments were calibrated following the procedures and frequency recommended in their respective user's guides. The calibration standards were glossy white ceramic tiles, with reflectance factor values assigned by the manufacturer, as appropriate for the instrument under test.

A follow-up test was performed using additional matte Color Curve[®][8] specimens and a series of flexible packaging inks printed on polyester film using a precision gravure proofer from Saueressig[18]. A set of 718 ink on film specimens was collected and readings taken on both a bi-directional (45° : 0°) and on a hemispherical diffuse (di : 8°) spectroradiometer as reported by Rich[19]. Figures 7 and 8 show the distribution of colors as measured on the hemispherical diffuse instrument.

The instruments tested in this study include CS-5 (6 instruments), CS-3 (6 instruments), 3890 (4 instruments), 3990 (4 instruments), Microflash 200D (2 instruments), Spectraflash 500 (2 instruments), Spectraflash 600 (1 instrument),

and an X-Rite 968 (1 instrument). The measurements required approximately one day per instrument per specular mode.

4 Results

Table 1 shows the average model parameters for the average CS-5 instrument tested against a Spectraflash 500. The statistical fitting parameters indicate an excellent fit. The numerical derivatives were calculated using both simple numerical methods and by differentiation of a cubic interpolation polynomial. The latter method tends to provide better fits on low and moderate chroma samples and the former on high chroma samples. The model is derived by weighted multiple linear regression of reflectance factor data in decimal format. The spectral reflectance factor values from the test instrument, the first derivative of those values and the second derivative of the values are regressed on to the reflectance factor values from the reference instrument at each wavelength of interest.

Figure 3 also illustrates the single largest problem with numerical derivatives: noise. The numerical derivative process amplifies the noise present in the signal. This effect varies from instrument to instrument and from software package to another. One would expect the noise level to vary across instruments but it was surprising to see the noise level vary across software packages. Further investigation revealed that the source lay in the manner in which the spectral values were stored. Some older packages truncated the spectral data to integers with no more than five significant digits in order to save on disk storage space. Other packages kept more or less of the original measurement data as floating point numbers. The more data kept, the lower the noise. Because of the noise problem, it was suggested that another method for generating the derivatives must be employed. The literature and commercial color software contain many applications of cubic spline interpolation algorithms to spectral data. They generally behave well but have been known to develop large erroneous predictions due to polynomial oscillations. These instances are most commonly observed at the transition points of high chroma colors, such as yellows, oranges and red. The required derivatives could be obtained by generating a cubic polynomial (spline) interpolating function, then analytically differentiating the polynomial as long as there were no oscillation events present. This was tried with excellent results. Models fitted with spline-generated derivatives always performed better than those fitted with numerical derivatives. A study of the spectral reflectance curves of some typical and difficult materials revealed that, in general, approximating the second derivative with linear functions (the assumption of a cubic polynomial) was a good first approximation and provided significant noise reductions. A paper in the literature by James Hyman[14] revealed an algorithm for switching between a natural cubic spline with linear second derivatives and a spline with quadratic second derivatives. Such an algorithm further reduces the ringing and oscillation often observed in spline routines. This type of spline is one of a class of spline interpolating functions known as “shape preserving” splines.

It can be seen from the data in Table 2 that the standard errors of the estimate are quite small. As has been shown by Marcus and Billmeyer[15] and by Billmeyer and Alessi[16] the statistical distribution of the measurement data is very narrow, more so than the Normal distribution would predict. Test statistics such as confidence intervals, F-tests, Student's t-tests, etc. based on the Central Limit Theorem actually over-estimate the size of the error range. The Normal statistical distribution indicates that about 67% of the measurements should occur within one standard deviation of the mean. Billmeyer and Alessi confirmed that for spectroradiometers, about 95% of the measurements will occur within one standard deviation of the mean. Table 2 shows an analysis of variance table for the example given above for a small number of wavelengths.

Another verification of the validity of the model is shown in Figure 9. Here we have plotted the coefficient of the second derivative term versus wavelength for the CS-5 to Spectraflash 500 data. The second derivative term is intended to model the difference in bandwidth between the reference and test instruments. The CS-5, which utilizes an OCLI CVF[17] circular variable interference filter, has a variable bandwidth which increases from around 10 nm at 400 nm to around 20 nm at 700 nm. As can be seen in the figure, the coefficients become increasingly more negative (indicating a reduction of the CS-5 bandwidth) as a function of wavelength. This linear trend in average

bandwidth is exactly what one would expect. The variations around the line are due to noise (this data was generated from numerical derivatives) and to the intrinsic behavior of the CWF, which does not vary uniformly, but oscillates along a line just as this data appears to do.

With 420 samples in the test set, a frequency analysis of the change in color difference between instruments before and after modeling can yield information about the efficiency of the model. Figures 10 and 11 show such frequency histograms, before and after correlation, for the CS-3 versus the Spectraflash 500. These two instruments are about as different as two integrating sphere instruments can be with different sized spheres and vastly different collection optics. As shown in the figure, the area of the distribution always decreases. The tails of the distribution are sometimes reduced and sometimes left unchanged. We have never observed a tail that has increased after modeling. One can analyze the histograms more rigorously and generate statistical distribution parameters. Table 3 shows the data for such an analysis on the two histograms above.

The numbers in the table describe what one sees in the histograms. All of the parameters get smaller except those related to the shape of the distribution. The skewness increases indicating a stronger tendency for the histogram to lean toward the zero point. The standard bell-shaped distribution has a skewness of 0. The kurtosis increases dramatically, indicating a sharper, more peaked distribution. The standard bell-shaped distribution has a kurtosis of 3. All of this leads one to conclude that the model is exhibiting a positive effect on the measurement of the color of the test samples. Table 4 shows a summary of results for the instruments included in this experiment using the Spectraflash 500 instruments as the reference instrument. The column labeled “ ΔE Reversals” indicates the percentage of the 420 samples in which the color difference between the reference and test instruments increased rather than decreased after applying the correction indicated by the model. In all cases where a reversal occurred it was a small increase or a very small increase in color difference.

Figures 12 and 13 show the histogram of results when the method is applied to the measurements of ink on film using a $0^\circ : 45^\circ$ and a $di : 8^\circ$ spectroradiometer as reported by Rich[19]. This should be the worst case situation for color measurements of ink as the flexible packaging inks are very glossy and the surface of the film is very smooth. The distribution of CIELAB color differences bears out this assumption, exhibiting an average difference of about 10 units with a maximum difference of 25 units. After modeling, however, the numbers drop to something much more reasonable as the average is now 0.5 units and the maximum about 2 units. These numbers are in line with the values reported by Verrill[20,21] in his round-robin inter-comparison studies in Europe comparing commercial and reference instruments of similar geometry.

5 Discussion

When evaluating the appropriateness of a statistical model, one is always concerned with the robustness of the model. In statistical modeling, a mathematical model is considered robust if it is tolerant of small violations of the underlying assumptions that were made in establishing the model.

In this model, we have assumed that the photometric scale is independent of the wavelength scale and of the surface character of the sample. We know these assumptions can be violated. There is a very definite confounding of the surface character of the sample with the photometric scale. If the model is built with only glossy tiles, such as the BCRA ceramic tiles, then matte materials, like textiles, are poorly modeled. If the model is built with textiles or with matte materials like Spectralon™ then the model performs very poorly with glossy samples like automotive finishes. By using a mixture of matte, semi-gloss and glossy materials one obtains a model that minimizes the surface effects on predictions by the model. Indeed, even if we were to randomly reduce the number of samples by ten or twenty percent, the resulting model would perform equivalently to what has been shown above.

In a similar manner, we have included a range of chromatic and achromatic samples in such a ratio as to optimize the modeling of the photometric scale without masking the wavelength scale. Figures 4, 5, 6 and 7 show that the neutral

Improving the inter-instrument agreement of spectrophotometers

axis is much more densely populated than any hue branch. Again, if we randomly reduce the number of chromatic and achromatic samples, the model remains constant. If we reduce the relative ratio of achromatic to chromatic samples, we will experience a shift in the model. Thus, even in the development of pairwise models, in the field, it is important to obtain a substantially larger number of achromatic samples than chromatic samples.

These two constraints on the robustness of the model contribute to the impracticality of developing multiple models for a given instrument or set of instruments, especially if the specular component included (SCI) to specular component excluded (SCE) change is manually executed. The differences between the SCI and SCE models are quite substantial. It was expected that if the surface contribution was eliminated then the results from instrument to instrument should be improved. It is very disappointing to observe that this did not occur. Most of the poorer results in SCE mode can be traced to two causes, calibration of SCE mode and integrating sphere illumination nonuniformities.

Some of the instruments that have been modeled have specular excluded reflectance factor values transferred to the instrument from a master standard tile. Other instruments calibrate the instrument in SCI mode and then apply the same scaling factors when the specular port is opened. The SCI and SCE readings will only be valid for perfectly matte materials. Only Spectralon™ comes close to fitting such a description. NIST has not recognized the measurement geometry SCE in the past due to a lack of a clear definition of this mode of measurement. The NPL and PTB recognize $d_i : 8^\circ$ and $d_e : 0^\circ$ but not a geometry of 8° excluded ($d_e : 8^\circ$). Some materials will have identical values for $d_e : 8^\circ$ and $d_e : 0^\circ$ but most do not. The NRC and HCL offer $d_e : 8^\circ$ and $d_e : 6^\circ$ calibrations respectively. The two geometries generally agree, but some samples can be found that exhibit geometry dependent differences that cannot be correctly modeled. This was and is an active area of research in diffuse reflectance metrology.

Another source of error in the integration sphere SCI to SCE correlation is the uniformity of the light within the sphere. The uniformity of the light flux is dependent on such uncontrolled variables as the reflectance of the sphere wall in the area around the specular port, the area of the specular port, the area of the sphere opposite to the lamp, and the angular sizes of the specular port and lamp port relative to the areas of illumination and view. If the flux in sphere is significantly non-uniform then the simple calibration model that separates specular and diffuse components is incorrect and the resulting scaling factors generate sample dependent errors.

Because of all of the above, it was recommended in the Rich and Martin[6] paper that one not use $d_e : 8^\circ$ or SCE models to relate instruments of significantly different geometries. This conclusion also implies that trying to correlate Large Area of View (LAV) with Small Area of View (SAV) or Very Small Area of View (VSAV) geometries should be done with great care and the measurements performed in parallel (two instruments with the same areas of view) and not sequentially (two instruments with different areas of view). Further studies, such as the comparison of the $0^\circ : 45^\circ$ to $d_i : 8^\circ$ instruments have shown most of these concerns to be valid. The only exception is when the surfaces of the specimens can be carefully controlled.

Then it was observed in the Rich reference that the inter-action of the specimen with the geometry can be modeled within the photometric scale modeling. It is speculated, though not evaluated, that it may be possible to model integrating sphere readings of a series of specimens of different surface character by adding a term to the model that is dependent on the difference between readings taken with the specular component included and excluded. Since this difference will be proportional to the glossiness of the specimen appearance it should follow the differences in specimen surface character. The downside is that this equation would not be reversible.

As one can imagine, the precision of this model is critically dependent upon the stability of the numerical derivatives. The stability of the derivatives is dependent on both the storage and data handling of measured reflectance factors. Some PCDOS based programs transform all floating-point reflectance factors to 16 bit integers for compact storage. When the reflectance factors are recalled for use, they are converted back to floating point.

By shifting the decimal point to the right, once for each significant figure, radiance factors up to 320% can be stored as two-byte integers. This process generates about two bits of noise or about $\pm 0.02\%$. While this is a small number in absolute magnitude, it is easily seen when compared to the data stored as four-byte floats.

One very fortunate set of results has been observed when applying the “average” model to an individual instrument. Not only do the individual instrument's readings move closer to those of the reference instrument, but also the relative differences between two or more instruments of the same family also seem to be slightly reduced. This is an unforeseen and as yet unexplained phenomenon.

6 Conclusions

The model has been developed in a clear, logical manner, using only spectral reflectance data. The model parameters are proportional to real, optical effects such as the optical zero, linear photometric scale, wavelength scale and bandwidth of the instrument. The model can be derived for either the average production instrument or for unique groups of customer instruments, using user-supplied specimens; thus providing for very close agreement within a user's factory or between factories.

The model cannot separate or compensate for several types of residual or confounded differences between instruments. The sources of these differences include specular port effects, non-uniformities in integrating sphere illumination, non-linearities in the photometric scale, and the translucent blurring effect found in translucent materials. The model described here is definitely and specifically not applicable to transmittance measurements, neither regular or diffuse transmittance. It may be possible however, to provide an LAV to SAV model for a specific type of material and specific pair or group of instruments. For example, one could create a model on a single instrument to improve the large area of view (LAV) to the small area of view (SAV) or very small area of view (VSAV) agreement on an instrument that was used both to create a colorant file (LAV) and to correct production samples (SAV) of a given ABS plastic. It is possible to relate specular component excluded (SCE) to specular component included (SCI) or bidirectional ($0^\circ : 45^\circ$) to SCI only for carefully controlled specimen surfaces.

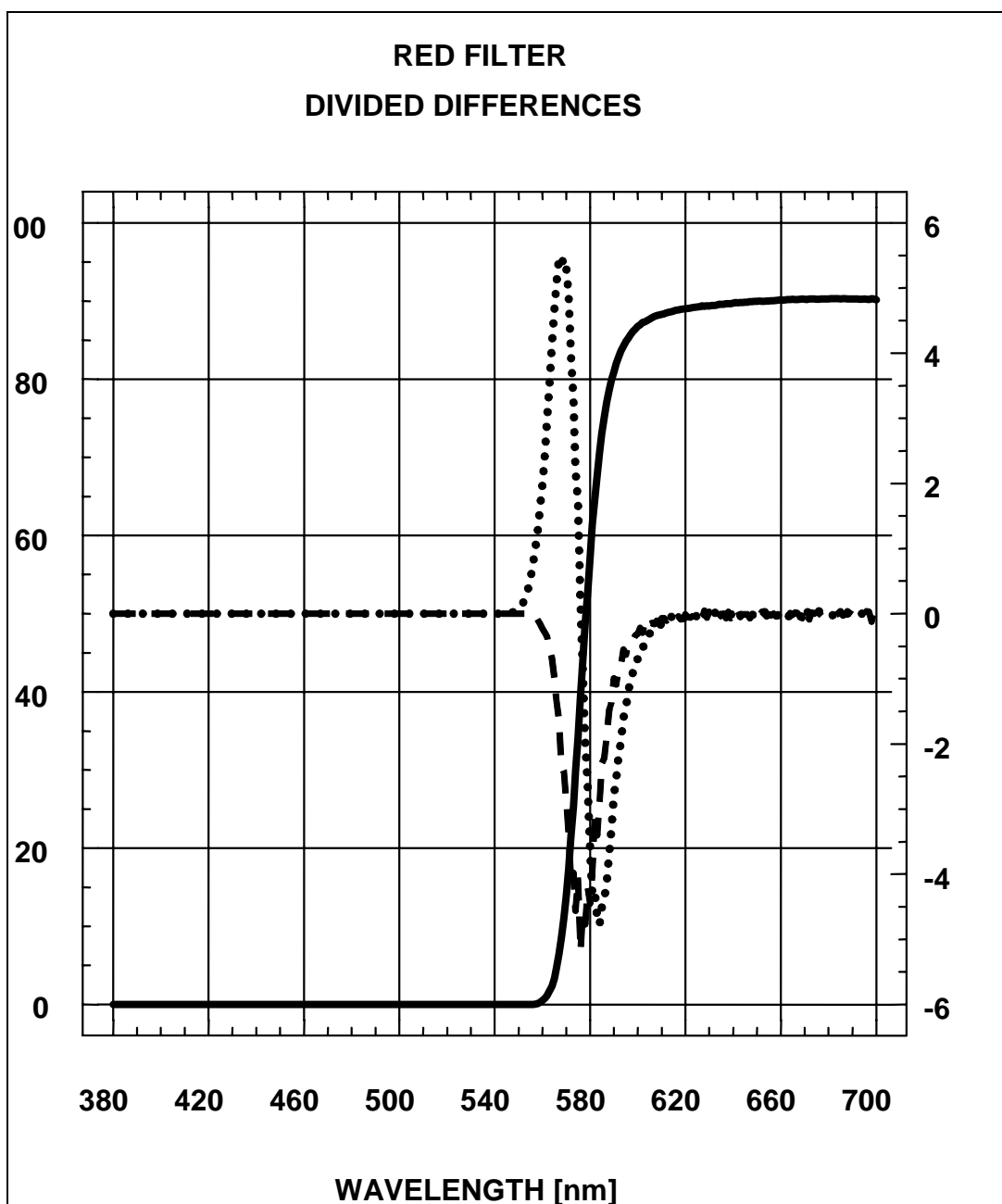


Figure 1 — Spectral transmittance of a red specimen (solid curve); divided differences between the original curve and a curve shifted by 1nm (dashed curve) and a curve broadened by a 10nm wide triangle (dotted curve)

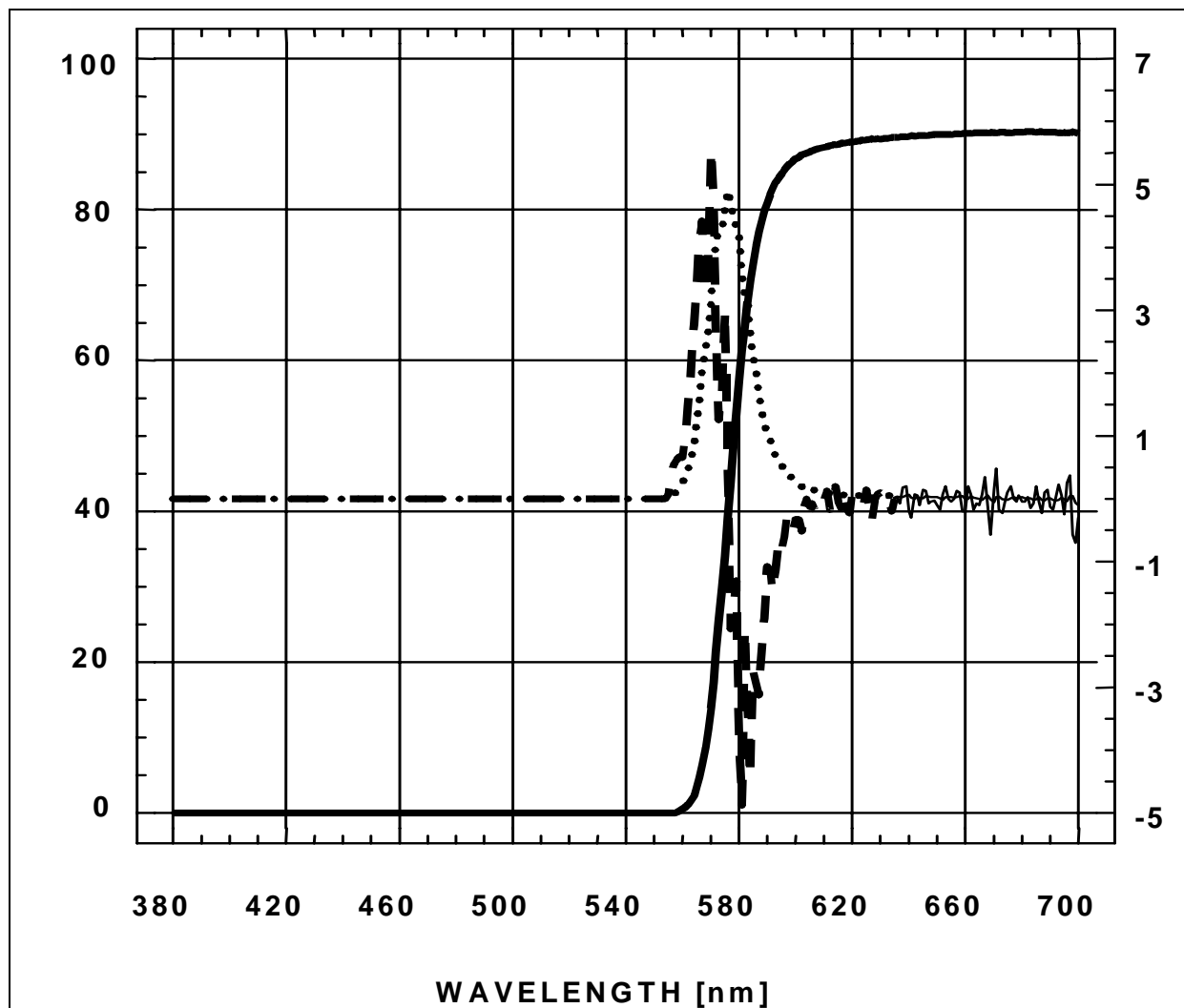


Figure 2 — Same red specimen as in Figure 1 (solid line) showing also the first derivative of the red specimen curve (dotted line) and the second derivative of the red specimen curve (dashed line)

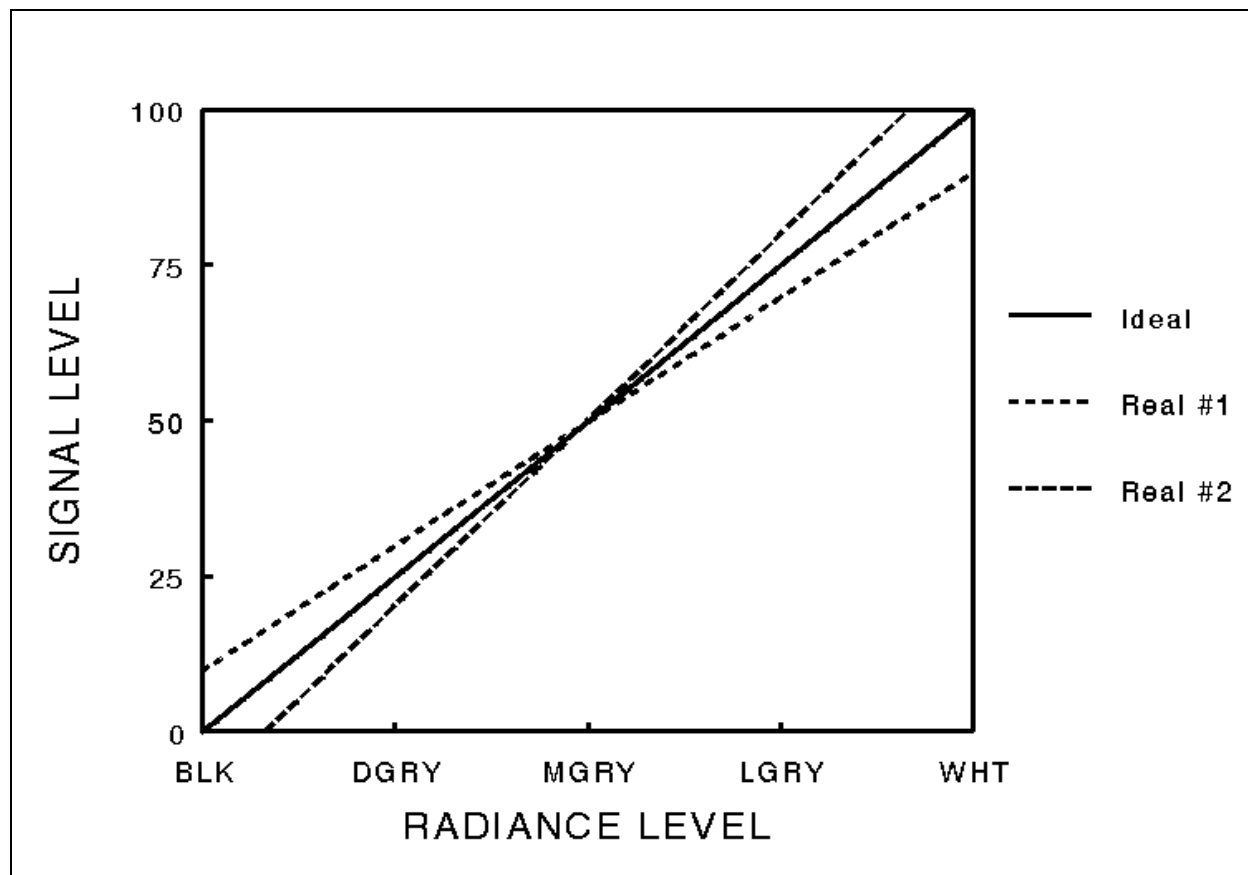


Figure 3 — Schematic illustration of an ideal linear reflectometer (solid line) versus two real instrument curves that exhibit a difference in slope and offset (white point and zero level)

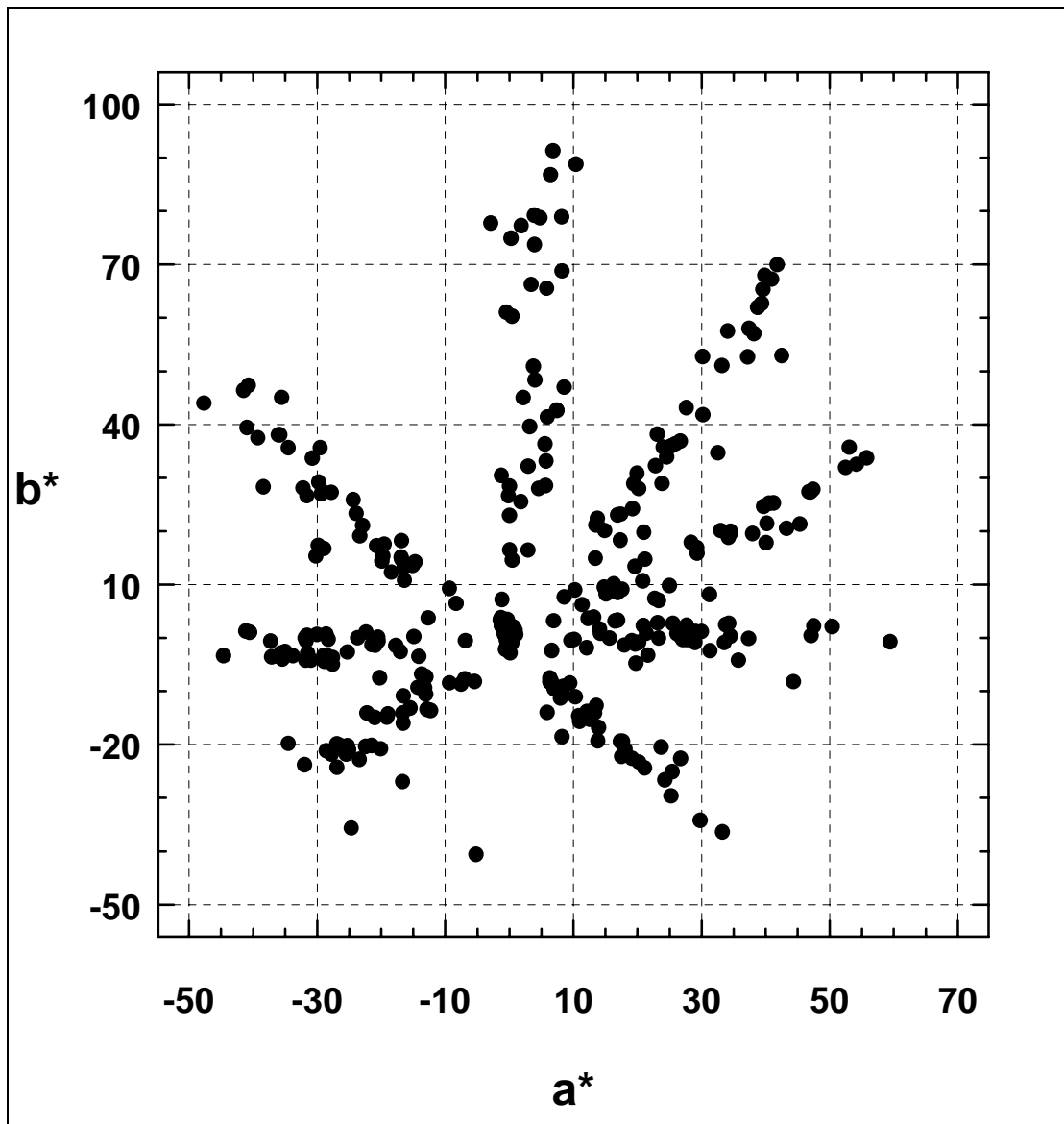


Figure 4 — Test sample set of 420 specimens displayed on CIELAB a^* , b^* plane

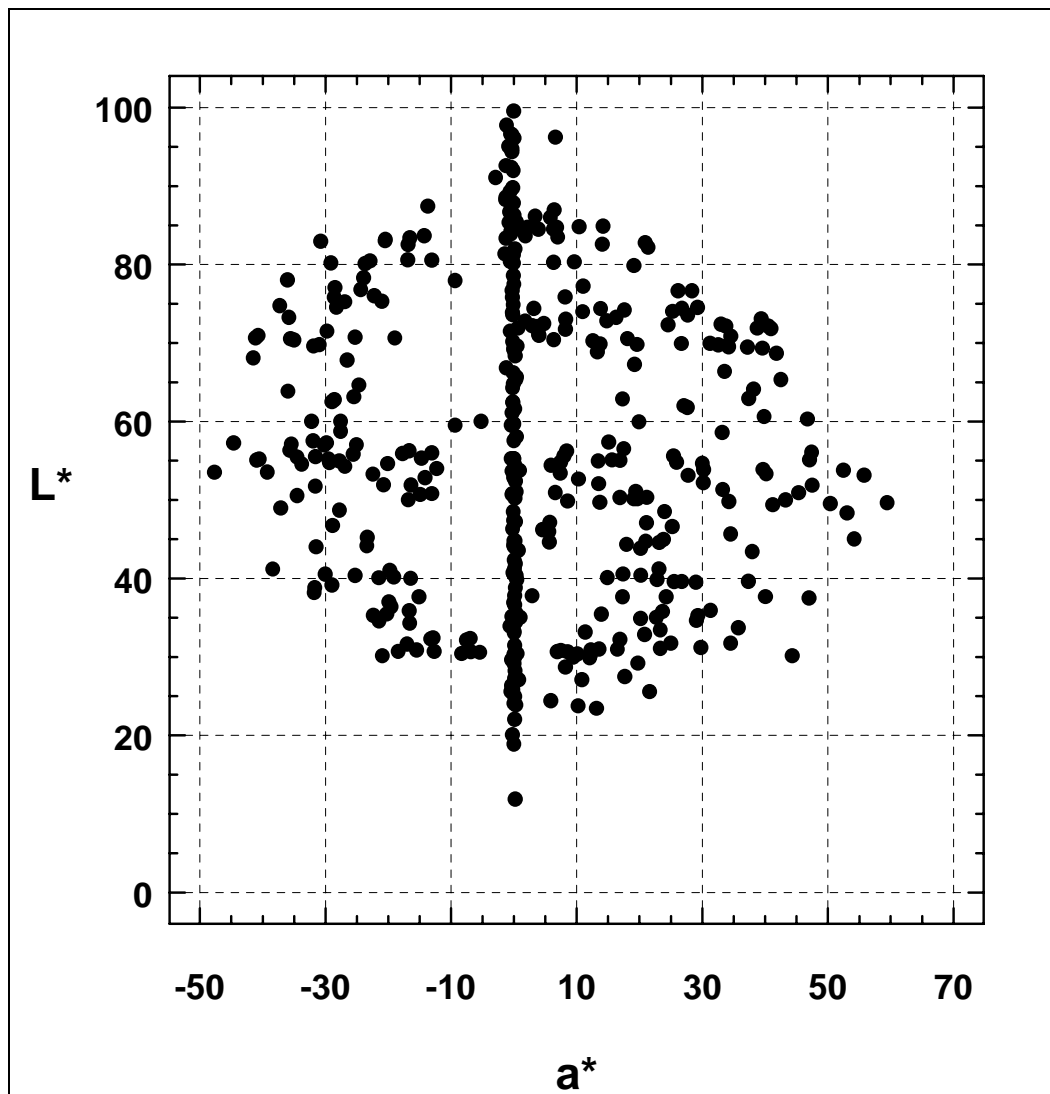


Figure 5 — Test sample set of 420 specimens displayed on CIELAB a^* , L^* plane

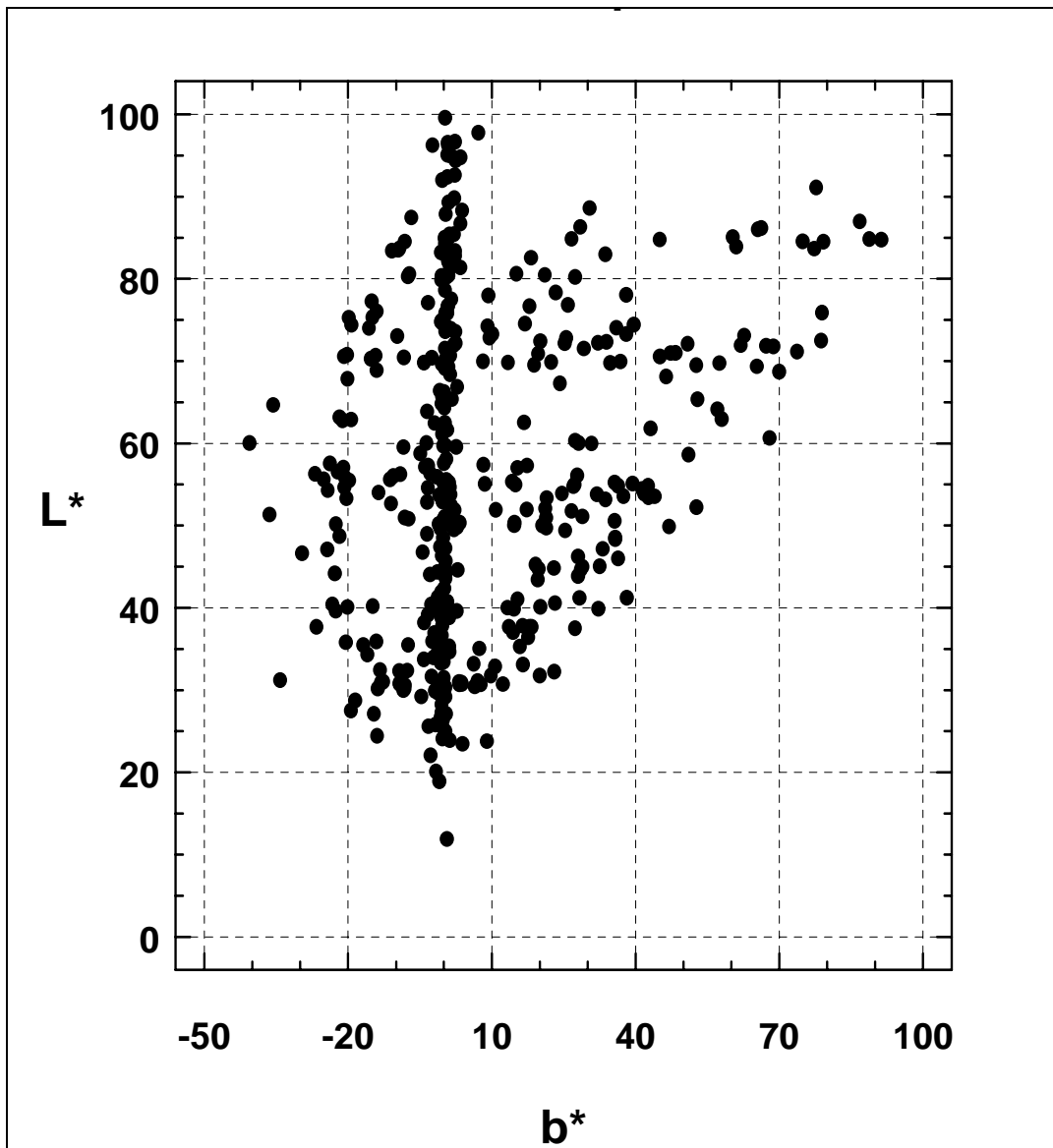


Figure 6 — Test sample set of 420 specimens displayed on CIELAB b^* , L^* plane

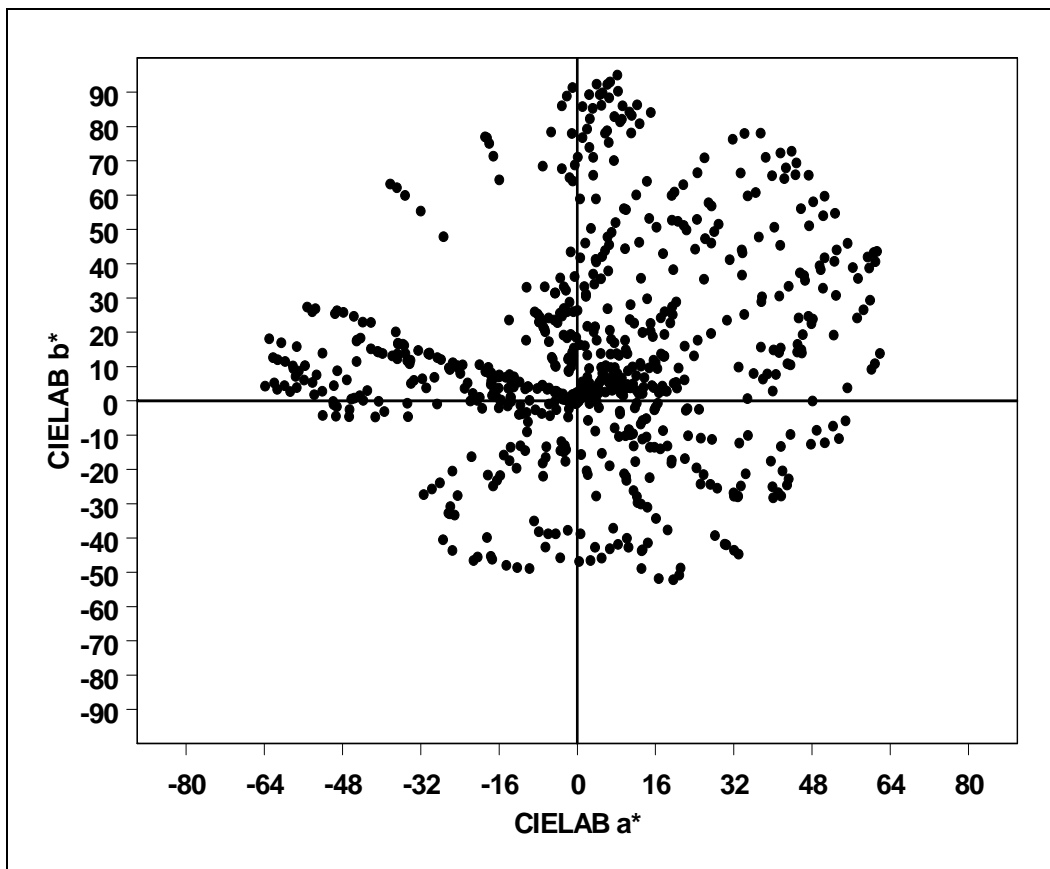


Figure 7 — Test samples of 718 ink-on-film specimens displayed on CIE a*, b* plane

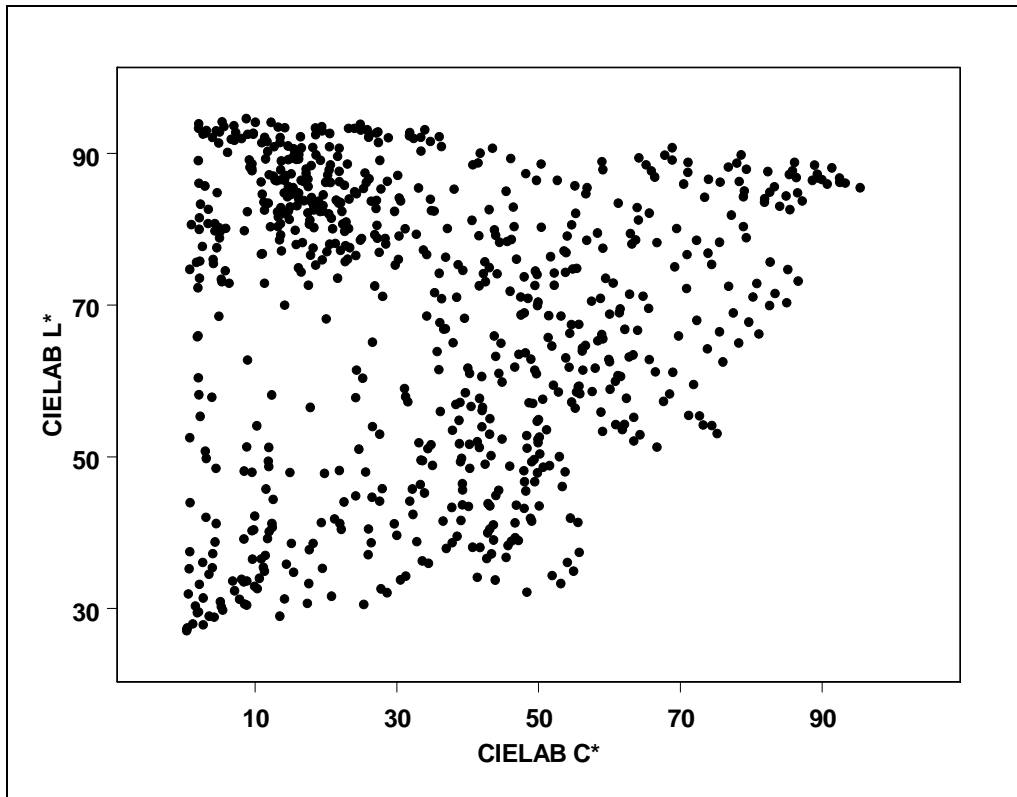


Figure 8 — Tests samples of 718 ink on film specimens displayed on CIE C*, L* plane

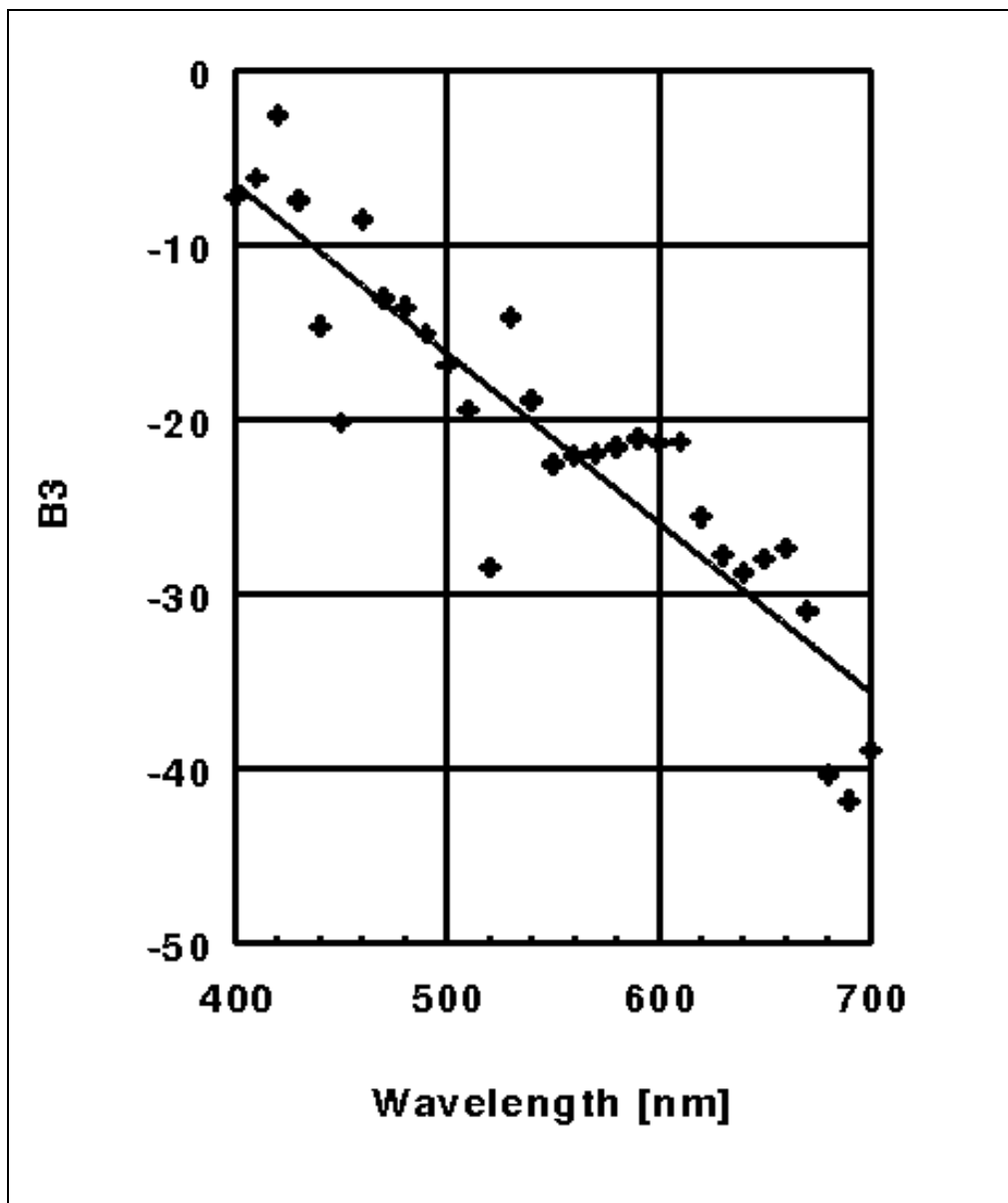


Figure 9 — Plot of model parameter B_3 versus wavelength showing the linear change in bandwidth for the ChromaSensor CS-5 correlated to a Spectraflash 500

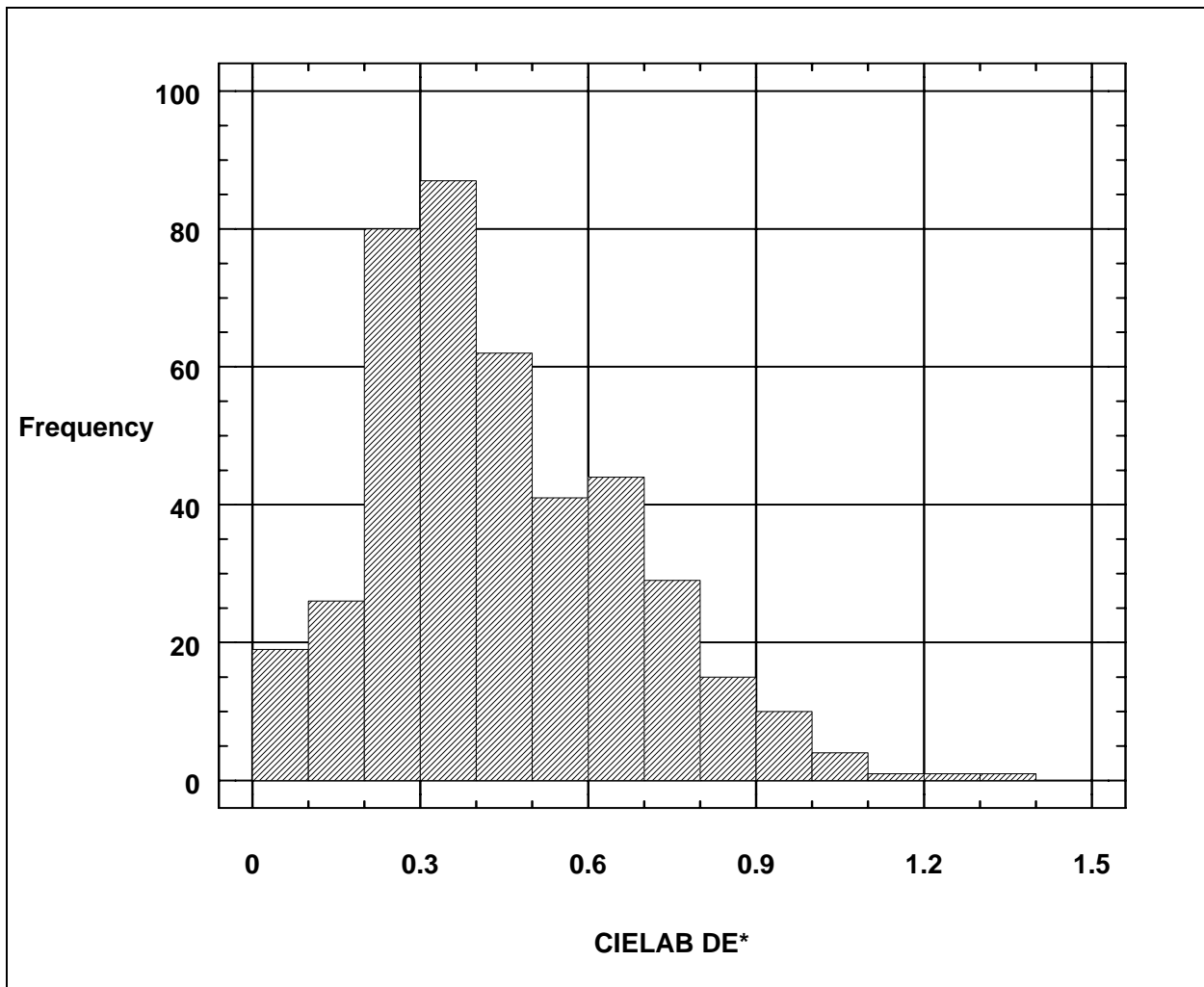


Figure 10 — Histogram of the 420 color differences between the ChromaSensor CS-3 and the Spectraflash 500 before application of the correlation model

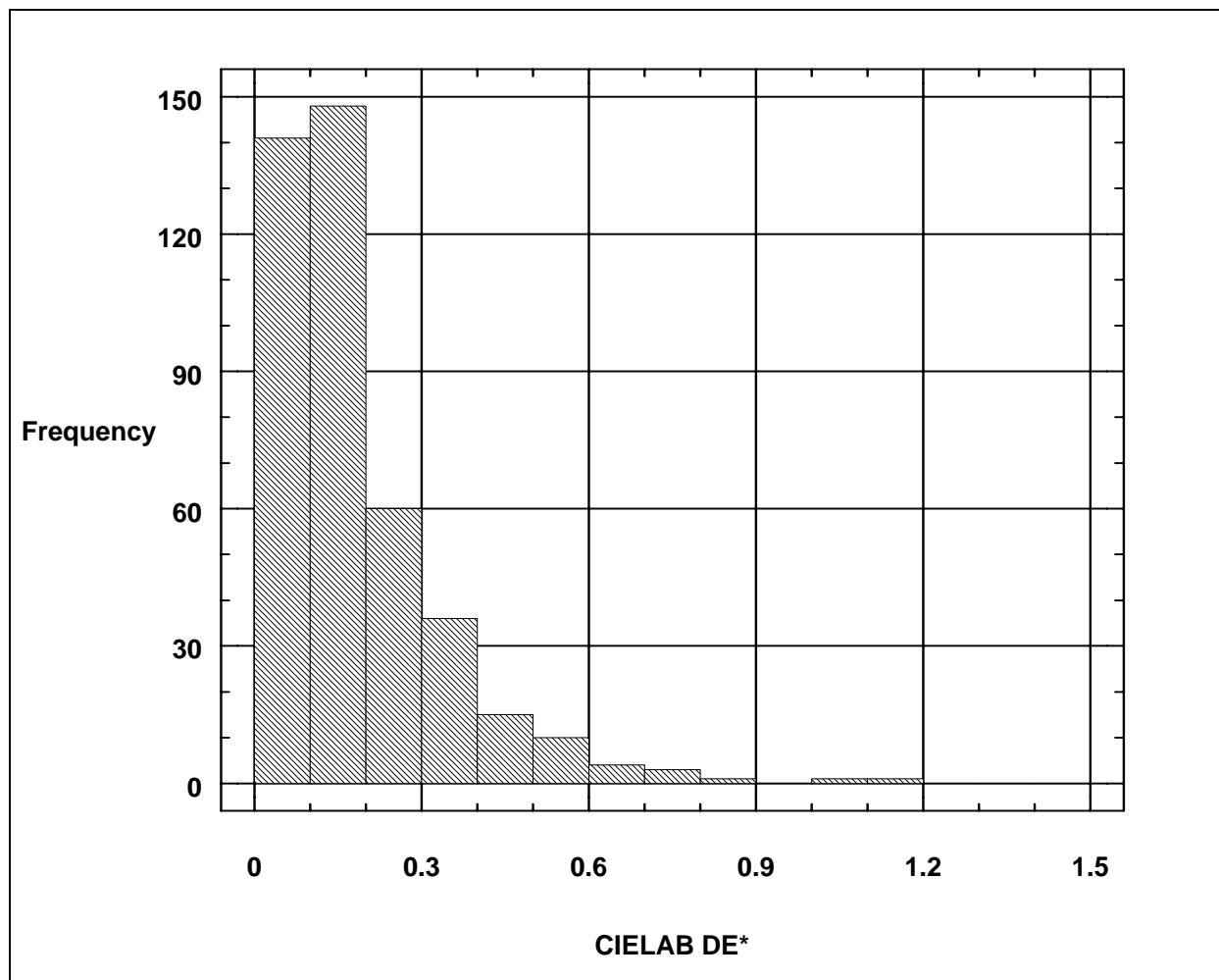


Figure 11 — Histogram of the 420 color differences between the ChromaSensor CS-3 and the Spectraflash 500 after application of the correlation model

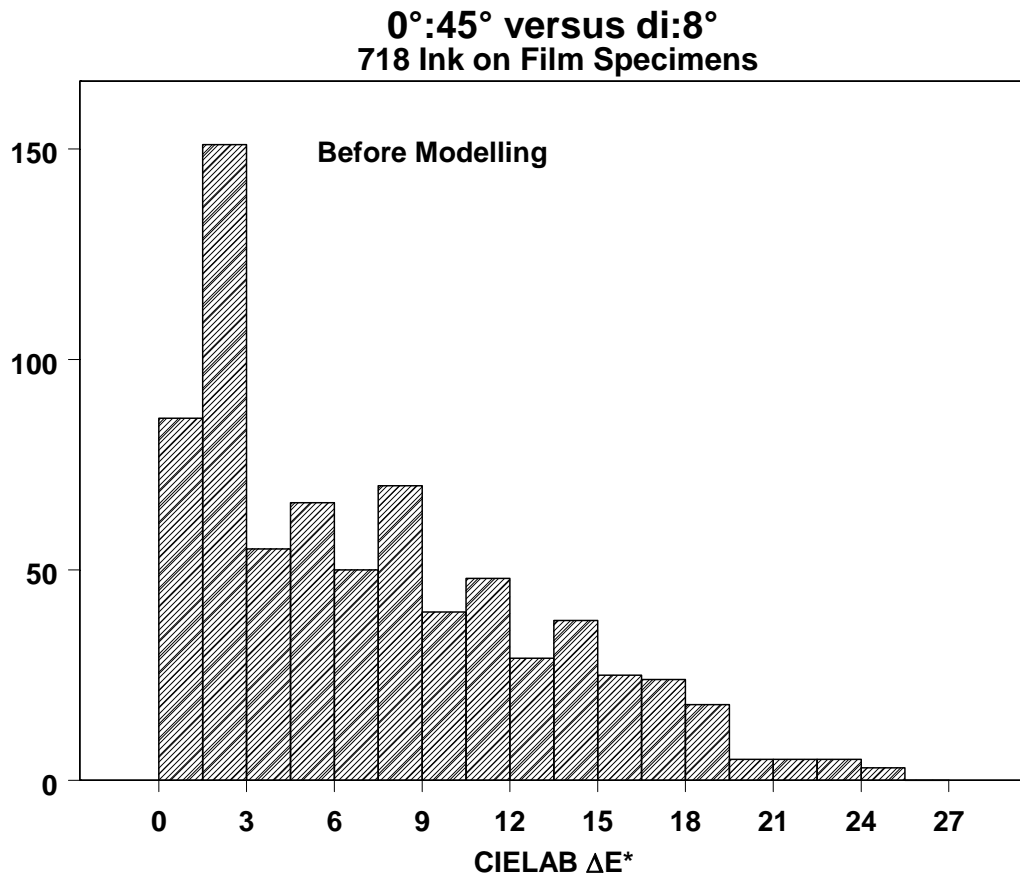


Figure 12 — Histogram of the 718 color differences between the 0°:45° and the di:8° before application of the correlation model. Average color difference is about 10 CIELAB ΔE^* units with a maximum of 25 CIELAB ΔE^* units

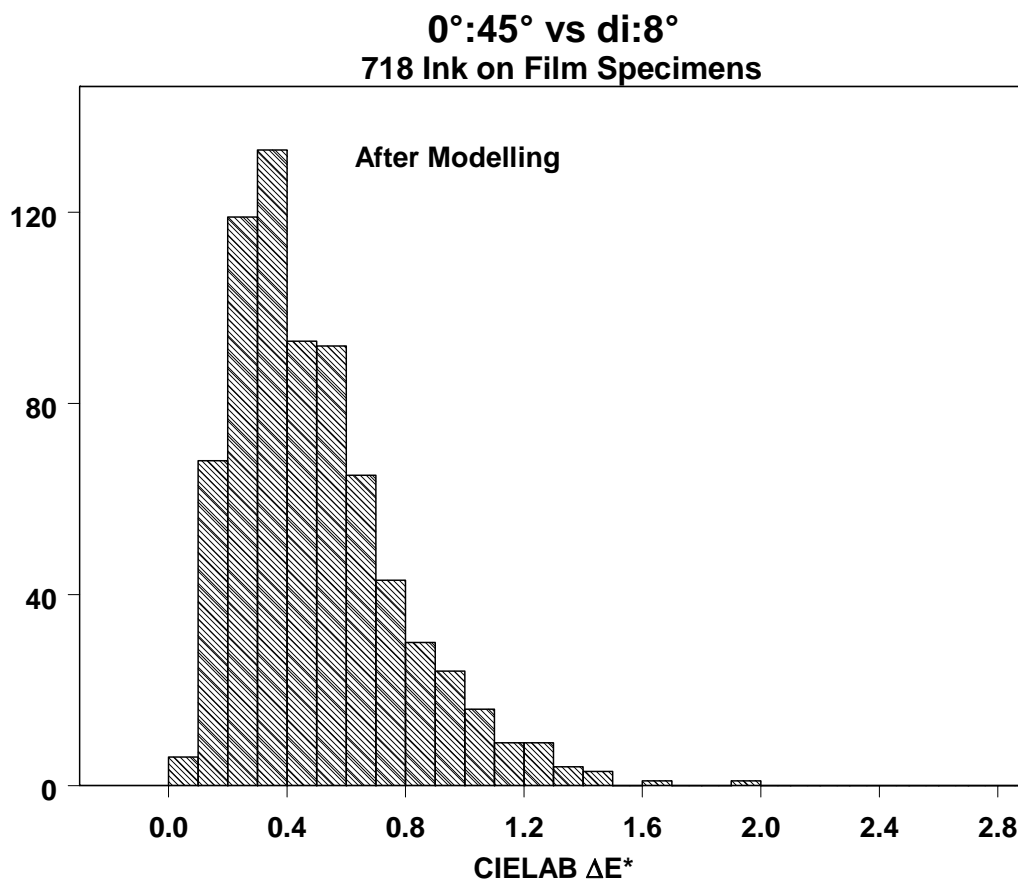


Figure 13 — Histogram of the 718 color differences between the 0°:45° and the di:8° after application of the correlation model. Average difference is 0.5 CIELAB ΔE^* with a maximum of 2.0 CIELAB ΔE^*

Table 1 — Regressing coefficients for CS-5 regressed onto Spectraflash 500, specular component included

Wavelength	Zero Level Scale	Photometric Scale	Wavelength Scale	Bandwidth Scale	Standard Error	Correlation Coefficient
400.0	-0.00180	1.01672	-0.63544	- 7.26143	0.00310	0.99970
410.0	-0.00162	1.01803	-0.02665	- 6.23576	0.00215	0.99989
420.0	-0.00116	1.01487	-0.85308	- 2.61149	0.00204	0.99991
430.0	-0.00056	1.01330	-1.27286	- 7.47763	0.00190	0.99993
440.0	-0.00056	1.01399	-1.08792	-14.69708	0.00216	0.99991
450.0	-0.00015	1.01037	0.11989	-20.20408	0.00194	0.99992
460.0	-0.00056	1.01152	-0.31178	- 8.55240	0.00191	0.99993
470.0	-0.00048	1.01204	-0.42914	-13.08351	0.00206	0.99992
480.0	-0.00046	1.01198	-0.18609	-13.64376	0.00199	0.99993
490.0	-0.00023	1.00991	-0.50509	-15.14576	0.00214	0.99992
500.0	-0.00054	1.01112	-0.73512	-16.90070	0.00189	0.99993
510.0	-0.00064	1.01149	-0.59646	-19.54718	0.00256	0.99988
520.0	-0.00016	1.00830	-0.58122	-28.45829	0.00225	0.99991
530.0	-0.00074	1.01255	-0.03944	-14.21137	0.00238	0.99990
540.0	-0.00046	1.01079	-0.30738	-18.85302	0.00176	0.99994
550.0	-0.00061	1.01102	-0.65530	-22.60042	0.00171	0.99995
560.0	-0.00049	1.01012	-0.39921	-22.12342	0.00181	0.99994
570.0	-0.00048	1.00950	0.07683	-22.03974	0.00167	0.99995
580.0	-0.00063	1.01008	0.12867	-21.56105	0.00180	0.99995
590.0	-0.00072	1.00900	0.40717	-21.13484	0.00185	0.99995
600.0	-0.00088	1.00932	0.06777	-22.44946	0.00198	0.99994
610.0	-0.00077	1.00916	-0.02181	-21.33063	0.00226	0.99993
620.0	-0.00064	1.00904	0.06393	-25.57795	0.00234	0.99993
630.0	-0.00062	1.00831	0.12300	-27.80753	0.00259	0.99991
640.0	-0.00063	1.00612	-0.05791	-28.79043	0.00322	0.99987
650.0	-0.00069	1.00727	-0.01629	-28.00580	0.00391	0.99981
660.0	-0.00063	1.00622	-0.39128	-27.40816	0.00420	0.99978
670.0	-0.00048	1.00523	-0.45879	-30.98296	0.00434	0.99977
680.0	-0.00026	1.00556	-0.10752	-40.40054	0.00429	0.99978
690.0	0.00008	1.00428	-1.01810	-41.85828	0.00444	0.99977
700.0	-0.00007	1.00493	-1.46124	-39.02821	0.00460	0.99976

Table 2 — Model fitting results

Model fitting results for WL=400					
Independent variable	Coefficient	Std. Error	t-value	Significance Level	
CONSTANT	0.002005	0.000286	-7.0133	0.0000	
WL400.R	1.018117	0.000983	1035.3330	0.0000	
WL400.R'	0.612584	0.095164	-6.4372	0.0000	
WL400.R''	6.485138	2.039632	-3.1796	0.0016	
Analysis of Variance for the Full Regression					
Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	18.6844	3	6.22812	414101	0.0000
Error	0.00625668	416	0.0000	150401	
Total (Corr.)	18.6906	419			
$R^2 = 0.999665$			Standard error of estimate. = 0.00387815		
R^2 (Adj. for d.f.) = 0.999663			Durbin-Watson statistic = 1.42163		
Further ANOVA for Variables in the Order Fitted					
Source	Sum of Squares	DF	Mean Sq.	F-Ratio	P-value
WL400.R	18.6822286	1	18.682229	1242162.49	0.0000
WL400.R'	0.0019715	1	0.001972	131.09	0.0000
WL400.R''	0.0001520	1	0.000152	0.11	0.0016
Model	18.6843521	3			
Model fitting results for WL=500 nm					
Independent variable	Coefficient	Std. Error	t-value	sig.level	
CONSTANT	-0.000466	0.000167	-2.7883	0.0055	
WL500.R	1.010823	0.000437	2313.4257	0.0000	
WL500.R'	-0.726686	0.035586	-20.4208	0.0000	
WL500.R''	-17.59105	0.721258	-24.3894	0.0000	
Analysis of Variance for the Full Regression					
Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	150401	3	10.4680	2033654.	0.0000
Error	0.00214131	416	0.00000514737		
Total (Corr.)	31.4060	419			
$R^2 = 0.999932$			Standard error of estimate = 0.00226878		
R^2 (Adj. for d.f.) = 0.999931			Durbin-Watson statistic = 1.23804		
Further ANOVA for Variables in the Order Fitted					
Source	Sum of Squares	DF	Mean Sq.	F-Ratio	P-value
WL500.R	31.3978488	1	31.397849	6099786.40	0.0000
WL500.R'	0.0029902	1	0.002990	580.91	0.0000
WL500.R''	0.0030619	1	0.003062	594.84	0.0000
Model	31.4039008	3			

Table 2 — Model fitting results (continued)

Model fitting results for WL=600					
Independent variable	Coefficient	Std. Error	t-value	sig.level	
CONSTANT	-0.000943	0.000179	-5.2635	0.0000	
WL600.R	1.009601	0.000414	2440.0609	0.0000	
WL600.R'	0.052642	0.032955	1.5974	0.1109	
WL600.R"	-21.974231	0.860595	-25.5338	0.0000	
Analysis of Variance for the Full Regression					
Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	40.3294	3	13.4431	2357251.	0.0000
Error	0.00237240	416	0.00000570288		
Total (Corr.)	40.3317	419			
R ² = 0.999941			Standard error of estimate = 0.00238807		
R ² (Adj. for d.f.) = 0.999941			Durbin-Watson statistic = 1.05686		
Further ANOVA for Variables in the Order Fitted					
Source	Sum of Squares	DF	Mean Sq.	F-Ratio	P-value
WL600.R	40.3251258	1	40.325126	7071009.29	0.0000
WL600.R'	0.0005173	1	0.000517	90.72	0.0000
WL600.R"	0.0037181	1	0.003718	651.97	0.0000
Model	40.3293613	3			
Model fitting results for WL=700					
Independent variable	Coefficient	std. Error	t-value	sig.level	
CONSTANT	0.000063	0.000462	0.1363	0.8917	
WL700.R	1.00483	0.000862	1166.2057	0.0000	
WL700.R'	-1.540242	0.129671	-11.8780	0.0000	
WL700.R"	-37.289572	4.078553	-9.1428	0.0000	
Analysis of Variance for the Full Regression					
Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	49.7077	3	16.5692	488126.	0.0000
Error	0.0141209	416	0.0000339446		
Total (Corr.)	49.7218	419			
R ² = 0.999716			Standard error of estimate = 0.0058262		
R ² (Adj. for d.f.) = 0.999714			Durbin-Watson statistic = 1.26816		
Further ANOVA for Variables in the Order Fitted					
Source	Sum of Squares	DF	Mean Sq.	F-Ratio	P-value
WL700.R	49.6974597	1	49.697460	1464077.00	0.0000
WL700.R'	0.0073727	1	0.007373	217.20	0.0000
WL700.R"	0.0028375	1	0.002837	83.59	0.0000
Model	49.7076699	3			

Table 3 — Statistical analysis of the histograms shown in Figures 9 and 10

Variable	DE*	
	Before modeling	After modeling
Sample size	420.000	420.000
Average	0.447	0.191
Median	0.400	0.150
Mode	0.320	0.170
Variance	0.053	0.023
Standard deviation	0.229	0.152
Minimum	0.030	0.020
Maximum	1.320	1.190
Range	1.290	1.170
Lower quartile	0.290	0.090
Upper quartile	0.605	0.240
Interquartile range	0.315	0.150
Skewness	0.719	2.259
Standardized skewness	6.015	18.897
Kurtosis	0.355	7.809
Standardized kurtosis	1.485	32.666
Coefficient of variation	51.259	79.647

Table 4 — Results of the average inter-instrument model in terms of CIELAB ΔE test instrument vs. Spectraflash-500, with the specular component included for the 420 total samples

Instrument	Avg. ΔE (Uncorrected)	Avg. ΔE (corrected)	Percentage of ΔE reversals
CS-5	0.57	0.21	9%
CS-3	0.45	0.19	9%
3890	0.57	0.29	14%
3990	0.38	0.18	9%
Microflash	0.47	0.32	15%
Dataflash 100 to CS-3	0.49	0.29	15%

REFERENCES

1. A. R. Robertson, "Diagnostic Performance Evaluation of Spectrophotometers", presented at *Advances in Standards and Methodology in Spectrophotometry*, Oxford, England (1986)
2. R. S. Berns and K. H. Petersen, "Empirical Modeling of Systematic Spectrophotometric Errors", *Color Research and Application*, **13**, (4), 243, (1988)
3. L. Reniff, "Transferring the 45/0 Spectral Reflectance Factor Scale", *Color Research and Application*, **19**, (6), 332, (1994)
4. R. S. Berns and L. Reniff, "An Abridged Technique to Diagnose Spectrophotometric Errors", *Color Research and Application*, **22**, (1), 51, (1997)
5. L. Reniff, Private communication, (1993)
6. D. Rich and D. Martin, "Improved model for improving inter-instrument agreement of spectrophotometers", *Analytical Chemica Acta*, **380**, 263-276, (1999)
7. Munsell[®] Book of Color is a tradename of Gretag-Macbeth, 405 Little Brittain Road, New Windsor, NY 12553, USA
8. ColorCurve[®] is a tradename of ColorCurve System Division of Colwell General, 200 Sixth Street, Fort Wayne, IN 46808, USA
9. NCS – Natural Color System[®] is a tradename of the Scandinavian Colour Institute, AB, P.O. Box 49022, S 100, 28 Stockholm, Sweden
10. SCOTDIC[®], Standard Color of Textile Dictionnaire Internationale de la Couleur is a registered tradename of SCOTDIC Colours, Ltd., 488 Seventh Avenue, 10C, New York, NY 10018, USA
11. FTS, Inc., 699 Isaqueena Trail, Clemson, SC 29633, USA
12. Ceramic Colour Standards (BCRA) tiles are a product of Ceram Research, Queens Road, Penkhull, Stoke on Trent, ST4 7LQ, United Kingdom
13. Spectralon[™] is a registered tradename of Labsphere, Inc., P.O. Box 70, Shaker Street, North Sutton, NH 03260, USA
14. J. Hyman, "Accurate Monotonicity Preserving Cubic Interpolation", *SIAM*, **4**, (4), 645-654, (1983)
15. R. T. Marcus and F. W. Billmeyer, Jr., "Statistical Study of Color-Measurement Instrumentation", *Applied Optics*, **13**, 1519-1530, (1974)
16. F. W. Billmeyer, Jr. and P. J. Alessi, "Assessment of Color-Measuring Instruments", *Color Research and Application*, **6**, 195-202, (1981)
17. OCLI CVF[®] is a registered tradename of the Optical Coatings Laboratory division of JDS Uniphase, 2789 Northpoint Parkway, Santa Rosa, CA 95407, USA
18. Colour Proofer C.P.90/60, Saueressig[®] Gutenberg Strasse 1-3, D-48691 Vreden, Germany

Improving the inter-instrument agreement of spectrophotometers

19. D. C. Rich, "Industrial Color Specifications: Reducing the Uncertainty Beyond the Limits of Absolute Reflectometry", presented at the CORM Annual Conference, Rochester, New York, (2000)
20. J. Verrill, "Intercomparison of Colour Measurements Synthesis Report", *Report EUR 14982 EN*, European Commission, Brussels, (1993)
21. J. Verrill, P. J. Clarke and J. O'Halloran, "Study of Colorimetric Errors on Industrial Instruments", *NPL Report COEM 2, Project 7, Reference MPU 8/36.3*, (1997)

Biographical Information for Dr. Danny C. Rich

Dr. Rich obtained his Bachelors degree in Physics from the University of Idaho in 1973. He received a Masters degree in Physics in 1977 from Virginia Polytechnic Institute and State University in Blacksburg, Virginia. His research involved laser optics and its application to light scattering experiments.

During his studies at VPI&SU he met Dr. Fred W. Billmeyer, Jr. who convinced him to transfer to Rensselaer Polytechnic Institute and work toward a PhD in Color Science. In 1980, he completed his program of studies by defending his dissertation entitled, "The Perception of Moderate Color Differences in Surface-Color Space".

He joined the Sherwin-Williams Company in the fall of 1980 where he directed work on optical properties of coatings, process control, computer modeling and statistical experiment design. In the fall of 1984 he became the Manager of Research for Applied Color Systems, Inc. in Princeton, New Jersey where he did research on color simulation, instrument design and optical metrology and calibration.

In 1989, Applied Color Systems, Inc. became part of the Eichhof Group of Lucerne, Switzerland and along with Instrumental Colour Systems in the United Kingdom and Datacolor of Switzerland, is now Datacolor International. Dr. Rich held the position of Manager of Advanced Colorimetry and Metrology for Datacolor International. In 1998 Dr. Rich joined the Sun Chemical Corporation to direct the Sun Chemical (GPI) Color Research Laboratory in the Daniel J. Carlick Technical Center, Carlstadt, New Jersey.

His current responsibilities include visual and instrumental tolerancing, corporation wide instrument reproducibility, standards and calibration and lighting engineering for Sun Chemical world-wide. Dr. Rich has published on all aspects of Color Science and Engineering, including visual perception, instrumentation, and mathematical modelling.