Colorimetric estimations for ISO 2846-3

John Seymour, 4 May, 2011

Abstract

The current version of ISO 2846-3:2002 provides a single target CIELAB value for each of the inks. Testing for this requires iteratively preparing dilutions of the ink that come closer to the target color value. This is a tedious process and is not absolutely necessary.

The most recent draft of 2846-3 (revision E) introduced an easier approach. Rather than define a target CIELAB value, the standard could define a set of CIELAB values along a curve. I will call this curve a trajectory. If a given dilution is within the color tolerance of some point along the trajectory, then the ink will be deemed suitable.

The idea was stated in slightly different words. The idea in this draft was that, for each ink, there would be an a*b* pair defined for each L* value. These a* and b* values would follow a simple linear trajectory through L*a*b* space. To determine if the measured ink is compliant, the L* value of the measured ink is used to compute target a* and b* values. The ΔE is then computed.

While the general idea is quite sound and practical, the mathematical formulation of the principle (as stated in the current draft) can be improved upon. This paper describes a preferable means of performing the calculation.

This paper is in three parts. The first part is my proposed change to section 5.2. This is a description of the modest calculations that one must go through to determine whether an ink complies.

The second part of this paper will come in handy when the committee revises the set of allowable target values for an ink. This part describes the more complicated calculations that went into determining the values in the proposed table. This might be suitable as an annex to the standard, or it may be appropriate to maintain this as a document internal to TC 130.

The third part gives the theory behind the method.

1. My proposed change

I suggest the following as a replacement for the second paragraph of section 4.2.1

The prints shall be colorimetrical measured according to 4.3. At least three of the test prints shall have values of $-10 < t_p < 10_p$, as calculated according to 5.2. If the extension series does not produce test prints showing the appropriate values of t_p, produce another series of specimens with different and more reasonable ratios of extender to ink concentrate. The extender shall not contain anti-penetration aides typically used for the extension of inks intended to be used on newspapers or improved newspapers.

I suggest the following for section 5.2 of 2846-3

5.2 Colorimetric values

To meet the specification for color, the color difference between the the measured CIELAB value and the "perifarbe" shall be less than 5 ΔE_{ab}, when printed as defined in 4.2.1 at three different lightnesses specified in table 1. The perifarbe is determined according to the following formulas.

Define the following:

L_{nom}, a_{nom}, and b_{nom} nominal CIELAB values of the ink, as listed in Table X

k_L, k_a, and k_b trajectory coefficients for the ink, as listed in Table X

It is presumed that an ideal ink could have its strength adjusted so as to reach any of the CIELAB values represented by the following linear formula that defines the nominal trajectory of the ink:

$$L_{nom} + tk_L, a_{nom} + tk_a, b_{nom} + tk_b$$

for *t* in the range $-10 < t < 10$.

Note 1: It has been determined that the trajectory of the ink is substantially linear within this range. It is not recommended to use this formula for values of t outside this range.

Note 2: The color difference between the nominal color and the color determined by this formula is $|t|$ ΔE_{ab}.

Note 3: The nominal CIELAB values are not to be confused with target values in the 12647 series. The nominal CIELAB value is a somewhat arbitrary anchor point on with to base the equation.

L_{test} , a_{test}, b_{test} CIELAB values of the ink under test

The perifarbe is defined as that CIELAB value on the nominal ink trajectory that is closest to the CIELAB value of the ink under test. The perifarbe (L_p, a_p, b_p) is computed by the following

$$t_p = k_L(L_{nom} - L_{test}) + k_a(a_{nom} - a_{test}) + k_b(b_{nom} - b_{test})$$

$$L_p = L_{nom} + t_p k_L$$

$$a_p = a_{nom} + t_p k_a$$

$$b_p = b_{nom} + t_p k_b$$

Table X – nominal CIELAB values and trajectory coefficients for each ink

Ink	L_{nom}	a_{nom}	b_{nom}	k_L	k_a	k_b
Cyan, type 1						
Magenta, type 1						
Yellow, type 1						
Cyan, type 2	56.9	-27.0	-39.9	-0.765	-0.010	-0.644

Magenta, type 2	49.3	71.6	-4.1	-0.530	0.562	0.634
Yellow, type 2	89.4	4.4	86.7	-0.116	0.246	0.962

2. Calculating the values in Table X

The values in Table X were computed from two sources. A set of measurements at different pigmentation levels and screens (provided to me by Danny Rich, from Uwe Berthold) was used for type 2. <<other data from Bob Hallam?>>

First, the average L* value, a* value, and b* value is computed. The resulting average L*a*b* will be the starting position for the trajectory, that is, the nominal CIELAB values.

$$L_{nom} = \frac{1}{n}\sum {L^*}_n$$

$$a_{nom} = \frac{1}{n}\sum {a^*}_n$$

$$b_{nom} = \frac{1}{n}\sum {b^*}_n$$

The differences in color components between this starting position and each of the measurements for a given ink are then arranged in a matrix, which I call A:

$$A = \begin{bmatrix} {L^*}_1 - L_{nom} & {a^*}_1 - a_{nom} & {b^*}_1 - b_{nom} \\ {L^*}_2 - L_{nom} & {a^*}_2 - a_{nom} & {b^*}_2 - b_{nom} \\ \vdots & \vdots & \vdots \\ {L^*}_n - L_{nom} & {a^*}_n - a_{nom} & {b^*}_n - b_{nom} \end{bmatrix}$$

The product of the transpose of this matrix with itself is computed so as to form a 3 X 3 symmetric matrix, which I call S:

$$S = A^T A$$

Next, we compute the eigenvectors of S. The first eigenvector is the vector of trajectory coefficients.

[Unfortunately, eigenvectors are not a standard function in Excel. I understand there are add-ons available. Alternately, the computation of the first eigenvector is a fairly straight-forward computation that could be done relatively easily using standard Excel functions. Personally, I use Mathematica.]

3. Theory behind the method

The function

The method described and the first possible resolution both suffered from bias because they did not treat L*, a*, and b* equivalently. In both approaches, one of the three is singled out as the basis. The proper way to solve the problem is to treat all three equally.

One way to do this is to write the functional relationship in parametric form. Rather than expressing one of the color coordinates in terms of the others, the values for L*, a*, and b* will all be expressed as a function of a fourth variable, which I shall call *t*. One may think of this as indicative of ink strength, that is, something akin to ink density.

I define the following linear relationship that can be used to approximate a trajectory:

$$\begin{aligned} \Delta L^* &= k_L t \\ \Delta a^* &= k_a t \\ \Delta b^* &= k_b t \end{aligned} \qquad (7)$$

The values of the trajectory coefficients k_L, k_a, and k_b are such that the resulting Δ L*a*b* vector points in the direction of the ink trajectory for that particular ink. They are scaled such that when $t = 1$, the color difference is 1. In other words, ${k_L}^2 + {k_a}^2 + {k_b}^2 = 1$.

Note that this set of equations merely gives an offset from a starting CIELAB value. For this method, the starting CIELAB value is the average of the CIELAB values from in the original test. This set of equations (7) is used to describe all CIELAB values that were achieved with that particular ink.

Linear versus quadratic

I have proposed a linear function. Is this justified?

The relationship for some of the terms could more accurately be portrayed as parabolic. For example, Figure 6 shows a* as a function of L* for cyan ink. It's clear that this relationship has some curvature to it. Statistically, it can be said that the second order term is very significant – above the 99.99% level.

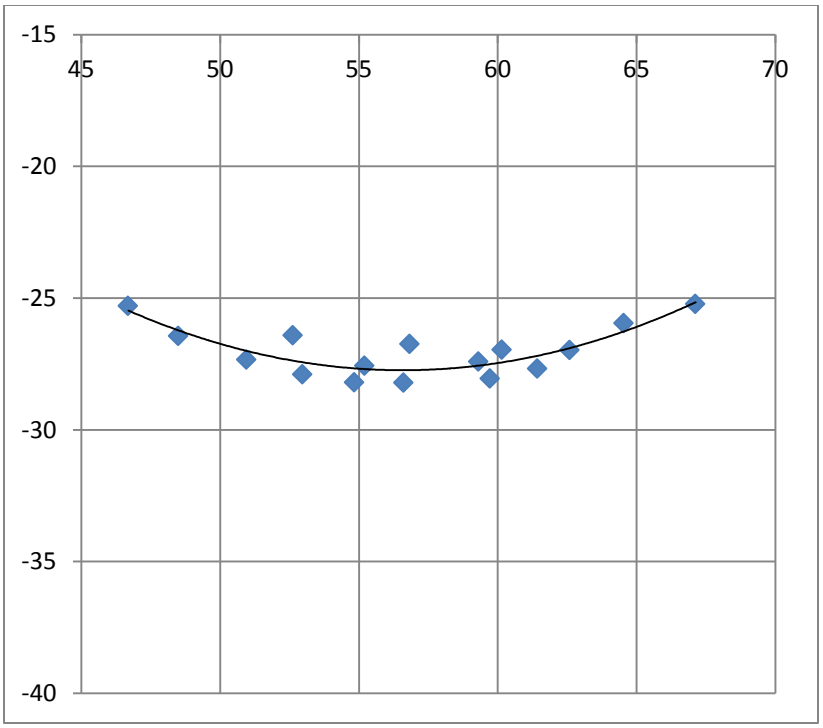

Figure 6 – a as a function of L* for cyan ink*

On the other hand, if the trajectory is defined as a parabolic segment, the math that is required of the user would be substantially more difficult. For my suggested draft of section 5.2, I used a very simple equation that determines the point on a line that is closest to a given point. I don't know off the top of my head the analogous equation for determining the point on a parabolic trajectory that is closest to a given point. I know that it must include in it the solution of a quadratic equation, and I know it must deal with the potential of having two solutions.

If the linear estimate is not good enough, I would want to consider using a method that is a bit more general to determine the trajectory. While Beer's law is not particularly good at estimating the effect of increased ink film thickness or concentration, it has proven useful for predicting spectra that could occur at some ink film thickness or concentration.

Given the spectrum of a solid patch at nominal ink film thickness, $S_1(\lambda)$, and the spectrum of paper, $P(\lambda)$, the following is the estimate of the spectrum $S_k(\lambda)$ of a solid patch with an ink film thickness *k* times that of the nominal density patch:

$$S_k(\lambda) = \left(\frac{S_1(\lambda)}{P(\lambda)} \right)^k P(\lambda)$$

This is a quite general formula that would be very accurate and direct. The solution of this equation would, however, require one to find the value of the exponent k that brings the L*a*b* value, computed from the spectra, closest to a given L*a*b* value. Unless we provide the user of

the standard with a program or spreadsheet to do this calculation, we can't expect them to solve this.

Is it worth the extra effort?

The bow in the curve is a few ΔE, and this is over a range of 20 ΔE. One would hope that the method would not be used to make a correction that large. If a couple of points at both extremes are eliminated, the resulting data starts to look pretty linear.

Given the added complexity of a parabolic function, I have decided to stick with a linear approximation to the trajectory.

Why eigenvectors?

The matrix called "A" has rows ("vectors") that are each deviations from the nominal CIELAB value. If the relationship between the L*, a*, and b* values are all exactly linear, then each one of these entries is exactly a factor of the others. The vectors all lie on a line. We want to find the set of three values that provides the best estimate of the ratio between these three, that is, the vector that comes closest to going through all these points.

Eigenvectors and eigenvalues are a way to do this sort of compression on a matrix, which is essentially a set of vectors. Each eigenvector of a matrix comes along with an eigenvalues that says how important the corresponding eigenvector is in fitting the set of vectors. If we find the largest eigenvalue, and then pick the corresponding eigenvector, we have found the required vector.

There is one little problem. The matrix *A* is not square. There are more than three data points. To get past this problem, we make use of the fact that when we multiply a matrix by itself ($A^T A$) the result is a square matrix that has the same eigenvectors that *A* had.

Limitations of the described method

The method described in the draft is as follows.

1. A measurement is taken of a single test patch, under conditions that are expected to produce a fairly decent match. It is expected that the color match could be improved upon using the normal methods available to a press crew, that is, by changing pigment concentration or dot size.
2. A linear equation is used to estimate what the a* and b* values would be at the target L* value.
3. This color value (the target L* value along with the estimated a*and b* values) is compared against the target L*a*b* value to determine a ΔE value. This is then compared against a threshold value.

I mentioned at the April 2011 TC130 meetings in Berlin that this method does not work well for yellow.

The following three subsections get increasingly more mired in mathematics. Should the reader decide to skip these three sections, I provide summary of the results. The first section basically reiterates the problem I raised in Berlin. This problem could be taken care of with the resolution that was suggested in the meeting.

The second section describes a problem that the method described, even with the simple resolution, would tend to overestimate the color difference. For cyan, this overestimate may be about 40%.

The third section describes an obscure problem that means that the method by which the formulas were derived is less than optimal.

The first problem with the current draft

The first problem with the described method may not be apparent from a graph of b* as a function of L*. Unless you tell Excel otherwise, it may give you a deceptive view of the relationship. Figure 1 is the default scaling that Excel gave me when I looked at b* as a function of L*. The plot can be deceptive in that the horizontal axis covers 6 points of color difference, whereas the vertical axis is 120. The relationship is very misleading.

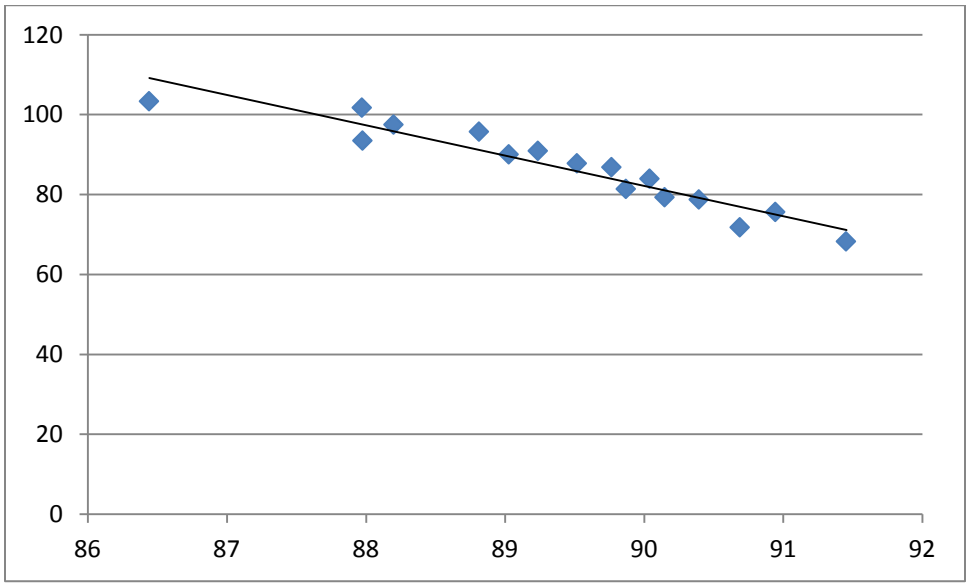

Figure 1 – b as a function of L* value for a yellow ink*

If on the other hand, one scales the graph to cover the same range in L* as in b*, then the difficulty in the relationship is more apparent. Say for example, a yellow sample has been measured at an L* value of 89, whereas the target L* value is 90. The formula shown in Figure 2 would suggest that the b* value be decreased by 7.5 points. A small measurement error in L* is multiplied into a large error in b*. It's just not a good thing to estimate b* from L*.

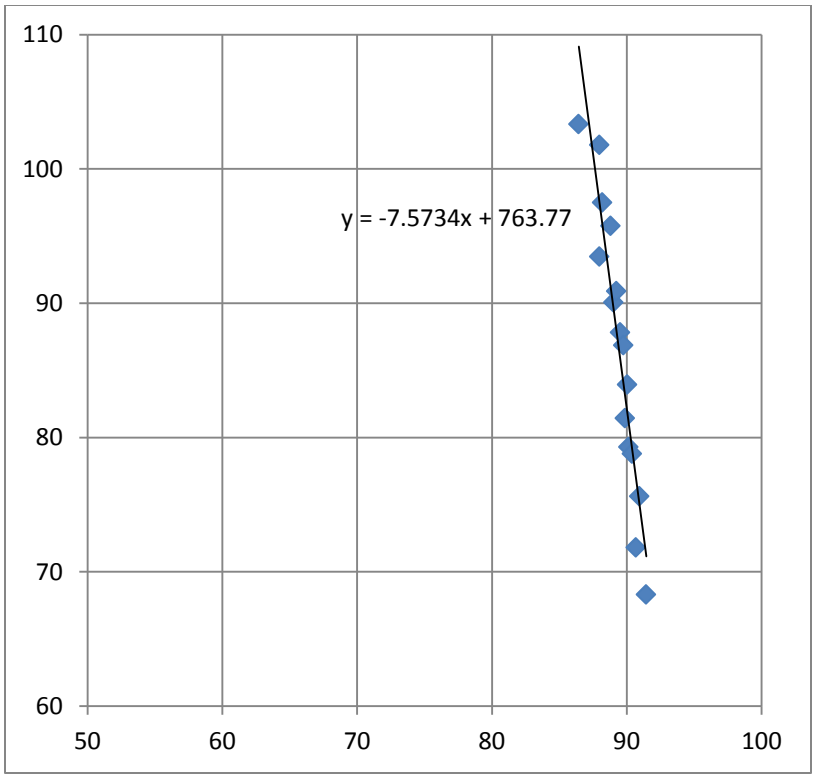

Figure 2 – Same data as Figure 1, but "properly" scaled

I said this in the meeting. My comment was (I think) generally understood. As I recall, the first resolution was to invert the computation for the yellow ink. Rather than using the L* value to estimate the b* value, it is far more stable to estimate the L* value from the b* value. (Indeed, without even knowing the b* value, you could estimate the L* value to be about 89 and not be far off!)

The technique of using L* to estimate a* and b* is more stable for the other inks, so during the initial discussion, it was left at that. Cyan, magenta, and black a*b* values would be estimated from L*, and yellow L*a* values would be estimated from b*.

The second problem with the current draft

As I pondered this further in the meeting, I realized there was another more general problem with the method as described, and I volunteered to describe a more stable solution.

The more general problem is illustrated by first understanding the gist of the method described in the April 2011 draft. Figure 3 shows a plot of b* as a function of L* for cyan ink. The blue diamonds are actual measured values. The black line is a line fit to those values. The large black dot is a hypothetical target of L* = 60, b* = -37.

Suppose a measurement is made where the L* value is 65 and the b* value is -40. This is illustrated by the red dot. According to the premise of this method, adjusting the strength of the cyan ink would move the color along the red line. In particular, it should be possible to move the ink to the correct L* value, as indicated by the solid red arrow and the green dot.

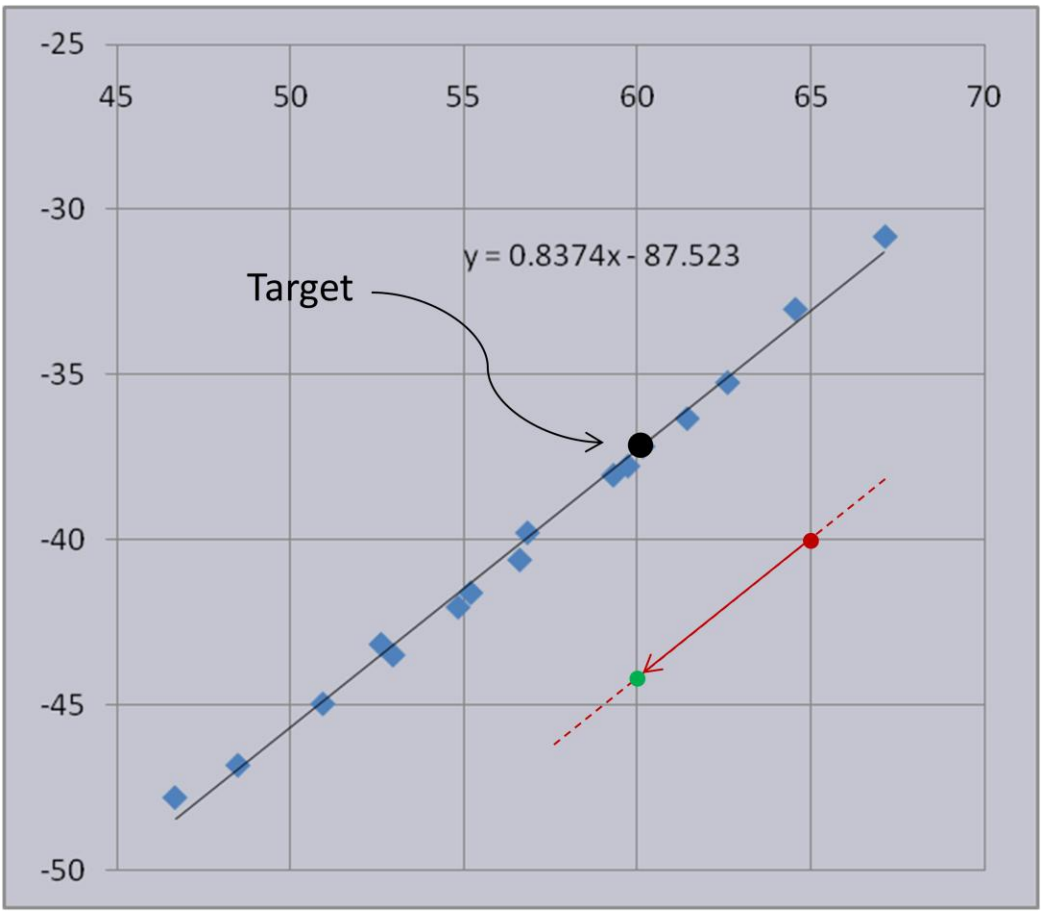

Figure 3 – illustration on the method from the draft

The green arrow in Figure 4 illustrates the color difference that is to be reported according to the current draft. The current draft calls for the color difference if the ink were to have been printed at the correct L* value. This is not, however, the closest color match that could be arrived at (the "perifarbe[1]"). The perifarbe is indicated by the orange point and orange line.

[1] Perigee means the closest approach of a satellite or the Moon to Earth along the orbit. Perihelion is the closest approach to the Sun. I coined this word to mean the closest approach along an ink color trajectory of the color of the ink to a target color. I chose not to call this the "perichrome" since this could be misinterpreted to mean the point where the difference in chroma is the least. I chose the German word for color (Farbe) as the suffix. I figured that sprinkling in a tiny amount of German would make me sound more intelligent, since everyone knows that Germans are smart people.

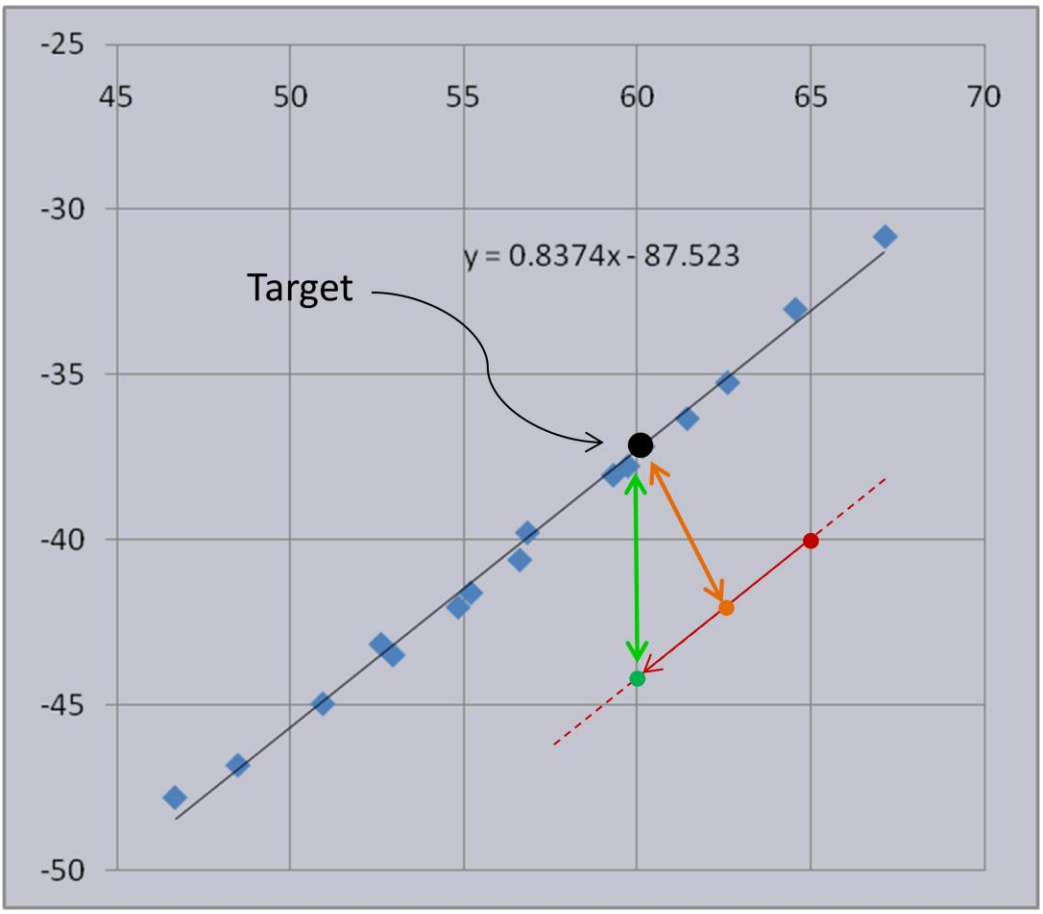

Figure 4 – comparison of the two definitions of closest color match

If the regression line in Figure 4 were to be at 45°, the length of the green line would overestimate the length of the correct answer by 40%.

The third problem with the current draft

The third problem is considerably less obvious. It did not occur to me until I was writing up this paper. It is based on an obscure fact about regression that I did not discover until a few years ago. (I have added this to my list of things that should have been taught to me in college.) It is based on the fact that the standard regression techniques assume that there is uncertainty in the y values, but no uncertainty in the x values.

If one is charting a baby's weight as a function of the number of days old, one can be pretty certain of the time the measurements were made, but there is uncertainty in the weight. This fits the model of the standard regression techniques.

If on the other hand, one is charting weight versus height, then there are uncertainties both in the x values and the y values. Standard regression techniques will not work as well. The standard techniques will make it appear that the line is closer to flat.

Contrary to the layman's description, standard regression techniques do not find the line that comes closest to the data points. Instead (subtle difference), the standard techniques minimize the distance *in y* between the line and the data points. The difference in x is ignored.

Linear regression is used to determine the coefficients m and b in the equation $y = mx + b$, when you know the values for x and y. Figure 5 illustrates the goal of linear regression, to squiggle the green line around so as to minimize a certain measure of the length of the red lines.

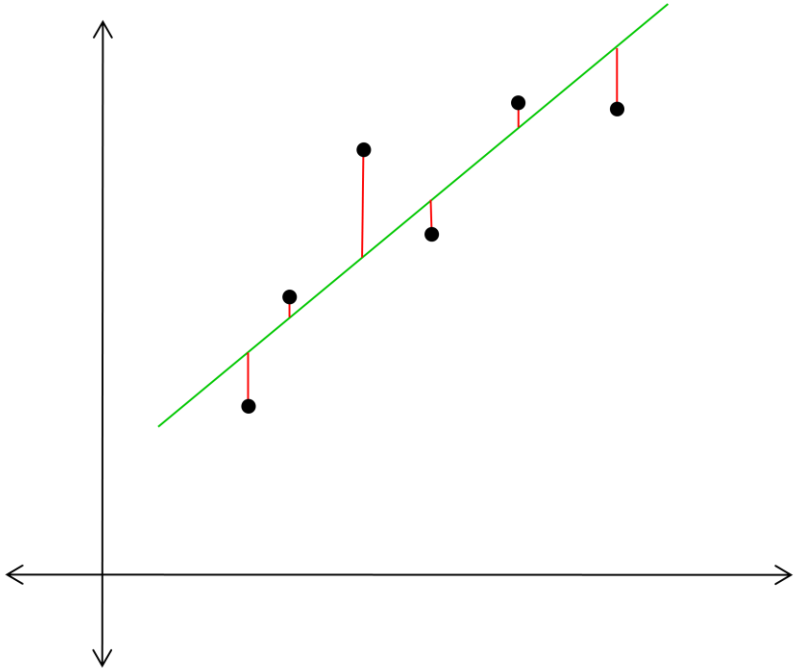

Figure 5 – Illustration of the task of linear regression

The line that comes closest to going through the points is illustrated in Figure 6. The algorithm to find this line is considerably more complicated.

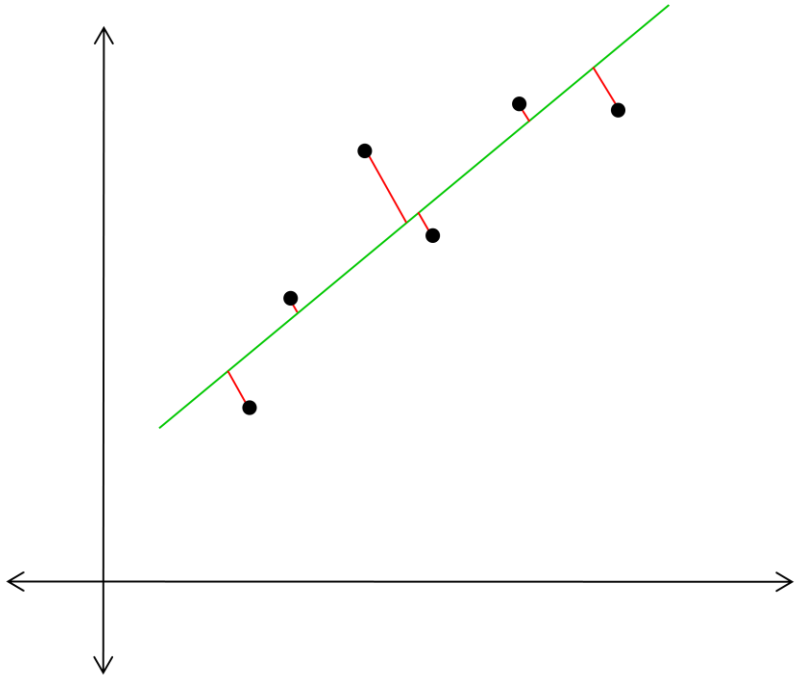

Figure 6 – Illustration of the preferred form of linear regression

When there is uncertainty both in x and in y, regression will report a slope that is biased toward zero. This can be illustrated by the data from Figure 2, which is plotted in an unusual way in Figure 7. The blue dots are the same as the data points from Figure 2, that is, the horizontal axis is L* and the vertical axis is b*.

For the red dots, the two axes are reversed. For the red dots, the horizontal axis is b*, and the vertical axis is L*. I have included the regression lines and equations that Excel provided.

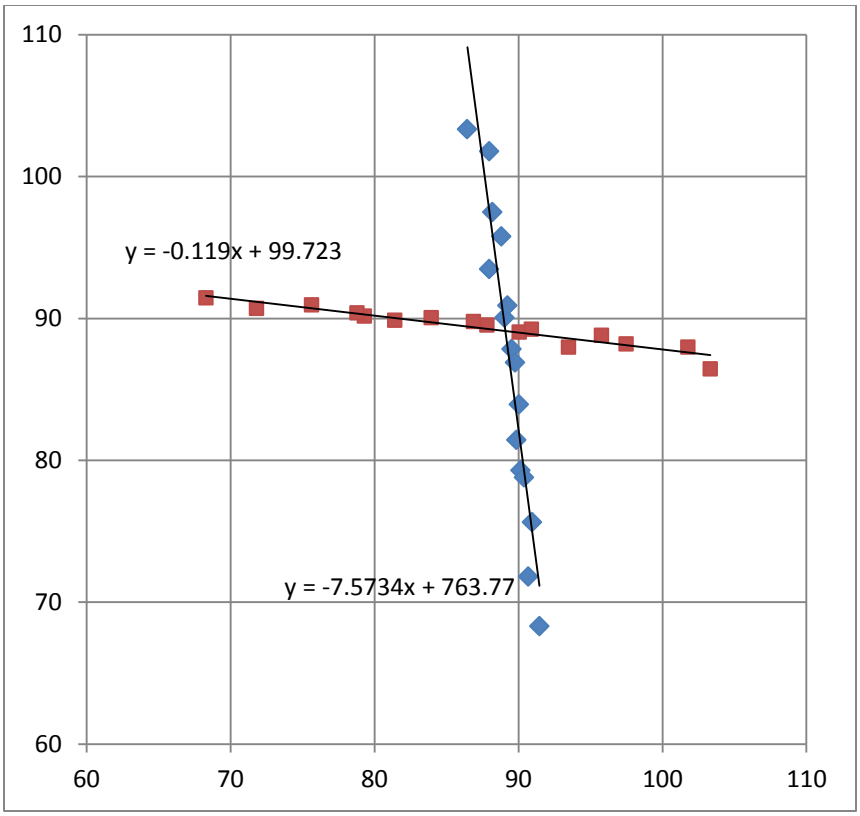

Figure 7 – example for showing the bias of standard regression techniques

One would think that the two equations would be inverses of each other. If I start with the equation for the blue data, I would expect that I could invert the equation to get to the equation for the red data.

$$b^* = -7.5734L^* + 763.77 \quad (1)$$

$$b^* - 763.77 = -7.5734L^* \quad (2)$$

$$-0.132b^* - 100.84 = L^* \quad (3)$$

Note that the slope from applying regression to the red points was -0.119 instead of -0.132. The regression to the red points estimated the slope as being closer to zero.

Similarly, we can do this the other way, that is, start with the regression line for the red points and invert this to get an estimate for the regression line for the blue points.

$$L^* = -0.119b^* + 99.723 \quad (4)$$

$$L^* - 99.723 = -0.119b^* \quad (5)$$

$$-8.403L^* - 838.01 = b^* \quad (6)$$

Note that the slope from applying regression to the blue points was -7.5734 instead of -8.403. Once again, the slope was estimated by regression as being closer to zero.

The difference arises because the regression to the blue data points minimizes the distance (between data points and the line) in b*. The regression to the red points minimizes the distance in L*. Since we have uncertainty in both L* and b*, neither approach is strictly proper. The unbiased answer lay somewhere between the two solutions.

This discussion may be considered pedantic, but it can lead to significantly biased results. Note that the estimations of slope in Equations 1 and 6 would lead to differences of almost 1 ΔE for every change in L* of 1.

Summary of problems

The first and third problems are perhaps not solved by the resolution suggested during the meeting, but are at least mitigated. The second problem, however, is not. The basic idea of determining a*b* from L* does not lead to the perifarbe. A different approach is needed. Hence, I developed the eigenvector method.