Statistical process control of color

John Seymour

Keywords: SPC, color

Abstract

The color difference (ΔE) has long been used as a metric for acceptance tolerances. Indeed, much of the impetus for developing better color difference formulas has been this very practical industrial problem. Tolerances objectively quantify the customer's requirements, in this case, for color fidelity. As such, a metric that quantifies our perception of color difference (for example, ΔE_{00}) is appropriate for customer tolerances.

It should come as no surprise that practitioners of statistical process control (SPC) have generally used color difference as a key metric whereby they benchmark their process. The cornerstone of SPC is to quantify the normal variation of the process to identify when the process starts to behave aberrantly. But, for reasons that are almost universally under-appreciated, the ΔE color difference is inappropriate for SPC.

The first part of the paper reviews deficiencies of the ΔE color difference for SPC.

The second part of the paper introduces *ellipsification* as a means to quantify a cloud of data points in color space. This is described graphically as fitting a three-dimensional ellipsoid to a set of data points. Mathematically, ellipsification is a generalization of the standard deviation to multiple dimensions, in this case, three.

The concept of the Z score is generalized to a three-dimensional metric, which is called Zc ("Z score for color") in the fourth part of this paper. If the original data is "three-dimensionally normal", then the Zc score will be chi distributed with three degrees of freedom.

Finally, these concepts are demonstrated on real world color data. It is seen that data sets from processes that are in good control will closely follow the theoretical distribution, and conversely, data sets from processes that are not in good control will not closely follow the theoretical distribution.

John the Math Guy, LLC

Further related methods are described including a metric to assess whether a process is in control, and metric to compare the variation in color of one process to another, and a way to estimate percentiles of color difference data.

Definition of the subject

To avoid any confusion, this section defines the closely related terms involved with making use of measurements of a manufacturing process.

The difference between quality assurance and process control

There are two basic purposes for making measurements of the final product of an industrial process: quality assurance (QA) and process control (PC). While these two applications of measured data are closely related and often overlapping, the focus is different. The purpose of QA is to verify that the final product meets certain tolerances, generally as dictated by the customer. The purpose of PC is to improve the process by which a product is made.

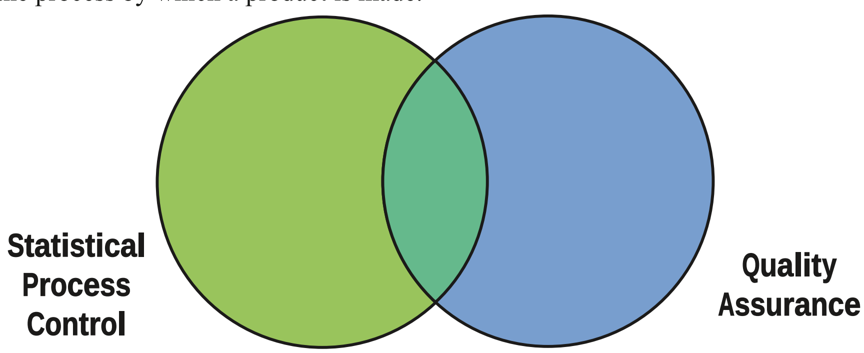

***Figure 1 –** The overlap of SPC and QA*

The ΔE color difference was developed specifically to assess tolerances which are correlated with our perception of the difference between two colors. CIEDE2000 (ΔE_{00}) is the most recent development of this tool. As such, ΔE_{00} is considered the best tool we have today to assess whether a color is in conformance with customer tolerances.

The purpose of this paper is not to challenge the use of any of the color difference formulas for QA purposes. It is first to demonstrate that color difference formulas are ineffective for SPC, and then to propose an effective alternative.

The difference between process control and statistical process control

There are many aspects to process control. It may involve the active adjustment of production equipment to better produce a desired result. When applied to the color produced on a printing press, this involves making color measurements during the press run and adjusting, for example, the opening of ink keys, the pressure between rollers, or the viscosity of the ink. These activities are the innermost loop of process control.

Process control also includes indirect control, which is control of parameters outside the actual manufacturing machinery. When applied to the color produced on a printing press, this refers to the formulation of ink, and often to the adjustment of digital image files through plate curves and ICC profiling to create printing plates that closer meet the color specifications. These activities are the next larger loop of process control.

SPC is a set of techniques that can be used to quantify the behavior of a process. As such, it can be employed in the innermost loop of process control to identify when the process is operating outside of its normal behavior. It can be used in the second larger loop of process control to identify production runs that are aberrant, or to compare one run against another. SPC can also be used at the outermost loop of process control as a technique for process improvement.

Deficiencies of ΔE color difference as an SPC tool

The graph below is a scatter plot of (a*, b*) values of a pink spot color during a production run from Company R. The green dots represent measurements that would be considered normal process variation. The red dots represent abnormal measurements. The red X is the target (a*, b*) value.

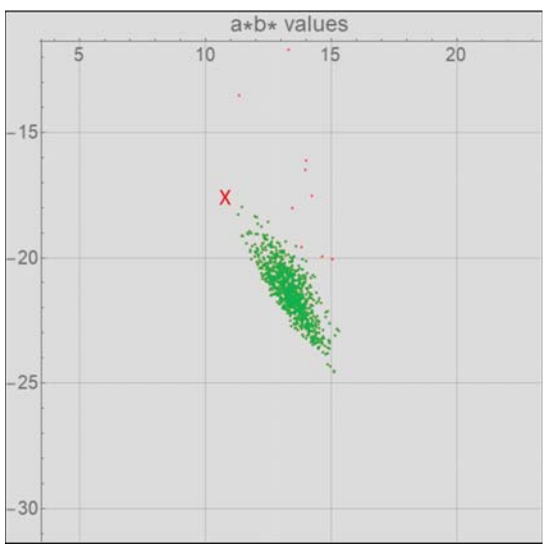

Figure 2 *– a*b* plot of a production run of a pink spot color*

Figure 3 is a different view of that same data. This is a runtime plot of the ΔE_{00} color difference data. The actual tolerance for this production run is unknown, but the red dashed line at a color difference of 3.0 ΔE_{00} is a plausible tolerance.

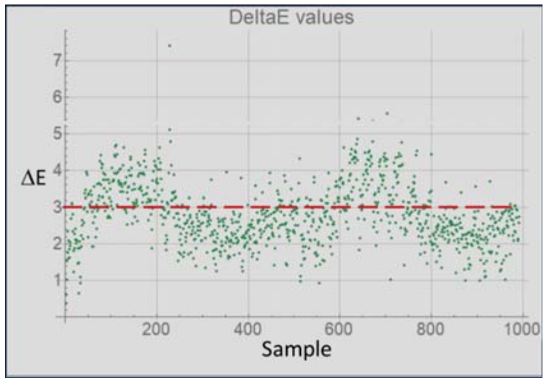

Figure 3 *– Runtime chart of color difference data in Figure 2*

At first glance, it would appear that this press run was woefully out of tolerance. Whether this is actually the case depends on precisely how the contract between printer and print buyer is worded. The contract may specify that all the measurements be within the specified tolerance. This is not recommended!

CGATS Technical Report 016 provides a somewhat different interpretation of how a tolerance should be interpreted:

> *The printing run should be sampled randomly over the length of the run and a minimum of 20 samples collected. The metric for production variation is the 70th percentile of the distribution of the color difference between production samples and the substrate-corrected process control aims.*

From this interpretation of what a tolerance means, the press run looks much closer to being in conformance. Whether 700 of the 1000 data points are below this line or not is a tough call, however.

This leads to one conclusion. A runtime chart of ΔE is useful for QA when the tolerance is an absolute tolerance, which is not recommended for printing. A runtime chart of ΔE is not useful for assessing a tolerance when the number of measurements that meet the tolerance is a percentile. Therefore, I do not recommend runtime charts of color difference for purposes of QA.

Now let's look at the same runtime chart (Figure 4) from the standpoint of SPC. In standard practice for SPC is to set and upper and a lower control limit based on the mean and standard deviation of measured data. (Those statistics may be determined from the data set that you are working with, or they may be determined from previous runs.) The control limits are generally the mean value, plus or minus three times the standard deviation.

When a measurement is inside the control limits, it is assumed that the process is running normally. When it steps outside of those control limits, it is likely that something in the process has changed. A single outlier may or may not be cause for concern, but if several occur close together, then it would be a good idea to check the machine.

A lower control limit for color difference data is not useful. If the color difference between the target color and the measured color is below, for example, 0.1 ΔE, this is cause for jubilation, not concern! So, Figure 4 below shows only the upper control limit, which in this case was about 5.2 ΔE_{00}.

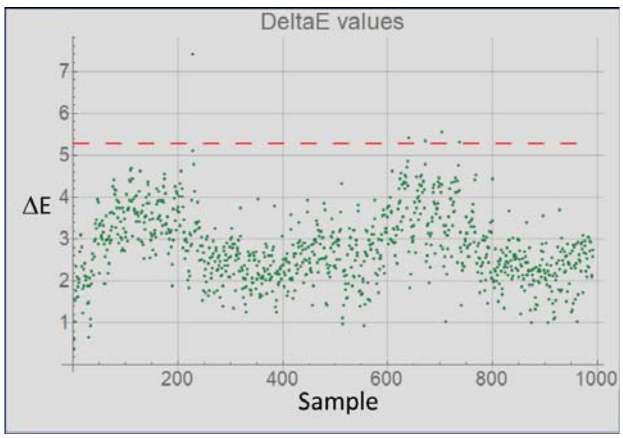

Figure 4 *– The ΔE runtime chart with an upper control limit*

The runtime chart in Figure 4 tells a story of the press run. There is one clear outlier between sample 200 and 250, and there are four borderline outliers between 650 and 750. Aside from that one point – which can readily be passed off as a misread or a hickey in the print – this run looks stable.
But this story is not particularly accurate. Figure 5 below is another (a*, b*) plot of this data, this time with the 5.2 ΔE_{00} ovoid upper limit shown in red. There are at least seven data points that are well outside of the cluster of data points. Note that this is one projection of the points. Additional outliers may be obvious from other directions.

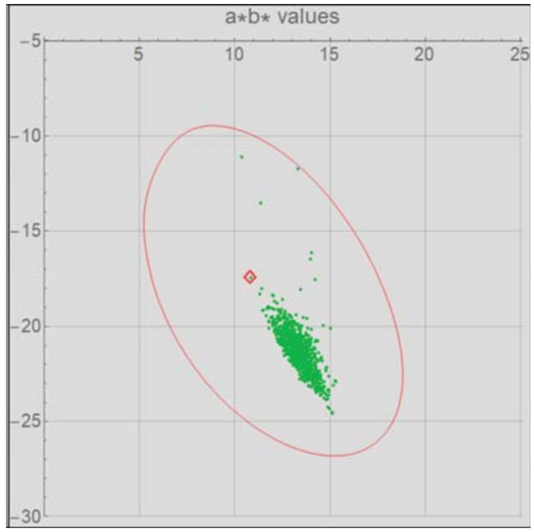

***Figure 5** – Classical upper control limit applied to ΔE_{00} values*
Everything within the red ovoid is considered "normal"

The fundamental premise of a runtime chart is that you use statistics (in this case, mean and standard deviation) to characterize your process. One would hope that the characterization is "tight", that is, unnatural changes in the characteristic that you wish to measure are outside of the bounds of the statistical characterization.

The parameter ΔE cannot be considered tight when used to assess the process variation of color. In the hypothetical graph shown in Figure 6, the color difference can be described as "the color difference is between 1.5 and 3.0 ΔE", but it is clear that there are a lot of colors that are outside the normal variation of the process, but yet fit this simple description.

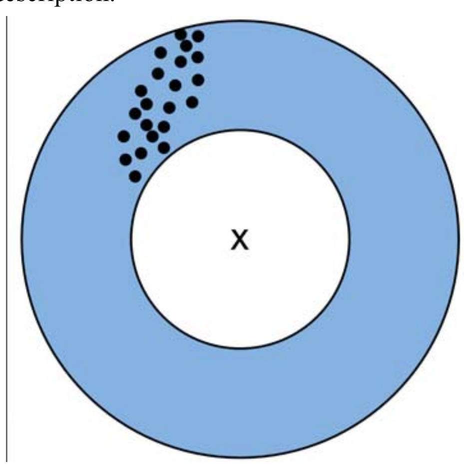

***Figure 6** – Any point within the blue annulus is within the range of 1.5 to 3.0 ΔE, but the majority of this annulus is outside the normal variation in the process*

Gamm and Silvestrini [2015] recognized the issue described here and discussed a simple variant on the color difference. Rather than determine the color difference between the measured samples and the target color, they investigated using the first sample of the production run as the target color. In this way, the ΔE tolerance ovoid is shifted so that it is centered on the data.

This is a slight movement away from pure QA. In some cases, an OK sheet at the start of the run defines the target color, but in general, the target color is specified before the press run.

This approach is an improvement, and it is certainly easy to implement. On the downside, one must be careful not to mistake this for a runtime chart of conformance to a target color. But a more important downside is that the shape and orientation of the ΔE_{00} ovoids is based on a model of the human visual system, and not on the particular characteristics of the dataset being analyzed. Hence the characterization of the color data is not as tight as it could be. The two images below demonstrate the 3 sigma ΔE_{00} ovoids when the target color is closer to the centroid of the data points.

Note: The right-hand graph in Figure 7 is a plot with C* (chroma) on the horizontal axis.

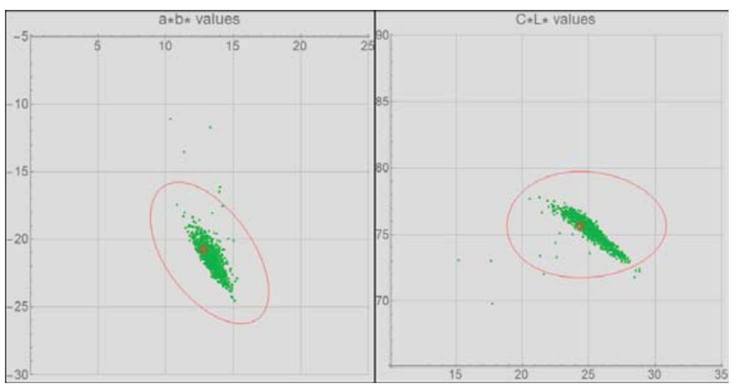

Figure 7 *– Similar plot to Figure 5, but with a point near the centroid chosen as the target color*

Thus, there is room for improvement in the statistical characterization of color data.

Ellipsification and the three-dimensional standard deviation

Motivation

The standard deviation and mean can be thought of as a way to compress a data set down to two parameters. A line segment from the mean minus three standard deviation units to the mean plus three standard deviation units represents the expected range of the underlying data. This is a one-dimensional object.

Color is three-dimensional: L*, a*, and b*, so in order to do SPC of color, we need a generalization of the standard deviation to three dimensions. As a first guess, we might consider using the mean and standard deviations of L*, of a*, and of b*. Thus, we get three means which together represent the centroid of the data, and three standard deviations, which together represent the range of the data in each of the three directions.

Figure 8 shows some hypothetical two-dimensional data, with means of 36 and 28, and with standard deviations of 6.5 and 2.4. The dotted lines show the plus/minus three sigma line segments for each of the two axes. The orange ellipse is the ellipse that has these two line segments as its axes. This is a reasonable characterization of the data for the purposes of SPC.

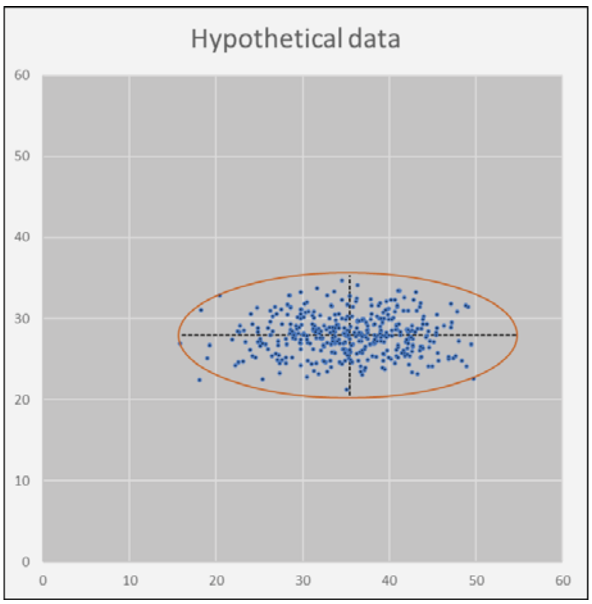

Figure 8 *– A hypothetical collection of x/y data*

Figure 9 shows that same data, only rotated about the centroid by -45°. The mean is the same as the data in Figure 8, but the standard deviations have changed. The standard deviation in the horizontal direction has gone down from 6.8 to 5.0, and in the vertical direction, the standard deviation increased from 2.4 to 5.3. The green ellipse is the ellipse determined from these new standard deviations.

The orange ellipse is the ellipse from Figure 8, rotated by -45° along with the data. It is readily seen that the orange ellipse is a much better characterization of the data. That’s the ellipse that we want for SPC characterization.

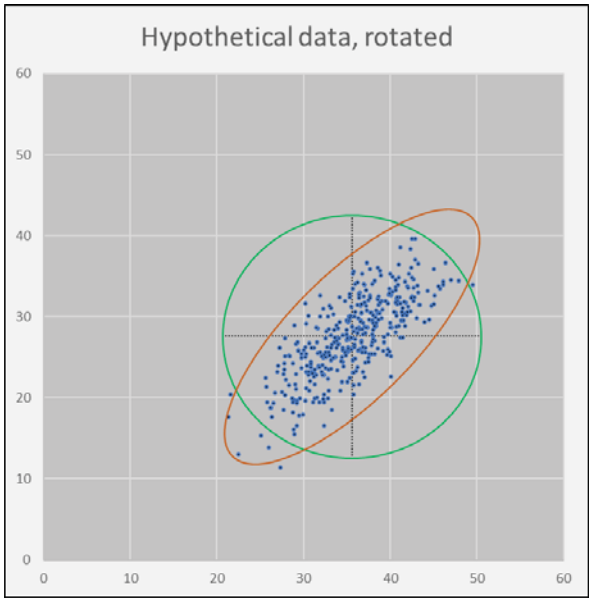

Figure 9 *– The data from Figure 8, rotated*

The orange ellipse (in this two-dimensional example) is defined by five parameters: two parameters define the center of the ellipse, two parameters define the major and minor axes of the ellipse, and a fifth parameter defines the rotation angle where the two axes are uncorrelated.

In three dimensions, we have nine parameters to define an ellipsoid. There are three parameters which define the centroid. There are three parameters which define the major, medial, and minor axes of the ellipsoid. Two parameters define the direction that the major axis points. These could be, for example, a latitude and longitude. Finally, there is another angle which defines the rotation of the ellipsoid about the major axis.

Ellipsification is a word coined by this author. While technically it means *the process of statistically characterizing a set of two-dimensional data with an ellipse, but the concept is not limited to two dimensions. To avoid using words like ellipsoidification and hyper-ellipsoidification, I have chosen to broaden the definition of ellipsification to mean the process of statistically characterizing a set of multi-dimensional data with an ellipse, ellipsoid, or hyper-ellipsoid of the same dimension as the original data.*

This process has been discussed elsewhere [Fotak, McDonald, Volz, Wyble and Laird,].

Zc score

The Z score is one of the basic statistical tests that is introduced in Stats 101. It is used to ferret out suspicious data points in a data set. The mean and standard deviation of the data set are first determined. Next, each data point is normalized by subtracting the mean, and then dividing by the standard deviation. The resulting numbers are unit-less, and represent the number of standard deviation units that a data point is away from the mean. If a Z score for any data point is outside of the range of -3 to +3, then the point is suspicious. If the assumption is made that the underlying distribution is normal, then the probability of such a value is one in 370. If Z is less than -4 or greater than +4, then it is quite likely to be an outlier. This corresponds to a probability of one in about 16,000.

Mathematically, this is the same test as is used in traditional SPC, except that the upper and lower limits for SPC are generally expressed in native units (non-normalized).

Figure 10 illustrates Zc, the three-dimensional generalization of the Z score. Zc stands for *Z score for color*. All of the points within the ellipsoid labelled "Zc = 1" have Zc values less than 1, etc.

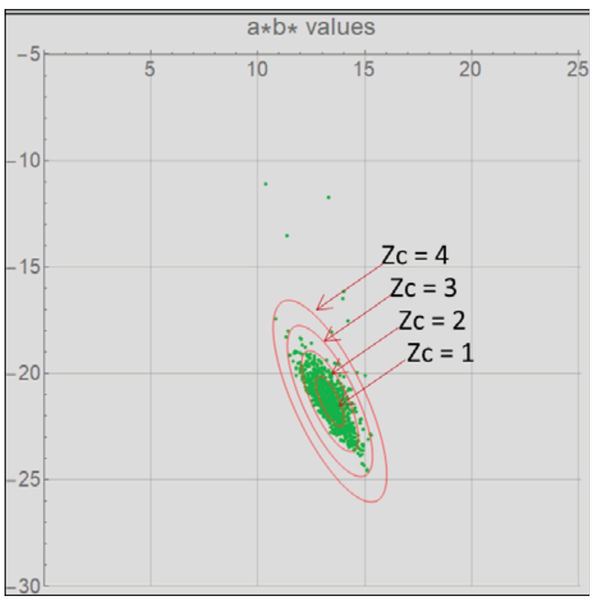

Figure 10 – *Illustration of the Z score on real data*

The Zc value for a data point is determined by first ellipsifying the whole data set to determine the nine parameters of the ellipsoid that fits the data. With this knowledge, the centroid is subtracted from the data point so that the data is centered on the origin. Next, the appropriate rotations are applied so that the coordinates of the data point are based on the axes of the ellipsoid. Finally, the individual coordinates of the transformed data point are divided through by the length of the corresponding axes, which is to say, by dividing by the corresponding standard deviation.

Described in different words, the L*a*b* data points are transformed by the transform that changes the ellipsoid into a unit sphere.

The Zc value for that data point is the distance from the origin for that transformed data point.

This is a generalization of the computation of the Z score for one-dimensional data.

Under the hypothesis that the variation of L*a*b* is trivariate normal, then Zc will have a chi distribution with three degrees of freedom. The generality of this hypothesis will be tested against a large data set in a subsequent section of this paper.

The traditional cutoff levels for Z are similar for Zc. The short table below compares the probabilities of common two-sided Z tests with the equivalent Zc tests.

Probability	Z value	Zc value
90%	1.645	2.501
95%	1.960	2.795
99%	2.576	3.368
99.73%	3.000	3.762

Figure 11 is another look at the L*a*b* data that created the runtime chart in Figure 4. The new runtime chart looks at Zc instead of ΔE_{00}.

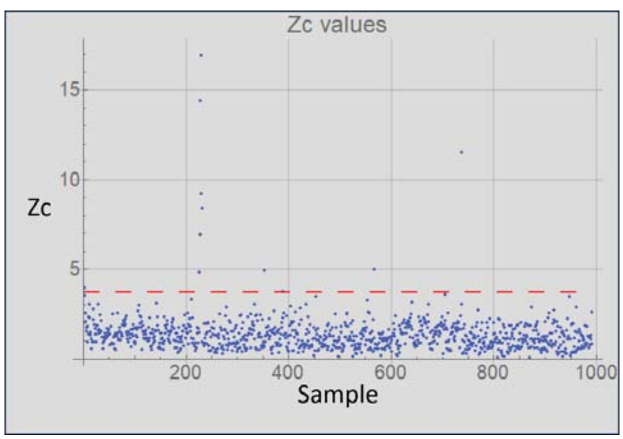

***Figure 11** – Runtime chart of Zc for the Company R pink spot color data*

Recalling Figure 4, there was one outlier identified near sample 210. This could easily be passed off by the operator as a transient anomaly. Perhaps the spectrophotometer was not positioned squarely on the patch? Perhaps there was a small defect on the paper within that small patch? The Zc chart clearly shows that this was not a one-off anomaly, since there were six measurements in a cluster that were outside the upper limit of 3.76 for Zc. Five of these six outliers were missed by the ΔE runtime chart simply because the direction of color variation was toward the target L*a*b*.

The Zc runtime chart identifies three other outliers which appear to be transient outliers. (I say *appear* to be since this data is sampled from a run. It may be that these were one-off, or it may be that these represented a string of outliers that was too short to show up in multiple samples.)

Note that the Zc chart does not show any outliers in the region between sample 650 and 750, as did the ΔE runtime chart. Thus, not only did the ΔE runtime chart fail to show true outliers, it identified non-outlying points as outliers.

Earlier, it was stated that a ΔE runtime chart is not useful for gauging whether a process is meeting a conformance criterion that is of the form "the *n*th percentile of color difference must be below *x* ΔE". We can add to that, the statement that a ΔE runtime chart is not recommended for SPC.

Figure 12 shows an example of a runtime chart that highlights a process that is not in control. This is data from another anonymous company which will be referred to as Company B. This company is a flexographic printer which regularly prints CMYK on films. They collect color measurements of CMYK solids and halftones for each of multiple press runs each day. Figure 12 is a Zc runtime chart of all the C50 patches printed over the course of a year.

The dashed red line is at Zc = 3.762. One would expect, if the underlying distribution were trivariate normal, that about 1 in every 370 Zc values would be above this line. Since there are roughly 3,700 measurements represented in the chart, there should be about 10 outliers. Clearly there are considerably more than this. Furthermore, there are clusters of outliers, indicating that at least twice (indicated by the ellipses) the process was behaving outside its normal behavior for days at a time.

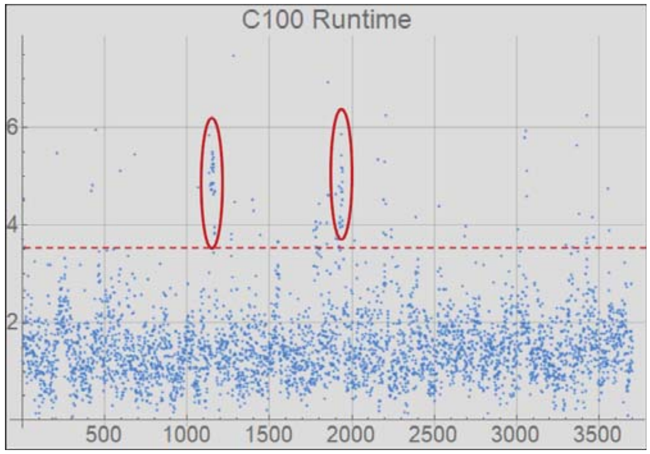

Figure 12 – Zc runtime chart of C50 measurements for Company B

An a*b* scatterplot of this data is shown in Figure 13. It is clear from this view that there are a large number of outliers in this data set.

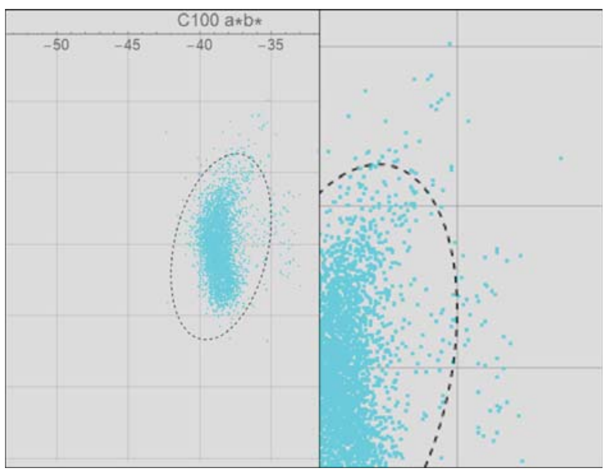

Figure 13 – *Scatterplot of the a*b* values from the data in Figure 12*

Relation between Zc and Hotelling's T^2

Hotelling introduced a statistic in 1931 which was an extension of Student's t statistic to multi-dimensional data. It is called Hotelling's T^2 statistic. Hotelling used this, for example, for process control of the alignment of bombsights [Hotelling 1947]. Hotelling's T^2 has been used on color data [Brown et al., Jackson, Fairchild, Nadal et al., ASTM 2214].

Zc and T2 are closely related, with the key difference being that Zc is in linear units, and T^2 is in squared units. From the standpoint of statistical inference tests, the difference is immaterial. The critical values are determined separately for each, so that any statistical conclusion made with Zc will be the same if made with T^2.

Zc has a simple intuitive explanation: just like the one-dimensional Z score, Zc is the number of standard deviation units that the data point is from the mean. Another conceptual advantage is shown by considering a plot like the ellipsoid plot in Figure 10, with concentric ellipses for Zc = 1, 2, 3, 4. The same plot for T^2 would not have equal gradations in ellipse size.

The runtime chart in Figure 11 illustrates a clear practical advantage for Zc. In this plot, the outlier forced a scaling on the plot up to Zc = 17. Because of this scaling, the data below the critical value of Zc = 3.75 is compressed to the lower 20% of the chart. If the same plot were to be made of T^2, the plot would be scaled to around $T^2 = 300$, with the majority of the data compressed into the lower 5% of the chart.

Travel

There is an interesting feature in the ΔE plot in Figure 4 (Company R pink spot color data) that deserves a closer look. Figure 4 is repeated below with a cyclical pattern drawn in blue. It appears that there is a slowly varying change in the process. This variation is not apparent in the corresponding Zc runtime plot (Figure 11).

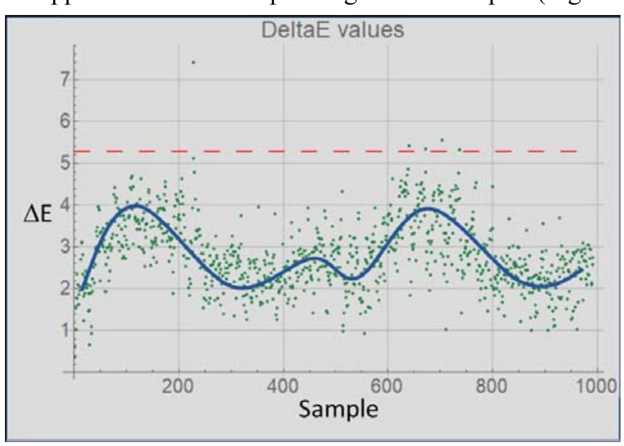

***Figure 13** – a reprise of Figure 4, with a cyclical variation highlighted*

The first question is why this trend in ΔE is apparent.

Since this anonymous data did not come with much in the way of explanation, the type of press this is unknown, the underlying cause is unknown. Note that the variation shown in Figures 2, 5, and 7 is generally toward and away from the white point, which is to say, along the ink trajectory. This is consistent with a change in ink strength, which could be the result of changes in pigmentation level of the ink or in ink transfer.

Note also that the target color (from which ΔE is computed) is close to the ink trajectory, well on the weak end of almost all the data points. Thus, changes in ink strength will directly translate to ΔE values.

This will not always be the case. If the target color were to be off to the side of the ink trajectory (perhaps a slightly different hue) then the change with ink strength will be less significant. If the data is centered on the target color, then (since ΔE is always positive) the relationship between ink strength and ΔE will be complicated by the fact that ink strength below the target strength will look much like ink strength above the target.

Zc exhibits this same behavior, since neither Zc or ΔE can distinguish the direction of the color change. I introduce a new SPC parameter, which I call *major axis travel*, shortened to just *travel*, to serve as a diagnostic. Travel is defined as the position of a color value along the major axis of the ellipsoid. When a point is at the centroid, travel is zero. As it moves toward one end of the ellipsoid or the other, travel gets either larger or smaller (larger in magnitude toward the negative end). If the original data is 3D normal, then travel is normally distributed with zero mean and standard deviation of one.

Those familiar with principal component analysis may recognize this as a principal component score [Gamm, Jackson 2003].

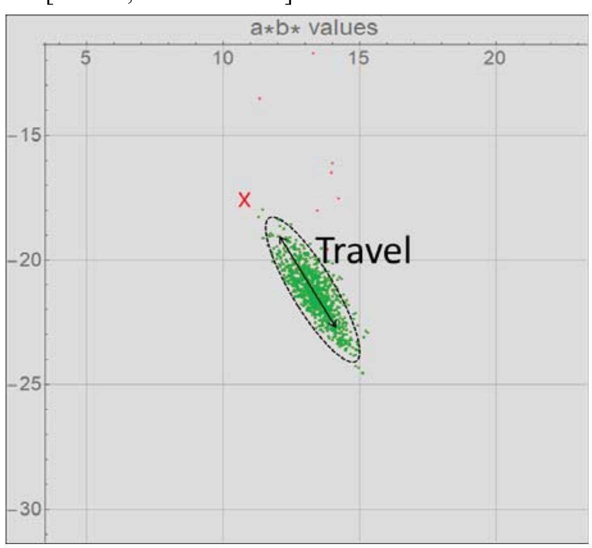

Figure 14 *– an illustration of the travel parameter*

Figure 15 shows a runtime chart of travel for the pink spot color data set from Company R, which appears in Figure 14 and others. This clearly shows the same cyclical trend that was seen in the ΔE runtime chart of the same data.

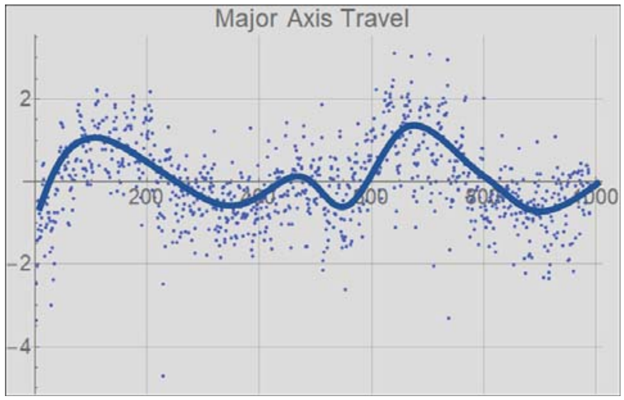

***Figure 15 –** A runtime chart of travel*

Unlike runtime charts of ΔE, travel will always show variation in the predominant direction of variation. This, of course, may or not be a change in ink strength. Further analysis is necessary to diagnose the actual source of predominant variation. This can be done by, for example, determining the expected direction of variation due to any particular physical cause, and then calculating the angle between that and the direction of the major axis of the ellipsoid.

The distribution of Zc

Measured production color data has been collected from eight sources, encompassing 3,005 data sets of lengths from 50 measurements to 7,000 measurements. The combined data set includes 559,449 color measurements.

Company B – Flexo CMYK solids and halftones, over 1 year
Company C – Flexo, 4 spot colors
Company K – Ektachrome photography
Company M – Toner-based digital printing of IT8 charts
Company N – Newspaper, 102 printers
Company P – Plastics
Company R – Two press runs of spot colors
Company S – 170 spot colors

Ellipsification was done on all 3,005 data sets. Extreme outliers ($Zc > 6.0$) were removed one at a time from each of the data sets before combining all the Zc values.

Figure 16 shows a cumulative probability density function of these one-half million data points, with Zc on the horizontal axis and the probability that a Zc value is less than that value on the vertical axis. For example, the probability that a Zc value from this set of data is below 2.0 is roughly 75%.

The purple line is actual data. The dashed blue line is the chi distribution with three degrees of freedom.

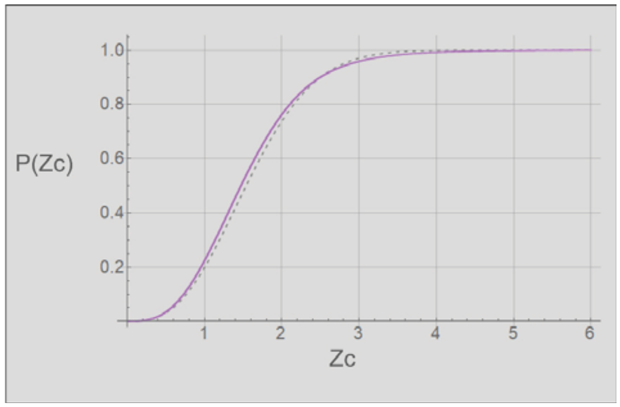

Figure 16 – *The statistical distribution of one-half million Zc values*

This is strong evidence that Zc values computed from real production color data is close to following the chi distribution with three degrees of freedom, or that color data is trivariate normal.

Chi versus chi-square distribution

The chi distribution is not to be confused with the more common chi-square distribution, which is the distribution of the sum of the squares of some number of normal variables with zero mean. It is often associated with regression analysis.

In the field of color, it has been noted in analysis of real data that the square of ΔE has approximately a chi-square distribution with three degrees of freedom [Dolezelak, McDowell 1997 and 2003, ASTM].

Viggiano has laid out the four criteria required for ΔE to have a chi-square distribution with three degrees of freedom. The three color components (L*, a*, and b*) must all: 1) have zero mean, 2) be normally distributed, 3) have the same standard deviation, and 4) be independent.

Nadal et al. hypothesize that since L*, a*, and b* are generally correlated, it may be appropriate to use less than three degrees of freedom. The use the number of degrees of freedom as a regression parameter and determine non-integer values for their sets of color difference data. (Their color differences are all determined from the centroid of the data set, so Viggiano's first criteria is always met.

Seymour [2012 and 2016] investigated numerous color difference data sets from various sources and concluded that Viggiano's criteria are rarely met in practice, and that every data set has its own distribution of ΔE. This is readily apparent from the data in Figure 2.

The fact that there is not one standard distribution for color difference data makes it problematic for use with SPC.

The chi distribution is much less well known than the chi-square distribution [Weisstein, Evan et al.]. It is the distribution of the square root of sum of the squares of some number of normal variables with zero mean.

Wonkiness

Naturally, all data sets do not closely follow the chi distribution. The degree to which the distribution fits correlates with the degree that the production run was "in control". If the production run was "wonky" then it is likely that the cumulative probability density function will deviate significantly from the theoretical.

The difference between the median of the data set and the median of the chi distribution (Zc = 1.53817) is a single number measure of the discrepancy in cumulative probability density functions. This is illustrated in Figure 17, which is the cumulative probability density function from the data set in Figures 12 and 13. From these figures, it was clear that the full production was not in control.

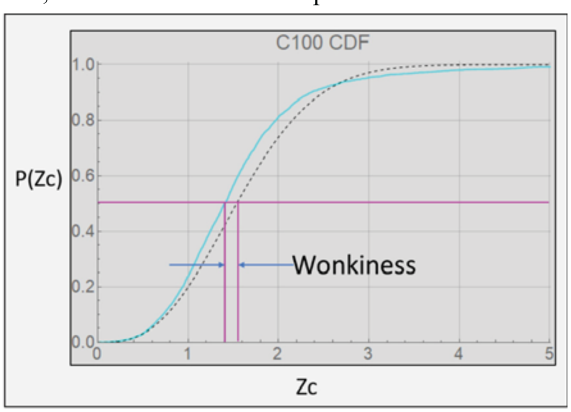

Figure 17 *– an illustration of wonkiness on a data set that is ill-behaved*

I define the wonkiness, W, to be the difference (1000 × (1.53817 - median (Zc data))) to be the wonkiness factor. In the example in Figure 17, W is roughly 200. From a cursory look at the data sets, it would appear that W > 100 is an indication of an ill-behaved production run. Further work needs to be done to establish threshold values for W. In particular, the calculation of W is less stable for smaller data sets (below 300 samples).

One curiosity is that for larger data sets, W is generally greater than 0, which is to say, the median falls below 1.53817, as in Figure 17. It can also be seen in Figure 17 that the Zc values of the data starts to catch up to the theoretical around the 90th percentile so that values of Zc > 3 are much more likely than one would expect. The prevalence of outliers is likely what drives W to increase.

Ellipsoid volume

The volume of the ellipsoid is a single-number measure of the color variation in units of ΔE. The volume is proportional to the product of the length of the three axes, that is, to the product of the three standard deviations.

If the product of the three standard deviations is multiplied by 65.44, then the volume can be interpreted as the volume of an ellipsoid which will theoretically contain 90% of the samples.

This may be used as a simple way to compare the variation of one process against another. For example, if a printer purchases a new piece of equipment or makes some change in the work flow, the ellipsoidal volume can be used to gauge whether the variation has improved.

If the calculations are done in ΔE_{00}, then it is possible to gauge the visual impact of the variation in a way that is independent of the position in color space.

Conformance and the three-dimensional capability index

The goal of conformance is to establish whether a certain percentage of samples is within a certain tolerance. This is a very important question, financially. Contracts, rebates, and re-runs of jobs depend on questions of conformance to established tolerances.

If color were a simple one-dimensional parameter that were normally distributed, conformance testing would be something on the order of "if the average is within the tolerance range and is more than x standard deviation units away from either end of the range, then conformance has been demonstrated."

In most industries, conformance testing is based entirely on the estimation of the mean and standard deviation. These two numbers are derived from the whole data set; every data point is represented in the estimation. As such, these are robust estimates.

Since color is not a simple one-dimensional, normally distributed parameter, and this is generally known, conformance testing is often based on a direct estimate of a predetermined percentile. Various ISO standards in printing have required the 68th percentile, the 70th, and the 95th.

Direct estimation of higher percentiles from small data sets is possibly less robust than estimates of average and standard deviation, since the estimate of percentile depends on just a few samples. For example, if the 95^{th} percentile is determined from 20 or fewer data points, the result depends critically on a single measurement. Measurements at the extreme end tend to have the most variability.

Ellipsification provides a means for estimation of the percentage within a tolerance window which has the same robustness as the method based on average and standard deviation.

The capability index from Six Sigma is a metric from traditional (one-dimensional) SPC which is used to gauge whether a process in its normal state is capable of consistently producing product which is within the customer's tolerance. As such, this lies at the intersection between SPC and QA. SPC defines the normal state of the process and QA defines the customer's tolerance.

At first thought, it would seem reasonable to "simply" generalize the formula for the capability index to three dimensions, as has been done with Zc. Unfortunately, for many reasons, this is neither useful, appropriate, nor possible.

1. The units of SPC (Zc) and QA (ΔE) are different.

2. The contractual statement of tolerances for color are different than in other industries.

3. In one dimension, the test is a simple position on a line segment, so simple subtraction is adequate. For three-dimensional color data, the test involves integrating the probability defined by the ellipsoid at each color within the ovoid color tolerance, as shown in Figure 18. There is no simple way to compute this!

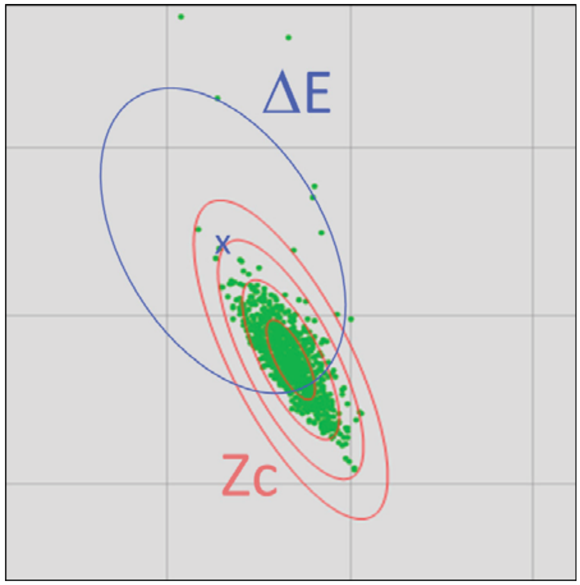

Figure 18 *– Illustration of the complication of the three-dimensional capability index*

Figure 19 provides motivation for how the probability of satisfying tolerances can be calculated. The orange bars are the histogram of ΔE_{00} values (from target color) for a set of 7,000 measurements of a spot color from Company S.

This data was ellipsified to arrive at the nine parameters of the ellipsoid. From these nine parameters, and the L*a*b* of the target color, the theoretical histogram was calculated. This is the dark blue curve in Figure 19. It matches the histogram reasonably well, which is to say, the production was in fairly good control.

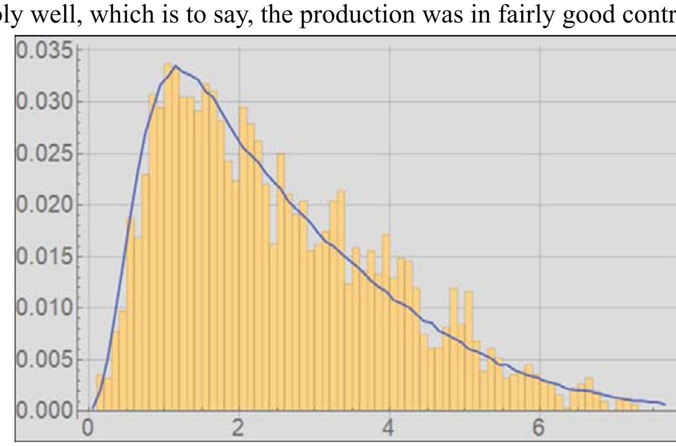

Figure 19 *– Comparison of real and estimated histograms*

This theoretical histogram can be used to determine the percentage of production units that will meet a given color difference tolerance.

In this case, where there are 7,000 data points available, there is no need for the complication of ellipsification. It would be relatively easy and accurate to just count the number of data points in the original data that are within the desired tolerance. Based on this, one can confidently make statements like "based on previous production runs, 98.2% of the product will be produced within 3.0 ΔE_{00} of the target color."

But in many cases, considerably less data is available. Unfortunately, this is often the case when conformance requirements are written into contracts. This is also the case during a press run, where it would be useful to have a runtime estimate of either the 95th percentile of the ΔE values or of the percentage of the run which is expected to be within the ΔE tolerance.

One conjecture is that the ellipsification method of estimating a percentile of color difference data will be more accurate for small data sets than the alternative brute force method of determining percentiles of a data set. Testing of this conjecture is part of ongoing work.

Conclusions

It has been demonstrated that a runtime chart of ΔE is not useful for monitoring of conformance.

Further, such a runtime chart is misleading for purposes of SPC, since it often misses flagging measurements which are outliers, and can also flag measurements which are not truly outliers. The flagging of outliers can be improved if the target color for ΔE calculation is the first sample of the press run.

An improved metric for identification of outliers, Zc has been proposed and demonstrated on real production data.

The Zc metric is based on a process called ellipsification, which is also introduced in this paper. If the underlying data is trivariate normal, then Zc will follow a chi distribution with three degrees of freedom.

It has been demonstrated with one-half million measurements of production runs that Zc will follow a chi distribution with three degrees of freedom.

A new metric, wonkiness, has been introduced, which is a measure of the degree to which a set of color measurements is 3D normal. This can be used as a test for whether a color process is in control.

Another new metric, ellipsoidal volume, has been introduced, which is a measure of the variability of a process. This can be used to compare one process to another.

Finally, a method has been introduced for estimating the percentage of color values from a production run which are within a color tolerance. It has been asserted that this will be a reliable method for providing this estimation from a small sample set. This assertion has not been tested.

Future work

This is an ongoing project. The results in this paper provide a proof of concept of several of the methods, in particular, the Zc metric for identification of outliers and the wonkiness metric for identifying processes which are not in control.

Other metrics introduced in this paper have not been tested. This is part of ongoing work.

The more substantial part of this ongoing project is taking these methods from proof of concept to proof of utility. The author is actively soliciting color practitioners from all industries to provide real world problems to test whether the methods can provide actionable information to production environments.

Acknowledgements

The author gratefully acknowledges the anonymous donations of data from Companies B, C, K, M, N, P, R, and S. Without this data, this work would not have been possible.

The author also gratefully acknowledges the careful review of an earlier draft by Brian Gamm, Danny Rich, and Steve Tiltman.

Appendix

Table of probabilities for the chi distribution with three degrees of freedom.

x	**P (Zc < x)**	**1 in *n* chance?**
0.25	0.00407859	1.00
0.50	0.0308596	1.03
0.75	0.0950391	1.11
1.00	0.198748	1.25
1.25	0.332078	1.50
1.50	0.477833	1.92
1.75	0.617911	2.62
2.00	0.738536	3.82
2.25	0.832723	5.98
2.50	0.899939	9.99
2.75	0.944026	17.9
3.00	0.970709	34.1
3.25	0.985657	69.7
3.50	0.993426	152
3.75	0.997179	354
4.00	0.998866	882
4.25	0.999573	2342
4.50	0.999849	6638
4.75	0.99995	20075
5.00	0.999985	64765
5.25	0.999996	222866
5.50	0.999999	817894
5.75	1	3.20064*10^6
6.00	1	1.3354*10^7

References

1. *ASTM E2214-02*, “Standard practice for specifying and verifying the performance of color-measuring instruments”, 2002
2. Brown, W. R. J., W. G. Howe, J. E. Jackson, and R. H. Morris, “Multivariate Normality of the Color-Matching Process”, *Journal of the Optical Society of America Vol. 46, Issue 1*, pp. 46-49 (1956)
3. *CGATS Technical Report 16*, “Graphic technology — Tolerance and Conformity Assessment”, 2012
4. CIE, “Improvement to Industrial Colour-difference evaluation”, *CIE 142*, 2001
5. Dolezalek, Friedrich, “Appraisal of production run fluctuations from color measurements in the image”, *TAGA 1994*
6. Evans, M.; Hastings, N.; and Peacock, B. “Chi Distribution.” *§8.3 in Statistical Distributions, 3rd ed. New York: Wiley*, p. 57, 2000.
7. Fairchild, Mark, and Richard Alvfin, “Precision of Color Matches and Accuracy of Color-Matching Functions in Cross-Media Color Reproduction”, Color and Imaging Conference, *3rd Color and Imaging Conference Final Program and Proceedings*, pp. 18-21(4)
8. Fotak, Veljko, “A Process Capability Index for Three-Dimensional
9. Data with Circular or Ellipsoidal Tolerances”, *Master’s Thesis RIT*, 2005
10. Gamm, Brian, and Rachel Silvestrini, “The Analysis Of Inline Color Measurements For Package And Labels Printing Using Statistical Process Monitoring Technique”, *TAGA 2017*
11. Green, Phil, “Precision and Bias of Spectrocolorimeters”, ICC White paper, 2006
12. Hotelling, Harold, “The Generalization of Student’s Ratio,” *Annals of Mathematical Statistics, Vol. 2 (1931)*, pp. 360-378.
13. Hotelling, Harold, “Multivariate Quality Control, Illustrated by the Air Testing Of Sample Bombsights”, Chapter 3, *Selected Techniques of Statistical Analysis*, editors Eisenhart, Hastay, and Wallis, MacGraw Hill, 1947
14. Jackson, J. Edward, “Some Multivariate Statistical Techniques Used in Color Matching Data”, *Journal of The Optical Society Of America, Volume 49, Number 6,* June 1959
15. Jackson, J. Edward, “A User’s Guide to Principal Components”, Wiley, 2003
16. Johnson, R.A. and Wichern, D.W., “Applied Multivariate Statistical Analysis”, *Pearson Prentice Hall, 6th Edition,* 2013
17. Roderick McDonald, “Industrial Pass/Fail Colour Matching. Part II - Methods of Fitting Tolerance Ellipsoids”, *JSDC Volume 96 August 1980*
18. McDowell, David, “Statistical distribution of DeltaE”, pdf dated Feb 20, 1997
19. McDowell, David, “Kodak Q-60 color input targets”, Kodak tech paper, June 2003

20. Nadal, Maria, C. Cameron Miller, and Hugh Fairman, “Statistical methods for analyzing color difference distributions”, *Color Research and Application,* June 2011
21. Seymour, John, “Assessing color difference data”, *John the Math Guy blog*, November 14, 2012 http://johnthemathguy.blogspot.com/2012/11/assessing-color-difference-data.html
22. Seymour, John, “Statistical process control of color difference data, part 4”, *John the Math Guy blog*, November 9, 2016 http://johnthemathguy.blogspot.com/2012/11/assessing-color-difference-data.html
23. Viggiano, J A Stephen, “Statistical distribution of CIELAB color difference”, unpublished, June 1999
24. Volz, Hans G., “Industrial Color Testing”, *2nd Ed. Wiley, 2001*, Chapters 2.4 and 6.3
25. Weisstein, Eric W. “Chi Distribution.” From *MathWorld--A Wolfram Web Resource*. http://mathworld.wolfram.com/ChiDistribution.html
26. Wyble, David, and Justin Laird, “Precision and Accuracy of Commercial Spectrophotometers”, *ISCC/SCAI*, 2003