A Measuring Protocol for Printed Matter, Using the X-Rite i1iO Scanning Table

Paul Centore

Abstract

This article suggests a protocol for spectrophotometric measurements of printed matter, when using the X-Rite i1iO Automatic Scanning Table (AST), equipped with an i1Pro2 spectrophotometer. In some sample measurements, about 10 percent of pages showed a significant number of outlying reflectance spectra. The mean colour difference from the mean (MCDM) is a standard measure of spectrophotometric repeatability; the outliers led to significantly higher MCDMs, even when averaged with other measurements. This finding motivated the protocol, which uses the median, rather than the mean, of reflectance spectra. The protocol also accounts for drying time and degradation of printed samples when measured repeatedly. Although the protocol is considerably more involved and time-consuming than a one-shot measurement, its increased accuracy reduced errors in an interpolation-based colour reproduction algorithm by 40 percent, even when the shade bank was half the original size. Practical implications of using, and not using, such a protocol are discussed.

1 Introduction

Spectrophotometric colour measurements are important for colour reproduction. Accurate measurements can make reproduction algorithms fast and accurate, while inaccurate measurements can make them unreliable and slow to converge. A "measurement" might not be a single reading of a measuring device, but rather the output of an involved protocol, perhaps incorporating multiple uses of a measuring device. This article suggests a protocol for measuring printed matter, in the form of uniform colour patches, with an X-Rite i1iO automated scanning table (AST), equipped with an X-Rite i1Pro2 spectrophotometer; the combination of the table and the spectrophotometer is referred to as the i1iO2.

This protocol was motivated by a colour reproduction algorithm for an inkjet printer. The algorithm aimed to print target colours, specified colorimetrically. Using a fixed paper, inkset, and printer settings, a large shade bank of colour samples was printed, with known red-green-blue (RGB) coordinates. These samples were measured, and their colorimetric coordinates were calculated. Linear interpolation was then used to estimate RGBs for the target colours. These RGBs were printed and measured, and their colour differences (or DEs) from the targets were calculated. The smaller the DE, the better the RGB match. During the early development, however, the calculated RGBs were not very good matches. In fact, nearest neighbor (NN) interpolation, which simply selects the sample with the smallest DE to the target colour, was performing nearly as well as linear interpolation. Measurement variability was a plausible cause for this situation. The measurement protocol described in this article was developed to eliminate this cause; with the new protocol, the DEs from linear interpolation were 60 percent smaller than NN DEs. In addition, the DEs achieved with the protocol's measurements were 40 percent smaller than the DEs achieved with other measurements, even though the size of the shade bank was reduced by 50 percent.

The protocol presented here accounts for three possible sources of error: drying time, degradation with repeated measurements, and spectrophotometer variability. The third source is the most important, and the most challenging. A spectrophotometer's variability, or more properly its *repeatability*, is typically measured by the mean colour difference from the mean (MCDM). It was found that one measurement of a page of samples occasionally differed greatly from other measurements of that page, yielding high MCDMs.

The average measurement could therefore be significantly in error, even if many repeated measurements were used. One possible method to avoid this problem is to calculate the MCDMs for all sample measurements, identify which page measurements lead to high MCDMs, and remeasure as needed. The average measurement would then be more reasonable. A second method, which the protocol uses, is to calculate the median rather than the mean of the reflectance spectra. The median automatically disregards outliers, without identifying them explicitly. A mild outlier might be different enough from the other reflectance measurements to be noticeable, but not enough to raise any red flags. Even a mild outlier, however, makes the mean a questionable estimator—the median, on the other hand, is not affected.

The protocol is as follows:

1. Print each sample three times, on three different sheets of paper. If n pages are needed for a single printing of the samples, then $3n$ pages are needed for the three printings. (Multiple printings avoid the degradation caused by repeated measurements.)
2. Let the samples dry enough to achieve colour stability.
3. Use the i1iO2 to measure each printed page three times. Make the measurements in spot mode, and set the "Averaging" parameter to five measurements per patch.
4. Each sample will yield nine measured reflectance spectra (three measurements for each of the three printings). Take the median reflectance spectrum as the final measurement.

While elaborate and time-consuming, this protocol greatly reduces measurement variability, and makes iterative colour reproduction algorithms converge more quickly and accurately, with considerably fewer initial samples. This article describes the protocol development in detail, quantifies the improvement seen in an example reproduction algorithm, and discusses some practical implications.

Figure 1: The X-Rite i1Pro2 Spectrophotometer

2 Description of Protocol

2.1 Description of Measuring Equipment

The measuring equipment used in this protocol consists of an X-Rite i1Pro2 spectrophotometer and an X-Rite i1iO automated scanning table (AST), which are pictured in Figures 1 and 2. The i1Pro2 is a handheld colour measurement device. When inserted into the i1iO, as shown in Figure 3, the combination is referred to as the i1iO2.

The i1iO2 measures a grid of colours on a test sheet, like the one shown in Figure 4. The i1iO moves the i1Pro2 row-by-row across the grid. The measurements can be made in either scanning mode or spot mode. In scanning mode, the i1Pro2 moves smoothly over a row of colours without stopping, measuring as it moves. In spot mode, the i1Pro2 moves to the row's first colour, stops and measures it, moves to the second colour, stops and measures, and so on. Both modes were tried. Scanning mode was faster when it worked, but frequently did not work. Apparently, the colours could not always be sufficiently well measured while the i1Pro2 was constantly moving. Often in scanning mode the i1Pro2 would scan a row of colours, and then stop, go back to the beginning of the row, and remeasure the row in spot mode, stopping at each colour.

Given that scan mode often remeasured in spot mode, it was decided to use spot mode exclusively. Spot mode actually offers two submodes: single spot (using just the M1 geometry) and dual spot (using M0, M1, M2, and OBC). Dual spot mode was chosen, to provide more data. Both submodes offer the option of averaging up to five measurements; the maximum of five was selected and used consistently. When averaging is used, the i1Pro2 can be seen moving slightly to different locations within the same patch, stopping after each movement to make a measurement. In timing tests, averaging over five measurements took

Figure 2: The X-Rite i1iO Automatic Scanning Table (AST)

Figure 3: The X-Rite i1iO2 (an i1iO AST, equipped with an i1Pro2 spectrophotometer)

Figure 4: Sample Sheet of Colour Patches

Figure 5: Initial Screen for i1iO

five times longer than averaging over one measurement. The i1iO2 consistently made just over 30 measurements per minute in dual spot mode. Using the settings recommended in this protocol, the i1iO2 takes about 28 minutes to measure the test page in Figure 4. Once a sheet has been measured, one can produce a text file that gives the average reflectance at each wavelength from 380 to 730 nm, in increments of 10 nm. Unfortunately, there seems to be no way to find the five measurements that went into the average.

Here are screen-by-screen instructions for the i1iO2, as used in the suggested protocol:

1. Figure 5 is the screen that appears when the i1Profiler.app software, which runs the i1iO AST, is launched. Select the circled entry, "Measure Chart," after making sure that the User Mode is "Advanced."
2. On the next screen (Figure 6), choose i1iO2 as the device. The grid in Figure 4 has 15 rows and 12 columns, so make those selections and then choose the "Measurement" option at the bottom.
3. On the measurement screen (Figure 7), select "Dual Spot" as the Measurement Mode, and five measurements per patch under "Averaging."
4. After measuring the sheet, click "Save" and choose a file name and location. A window of savings options (Figure 8) will appear; select the options as indicated in the figure.

The result will be three .txt files, one for each of the measurement modes (M0, M1, and M2). After some header information, each row of a file is a text listing of reflectance fractions for one colour patch. A programming language such as C or Octave can read in this data for further calculations.

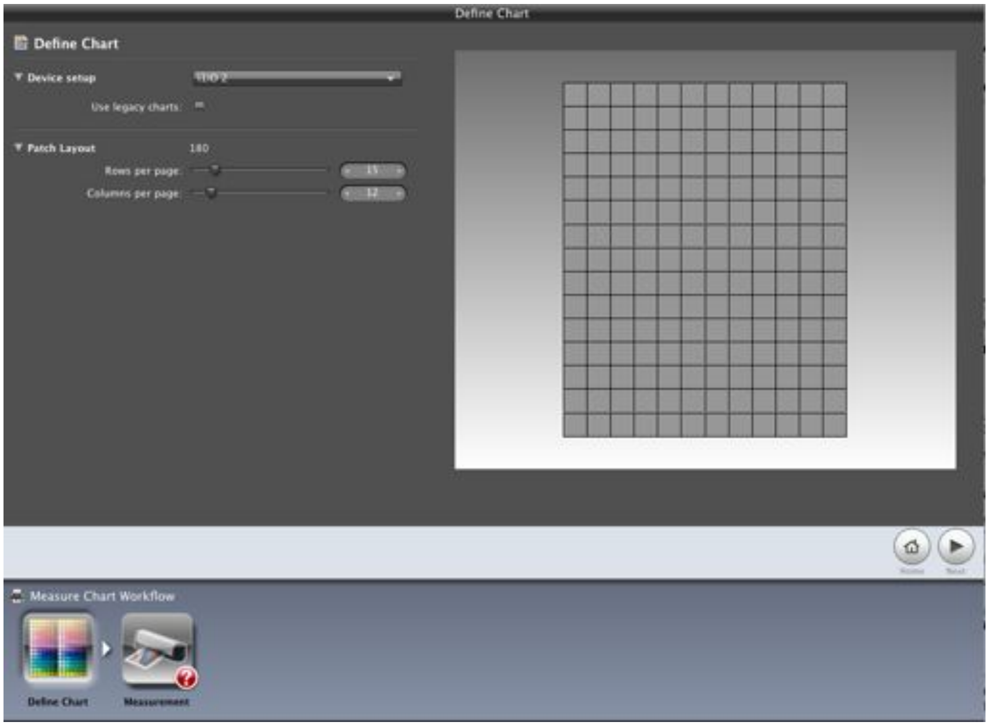


Figure 6: Custom Grid Definition Screen for i1iO

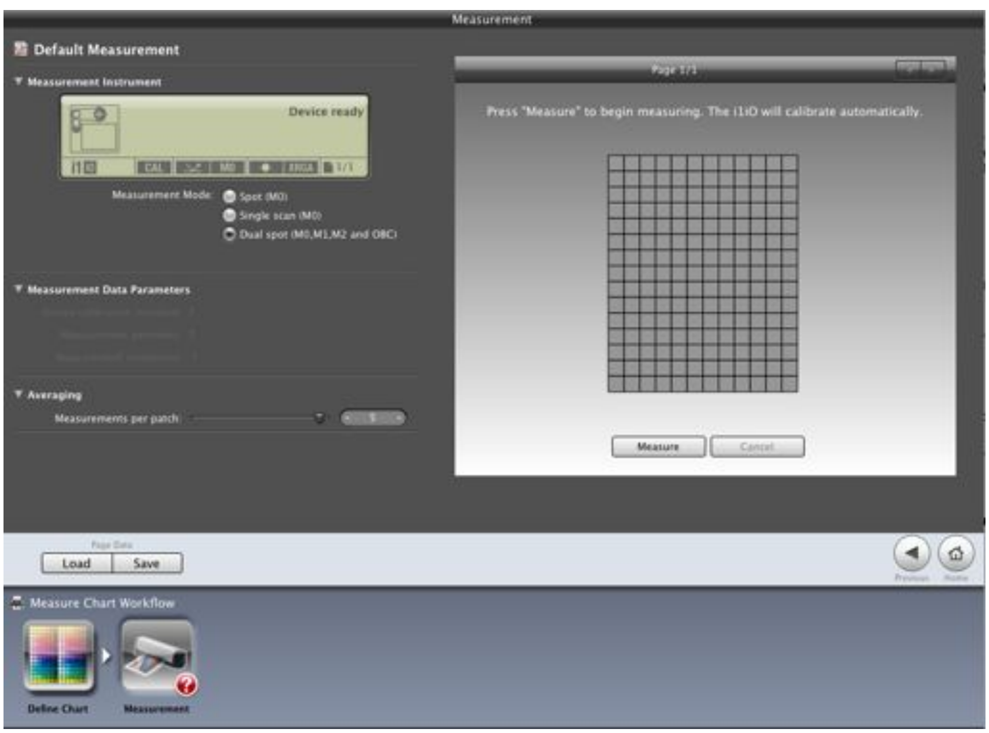


Figure 7: Measurement Screen for i1iO

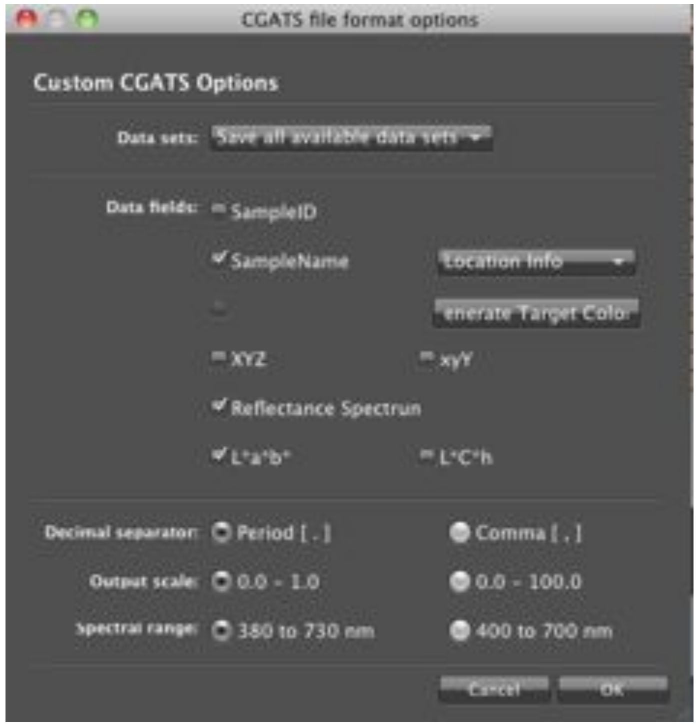


Figure 8: Measurement Saving Options for i1iO

2.2 Sample Degradation

A particular sheet can be measured repeatedly, and the measurements combined statistically. Repeated measurements, however, can degrade a sheet of samples, because the i1iO drags the measuring device across the paper. Although a teflon ring reduces friction, the dragging eventually scars the paper surface, and smudges colours. Figure 9 shows a test sheet after about 10 repeated measurements. The red ovals indicate ink that has been trailed across the paper, as the i1Pro2 moved over the colours. While the effect of repeated measurements was not quantified, it was decided to measure a sheet no more than three times, to avoid degradation. If more than three measurements were needed, then additional sheets would be printed. The protocol presented here uses nine measurements for each sheet, so three printings were needed.

2.3 Drying Time for Colour Stability

A printout can require some drying or curing time, even if the printing ink is already dry to the touch. Framing a print before it has dried can cause "ghost" images on the frame's protective glass. While photographers often think of drying time in terms of ghost images, the drying time needed for colour stability can be a different time. The concern for colour evaluation is that a printout's colours might change slightly over time. Therefore, the measurement protocol should not be started until the printed matter has dried sufficiently that colours will not change further. The following method, which requires no special equipment, is suggested for determining such a time:

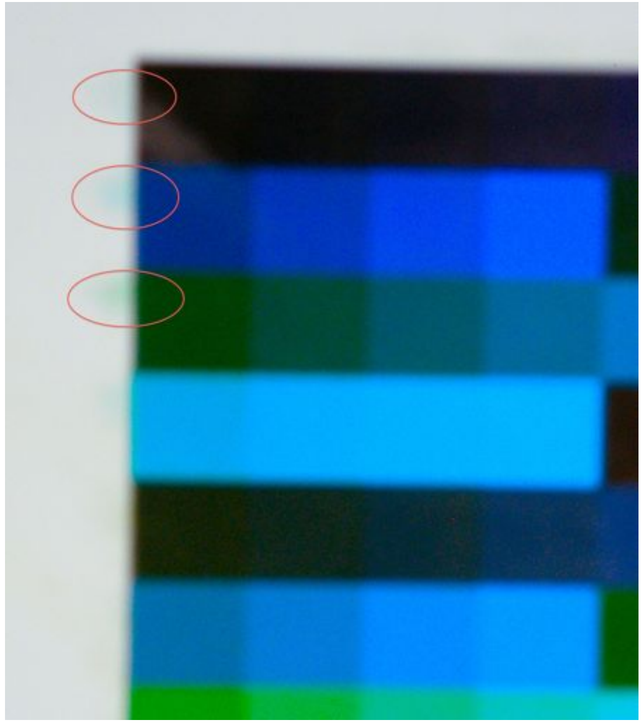
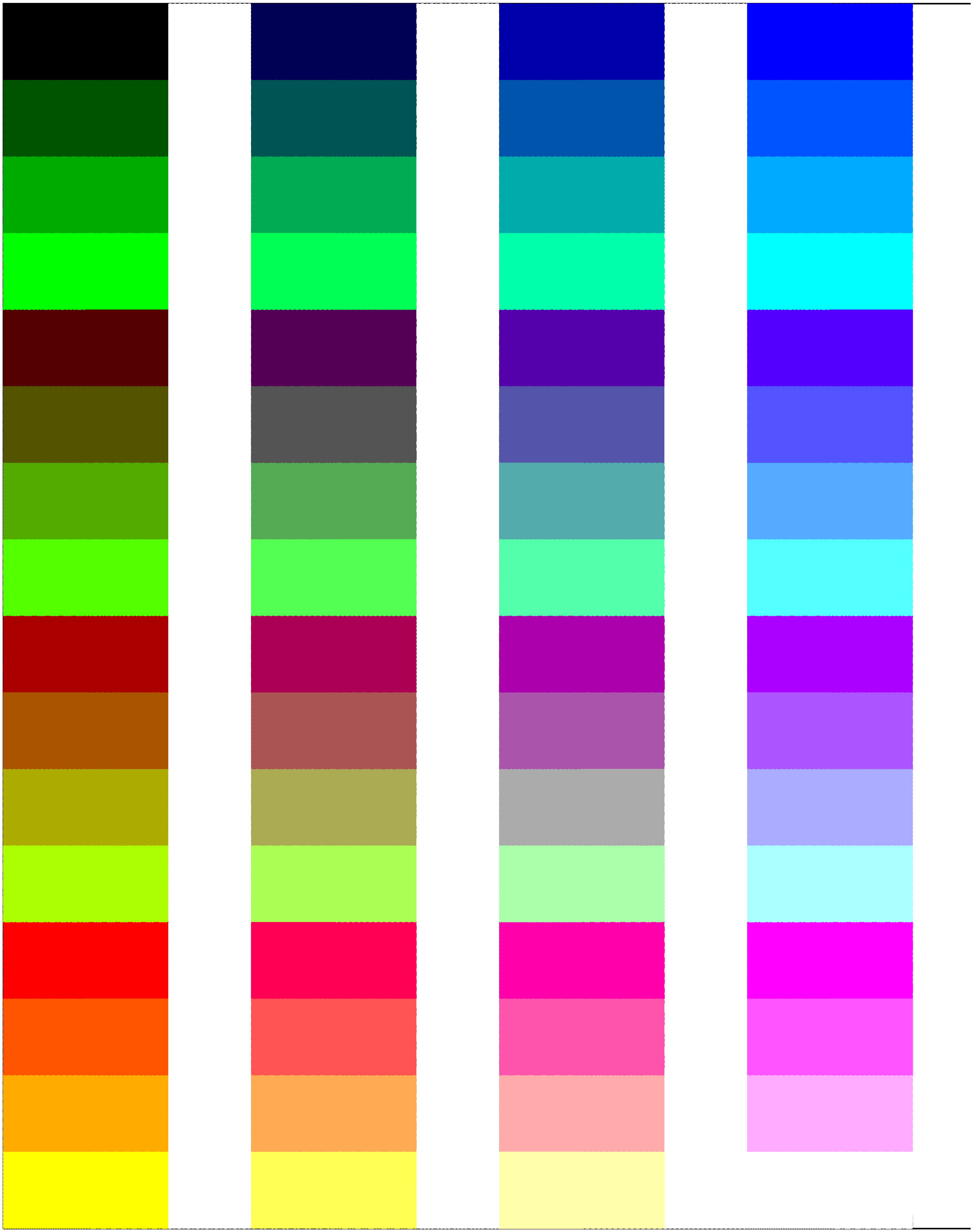


Figure 9: Degradation of Samples after Repeated Measurements

1. Print out the test sheet on the next page. At the bottom, record the date and time of the printing.
2. Set this sheet aside for a few days.
3. After the set-aside period, print out a fresh version of the test sheet. At the bottom, note the date and time. Immediately upon printing, use a pair of scissors to cut sections out of the fresh version as shown in Figure 11.
4. Lay the cut version directly on top of the previous printing. Even slight differences in colours will now be obvious.
5. Every 10 or so minutes, check the colour differences again. When no more colour differences are visible, drying is complete enough for visual purposes. The time that has elapsed since printing the fresh version is the drying time.

The test sheet contains 64 colour patches, specified by red-green-blue (RGB) coordinates. The set of sixty-four is the Cartesian product of four evenly spaced values of the R, G, and B coordinates individually. This choice insures that the colours span the printer gamut. The setting-aside time in Step 2 relies on the assumption that any colour changes will occur within a few days, after which the colours will be constant; the waiting period insures that the original test sheet will be the "final" colours. Over several years, of course, even the final colours can fade or otherwise alter. The above method could easily be adapted to assess fading: simply compare a test sheet that is a few days old to a test sheet that is a few years old.

This method works because the human eye, although poor at measuring absolute colour quantities, is excellent at detecting very slight colour differences. Furthermore, metamerism,



Printer: Paper: Date/Time: Settings:

Figure 11: Cut-UpTest Page

which is the usual obstacle to obtaining perfect matches, would likely not be present for two printouts from the same printer. Metamerism occurs when two colours have different reflectance spectra, but appear the same if viewed under a particular illuminant; a slight change in the illuminant can then destroy the match. A printer would likely consistently print a particular RGB with the same pattern and quantity of inks, so the reflectance spectra for two printouts of the same colour should be identical. In such comparisons, the illuminant would be irrelevant.

Fine visual comparison requires the colours being compared to be directly side by side, with no space between them. Even on a plain, unmodulated background, two colours that are very close can appear identical when placed a few inches apart; laying them side by side will make their difference apparent. Sometimes one is tempted to compare the colours in two printouts by holding one printout in each hand and looking from one to the other. That method will identify obvious differences, but not the fine differences that are of concern here.

Apart from its ease of implementation, visual assessment is conceivably more accurate than a spectrophotometer in this case. We will see later that the i1iO2 produces occasional outliers, large enough to make two identical colours seem different. This fact could make a spectrophotometric assessment much more suspect than a visual assessment. Furthermore, determining drying time with the i1iO 2 would require repeated measurements of the same printout, which would likely alter the printout. In fact, the degradation of the sheet in Figure 9 resulted when it was measured repeatedly in an attempt to estimate drying time.

2.4 Combining Multiple Measurements of Samples

Any one measurement is prone to variability, sometimes significant variability. A standard statistical technique to reduce measurement variability is to make multiple measurements of a quantity of interest, and then combine them into a single estimate of the quantity's true value. This section examines empirically the variation that occurs when measuring printed matter with the i1iO2. As a result of this examination, the protocol recommends the *median* of repeated measurements, rather than the more commonly used *mean*. Another result is an optional procedure that provides a close, intuitive look at measurement variability, and suggests where remeasuring would be helpful.

The mean colour difference from the mean (MCDM) is a commonly used measure of spectrophotometric measurement variability. More precisely, the MCDM evaluates *repeatability*, the extent to which a spectrophotometer gives identical readings, when applied to the same sample under the same conditions, but at different times. To calculate the MCDM for a particular colour sample, first make N measurements of that sample's reflectance spectrum. Next, calculate the mean reflectance spectrum and the mean spectrum's colorimetric coordinates. Similarly, calculate coordinates for each of the N original spectra. Then find the colour differences between the mean spectrum and each of the N original spectra. The mean of these N colour differences is the MCDM.

A high MCDM can also be seen visually in a reflectance plot, such as occurs in Figure 12. This figure shows nine reflectance spectra measurements for a particular sample. The reflectance spectra are numbered. One clear outlier, from measurement 2, is present, along with one mild outlier, from measurement 1. The remaining measurements cluster closely together at the bottom. Ordinarily, the mean reflectance spectrum would fall along the bottom cluster. Since the mean is sensitive to outliers such as measurement 2, however, it actually falls somewhere slightly above the cluster. The colour difference between the mean and measurement 2 is very high, so the MCDM is also very high. When the MCDM is low, on the other hand, all the reflectance spectra fit in a narrow band.

Since an MCDM can be calculated for each sample, a revealing plot is just the sequence of MCDMs for all the samples in a set. Figure 13 shows the MCDMs for a set of 1023 samples, from which Figure 12 is taken. The samples were printed on six pages, as shown in Figure 13; the first five pages contained 180 samples each, while the last page contained the remaining 123. The figure shows that some pages cause disproportionately many high MCDMs. While the first three pages give mostly low MCDMs, with a few slightly higher ones, the fourth page, which contains Sample #714, contains many very high MCDMs.

Closer analysis of the high MCDMs from the fourth page reveal an interesting consistency. Figures 14 and 15 show two further reflectance spectra plots from that page. In both cases, measurement 2 produces an outlier—and an excessive outlier in Figure 15. Plots for another few dozen high MCDMs from Page 4 revealed that measurement 2 always produced outliers. The obvious conclusion is that something questionable occurred when performing measurement 2.

One natural solution to this problem would be to perform a new measurement for Page 4, and replace measurement 2 with the new measurement. The new measurement should cause the Page 4 MCDMs to collapse until they are on a par with the first three pages. While effective, this solution has some disadvantages. First, it requires plotting and examining MCDM sequences, and reflectance spectra, probably multiple times. Second, reflectance plots can show not only obvious outliers, but also mild outliers. In Figure 12, for example, measurement 2 is an obvious outlier, but measurement 1 is a mild outlier. It is not always clear whether a mild outlier is serious enough to warrant remeasuring, or just a result of

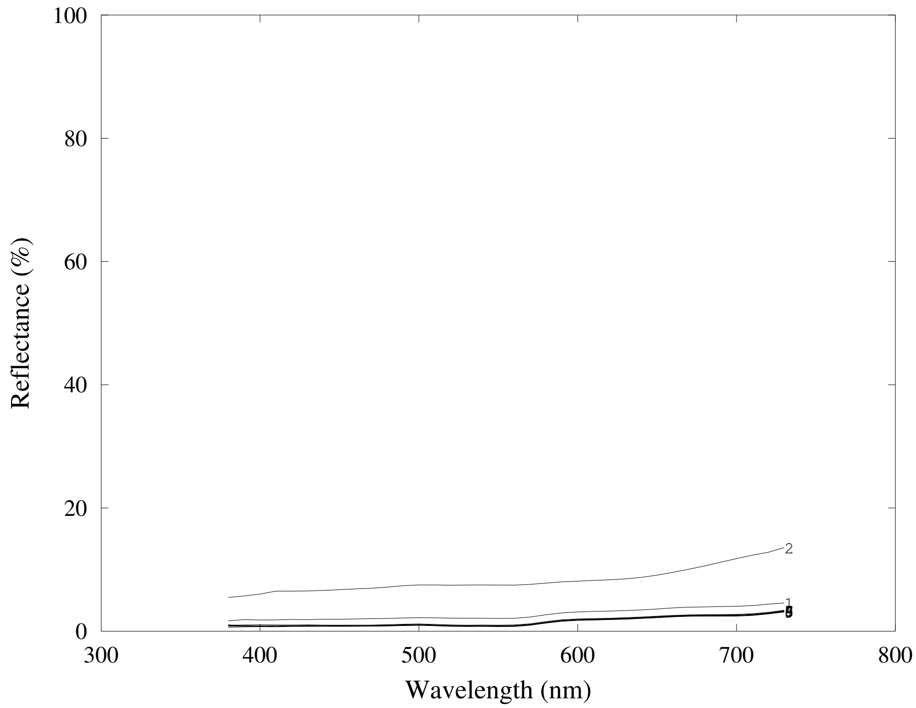


Figure 12: Repeated Measurements of Reflectance Spectra (Sample #714)

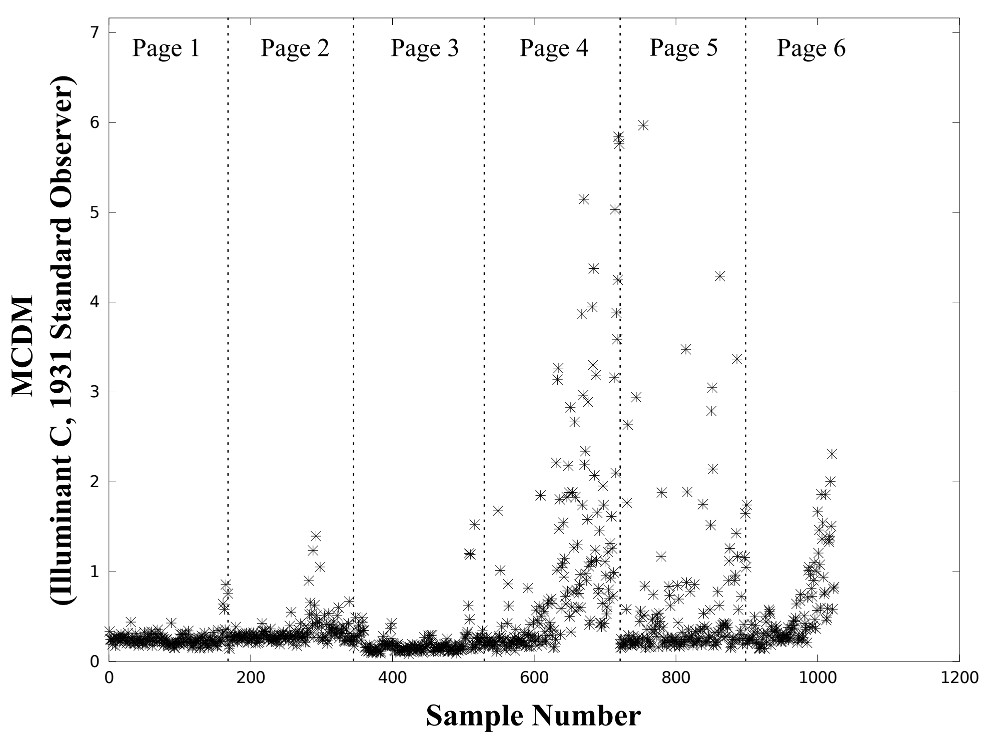


Figure 13: MCDMs for all 1023 Samples in Set

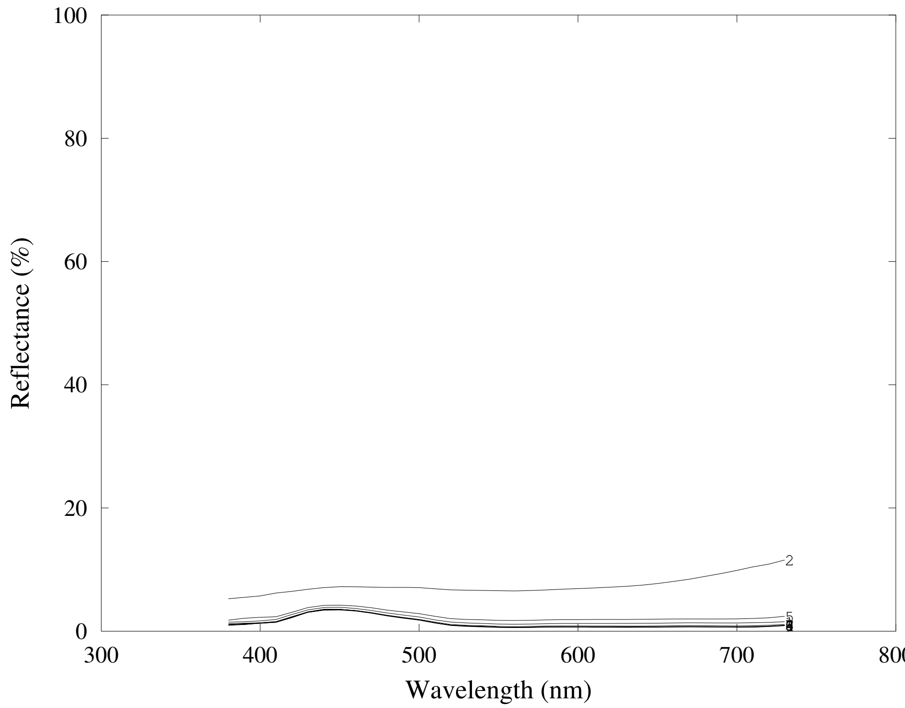


Figure 14: Repeated Measurements of Reflectance Spectra (Sample #720)

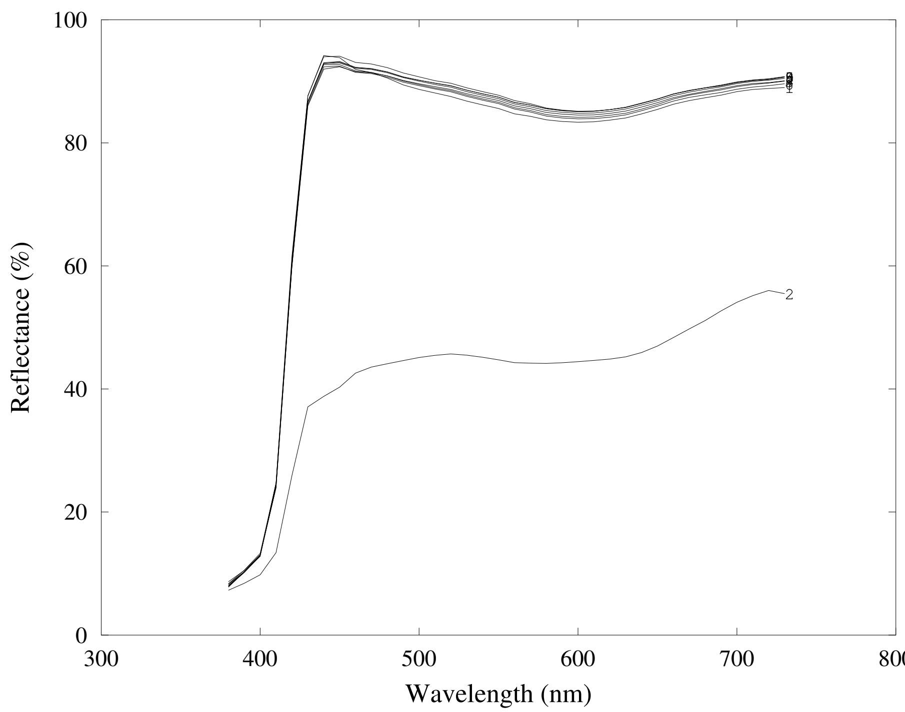


Figure 15: Repeated Measurements of Reflectance Spectra (Sample #676)

normal variation.

A more elegant solution, which avoids both these drawbacks, is to use the median of the reflectance spectra, rather than the more commonly used mean. The median will automatically disregard outliers, whether they are extreme or mild, without having to identify them explicitly as outliers. Since a total of nine measurements are being combined, the median is robust enough to be unaffected by one, two, or even three outliers. Even if a sample has a high MCDM, the median should fall within the common cluster of reflectance spectra. It will thus provide an accurate estimate of the true reflectance, despite a few questionable measurements.

For these reasons, the protocol recommends the median of a fairly large number of measurements. Of course, if finer information is desired, the MCDMs and reflectance spectra can be plotted and examined, with remeasuring as indicated; this procedure, however, is optional, and likely offers no advantage over the median.

3 Empirical Evaluation of Protocol

The need for a protocol such as the one presented became obvious while developing a colour reproduction algorithm for an inkjet printer. The algorithm aimed to use a printer to produce a target colour, given in colorimetric coordinates relative to a specified illuminant and observer. Using a fixed paper, inkset, and printer settings, a large shade bank of colour samples, with known red-green-blue (RGB) coordinates, was printed. These samples were measured, and their colorimetric coordinates were calculated. Linear interpolation was then used to estimate RGBs for the target colours. These RGBs were printed and measured, and their colour differences (in terms of the DE2000 expression) from the targets were calculated. An RGB was considered a sufficiently accurate match if its corresponding DE was below a certain threshold.

During the development, however, it was found that the calculated RGBs were not very good matches. In fact, a nearest neighbor (NN) interpolation, which simply selects the sample with the smallest DE to the target colour, was performing nearly as well as the linear interpolation. Some analysis suggested measurement variability as an explanation: if the colorimetric coordinates of the RGBs contained significant error, then linear interpolants of them could also contain significant error, perhaps enough to make linear interpolation ineffective. This section describes some tests of these hypotheses, and shows that the new protocol makes the reproduction algorithm considerably more effective. In addition, the protocol makes linear interpolation give considerably better matches than nearest neighbor interpolation.

3.1 Nearest Neighbor vs. Linear Interpolation

A printer's gamut, for a fixed paper type, inkset, and settings, consists of the printed colours resulting from all possible RGBs. Each of the R, G, and B coordinates can vary between 0 and 1, so the set of all RGBs can be visualized as a cube. While there are too many RGBs to print and measure all of them, a representative shade bank of n RGBs can be printed and measured. If this shade bank is chosen to cover the gamut approximately evenly, then any colour in the gamut should be fairly near a few points in the bank, and those points can be used to estimate an RGB that produces that colour.

Two basic estimation methods are nearest neighbor (NN) interpolation and linear interpolation. In NN interpolation, the target colour is compared to each of the n colours in the representative shade bank, and n DEs are calculated. The RGB that produces the representative colour with the smallest DE is chosen as the estimator for an RGB that produces the target colour. NN interpolation is easy to understand and implement, but is limited because its estimates are always taken from the same shade banks of n RGBs.

Linear interpolation is more sophisticated than NN interpolation, and can produce estimates anywhere in the RGB cube. In linear interpolation, the colorimetric coordinates of the shade bank's gamut points are tessellated into a set of tetrahedra that covers the gamut. Any target point in the gamut is in a unique tetrahedron, and the RGB coordinates of the tetrahedron's four vertices are known. The target point can be expressed in barycentric coordinates, referred to the tetrahedron's vertices. The RGB estimate is a linear combination of the vertices' RGBs; the coefficients in the linear combination are those barycentric coordinates.

Linear interpolation should be more accurate than NN interpolation. One can think of the n measured RGBs as a finite set of points in the function f that assigns colorimetric coordinates to RGBs. NN interpolation models f as a discrete step function, with steps centered on the known points. Linear interpolation models f as a piecewise linear function, which goes through the same known points. As long as f is fairly well-behaved, the piecewise linear approximation should be more accurate than the discrete step approximation. This statement could break down, however, if the "known" points were not known very accurately. In that case, linear interpolants would have a strong random component. They would be somewhere near the vertices, just as NN interpolants are, but they might not be any closer to the target colours.

3.1.1 The Limits of Nearest Neighbor Interpolation

The larger a shade bank is, the more accurate its nearest neighbor interpolants should be, although errors will never be consistently less than measurement error. This section quantifies the effect of increasing a shade bank, and shows that NN interpolation suffers from diminishing returns: after a certain point, additional colours in the shade bank only minimally improve the accuracy. A practical example is investigated, and the relationship to colour measurement accuracy is analyzed.

An initial shade bank of 4096 RGBs was chosen, by choosing 16 equally spaced values in each of the R, G, and B components, and taking all possible combinations. To avoid very flat tetrahedra, RGBs in the interior were jostled randomly, so that they would not form planes. These 4096 RGBs were printed out three times on Kirkland Signature Glossy Photo Paper, using an Epson Stylus Photo R2880 printer. Each printing required 23 pages, so in all 69 pages were printed. After being left to dry for a few days, each page of each printing was measured three times, giving 207 page measurements. In accordance with the protocol, the reflectance spectrum for each patch was taken to be the median of the nine reflectance spectra.

To evaluate the effectiveness of NN interpolation, 540 RGBs were chosen from a random

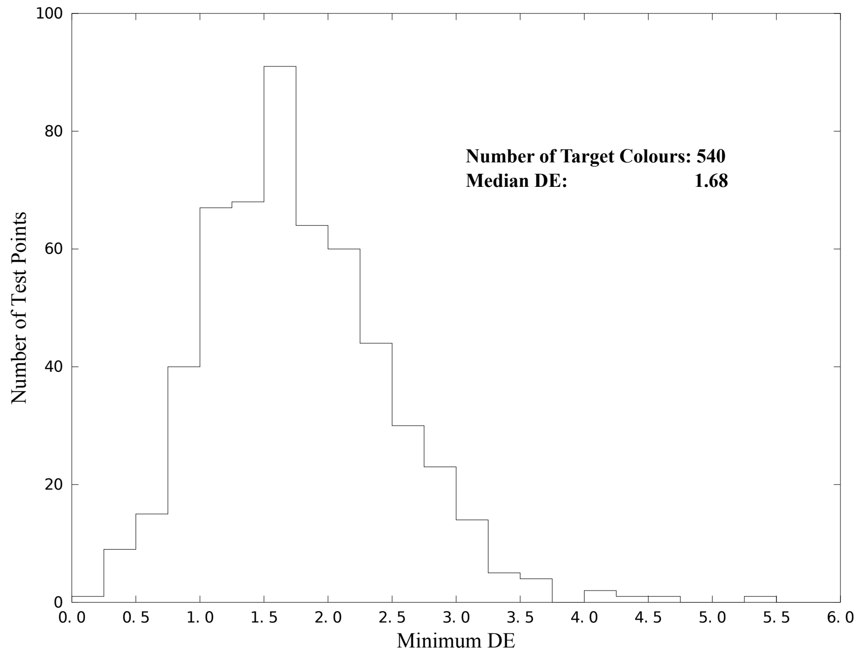


Figure 16: DEs to Nearest Neighbors in a Set of 4096 RGBs

distribution, which was uniform in each of the R, G, and B coordinates. The 540 RGBs were printed out and measured, and their colorimetric coordinates were calculated. Ordinarily, a target colour is given in colorimetric coordinates, and there is no guarantee that the target is within the printer's gamut. Starting with RGBs guarantees that the target colours are in fact within gamut; once the target colorimetric coordinates were calculated, the RGBs were discarded. For each target colour, the nearest neighbor in the 4096 points was found, by an exhaustive check. Figure 16 is a histogram of the 540 DEs from the target colours to their nearest neighbors. The median DE is about 1.68.

Now suppose that, to reduce the NN DEs, the initial shade bank was doubled, by adding another 4096 colours. Then the distribution of the augmented shade bank can be calculated from the data in Figure 16. Figure 16 can be seen as an (un-normalized) estimate for the probability density function (PDF) of DEs for the original shade bank. Denote this PDF by $f(\mathrm{DE})$. Figure 17 displays the same data as a cumulative distribution function (CDF), which we will denote $F(\mathrm{DE})$.

The doubled shade bank is the union of the initial shade bank and additional shade bank, each with 4096 colours. The CDF for the additional shade bank should then be approximately identical to the CDF for the initial shade bank, and both would be represented by $F(\mathrm{DE})$. The NN DE for a particular target colour would be the minimum of the NN DE for the initial shade bank, and the NN DE for the additional shade bank. Let $F_2(\mathrm{DE})$ denote the CDF of the minimum NN DE for the doubled shade bank. The chance that this minimum would be δ or less is

$$
\begin{aligned}
P\left(\text{a particular minimum NN DE } < \delta\right) &= F_2(\delta) && (1)\\
&= 1-(1-F(\delta))\,(1-F(\delta)) && (2)\\
&= 2F(\delta)-F^2(\delta). && (3)
\end{aligned}
$$

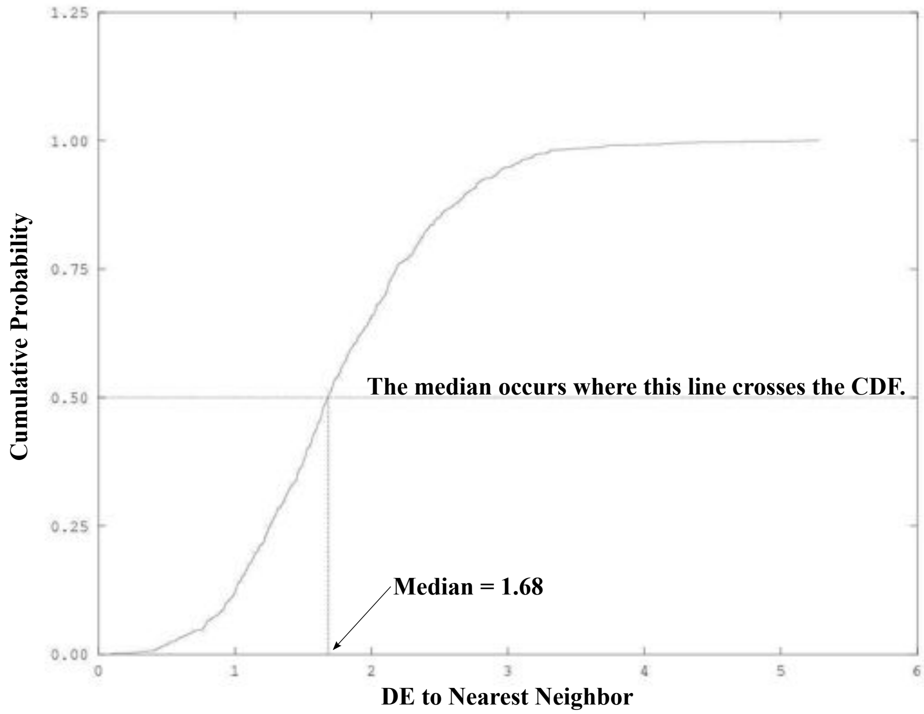


Figure 17: CDF of DEs to Nearest Neighbors in a Set of 4096 RGBs

Equation (2) is the complement of the probability that the target colour's NN DE is greater than δ for both shade banks. This equation assumes that the two shade banks are chosen independently, which is likely to be the case in practice. If k shade banks of 4096 colours were combined, then Equation (2) could generalizes to

$$F_k(\delta) \;\; = \;\; 1 - (1 - F(\delta))^k \,, \tag{4}$$

where F_k is the CDF for a combination of k shade banks.

While a model for the PDF in Figure 16 might lead to an analytic expression for Equation (4), such a model is elusive, and in any event is not necessary in this case. Instead, we can use the empirical approximation to the CDF that is generated by the vector of the measured DEs. Equation (4) can then be evaluated exhaustively from this data. Figure 18 shows the CDFs that result when k varies from 1 to 8. When $k = 1$, the CDF is just the CDF that appears in Figure 17. The median of a CDF occurs when the ordinate takes on the value 50%. Those medians have been indicated, along with the DEs where they occur. Table 1 lists the medians for all values of k.

As expected, the median DE decreases as the size of the shade bank increases. The improvement, however, suffers from diminishing returns. The first addition of 4096 colours decreases the median by a significant 0.33. Subsequent additions of the same number of colours have progressively smaller effects. By the time there are 20,000 colours, adding another 4000 colours only decreases the median DE by about 0.04. From a practical point of view, Table 1 shows that the nearest neighbor approach has limited value. Adding colours to the shade bank produces noticeable improvement until there are about 10,000 colours. Further additions produce much smaller improvements. The CDFs in Figure 18 show the

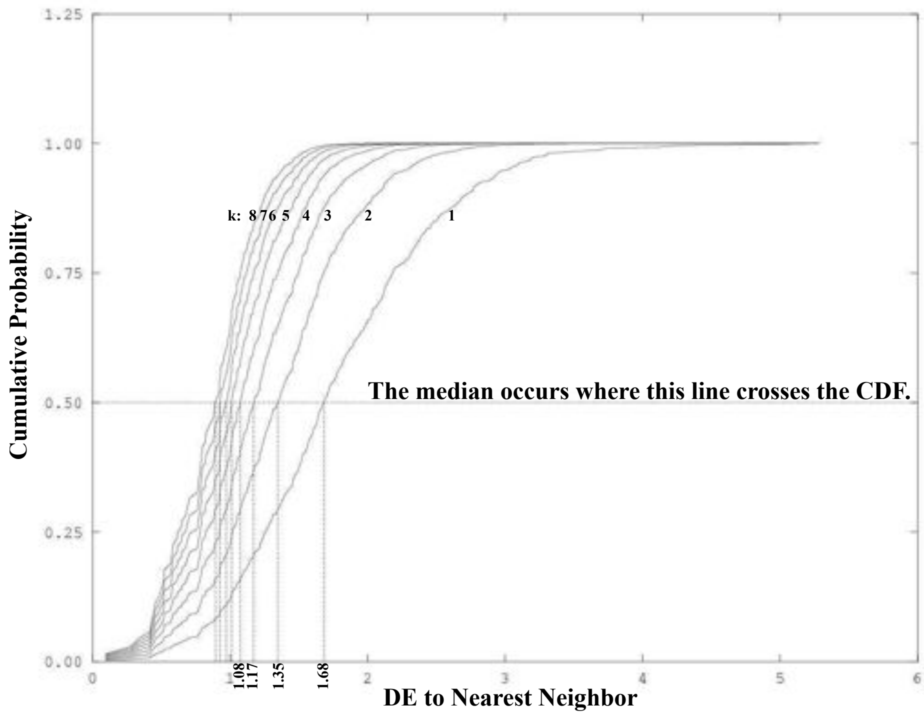


Figure 18: CDFs of DEs to Nearest Neighbors in a Set of 4096k RGBs

k	Colours in Shade Bank	Median
1	4,096	1.68
2	8,192	1.35
3	12,288	1.17
4	16,384	1.08
5	20,480	1.01
6	24,576	0.97
7	28,672	0.93
8	32,768	0.89

Table 1: Medians for Combined Shade Banks

reduction visually: the plots become closer together, and almost indistinguishable, as k increases.

These results are relevant because nearest neighbor estimation confounds colour measurement errors and interpolation errors in. Even if the shade bank contained large, random measurement errors, the histogram in Figure 16 would not change appreciably, because the error-laden shade bank would still cover the printer gamut just as closely. Likewise, even if the shade bank contained only perfect measurements, Figure 16 would not change. Any improvements in measurement accuracy will therefore not be noticed when analyzing the results of NN interpolation.

One reason for this situation is that NN interpolation does not make testable predictions; the effects of measurement error can therefore not be observed. We will see later that linear interpolation, which usually produces an RGB that is not in the shade bank, can be tested—simply print and measure the new RGB to see how closely it matches the target colour.

3.2 Empirical Testing of Effects of Measurement Accuracy

The issues just described seemed to occur as the reproduction algorithm was being developed. A tetrahedral tessellation was found for a large shade bank, and target RGBs were estimated using linear interpolation, but the DEs did not seem much better than would have resulted from NN interpolation. The colour measurements at that time were the mean values of three spectrophotometric measurements of each patch, and it was suspected that this method's results were too variable. This section analyzes a similar situation, using the new measurement protocol rather than the mean of three measurements, and shows that linear interpolation gave DEs about 60% smaller than NN DEs, a clear improvement.

To begin with, the MCDMs for the last section's shade bank of 4096 colours were calculated and plotted as in Figure 13. About 20 cases were found like those in Figures 12, 14, and 15, in which a page had to be remeasured. After all the remeasurements, the median MCDM was about 0.32. Since 20 page measurements out of about 207 had to be redone, and since the initial set was large and comprehensive, it might be taken as a rule of thumb that about 10 percent of page measurements will result in outliers like measurement 2 in Figures 12, 14, and 15. Without careful checking, then, a substantial proportion of page measurements by the i1iO2 might turn out to be unreliable.

To compare the effects of measurement accuracy on different kinds of interpolation, the results of linear interpolation was also investigated. In this case, a pre-existing data set was used. The set consisted of 433 target colours from artist's pastels manufactured by Sennelier. These 433 pastels, chosen from the complete set of 525, were known to be within the printer's gamut. Furthermore, they spanned a wide array of hues, values, and chromas, so they filled the gamut fairly well. The shade bank of 4096 printed colours was tessellated tetrahedrally in CIE Lab space, with respect to Illuminant C and the 2° standard observer. Each target colour was similarly expressed in Lab space, and the unique tetrahedron containing it was found, along with the barycentric coordinates of the target in terms of the four vertices, whose RGBs were known. The estimated RGB was a linear combination of the vertices, using the barycentric coordinates as coefficients.

The estimated RGBs for the targets were then printed out and measured, using the protocol described in this article. Ideally, the printed colours should match the target colours.

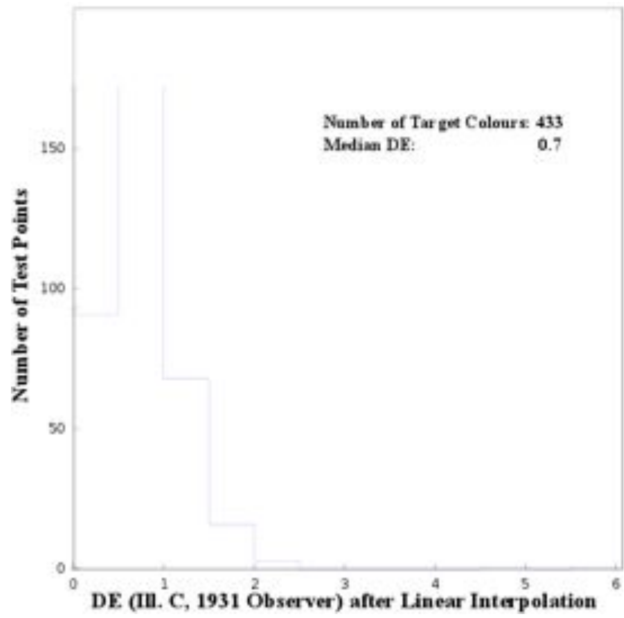


Figure 19: DEs to Linear Interpolants over a Set of 4096 RGBs

Figure 19 is a histogram of the 433 DEs between the printed colours and the targets. We can see that the median DE is about 0.7, which is a 60 percent reduction from the median DE of 1.7 achieved by nearest neighbor interpolation.

Such an improvement is to be expected—as long as colour measurements are sufficiently accurate. The proposed protocol was designed to provide such accuracy. When measurements are made without such a protocol, linear interpolation might prove to be not much better than guessing. An example was an early stage of the reproduction algorithm. At that time, an initial set of 8000 uniformly distributed RGBs were used, about twice as many as were used in the most recent example. The measurement protocol, however, was a single measurement per patch, using the X-Rite ColorMunki Design. While MCDMs for the ColorMunki are about on par with MCDMs for the i1Pro2, there was no combination of multiple measurements, and no attempt to eliminate outliers. The Munsell renotation provided the set of target colours. About half the target colours were outside the printer gamut; linear interpolation was used to match the other half.

The median DE was about 1.2. The current median DE of 0.7 is 40 percent smaller, and was achieved with a data set of half the size. The main difference in the two cases is the measuring protocol, which provides more accurate estimates of printed colours. The median DEs of 0.7 and 1.2 both occur when linear interpolation is used. Table 1 shows that a shade bank of 8000 points would lead to a median NN DE of just over 1.35. Since the median DE for linear interpolation was 1.2, there is not much advantage to using linear over nearest-neighbor interpolation—as long as measurements are too variable. Reducing the measurement variability, however, makes linear interpolation a definite improvement.

In summary, then, we have seen a practical case in which the suggested measurement protocol significantly improved a reproduction algorithm's accuracy, while simultaneously reducing the size of its shade bank. Some analysis has shown that this improvement occurs because the algorithm uses linear interpolation; a similar improvement would not be expected with nearest neighbor interpolation.

4 Practical Implications

Colour measurement is an important step in colour management and reproduction. Inaccurate measurements can eventually result in questionable colour matches or inconsistent print quality. On the other hand, the examples in this article have shown that accurate measurements can reproduce colours more quickly and accurately, even with less data. In terms of results, the suggested protocol has much to recommend it. In terms of time and attention, however, the protocol has much to recommend against it. It requires three separate printouts, nine measurements, and possibly further analysis; furthermore, much commercial software requires input in a certain format, which the protocol does not provide.

One common use for a scanning table is printer calibration. A print shop might calibrate its printers every morning. A typical calibration procedure starts by printing out a standard test sheet, which the i1iO2 then measures. The measurements are automatically input to some printer software, which adjusts the printer to conform with its ICC profile. The examples in this article show that colour measurements in this procedure could occasionally be highly variable. In particular, about every tenth sheet could contain significantly many patch measurements that are very much in error. Similar errors and problems would arise when profiling a paper.

Unfortunately, there seems to be a dilemma: either make a single measurement, which is quick and easy, but occasionally very wrong, or make many measurements, using the protocol to get accurate results, but at a high cost in time and energy. This dilemma will likely limit the protocol to research or development, where time constraints are not a major concern. Algorithm development, such as in the examples in this article, is an obvious application where the protocol is worth its costs. A manufacturer developing profiles or printing equipment would likewise find the protocol worth the trouble. For many print shops, especially those with high volume and quick turnaround, however, the protocol could simply prove too burdensome for regular use. Nevertheless, an understanding of the limitations of one-shot measuring methods is itself helpful, if only to identify causes of unexpected colour problems. In some cases, a scaled-down version of the protocol might improve on current practice, without too much cost in time.

Since very inaccurate measurements are uncommon, but not rare, they likely indicate a "special cause," in the terminology of process or quality control. For example, they might only occur when equipment is overheated, or there are many vibrations in the environment. Ideally, special causes should be tracked down and eliminated. An immediate increase in measurement stability would result, as the measurement process is brought back into "control." Once in control likely only a few measurements, perhaps just one, would be needed, rather than nine, which would immediately simplify and shorten the protocol. Given the likely improvement, it would be of great interest to identify any special causes, whether they are in the equipment or the environment. Currently there are no indications of what causes could be at work, and any suggestions would be welcome. Until there is a better understanding, an elaborate protocol like the one presented here will likely prove useful.

