The Coefficient of Variation as a Measure of Spectrophotometric Repeatability

Paul Centore

Abstract

The mean colour difference from the mean (MCDM) is a standard measure of spectrophotometric repeatability. This paper proposes a supplementary measure: the coefficient of variation (CoV), which is the ratio of the standard deviation of the measured reflectances to their mean. The CoV is calculated from the same repeated sample measurements as the MCDM. Unlike the MCDM, the CoV depends only on physical quantities, and not on perceptual quantities; furthermore, a CoV is defined for each wavelength. This paper analyzes data from six different spectrophotometer-sample combinations. An important empirical result is that the CoV is nearly constant across wavelengths, except when the reflectance at a particular wavelength is less than about five percent, in which case measurement variability is dramatically greater. Since the MCDM tends to lose this fact through averaging, the CoV is recommended as an adjunct to the MCDM for spectrophotometer analysis and development. The CoV analysis also provides evidence that samples' surface geometry is a major factor in measurement variability.

1 Introduction

Spectrophotometric repeatability[1] means that repeated reflectance measurements of the same colour sample, made by the same spectrophotometer, using the same protocol and conditions, should agree, even if the measurements are at different times. In practice, of course, agreement is not perfect. A standard metric of repeatability is the mean colour difference from the mean (MCDM). 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. Choose an illuminant and observer[2] with which to calculate the mean spectrum's colorimetric coordinates. Similarly, calculate coordinates for each of the N original spectra. Then use an expression such as ΔE_{00} to 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. For colour practitioners, the MCDM provides an intuitive gauge of measurement variability in a particular situation.

While the MCDM is suitable for practical situations, this paper proposes the coefficient of variation (CoV) as a supplementary repeatability metric, that provides a finer understanding in more analytical situations. Like the MCDM, the CoV is calculated from repeated reflectance spectrum measurements of a particular colour sample. Unlike the MCDM, there is a CoV for each wavelength λ. The mean reflectance for λ is the average of the individual reflectance measurements at λ. In addition, the sample standard deviation can be found for the reflectances at λ. The CoV for a probability distribution is defined as its standard deviation, divided by its mean. We will use the following standard estimator[3] for the CoV at λ :

$$\mathrm{CoV}(\lambda) = \left(1 + \frac{1}{4N}\right) \frac{\text{sample standard deviation of reflectances at } \lambda}{\text{average reflectance at } \lambda}. \tag{1}$$

Leaving aside the technical distinction between an estimator and an underlying distribution, we will simply refer to Expression (1) as *the* CoV. For ease of understanding, this paper will express CoVs as percentages.

While the CoV gives a value at each wavelength, the MCDM gives one value for the entire visible spectrum. In effect, the MCDM, like the human visual system, integrates over all wavelengths to produce one answer. Integration can smooth over and obscure interesting features that occur on a smaller scale. This paper will give some empirical instances in which the CoV reveals some important structure that the MCDM obscures. The most important result is the dramatically increased variability of spectrophotometric measurements when reflectance is low, less than about 5%. Figure 1 shows an example that will be explained in detail later. The vertical axis is the mean reflectance, for individual wavelengths, of the N measured reflectance spectra, and the horizontal axis is the CoV. The striking horizontal spike at the bottom indicates that spectrophotometric reliability degrades significantly at wavelengths where reflectance is very low.

Each reflectance on the vertical axis refers to one particular wavelength, so one dot might give a CoV that occurs at, for example, 560 nm. A typical reflectance spectrum has low reflectances for some wavelengths and high reflectances for others, and different wavelengths can show different degrees of measurement variability. The MCDM produces a single variability assessment by averaging out the very different contributions of the various wavelengths. The CoV, on the other hand, considers wavelengths individually, and thus captures important information.

The CoV can also shed some light on the sources of measurement error. The magnitude of measurement error depends on the spectrophotometer, the samples being measured, and the measuring procedure. Empirical results will show that the CoV is largely wavelength-independent for a particular spectrophotometer-sample combination, except when a wavelength's reflectance is very low. The sources of error are therefore also presumably wavelength-independent. Geometric effects, in which a sample's surface structure causes a random sequence of absorptions and reflections, are thus plausible candidates. The paper will show mathematically that geometric explanations are consistent with the observed CoV behavior.

The empirical data used in this paper consists of measurements from six different spectrophotometer-sample combinations. Two spectrophotometers were used: an X-Rite ColorMunki and an X-Rite i1Pro2. The samples cover various media (artist's pastels, acrylic paints, inkjet prints) and substrates (artist's canvas, pastel paper, card stock). Each combination used somewhere between 24 and 65 samples, each of which was measured 10 or 12 times. Repeatability is typically classified as short-term (on the order of seconds or minutes), mid-term (on the order of hours or days), and long-term (on the order of weeks or

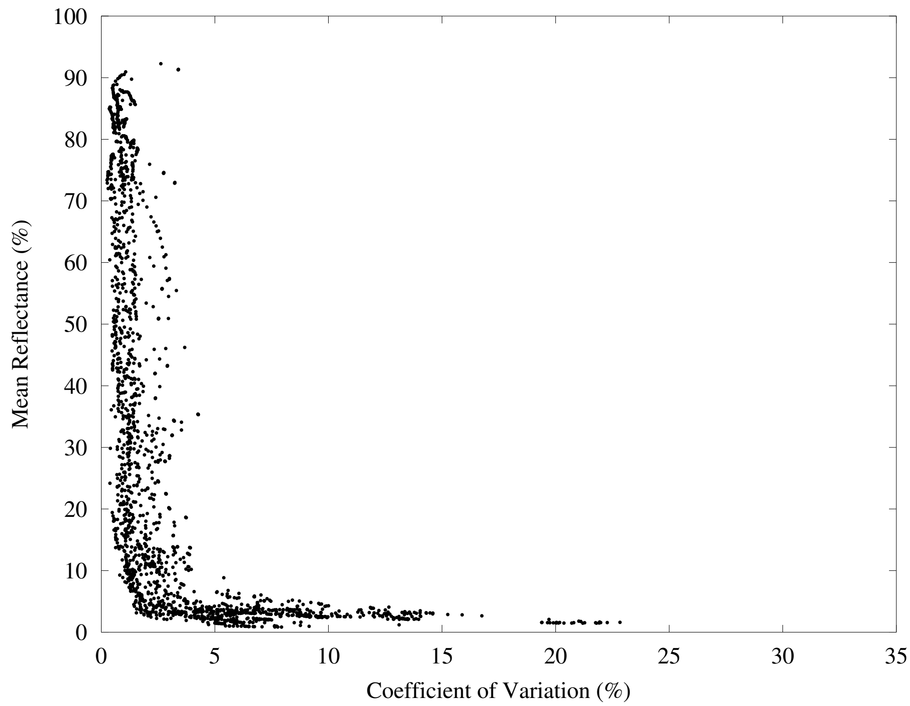

Figure 1: CoV vs Mean Reflectance for Case 4

longer). The data in this paper is all mid-term; the measurements for each combination were made within the span of a few hours. The mathematical analysis, however, would be identical regardless of term, and there is no reason to expect different mathematical behavior for different terms. Repeated measurements are also typically classified as *with replacement* (the sample and spectrophotometer are physically separated after each measurement, and reunited for the next) or *without replacement* (the sample is not physically separated from the spectrophotometer between measurements). All the measurements in this paper were with replacement.

The paper is organized as follows. First, the necessary mathematics behind the coefficient of variation is outlined. Next, the measurement experiments are described. Then, a visual examination of some experimental data is used to motivate the CoV. The CoVs are calculated for the data and plotted. An analysis of wavelength-level data leads to the paper's main result: the CoV indicates, much more clearly than the MCDM, that spectrophotometric measurements vary dramatically when reflectance is less than about 5%. Further wavelength-level analysis provides evidence for the hypothesis that samples' surface geometry causes much of the observed variability. The CoV and MCDM are then compared, and appropriate situations for the use of each measure are identified.

2 The Coefficient of Variation

The coefficient of variation (CoV) of a random variable is that variable's standard deviation, divided by its mean. The CoV is conveniently expressed as a percentage. It can be thought of as a quantity's expected variation, as a fraction of that quantity's average value. The CoV is useful when comparing quantities that differ by orders of magnitude. A classical example is comparing the length of mouse tails to the length of elephant tails: while the means and standard deviations differ greatly, the CoV could plausibly be the same for both. In the current paper, the random variable of interest will be the reflectance measurements of surface colours. Reflectances can typically vary by about two orders of magnitude, from lows just under 1% to highs that are over 90%. Despite this wide set of values, we will see that the CoV of reflectance measurements is nearly constant, except when reflectances are below about 5%.

The term *coefficient of variation*, like the term *standard deviation*, is often applied in two distinct contexts, population statistics and sample statistics, without making the distinction explicit. Population statistics assumes that a probability distribution of interest is known completely. If the distribution only takes on N discrete values x_i, then its population standard deviation σ_P is given by

$$\sigma_P = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \tag{2}$$

where μ is the mean value:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{3}$$

Unlike population statistics, sample statistics does not assume that a distribution is known. Instead, there is a set of samples, or measurements, from an underlying but unknown distribution. Summary quantities like the population standard deviation are still of interest, but can only be *estimated* from a sample. The sample standard deviation σ is given by

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}. \tag{4}$$

Apart from the denominator $N-1$ versus the denominator N, the expressions for the population and sample standard deviations are identical. The denominator adjustment insures that the expected value of the sample standard deviation is in fact the population standard deviation.[4] An estimator for a quantity, whose expected value is that quantity, is said to be *unbiased*, so the sample standard deviation is an unbiased estimator for the population standard deviation. The distinction between populations and samples is often not made, or only made implicitly. Frequently Equation (4) is referred to as *the* standard deviation, because, in most practical situations, only measured samples are available. In fact, however, Equation (4) is only an estimator.

Unlike the standard deviation and the mean, no general unbiased estimator is known for the coefficient of variation. One common estimator,[3] which will be used in this paper, is

$$\mathrm{CoV} = \left(1 + \frac{1}{4N}\right)\frac{\sigma}{\mu}. \tag{5}$$

This estimator is generally considered adequate when the underlying population follows a normal distribution.

In spectrophotometry, a practitioner is interested in the reflectance $\rho(\lambda)$ of a sample at a wavelength λ. While the reflectance itself is presumably an unvarying number, spectrophotometric measurement is a random process, so the reflectance measurements follow a probability distribution. A calibrated spectrophotometer is usually unbiased, meaning that the expected value of its measurements is the true reflectance. Furthermore, even though the measurements have some spread, they tend to cluster symmetrically around their expected value. These properties make a normal distribution a reasonable approximation to the distribution of reflectance measurements. Equation (5) is then a reasonable estimator of the population CoV, which is the rationale for using it here.

A naive estimation approach would simply have divided the sample standard deviation by the sample mean, giving

$$\mathrm{CoV} = \frac{\sigma}{\mu}. \tag{6}$$

Equation (6) is just Equation (5) with a coefficient of 1 instead of the coefficient $1 + 1/(4N)$. In practice, the adjustment of $1/(4N)$ is usually inconsequential. With ten measurements for example, $1/(4N)$ is only 0.025, so the coefficient in Equation (5) is 1.025 instead of 1, a difference of just 2.5 percent. As the number of measurements increases, $1/(4N)$ goes to 0 asymptotically, so its influence is even smaller. The choice between the estimators in Equations (5) and (6) is therefore not of much import.

3 Measurement Data

3.1 Experiments

This paper's conclusions about the CoV are based on reflectance measurements from six different combinations of spectrophotometers and samples. Only two combinations were measured specifically for this paper; the rest were measured in the course of other projects. In fact, the idea of using the CoV as a repeatability metric arose while calculating MCDMs from these measurements: a close look at wavelength-level data revealed some regularities that the MCDM was overlooking.

Table 1 lists the spectrophotometers and samples. In Cases 2 and 3, the X-Rite ColorMunki and X-Rite i1Pro2 measured the same chart. The two devices were alternated (first 24 measurements by the ColorMunki, then 24 by the i1Pro2, then 24 by the ColorMunki, and so on), to make the comparisons as similar as possible. Cases 4 and 5 used different selections of paint samples. In both cases, the paints were mixtures of Golden acrylic paints, along with a few paints straight from the tube. The data sets in general are from different projects, so contain different colours. In all six cases, however, the samples spanned a wide gamut of hues, values, and chromas, to avoid biases. Both spectrophotometers use a 45°/0° measuring geometry, but the i1Pro2 uses ring illumination while the ColorMunki uses unidirectional illumination. The i1Pro2 offers a choice of measurement conditions; in all cases, M2 (UV excluded filter) was selected. All the measurements were *with replacement*:

	Device	Description of Samples	Number of Samples	Measurements of Each Sample	Mean MCDM	Median CoV (%)
1	i1Pro2	Coated GoeGuide samples from Pantone fan deck	65	10	0.15	0.57
2	ColorMunki	ColorChecker reproduction, printed on Kirkland Signature paper	24	12	0.19	1.04
3	i1Pro2	ColorChecker reproduction, printed on Kirkland Signature paper	24	12	0.27	1.24
4	ColorMunki	Golden Acrylics, applied by handheld brush on Dick Blick canvas	56	10	0.34	2.03
5	i1Pro2	Golden Acrylics, applied by handheld brush on Dick Blick canvas	65	10	0.38	2.01
6	i1Pro2	Sennelier pastels, applied by hand on Somerset Radiant White paper	56	10	1.03	5.38

Table 1: Measurement Experiments

the spectrophotometer and sample were physically separated between measurements, and reunited for each measurement. The measurements for each case were all completed within the span of a few hours, so these are *mid-term* repeatability studies.

The second last column in Table 1 reports the mean MCDM. For any one case, each sample has a mean measured reflectance spectrum. The first case, for example, has 65 mean spectra, each an average of 10 individually measured spectra. Each sample in the first case produces 10 colour differences; each difference is between the mean spectrum and one measured spectrum. All colour differences were calculated with the ΔE_{00} expression,[2,5] assuming Illuminant C and the 1931 Standard Observer. The average of these 10 colour differences is the MCDM for that sample, so there are 65 MCDMs. The mean MCDM shown in the table is the average of the 65 MCDMs. Details of the CoV entries in the last column will be given later.

The MCDM entries can be compared with some previous work. Wyble and Rich[6,7] earlier attained mid-term MCDMs of between 0.10 and 0.27 (see Table II of Ref. 6) for four handheld spectrophotometers (the i1Pro2 and the ColorMunki are both handheld) when measuring disks of pressed polytetrafluoroethylene (PTFE). These disks are more similar to the first three cases in Table 1 than the last three cases. Since the first three cases found mean MCDMs of between 0.15 and 0.27, the results presented here are on par with their results, providing a sanity check.

3.2 Coefficient of Variation Calculations

The CoV was motivated by a visual examination of the measured reflectance spectra. As an example, Figure 2 shows all 10 reflectance measurements for Case 5, for the first of the 65 samples. The i1Pro2 returned reflectance measurements at wavelengths between 380 and 730 nm, in increments of 10 nm. Figure 2 shows the measured spectra as solid lines, while Figure 3 shows them more suggestively as isolated data points. Over the wavelength 600 nm, for example, there are 10 reflectance percentages, one for each spectrophotometric measurement of the sample. The spread in these ten percentages can be estimated visually

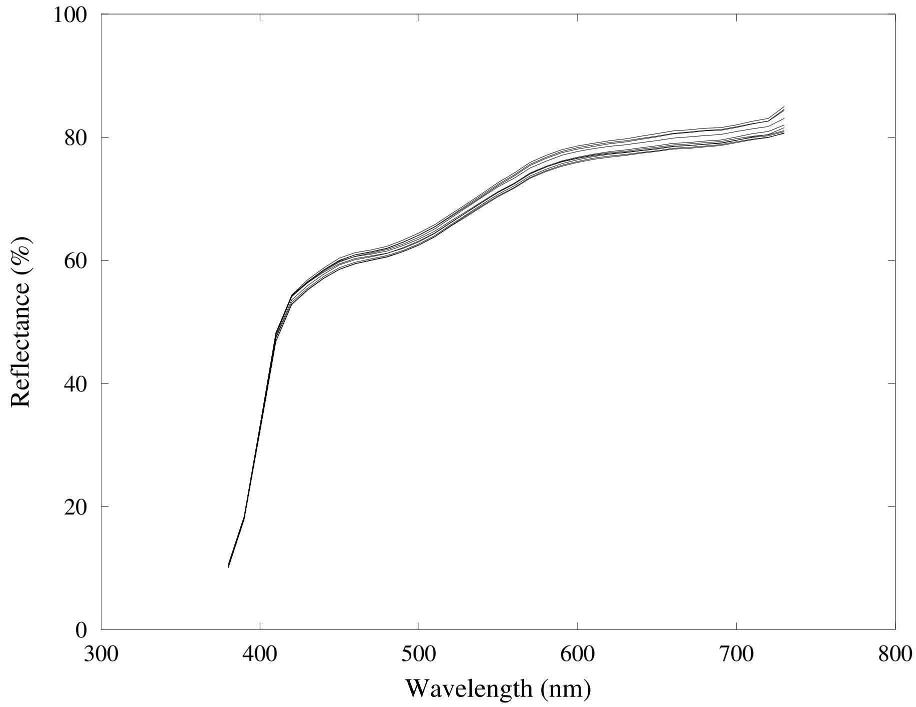

Figure 2: Measured Reflectance Spectra of First Sample in Case 5, as Lines

from the vertical extent of the data. In this example, the higher reflectances, at the right of the plot, have a greater vertical extent than the lower reflectances, at the left of the plot.

This observation motivates the CoV. The reflectance at a particular wavelength is best estimated by averaging the repeated reflectance measurements at that wavelength; these averages, of course, yield the mean reflectance spectrum. The spread of the reflectances at a particular wavelength is estimated by the sample standard deviation of those reflectances. Visually, it seems that higher reflectances show greater spread, so one naturally hypothesizes that the spread is a constant fraction of the mean. From the mathematical discussion, the CoV is just this fraction, i.e. the standard deviation divided by the mean. Figure 4 plots these CoVs as percentages, for the data in Figure 3.

With the exception of the endpoints (where measurements are likely unreliable), the CoVs are fairly consistent across wavelengths, at slightly over 1 percent. For many centrally clustering distributions, such as occur with repeated measurements, nearly all observed values are within three standard deviations of the mean. If the mean is 50 percent (as occurs at 410 nm), and the CoV is 1 percent, then the standard deviation is 0.5 percent (that is, 1 percent of 50 percent), so almost all the observed values would fall between 48.5 and 51.5 percent. With a mean of 80 percent (as occurs in the right half of the plot) and the same CoV, the observed values would fall between 77.6 and 82.4 percent. The total spread is 3 percentage points at 50 percent, and 5 percentage points at 80 percent. These predictions are consistent with Figure 3.

Once this example and many similar ones validated the CoV as a measure of spectrophotometric repeatability, CoVs were calculated for each sample, at each wavelength, in each of the six cases in Table 1. Figures 5 through 10 plot the CoVs. In each figure, all the

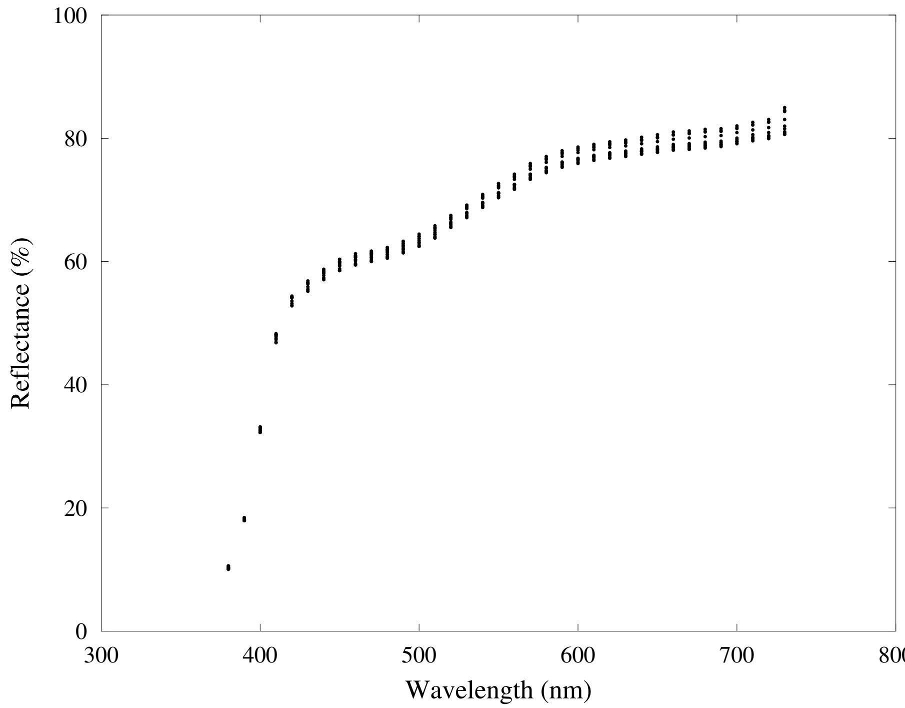

Figure 3: Measured Reflectance Spectra of First Sample in Case 5, as Points

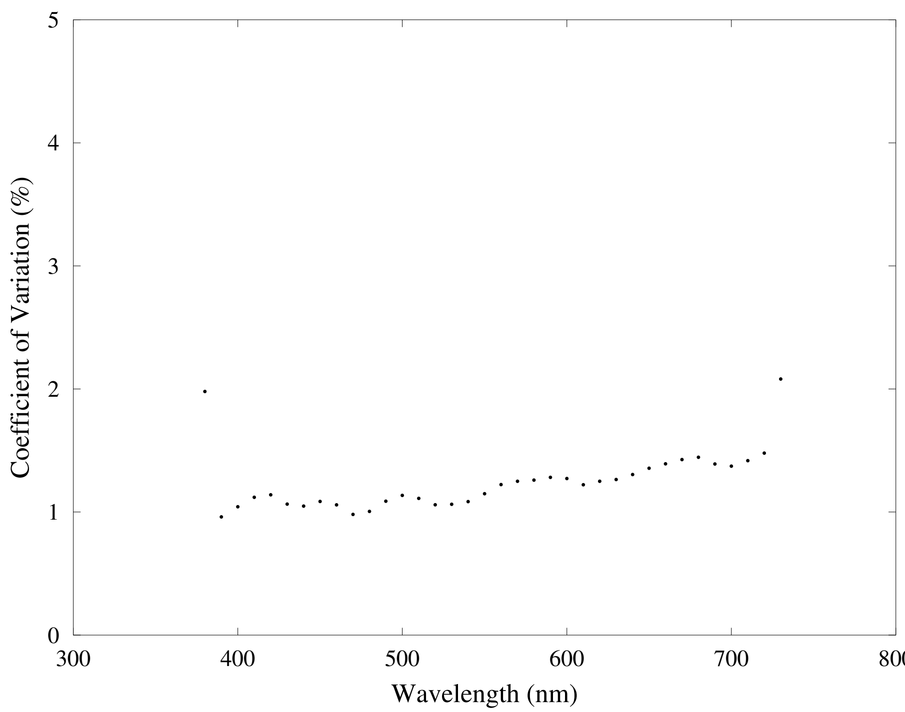

Figure 4: Coefficients of Variation for First Sample in Case 5

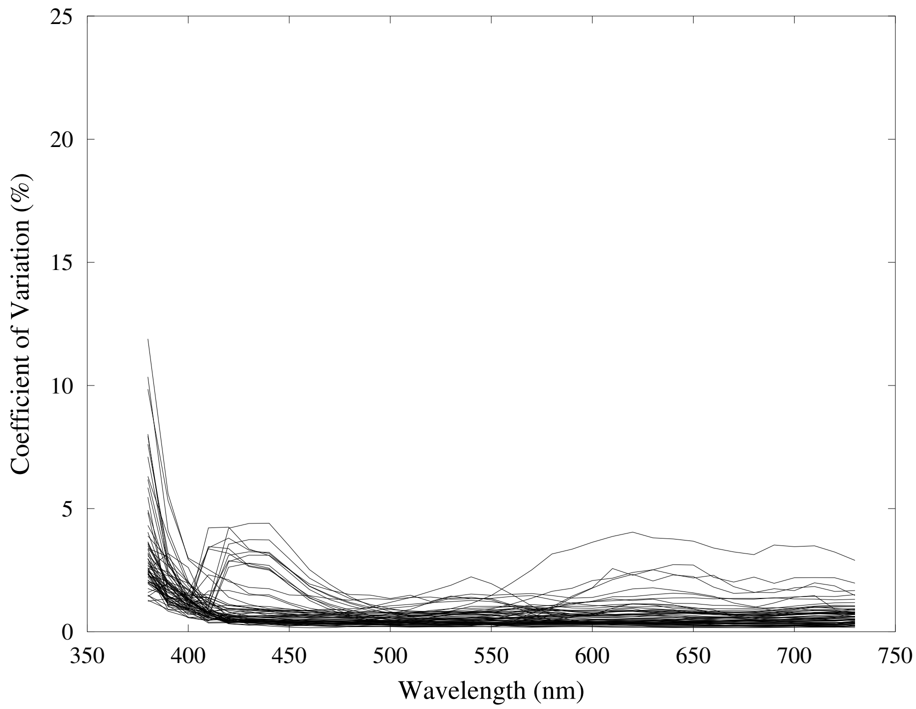

Figure 5: Coefficients of Variation for Case 1 (GOE Guide Measured with i1Pro2)

CoVs from one sample are plotted as a solid line that spans all the measured wavelengths. The total number of CoVs in each figure is the number of samples times the number of wavelengths. For each figure, the median of the set of all CoVs was calculated. This median appears in the last column of Table 1.

4 Analysis

4.1 Wavelength-Independence

The CoVs in Figures 5 through 10 tend to cluster. In Figure 8, for example, the bulk of the 56 samples cluster near 1%. Perhaps 10 to 15 samples are noticeably outside this clustering; without those samples, all the plots would form one thick line. (A later section will explain the samples outside the cluster in terms of low reflectance.) The important observation here is that, for any one case, the CoV only depends minimally on the wavelength. Given that, the median CoV in Table 1, even though it was calculated over data from all 36 wavelengths, would apply equally well to any one wavelength. In fact, the median was chosen over the mean because of robustness—the small percentage of outliers will not change the median's value, even if the outliers are very large. As a result, the median accurately captures the value which the CoVs cluster around.

Though the CoV does not depend significantly on wavelength, it does vary significantly from case to case. Table 1 shows that CoVs span an order of magnitude over the six cases. Since behavior at different wavelengths is not a factor, there must be other factors, and those

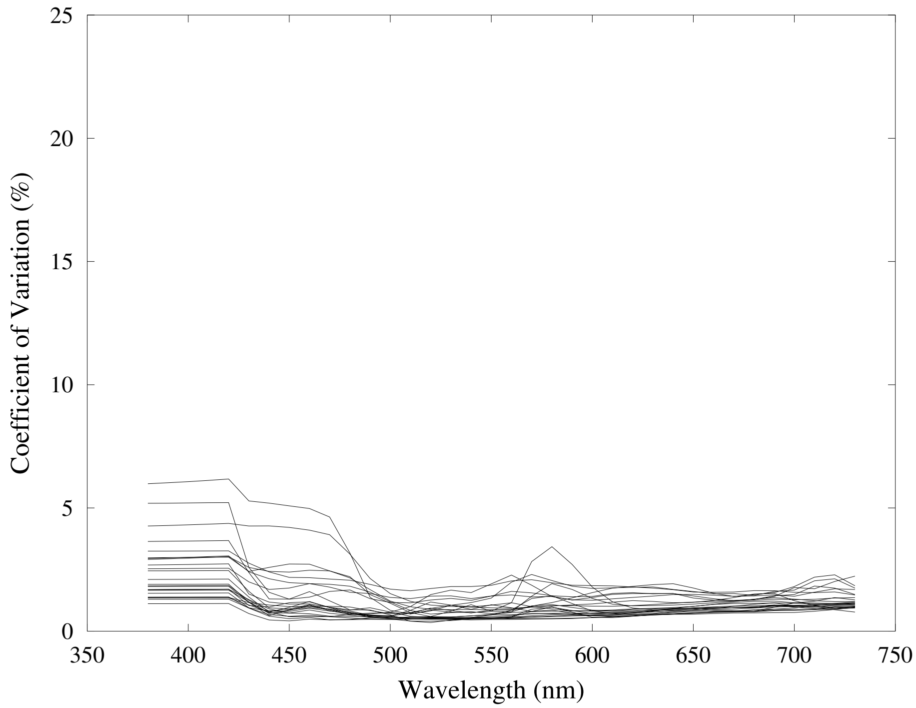

Figure 6: Coefficients of Variation for Case 2 (ColorChecker Measured with ColorMunki)

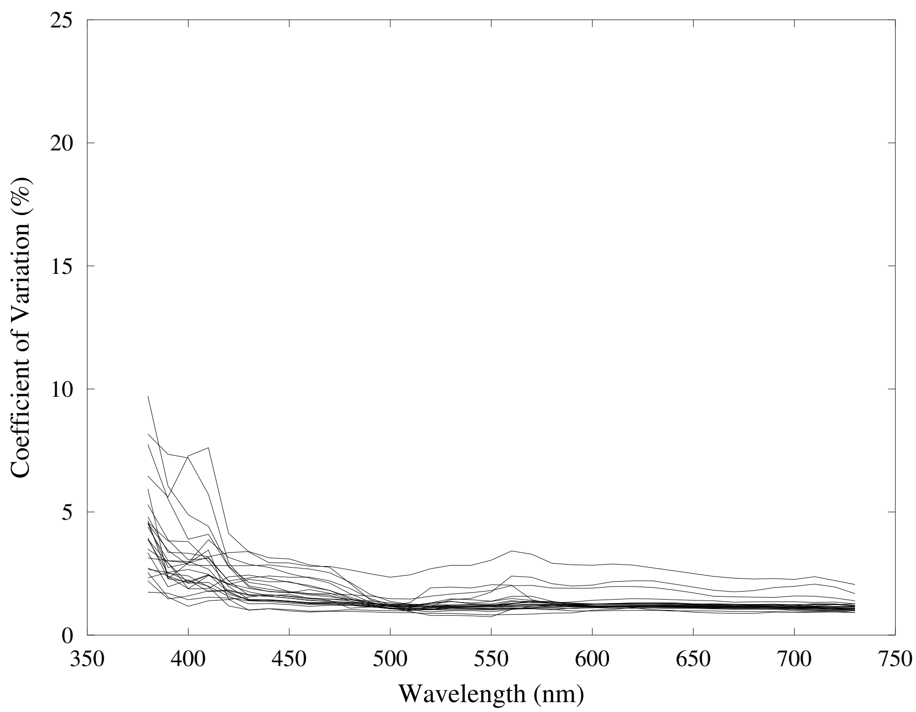

Figure 7: Coefficients of Variation for Case 3 (ColorChecker Measured with i1Pro2)

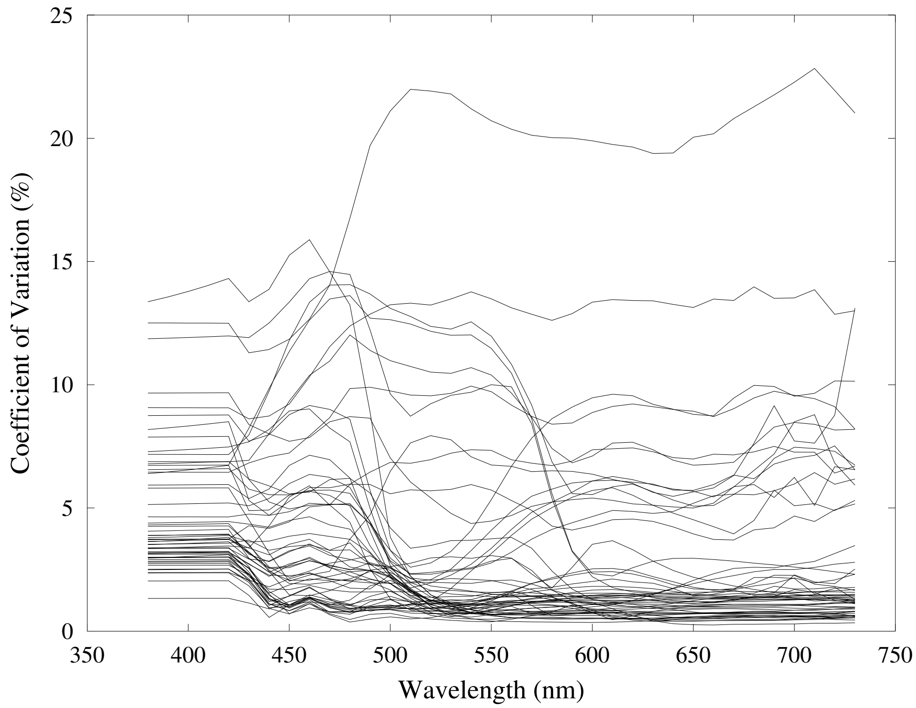

Figure 8: Coefficients of Variation for Case 4 (Acrylic Paints Measured with ColorMunki)

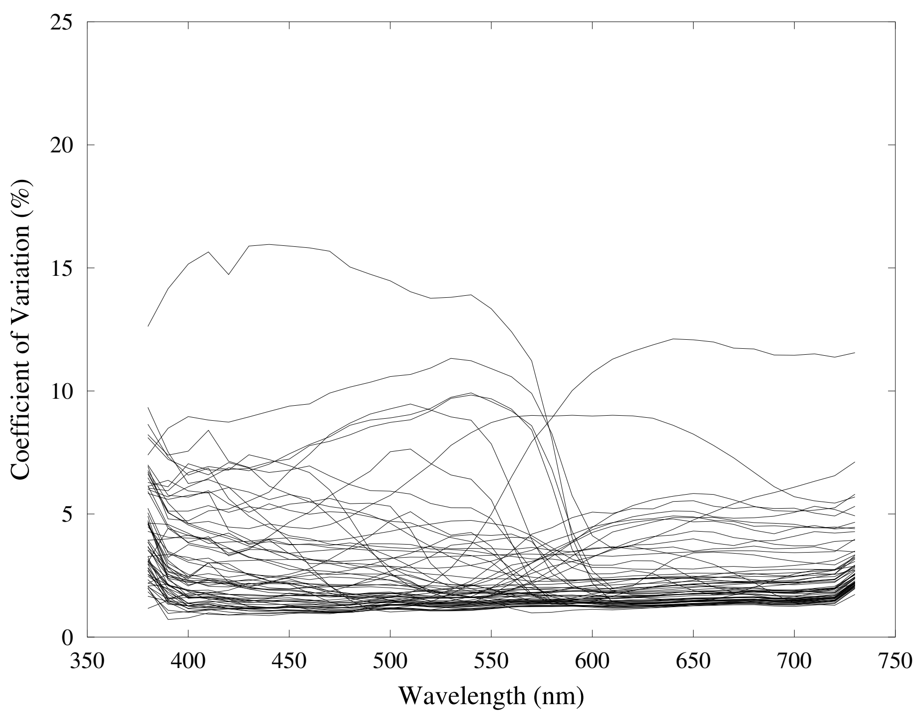

Figure 9: Coefficients of Variation for Case 5 (Acrylic Paints Measured with i1Pro2)

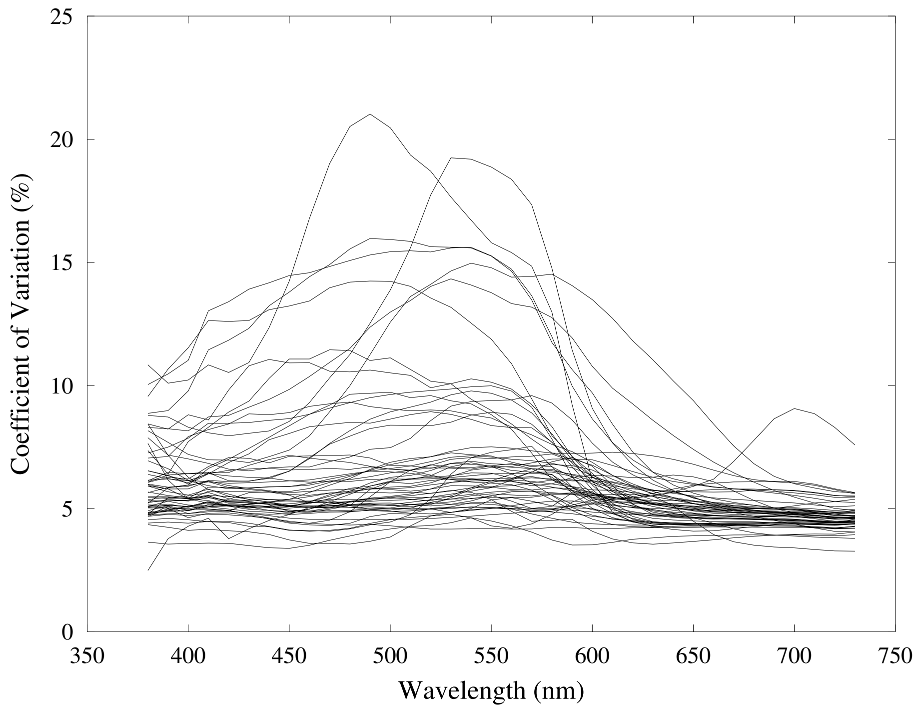

Figure 10: Coefficients of Variation for Case 6 (Sennelier Pastels Measured with i1Pro2)

other factors must be independent of wavelength. A later section will make the case that surface geometry is such a factor.

4.2 Low Reflectances

A closer look at the 10 to 15 outlying CoV curves in Figure 8 revealed an interesting feature: the high CoVs seen there are physically correlated with low reflectances. Figure 1 plots the CoV against the mean reflectance for all the samples and wavelengths in Figure 8. Each point refers to a particular sample and wavelength; for example, one point is the results for the 17$^{\text{th}}$ sample at wavelength 470 nm. There are 56 samples in Case 4, and each is measured at 36 wavelengths, so Figure 1 contains 2016 points in all. The figure shows that high variability occurs only when the reflectance is below about 5%, and that the variability increases dramatically there. Furthermore, the variability in CoV appears nearly constant when reflectances are between 5% and 90%. If reflectances below a few percent were removed from Figure 8, the plot would collapse to a fairly compact thick line. A low reflectance is therefore nearly a perfect predictor of a high CoV. In any event, the sudden spike in CoV seen in Figure 1 as reflectances approach 0 is too striking to ignore.

Perceptually, low reflectances correspond to dark colours. In fact, the prominent upper line in Figure 8, with CoVs hovering around 20 percent, is a very dark blue, almost black. The other isolated lines in the figure are also dark colours. This relationship explains a discrepancy in the figures: Figures 5 through 7 show few isolated lines, while isolated lines are plentiful in Figures 8 through 10. The last three figures consisted of artist's pastels and artist's acrylic paints, and presented a wide range of colours, including some very dark

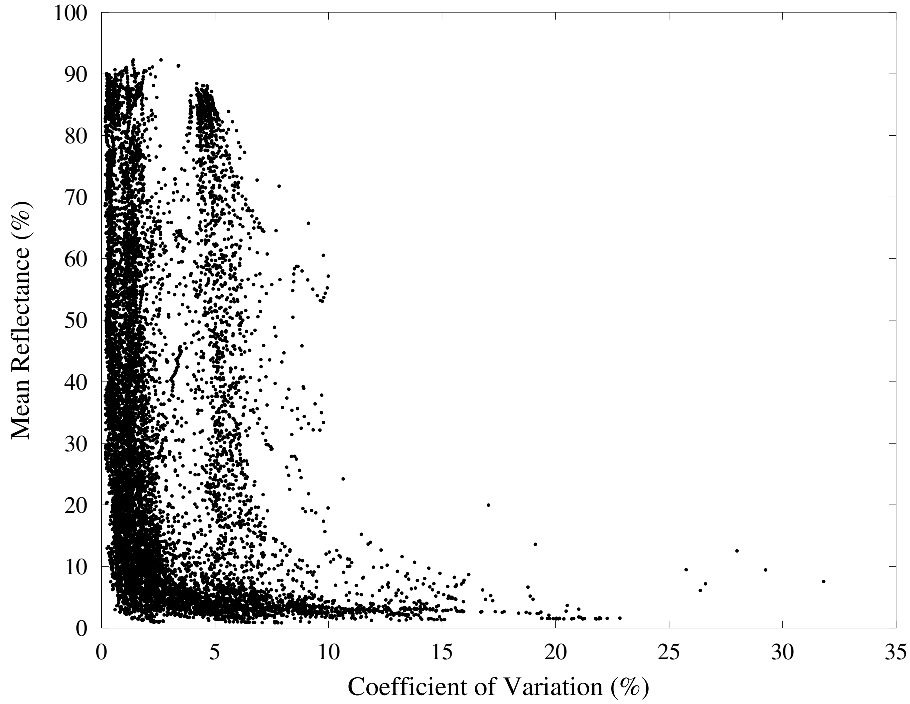

Figure 11: CoV vs Mean Reflectance for All Cases

colours. The second and third figures consist of measurements of a ColorChecker reproduction, which contains some darker colours, but not the extreme darks seen in the last three cases. Finally, the selection of colours from the GoeGuide were taken from a "Summer" fashion palette, with a paucity of dark colours. Had the first three cases contained some very dark colours, then they would likely have also exhibited the CoV outliers seen in the last three cases.

To make sure that Case 4 is not an anomaly, Figure 1 was augmented to include a point for every sample, at every wavelength, in all six cases, for a total of 10,440 points. Figure 11 shows the result. Visually, the same prominent spike appears along the bottom, indicating that the correspondence of high CoVs and low reflectances occurred consistently.

A random additive noise factor partially explains the results in Figure 11, but other factors are needed, too. Suppose that the spectrophotometer was unbiased at each wavelength, and its measurement error was an unbiased normal distribution with a standard deviation σ. σ would be measured in percentage points, and, because it is additive, σ would be the same at all reflectances, high or low. Figure 12 shows the CoVs, as functions of reflectances, that would result if measurement variability was due to additive noise alone. Curves are drawn for σ's from 1% down to 0.01%. A horizontal line at 5% reflectance is shown, for comparison with Figure 11.

Additive noise by itself would cause a spike at low reflectances, as Figure 12 shows. If additive noise were the only error, then all the data in Figure 11 would fall along one of the curves in Figure 12. The strong vertical column in Figure 11, however, when the CoV is just over 5%, falsifies this possibility. While the curves in Figure 12 could explain much of the behavior seen in Figure 11, it is also clear that other factors must be at work.

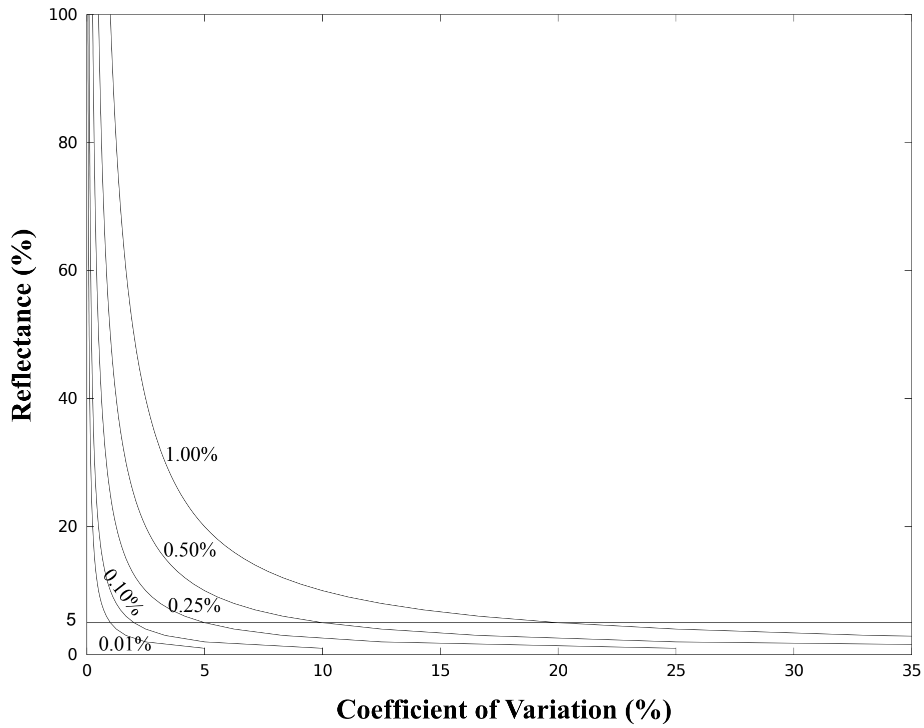

Figure 12: CoV vs Reflectance, When the Only Measurement Error is Additive Noise

4.3 Surface Geometry

The CoVs in Table 1 span an order of magnitude, from 0.57 to 5.38. Since CoVs are largely wavelength-independent, any explanation for this wide span must also be wavelength-independent. One natural explanation is the samples' texture or surface structure. The artist's pastels in Case 6, in particular, are very dusty. They were applied to the paper by rubbing the pastel over a square area, and then smoothing with a finger. A cloud of particles can easily be produced by shaking the samples or blowing on them. Glossy samples, such as the printed ColorChecker and coated GoeGuides, produced the lowest CoVs, while the slightly bumpy canvas texture produced intermediate CoVs. This section shows that surface geometry is a plausible explanation for the wide span of CoVs, and is consistent with the observed wavelength-independence.

Roughly speaking, spectrophotometric measurement error originates from three sources: the spectrophotometric device, the samples, or the measurement protocol. The six cases offer some controlled experiments, in which only one factor is varied. These experiments will allow us to conclude that the second source, sample irregularity, is causing most of the wide CoV span. The first source of error, the spectrophotometers, can easily be eliminated: Cases 1, 2, 5, and 6 all involve an i1Pro2 spectrophotometer (in fact, the very same physical device), yet yield very different CoVs. Since instrument instability would have made all the CoVs uniformly high, the instrument itself must be reasonably stable. A cross-model comparison is also avaiable. In Cases 2 and 3, the i1Pro2 and ColorMunki measured the same set of samples, and gave similar CoVs of 1.04 and 1.24. In Cases 4 and 5, the i1Pro2 and the ColorMunki again measured similar (though not identical) samples, giving CoVs of

2.01 and 2.03. Since these differences are not large, we can eliminate the first source of error. Since the measuring protocol was identical for all spectrophotometers and samples, the third source of error, measuring protocol, can also be eliminated.

The remaining source of error is then the samples themselves, and wavelength-independence suggests their geometric surface structure as a plausible explanation. The rationale is that an irregular surface leads to randomness in the number of times α that an incoming light ray interacts with the sample, and that this randomness explains CoV variability. At each interaction, some percentage p_A of light will be absorbed, and the rest will be scattered. The total percentage of absorbed light would then be a function not only of p_A, but also of the random variable α. If irregular geometry makes α highly variable, then the total absorption percentage is also highly variable. The percentage of light that is absorbed or scattered at each interaction is a function of wavelength, but α is not a function of wavelength.

The interaction sequence would be shorter for smooth surfaces and longer for uneven surfaces. A light source would interact only once with a perfectly smooth sample. Some light would be absorbed, and some would be scattered, either as a first-surface reflection component or as a diffuse, approximately Lambertian, component. The scattered light would not re-enter the sample, so there would only be a single interaction. The smooth surfaces in Cases 1 through 3 likely produce such short interaction sequences, and all produce lower CoVs. Now suppose that the sample's texture is not smooth. Perhaps, like the Torrance-Sparrow model,[8] it contains microscopic specular facets at various orientations. It could be even more complicated: given the pastel samples' dustiness, some pastel particles could be lying on the surface. With such uneven textures, scattered light could re-interact with the sample, likely multiple times. This long sequence would cause α to be high, and, more importantly, to be highly variable. As a result, CoV would be consistently high, as it is in Case 6. The interaction sequence would also be independent of wavelength, so the CoV would be the same at all wavelengths, as we observed.

4.4 Comparison of CoV and MCDM

The MCDM is already a standard measure for spectrophotometric repeatability, while the CoV is a proposed measure. This section compares and contrasts the two measures, and concludes that the CoV can be a helpful supplement to the MCDM, especially in analysis and research.

Both measures quantify the variability seen in spectrophotometric observations, and the two measures correlate well. The last two columns of Table 1 list the CoVs and MCDMs for the six cases studied. With the exception of the transposition of two nearly equal numbers, both quantities are listed in ascending order. The conclusion is that CoV tends to increase when MCDM increases, and vice versa. Either measure, then, gives an estimate of repeatability. The MCDM, however, uses the familiar DE scale, making it easy for a practitioner to grasp the result perceptually, and to estimate its effect in a certain situation. This ease of interpretation recommends the MCDM for practical situations

In scientific situations, on the other hand, the MCDM has some shortcomings. First, it conflates data from different wavelengths. Second, it requires a choice of illuminant. Third, it uses non-physical quantities such as a colour difference expression. From a practical point of view, of course, these shortcomings are strengths, because they account for perceptual

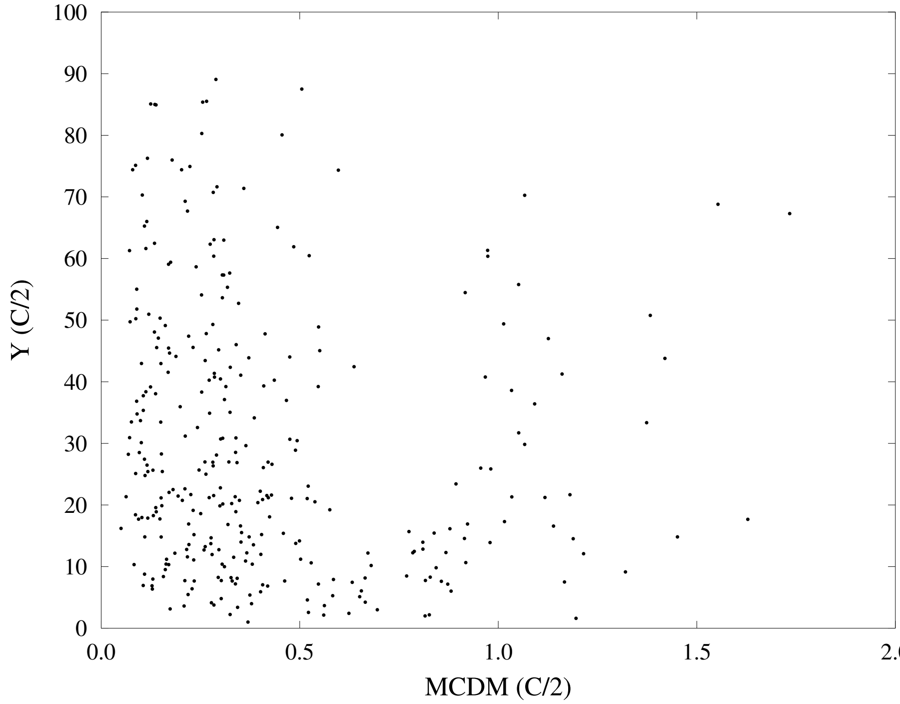

Figure 13: MCDM vs CIE Y for All Cases

factors: human vision also averages across wavelengths and assesses colour differences. The main disadvantage of accounting for perceptual factors is that valuable analytic and engineering knowledge is overlooked.

As an example, let us recalculate the plot in Figure 11, using perceptual quantities instead of physical quantities. In particular, we will replace the CoV (on the horizontal axis) with the MCDM. A perceptual correlate of reflectance (on the vertical axis) is CIE lightness Y, which is calculated from the photopic luminous efficiency function.[9] Y takes values from 0 to 100. A value of Y of 10 corresponds approximately to a Munsell value of 1, which would result from an ideal grey whose reflectance is 1% at every visible wavelength. Y was calculated for all 290 samples in the six cases. Each sample also had an associated MCDM. Both Y and the MCDM require an illuminant and an observer.[2] For consistency, Illuminant C and the 1931 2° standard observer were used for all calculations. The resulting scatterplot in Figure 13 is a perceptual analogue of the physical data in Figure 11.

While Figure 13 lacks the obvious spike seen at the bottom of Figure 11, the same relationship is present in a muted form. Most of the MCDMs above 0.5 occur when Y is less than 20 percent, which corresponds to a reflectance well below 5 percent, and only a few occur when Y is above 50 percent, which corresponds to a reflectance just below 20 percent. Figures 11 and 13 therefore exhibit the same trend, though it is not as pronounced in Figure 13.

Without the benefit of Figure 11, though, the trend in Figure 13 would likely be missed. While sophisticated statistical tests might reveal that some relationship is present, likely no engineer or analyst would think to apply such tests, because they would not observe the phenomenon being tested for. Even if a statistically significant relationship were found, it would likely be dismissed as unimportant. For engineers in particular, the CoVs in Figure 11, even without any statistical analysis, make it clear that development effort should be directed to low reflectances. Furthermore, test samples can be obtained easily, because the samples only need low reflectances at a few wavelengths; without this insight, much time could be wasted looking for very black samples that have low reflectances across the entire visible spectrum. Once samples are chosen, engineering improvements to spectrophotometers can be easily evaluated, using plots like Figure 11.

The CoV is also more useful than the MCDM when analyzing bi-directional reflectance distribution functions (BRDFs) for semi-reflective surfaces. Rather than measuring one overall reflectance curve, the BRDF measures a reflectance curve for each pair of directions. The first direction in the pair specifies the ray (in three-dimensional space) followed by light impinging on the sample, and the second direction specifies the ray along which reflected light leaves the sample. The BRDF is useful for glossy or semi-reflective surfaces, where the exiting light is a mixture of spectrally reflected incoming light (which could be of any colour) and the surface's local colour. For a pair of directions in which gloss is a factor, the MCDM can be determined but is not very informative because it is not clear what the "true" colour should be. The CoV, on the other hand, which only looks at individual wavelengths, does not require a true colour to estimate measurement variability.

A major analytic benefit of the CoV over the MCDM, then, is that the CoV provides finer physical information about the measurement process, and the factors that influence spectrophotometric variability. For this reason, analysis and research should use not just the MCDM, but also the CoV.

5 Summary

This paper has proposed the coefficient of variation (CoV) as a supplementary measure of spectrophotometric repeatability, at least for analysis and development; the long-established mean colour difference from the mean (MCDM) is likely more suitable for practical applications. A set of 290 colours, of various media and substrates, measured with two different spectrophotometers, provided data on which the paper's conclusions are based.

A major empirical result was that the CoV for a particular kind of sample and spectrophotometer was nearly constant across all wavelengths, except when the reflectance at that wavelength when less than about 5%. When reflectances were very low, the CoV increased dramatically. Furthermore, dramatic increases were always associated with low reflectances, suggesting that we have isolated the problematic conditions for measurement variability. While a particular spectrophotometer-sample combination would have the same CoV for all wavelengths, a different combination could have a different CoV; the CoVs in the data set, in fact, spanned an order of magnitude. The samples ranged from glossy, printed colours to dusty artist's pastels, and the more uneven the surface, the higher the CoV. From this correlation, combined with the CoV's wavelength-independence, it was inferred that samples' surface geometric structure explained much of the observed spectrophotometric measurement variability.

References

1. Roy S. Berns, *Billmeyer and Saltzman's Principles of Color Technology, 3rd ed.*, John Wiley & Sons, 2000.
2. CIE, *Colorimetry, 3rd ed.*, CIE Publication No. 15:2004, Vienna, 2004.
3. Robert R. Sokal, F. James Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research, 4th ed.*, W. H. Freeman and Company, 2012.
4. Dennis D. Wackerly, William Mendenhall III, & Richard L. Scheaffer, *Mathematical Statistics with Applications, 5th ed.*, Duxbury Press, 1996.
5. Gaurav Sharma, Wencheng Wu, & Edul N. Dalal, "The CIEDE2000 Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations," Color Research and Application, Vol. 30, Number 1, February 2005, pp. 21-30.
6. David R. Wyble & Danny C. Rich, "Evaluation of Methods for Verifying the Performance of Color-Measuring Instruments. Part I: Repeatability," Color Research & Application, Vol. 32, No. 3, pp. 166-175, June 2007.
7. David R. Wyble & Danny C. Rich, "Evaluation of Methods for Verifying the Performance of Color-Measuring Instruments. Part II: Inter-Instrument Reproducibility," Color Research & Application, Vol. 32, No. 3, pp. 176-194, June 2007.
8. K. E. Torrance & E. M. Sparrow, "Theory for Off-Specular Reflection From Roughened Surfaces," JOSA, Vol. 57, No. 9, pp. 1105-1114, September 1967.
9. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd ed.*, John Wiley & Sons, 1982.

