Measurement of Macro-uniformity: Streaks, Bands, Mottle and Chromatic Variations

D. René Rasmussen, Edul N. Dalal and Kristen M. Hoffman
Xerox Corporation, Webster, New York

Abstract

Macro-uniformity problems in an printed image are visible in the form of color variations which may be one-dimensional (streaks and bands) or two-dimensional (e.g., mottle). They may be limited to luminance variations or may include chromatic contributions as well. This paper examines issues related to quantification of the overall subjective impression of Macro-uniformity, and illustrates how several previously published measurements techniques can be combined to provide measures of overall Macro-uniformity.

Introduction

The overall image quality of output from color printers is influenced by many different factors. One method to describe the different aspects of color printer image quality is in terms of technology-independent attributes such as those used in the Xerox DAC system,[1] and those now being developed under the NCITS W1.1 workgroup on standards for perceptual image quality.[2] In this paper we will discuss issues related to measurement of one of the DAC color printer image quality attributes: Macro-uniformity. The W1.1 Macro-uniformity attribute has at the time of writing not been defined, and in the remainder of this paper we are exclusively referring to the *DAC* Macro-uniformity attribute.

The Macro-uniformity attribute

The ultimate purpose of all of the DAC attributes is to quantify quality in a manner that relates to how end-users judge the image quality of "real-world" images. Certain real-world images are more sensitive to one attribute than to others, for example, the real-world image in Figure 1b is quite sensitive to spatial color variations due to the large uniform background.

The Macro-uniformity attribute quantifies the *appearance* of spatial uniformity on a macroscopic scale. The attribute takes into account all spatial non-uniformities *except* those that are clearly perceptible when viewed through a 6mm-diameter aperture (those non-uniformities are characterized as *Micro*-Uniformity). The evaluation of Macro-uniformity uses several different test patterns, the most important being a letter-sized nominally uniform, light gray. Figure 2 gives examples of different types of Macro-

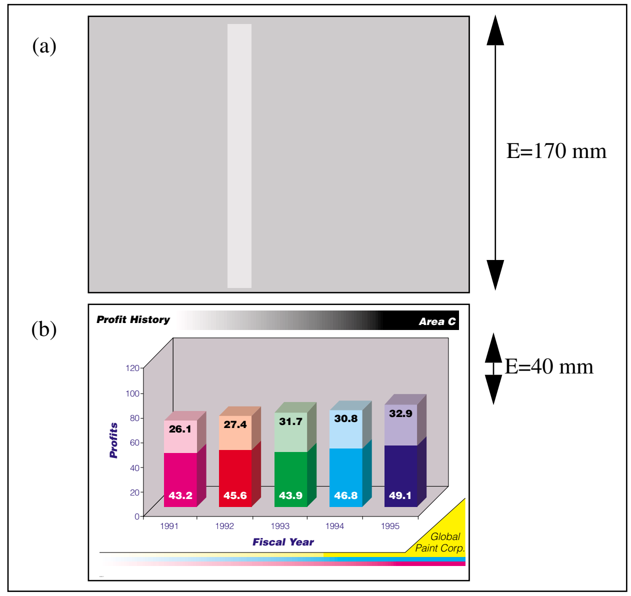

Figure 1 (a) test pattern of a letter-sized nominally uniform gray, showing schematically a single vertical streak defect. (b) A "real-world" letter-sized image, with nominally uniform background color.

uniformity defects on this test pattern, including many types of spatial lightness (L*) and chromatic (a*,b*) variations. Samples are rated by an "expert panel" by visual examination and comparison to a fixed scale of already rated samples. The visual examination takes place at a normal reading distance.

The key point is that only defects that appear as spatial variations in color are taken into account. A perceptible defect in a "real-world" image that is caused by a spatial color variation, but which does not give rise to the sensation of a spatial color variation, is not intended to be covered by the Macro-uniformity attribute. For example, a real-world pictorial image of a face might be very sensitive to the absolute colors used to render the face. That color is influenced by both the nominal color rendering of the printer system and by the spatial and temporal uniformity of the printer system's color rendering. However, even if the color of the face is rendered wrong due to a spatial nonuniformity across the image, the response of an observer will typically be a sensation of "wrong color" rather than "spatial uniformity problems", simply because the spatial nature of the problem is masked by the content of the image. The Macro-uniformity attribute does not address the accuracy of the color rendering. Thus, a slight hue change from one side of a page to the other would typically be judged as a small defect in terms of Macro-uniformity, but might be a more substantial problem in terms of color rendition.

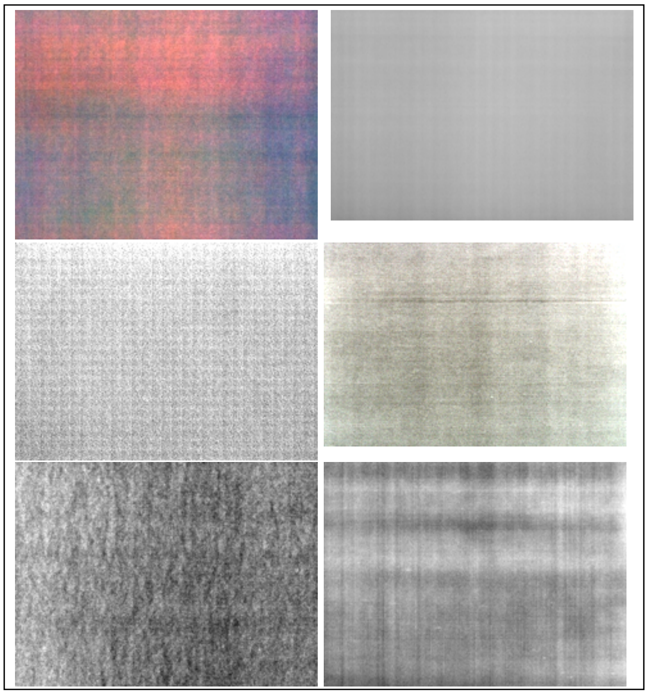

Figure 2 Examples of some of the letter-sized print samples used in this study. The contrast has been strongly enhanced to clearly show the defects. A wide range of "looks" are seen: chromatic, periodic, isolated streaks, random 2D, random streaks.

Later in this paper we will see that not only does the absolute objectionableness of a streak depend on the streak extent, E (see Figure 1), but the *relative* objectionableness of streaks with different characteristic spatial frequencies depends on E. This implies that a (hypothetical) measure that perfectly correlates with the objectionableness of vertical streaks on the image in Figure 1a, *must fail* to perfectly correlate with objectionableness of vertical streaks on the image in Figure 1b. Then what is the value of an attribute such as Macro-uniformity? There are two parts to the answer. Firstly, a test pattern such as in Figure 1a provides an upper bound on the stressfulness of real-world images. Secondly, the Macro-uniformity evaluation can be taken as a weighted average over several test patterns which together can predict the quality, not of a single real-world image, but of a population of different real-world images.

The advantages of measuring image quality in terms of an appearance based attribute, rather than exclusively by defect-specific metrics, are many. If market requirements are understood at the attribute level, then they can easily be translated to specifications for a specific (new) defect-type that may show up during product development. For example, some electrophotographic systems have a "ghosting" defect, that is, a weak, positive or negative version of a previously printed image that appears overlaid on the current image. When such a ghosting defect shows up, it may seem like a new, unique defect that calls for new, unique measurements, but in fact it affects the appearance of real-world images through a couple of high-level appearance attributes, most notably Macro-uniformity, and requirements for ghosting levels can be derived from Macro-uniformity requirements.

In summary, to define a procedure to assess Macro-uniformity of a printer, we need to specify 3 items:

- A set of images that will be evaluated.
- The scope of the attribute (i.e., appearance-characteristics of defects that are / are not taken into account).
- A defined set and sequence of printing test patterns, possibly including test patterns beyond those that will be evaluated, in order to provoke Macro-uniformity defects.

The discussion in the remainder of this paper is limited to assessments of single, nominally uniform images, such as the test pattern shown in Figure 1a.

From physical image to objectionableness

Much work has been published on the use of human visual system (HVS) models to predict "visual differences" and image quality.[6,7,8,10] The method we will use here,[3] consists of 2 steps: a HVS model that transforms the physical image into a "perceived" image, followed by a space-domain analysis that attempts to judge the objectionableness based on the perceived image. While the "perceived image" is reached

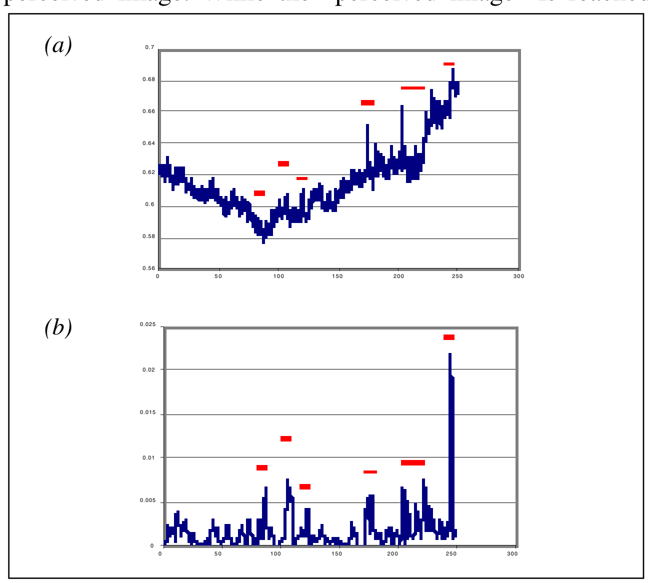

Figure 3 Horizontal profiles across print samples of test pattern Fig. 1a. The values are reflectivity relative to the average over the page. (a) The physical image (profile). (b) The perceived image (profile) further processed by calculating the square of the deviation from the average. The horizontal (red) marks indicate where human subject marked the presence of the most objectionable streaks on the print sample.

within the visual cortex, the interpretations and judgements that leads to a statement of objectionableness presumably take place outside the visual cortex, and involve consciousness about all the significant features (defects) in the image.

A simple HVS model[3,4] consists of applying a visual transfer function (VTF) to the physical data. This simple model can be extremely successful in filtering out all the visually insignificant information in the physical image, as illustrated in Figure 3. A single print sample of the test pattern in Figure 1a was recorded with a calibrated scanner, then for each horizontal position the photopic reflectance was averaged in the vertical direction, producing the profile in Figure 3a. Several observers were asked to look at the print sample and mark the location of the 5-10 most objectionable vertical streaks. The horizontal (red) marks on the figure show the consensus locations, and as expected there not an obvious relationship to the profile of photopic reflectance. Next the reflectance profile was modulated with a bandpass VTF, the mean value was subtracted, and the remaining signal squared. The result is in Figure 3b, showing a very good agreement between the signal and the visual assessment of objectionable streak locations.

Assuming for now that the HVS model is adequate, an important question remains: how do we go from the "perceived image" of Figure 3b, to a single number that correlates with the overall objectionableness? In general, the feasibility of this depends on the degree of variability of the "characteristic look" of the defects for which the correlation should hold. Table 1 defines four types of defect variability. A set of print samples falls under Type I if the defects of all the samples have (within a certain approximation) the same spatial pattern which varies only in amplitude. An example would be a single streak with a given width and shape, varying only in amplitude. In this case, a correlation between measurements and objectionableness is almost guaranteed, even if the HVS model is omitted. A set of Type II samples contains many different spatial patterns, but the *perceived images* can all be described by a single statistical distribution, and only the overall amplitude varies among the samples. An example would be samples characterized by 1/f noise[5] as long as the frequency range remains constant and the amplitude sufficiently high that a large number of streaks are perceptible. As in the case of Type I this leads to a single characteristic "look" and a straightforward correlation between objectionableness and *any* measure that tracks with the amplitude, even if the HVS model is omitted. Type III is characterized by containing samples that span many different "looks." The samples cannot be described statistically by a single distribution of defects, for example, one perceived image may be dominated by a single very wide streak, while another perceived image is dominated by multiple narrow streaks. In this case the HVS model is necessary, but not sufficient, unless it was expanded to account for the interpretation and judgement that can weigh different "looks" against each other.

The distinction between Types IIIa and IIIb is only a reflection of our (current) ability to predict objectionableness based on the perceived image. If a method exists to "integrate" the variation in the perceived image to a single numerical value that correlates with objectionableness, then we will say the sample set is Type IIIa, otherwise it is Type IIIb. Type IIIa implies that the defect variability has been reduced to a single dimension.

Table 1 Categories of defect variability

Type	Spatial pattern	Distribution	"Looks"	Dimensions
I	One	N/A	1	1
II	Many	One	1	1
IIIa	Many	Many	Many	1; Simple integration
IIIb	Many	Many	Many	> 1

The "tentpole function" for objectionableness

The sample set we are considering in this study is Type III, even if we consider only the 1-dimensional, horizontal L* variation, such as shown in Figure 3. In fact, the sample set was characterized by many samples where only a few (1-4) individual defects (say streaks) "stood out," and these individual defects had quite different spatial patterns.

Let us consider options for "defect-integration" of the perceived image in cases such as Figure 3 into a numerical objectionableness value. The graph shows that many distinct defects are perceptible, so how do they add up? The "tentpole effect" is often referred to in image quality, the extreme interpretation being, that when an observer judges image quality of a sample that has multiple defects, only the worst defect matters. Once the worst defect is removed, the second-worst is considered, and so forth. The characteristics of potential defect-integration methods are:

1. Should increase with defect amplitude;
2. Should capture the essential part of the tentpole effect, namely that the worst defect dominates;
3. The larger the number of defects in an image, the more objectionable it is (the pure tentpole effect violates this);
4. Given a localized defect, the objectionability should be relatively insensitive to the size of the surrounding (perfect) image.

Let *D'(x)* denote a one-dimensional profile representing the perceived image in cases such as shown in Figure 3, and let *D(x)* be the absolute value of the deviation from the average: $D(x) = |D'(x) - <D'>|$. We can interpret *D(x)* as the level of defect at location *x*.

The pure tentpole effect corresponds to using the maximum *Max{D(x)}*. This clearly does not satisfy all the requirements, since this value is not affected by the number of defects.

The average <*D*> does scale with the number of defects present, but it completely ignores the tentpole effect by assigning equal weight to all defects.

Other norms can be used to obtain compromises between the extremes of Max and Average. In general:

$$Obj = \left(\frac{1}{L} \int_0^L D(x)^p \, dx \right)^{1/p}$$

where $p=1$ corresponds to the average, and $p=\infty$ corresponds to the maximum. For values of p between 1 and ∞, all defects are taken into account and the worst defects will count with relatively higher weights, however, in cases where there are many identical, isolated defects, the integral is proportional to the number of defects, which is not compatible with the idea of the tentpole effect.

The method of "defect integration" we have used in this study is a generalized "tentpole function." The tentpole function requires that the continuous signal *D(X)* is first converted into a discreet, possibly infinite, series: d(i), where *d(1)* denotes the amplitude of the worst defect, *d(2)* the amplitude of the second-worst defect, and so forth. We define the tentpole function as

$$T = \sum_{i=1}^{N} \frac{d(i)}{p^{i-1}}$$

where *N* is the number of defects, and *p* is a tentpole parameter that must be larger than 1, and which defines how quickly the objectionableness, *T*, saturates when an increasing number of identical defects are present. In the limit of an infinite number of defects with equal amplitude, *d*, *T* converges to *dp/(p-1)*.

Although this function satisfies the four criteria stated above, there is of course no guarantee that it actually describes the proper defect integration. We are currently conducting psychophysical experiments to test this hypothesis and to determine the tentpole parameter *p* for the case of L^* streaks. We can expect this type of integration to be suitable when the perceived image is characterized by relatively few (say < < 100) distinct defects, but not in cases, such as graininess, where the visual impression is that of a single, spatially dispersed defect.

Determination of the VTF for L* streaks

We will now discuss the determination of the human VTF that was used for the analysis of L* streaks. The literature has plenty of data on human contrast sensitivity functions for spatial luminance variations,[4,6,7] and there is general agreement that the visual transfer function is a bandpass filter that peaks somewhere between 0.1 and 1 c/mm. Here and in the remainder of this paper we will describe spatial frequencies in cycles/mm, assuming a viewing distance of 0.40m. The details of the VTF depends however, on a number of conditions, including the average luminance level, the number of cycles that are visible and, as we shall see, on the extent (*E*, in Figure 1). Furthermore, most data in the literature concerns luminance variations close to the perceptibility threshold, while the current study also involves variations significantly above that threshold. For this reason we have here determined a VTF that is tuned to the viewing conditions and defect types that we wish to analyze.

Letter-sized print samples were created with controlled luminance variations around a light gray (L*=75), using an ink jet printer. Each sample contained a single vertical streak with a profile equal to a single period of a sinusoid. A schematic example is shown in Figure 4. Samples were created with widths of 1, 2, 4, 8, 16, 32, 64, and 125mm, and several different amplitudes for each width. The samples had an extent, E, of 170mm, and white masks were used to limit the visible extent to 20, 30, 45, 70, 106, and 170mm. 20 different observers rated the samples for quality on a scale from 1 (worst) to 10 (best) using anchor samples for the end points. The median of the 20 ratings was used as a measure of the sample quality. The samples were recorded with a drum scanner and the data calibrated to photopic reflectance.

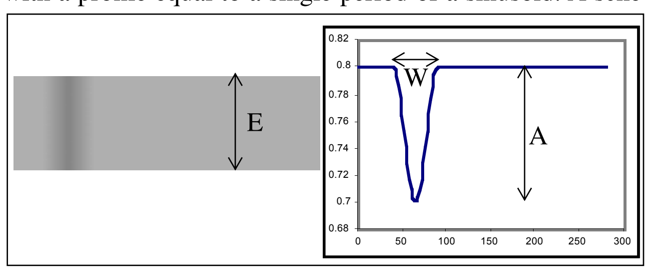

Figure 4 Schematic of a simulated streak.with width W, amplitude A, and extent E.

For each extent, *E*, a VTF was determined as follows. The luminance profiles were filtered with an initial VTF and the filtered amplitudes were plotted against the quality ratings. Typically this led to a set of different curves, one for each streak width, *W*, as shown in Figure 5a. Had the VTF been correct the curves would coincide. In an iterative manner the VTF was optimized to provide the best agreement between the curves. In this way we are guaranteed that at least for isolated streaks, the amplitude of the VTF-filtered signal is a measure of objectionableness, which is a requirement for being able to integrate multiple defects with the tentpole function. Two of the resulting VTFs are shown in Figure 6. The main conclusion is that perception is significantly affected by the extent, especially for low-frequency variations. For this study we considered letter-sized test patterns, therefore the VTF corresponding to the largest extent, *E*=170mm, was used for the subsequent analysis.

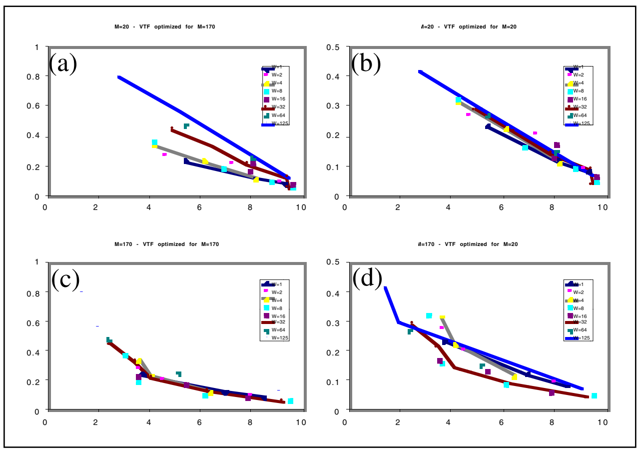

Figure 5 VTF-filtered amplitude versus experimental quality ratings in a matrix of two E values and two different VTFs. (a) and (b) are for E=20mm, while (c) and (d) are for E=170mm. (a) and (c) use the VTF optimized for E=170mm, while (b) and (d) use VTF optimized for E=20mm.

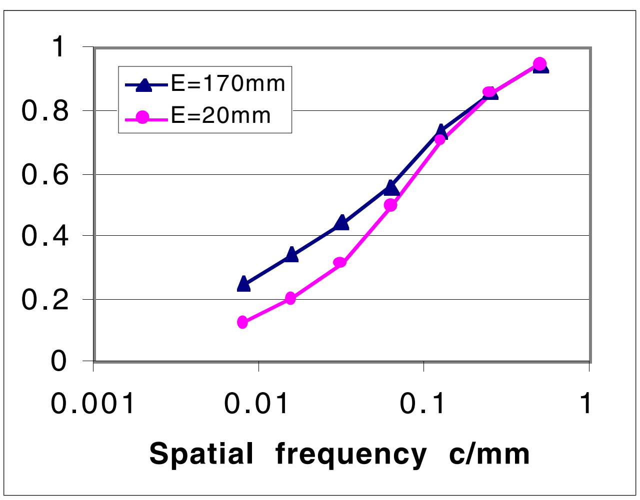

Figure 6 VTFs determined for two different extents, E=20mm and E=170mm.

Overall Macro Uniformity

Around 40 print samples were used to study the overall Macro-uniformity. The samples were letter-sized, nominally uniform, in most cases 20%CMY, and in some cases 20% K. The samples covered a wide range of printing technologies, including lithography, thermal and solid ink jet, electrophotography, and dye sublimation. A visual evaluation of Macro-uniformity was performed by an "expert panel", and the samples were recorded with a calibrated drum scanner.

The analysis of the images was separated into 3 parts: 1-dimensional horizontal or vertical luminance variations (e.g, streaks and bands), 2-dimensional luminance variations (e.g., mottle) and chromatic variations.

1-dimensional luminance variations

1-dimensional profiles of luminance variations were first filtered with the VTF as determined above. For an image with spatially isolated streaks the peak amplitudes could be used directly as input to a tentpole function, but in general the streaks may overlap as shown in Figure 7. To properly account for all such streaks the VTF-filtered profile is separated into 3 frequency channels, analogous to multichannel vision models.[9,10] The amplitudes of peaks in the resulting 3 profiles are taken as input to a tentpole function, with parameter $p=2$, and the result, T_1, is a measure of the objectionableness of the 1-D luminance variation.

2-dimensional luminance variations

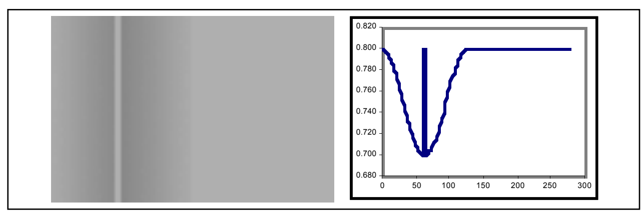

Figure 7 Streaks that are spatially overlapping but with widely separated frequency content must be counted as separate defects in the tentpole summation.

The 2-dimensional luminance variation is first filtered with a bandpass VTF, then separated into 3 frequency channels. The sum of the standard deviations of each channel was taken as a measure, T_2, of the overall objectionableness of the 2-D luminance variation.

Chromatic variation

Our perception of chromatic spatial variations is quite different from our perception of luminance variations. Most data indicate that the dominating effect is a low-pass filter.[11] We therefore chose to simplify this measurement to spectrophotometer measurements with a 3 mm aperture, which is roughly consistent with the perceptual frequency cutoff for chromatic variations. The measurements were performed on a 2mm x 2mm grid across the entire page. At each point the deviations Δa^* and Δb^* from the page-average were calculated, as well as the quantity $\Delta K = \sqrt{\Delta a^{*2} + \Delta b^{*2}}$. The root-mean-square of ΔK across the page is taken as a measure, T_3, of the chromatic variation.

Model of overall Macro-uniformity

The overall Macro-uniformity is modeled from the three components as:

$$M = \left(\sum_{i=1}^{3} a_i \tilde{T}_i^q \right)^{1/q}$$

where the power, q, and the coefficients, a_i, are fitted parameters.

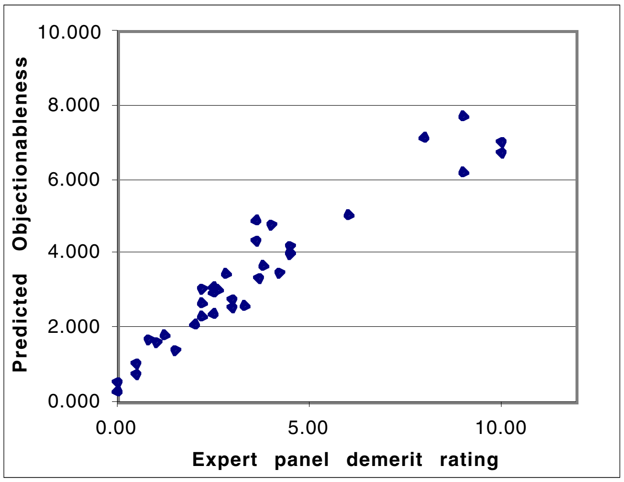

Figure 8 Predicted objectionableness versus the visual assessment by an expert panel.

eters. $\tilde{T}_i = \mathrm{Max}(0, T_i - t_i)$, where t_i is a threshold. The result of the best fit is shown in Figure 8. Future work will attempt to improve on this model. While we have good confidence in the assessment of 1D luminance variations, the separation into 1D, 2D, and chromatic variations may not be optimal. The main reason to separate between 1D and 2D variations is that for typical sample sets we often have a strong visual impression of 1D variations. From a fundamental point of view, however, the transition from 1D to 2D is not sharp. Similarly, there is no *fundamental* reason to separate purely chromatic variations from L* variations.

Acknowledgements

The authors would like to thank Michele Lochner for creating simulated images, and Franco Medina for scanning print samples.

References

1 E. N. Dalal, D. R. Rasmussen, F. Nakaya, P. A. Crean and M. Sato, Evaluating the Overall Image Quality of Hardcopy Output, Proc. PICS, pg. 169 (1998).
2 N. Burningham and E. Dalal, Proc. PICS, 2000.
3 D. R. Rasmussen, P. A. Crean, F. Nakaya, M. Sato and E. N. Dalal, Image Quality Metrics: Applications and Requirements, Proc. PICS, pg. 174 1998.
4 T. N. Cornsweet, Visual Perception, Academic Press, 1971.
5 H. Mizes, N. Goodman and P. Butterfield, The Perceptibility of Random Streaking, Proc. PICS, pg. 89, 2000.
6 J. Lubin, The Use of Psychophysical Data and Models in the Analysis of Display System Performance, in Digital Images and Human Vision, ed.A. B. Watson, MIT Press, 1993.
7 S. Daly, The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, in Digital Images and Human Vision, ed.A. B. Watson, MIT Press, 1993.
8 X. M. Zhang and B. A. Wandell, A spatial extension to CIELAB for digital color image reproduction, SID Symposium Proc., 1996.
9 B. A. Wandell, Foundations of Vision, Sinauer, 1995.
10 W. Wu, Z. Pizlo, and J. Allebach, Color Image Fidelity Assessor, PICS-2001
11 K. T. Mullen, The contrast sensitivity of human colour vision to red-green and blue-yellow, J. Physiology,

Biography

D. RENÉ RASMUSSEN received his Ph.D. in physics from the Niels Bohr Institute, Copenhagen, in 1990. He is Principal Scientist at Xerox Corporation and leads a project on development of Xerox internal standards for color image quality assessment. He has been with Xerox Corporation since 1992.

EDUL N. DALAL has a Ph.D. in Materials Science from the University of Utah. Since 1983 he has been employed by Xerox Corporation, where he is currently a research fellow and technical manager in the Wilson Center for Research and Technology in Webster, NY. His research interests include color science, image quality, polymer physics and surface science.

KRISTEN M. HOFFMAN received a B.S. degree in Imaging Science from the Rochester Institute of Technology in 1998. She is presently a Member of the Research and Technical Staff at Xerox Corporation, Office Systems Group. She has performed research in many areas of color materials, color modeling techniques, image quality and color image preference over the last ten years. This year the focus of her research has shifted to designing and setting image quality specifications for small office printing systems.