Using Metrics to Assess the ICC Perceptual Rendering Intent

Kristyn Falkenstern[a,b], Nicolas Bonnier[a], Marius Pedersen[c], Hans Brettel[b], Françoise Viénot[d,e]

[a]Océ Print Logic Technologies S.A., France;
[b]Institut TELECOM, TELECOM ParisTech, LTCI CNRS, France
[c]Gjøvik University College, Norway
[d]Muséum National d'Histoire Naturelle, France
[e]Centre de Recherche sur la Conservation des Collections, France

ABSTRACT

Increased interest in color management has resulted in more options for the user to choose between for their color management needs. We propose an evaluation process that uses metrics to assess the quality of ICC profiles, specifically for the perceptual rendering intent. The primary objective of the perceptual rendering intent, unlike the media-relative intent, is a preferred reproduction rather than an exact match. Profile vendors commonly quote a CIE ΔE^*_{ab} color difference to define the quality of a profile. With the perceptual rendering intent, this may or may not correlate to the preferred reproduction.

For this work we compiled a comprehensive list of quality aspects, used to evaluate the perceptual rendering intent of an ICC printer profile. The aspects are used as tools to individually judge the different qualities that define the overall strength of profiles. The proposed workflow uses metrics to assess each aspect and delivers a relative comparison between different printer profile options. The aim of the research is to improve the current methods used to evaluate a printer profile, while reducing the amount of time required.

Keywords: Color Management, Printer Profile, ICC, Perceptual Rendering Intent, Image Quality Metrics

1. INTRODUCTION

Color management systems are used in industry to obtain accurate and repeatable color transformation workflows. The specifications of an International Color Consortium (ICC) color managed workflow are well documented.[1,2,3,4] An ICC profile stores the Look-Up-Tables (LUT) that are used to transform the color data between the device color and the Profile Connection Space (PCS), an independent color space. The workflow is divided into two parts, one going forward from device color to the PCS (A2Bn) and the inverse transformation from the PCS to the device (B2An).

A successful reproduction may be different depending on the user's objectives. Common color reproduction goals may include: a spectral reproduction, an exact reproduction, an accurate reproduction, and a preferred reproduction.[5] These goals may not be possible with some workflows and they may not be mutually exclusive. The ICC has proposed different intents to handle these different goals. These rendering intents are not explicitly mathematically defined by the ICC,[6] which allows the software developer to define the algorithms that best fit each intent. This freedom results in profiles that may give large differences between different manufactures. Four different rendering intents are defined within the ICC profile format: absolute colorimetric, media-relative colorimetric, perceptual, and saturation.[1] The two colorimetric rendering intents (absolute and media-relative) are used to re-encode or re-target the image data while maintaining a current image state.[2] This work focuses exclusively on the perceptual rendering intent, which renders the colors to achieve a preferred color appearance within the reference or device constraints, also known as re-purposing. It is distinctly different from re-targeting, the reproduction goal is to produce the best possible reproduction. This intent is subjective, dependent on the user's goals, and may not be the closest possible color reproduction to the original.[2]

The assessment of an ICC profile is a complex issue that involves both color science and psychophysics. Profiles can be evaluated both subjectively and objectively.[7] They are commonly objectively evaluated using a

Send correspondence to Kristyn Falkenstern. E-mail: kristyn.falkenstern@oce.com

single metric test that calculates the average color difference between an original and reproduction of a digital target, CIE ΔE^*. A small color difference average is not adequate for the perceptual rendering intent, a smaller ΔE^* does not explicitly imply better or worse perceptual quality. Additional more complex metrics, Image Quality Metrics (IQMs), can be used to strengthen a profile evaluation by simulating the human visual system. IQMs often return a single value and a quality map. The quality assessment performed by an IQM can be full, reduced or no-reference.[8] The performance of an IQM can be measured through correlation statistics to subjective observer evaluations.

The main goal of this work is to create a thorough profile evaluation tool that assesses the performance of a set of profiles to help a user determine the most optimal profile for their printing objectives. Before discussing the proposed workflow, the paper starts with a short description on the creation of the profiles, Section 2. We propose a workflow that breaks the assessment down by aspect and finds a metric to evaluate each aspect separately. The workflow starts with the objective evaluation, in Section 3. This assessment uses a set of quality aspects and additional profile performance markers to summarize the behavior of each of the profiles, called the profile fact sheet. The aspects included in the objective assessment are: colorimetric accuracy, color primaries, smoothness, grayscale reproduction, black point compensation, gamut mapping, gamut volume and invertibility. Many of the aspects and the methods used to assess them come from previous work.[9] Next in the workflow is the subjective perceptual Image Quality (IQ) assessment, Section 4. The subjective testing is divided by preference aspects: overall, color, contrast, lightness, and sharpness.[10] A set of IQMs are chosen to test the performance of each of these subjective-aspects (sub-aspects). The sub-aspects are evaluated with a set of psychometric tests. Also reported are the correlation statistics used to assess the performance of the IQMs. The workflow delivers a relative comparison of the profiles, in Section 5. Finally, the conclusions and future work are in Section 6.

2. PROFILE CREATION AND PRINTING

The three ICC v2 printer profiles were created by printing a CMYK reference file with the Onyx ProductionHouse RIP-Queue x10 and measured with an X-Rite eye-one iO using GretagMacbeth's ProfileMaker 5.0.8 Professional Measure Tool. All of the profiles were applied in MathWorks MATLAB R2007a using the perceptual rendering intent and 16 bits per pixel image processing. The color managed CMYK files were printed with the Onyx RIP and a Canon image PROGRAF iPF700 CMYK inkjet printer on Océ premium coated matte paper (IJM113). All targets and test documents were printed using the same color management workflow.

3. OBJECTIVE ASPECTS - PROFILE FACT SHEET

The evaluation of a printer profile is specific to the user's needs and expectations. We have listed below the different aspects that can be used to objectively assess the quality of an ICC printer profile. The aspects are each defined, sections 3.1-3.9, with a common use, how it impacts the quality of a profile, the metric used to assess the aspects, and the results of the metrics. Table 1 gives an overview of the objective aspects and metrics used to assess them. A summary of the metric results (profile fact sheet) can be found at the end, Section 3.10. Some of the objective metrics are testing the profiles using the same target. The colorimetric accuracy and the colorfulness aspects both use a 625 patch target generated in GretagMacbeth's target generation software. The grayscale reproduction, Section 3.5, and both parts of the Black Point Compensation (BPC), Section 3.6, aspects use the same CIE L^* target, a 16 bit CIE LCh^* file that increments in the L^* channel from 0 to 100.

The objective aspects may have a direct relationship to perceived IQ and preference but not necessarily. The role of this set of metrics is to characterize the profiles and not determine if the reproductions will be more pleasing. Perceptual IQ is reserved for Section 4. The documents used in the assessment of the objective aspects are targets. With the exception of the round-trip test, none of the metrics require the targets to be printed, but with some tests a print may add to the profile assessment.

3.1 Colorimetric accuracy

Colorimetric accuracy describes how close, in terms of color, the reproduction is to the original under a given set of viewing conditions. It often returns the average color difference of a target between an original and a reproduction for a set of viewing conditions, commonly expressed with CIE ΔE^*_{ab}, ΔE^*_{94}, or ΔE^*_{00}. The

Table 1. Objective aspect overview: aspect, section, the details of the test used for the evaluation, and the metric.

Aspect	**Section**	**Test details**	**Metric or Tool**
Colorimetric accuracy	3.1	625 patch target round-trip test (in gamut)	CIE ΔE^*_{94}
6 Primaries	3.2	6 color primaries (out of gamut)	CIE ΔLCh^*
Colorfulness	3.3	625 patch target	Cui's colorfulness[11]
Smoothness	3.4	CIE ΔE^*_{94}, 30 targets	2^{nd} Derivative[12]
Grayscale reproduction	3.5	Deviation from the neutral axis	CIE $C^*\times\Delta h^*$
BPC shadow details	3.6	L^* target for values (≤ 20)	ΔL^*_{STDV}
BPC contrast	3.6	Slope of CIE L^* curve (≤ 20)	CIE ΔL^*
Gamut mapping	3.7	Out of gamut differences maintained	$\Delta LCh^*_{\text{STDV}}$[9]
Gamut volume	3.8	Size of the profile's gamut	ICC3D[13]
Invertibility	3.9	Software round-trip test, A2B0 LUT (1120 patches)	profileQA[14]

rendering intent is quite relevant for this aspect. As discussed, the vendors may aim for more pleasing colors rather than a close color match, a close colorimetric match may not result in the best looking reproduction. We have included this aspect as a way to characterize each of the three profiles. The metric describes whether or not the in gamut colors are re-rendered, indicated by the average color difference of the round-trip test.

The three profiles were applied to a 625 patch CIELAB target (original). The targets were printed, measured, and new targets were created from the measurement data (round 1). The color difference between the original and round 1 are reported in Table 2. The profiles were applied again, each to their corresponding CIELAB measurement values. The three new targets were printed and measured (round 2). The difference between the round 1 measured values and the round 2 measured values are calculated. The round-trip test ensures that the entire target is within the profile gamut.[7,15] The color difference average, standard deviation (STDV), minimum (min), and maximum (max), for round 1 and round 2 are reported in Table 2.

Table 2. Profile 1 has the largest difference for round 1 of the colorimetric accuracy aspect. With only colors that are in gamut, round 2, profile 1 has the smallest average. Profile 1 does the least amount of color re-rendering once the colors are within gamut. Profiles 2 and 3 both render the in gamut colors, indicated by the color differences in round 2.

Stats	**Round 1**			**Round 2**		
	Profile 1	Profile 2	Profile 3	Profile 1	Profile 2	Profile 3
Mean	5.01	4.88	4.86	0.74	3.84	3.98
STDV	1.56	1.59	1.43	0.30	1.25	1.09
max	11.74	11.06	11.77	1.77	9.40	9.23
min	0.47	0.17	0.84	0.04	0.09	0.65

Profile 1 has the largest color difference average after the first round. However, the small color difference average after the round-trip indicates that profile 1 preserves the in gamut colors more than the other vendors, after the initial mapping which brings them into the profile gamut, see Table 2.

3.2 6 color primaries

Whenever transitioning from a larger gamut to a smaller gamut, there are trade-offs. This aspect is used to characterize these trade-offs, if the profiles preserve one color channel over another. The six sRGB color primaries, (red, green, blue, cyan, magenta, and yellow) are used because of their high saturation, they will probably be out of gamut, and they represent several different hue angle regions. The red, green and blue primaries are likely to be the very most saturated colors with a printer profile color transformation. The primaries may also illustrate mapping trends at the different hue regions, if profiles intentionally shift the hues, which hues, and in which direction. For this aspect, we have looked at the CIE ΔLCh vectors for each of the primaries, see Figure 1.

The lightness changes are similar between the profiles, the largest differences occur with green and blue. Profile 3 has a larger change in lightness, except for green. For all profiles there are large chroma declines. Profile 3 has the largest chroma difference for all colors. Profiles 1 and 2 have similar differences, for profile 1

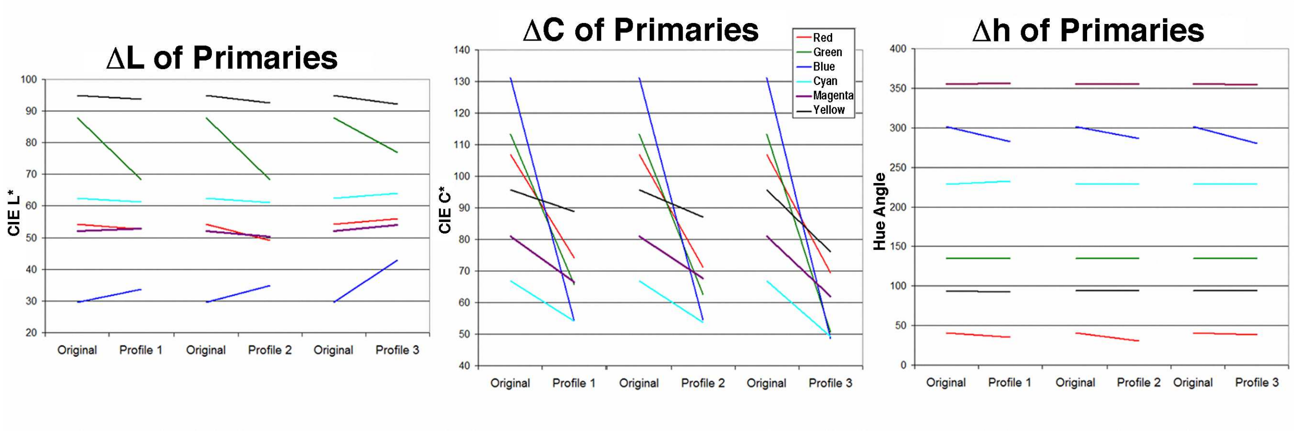

Figure 1. The 6 primaries: CIE ΔLCh vectors are plotted. The lightness differences are plotted on the left, chroma is in the center and hue to the right. The vectors are similar between the profiles for most colors. In general, profile 3 has the largest compression of lightness and loss of chroma. All three profiles have the greatest change in chroma. There are slight hue shifts, in the counter-clockwise direction, for red (profiles 1 and 2) and blue (all) and in the clockwise direction for cyan (profile 1).

has slightly smaller differences for most colors. The hues are mostly maintained, with the notable exception of red and blue. These colors have a slight shift in the counter-clockwise direction. Also cyan is slightly shifted in the clockwise direction with profile 1. The gamut mapping vectors, in general, are very similar. The profiles all loose more chroma than lightness or hue. The lightnesses are mapped towards a L* value of approximately 50. For most colors, the vectors from Profile 3 go towards the neutral axis more than the other profiles. Additional out of gamut colors are addressed in the gamut mapping aspect. A comparison of the vectors may be very useful to a user that is most interested the preservation of a particular region of the gamut, for example if the majority of their document is composed of a single color.

3.3 Colorfulness

Colorfulness is a visual sensation that indicates whether an area appears more or less chromatic. Colorfulness is open-ended with a zero origin and affected by luminance.[16] It is an aspect that is often described and quantified by CIE C^*. Cui's metric has been used to assess the colorfulness aspect, the metric uses both chroma and lightness for the assessment.[11]

The metric results in Table 5, show that profile 2 is the most colorful. Profiles 1 and 3 have similar levels of colorfulness. Again, this aspect characterizes the behaviors of the profiles with respect to colorfulness and does not assess perceived IQ or preference. If the observer's reproductive objectives include high saturation, profile 2 may be the best option.

3.4 Smoothness

Smoothness is perceived when transitioning between colors, if the increments are of equal color difference steps, the transition will be seen as smooth.[17] In IQ evaluations, smoothness often given a high priority and is considered a desirable aspect.[18] Many studies have evaluated the perceptability of color contouring for both hard and soft copy, and color transforms.[19,20] Artificial targets such as the Granger Rainbow have been used for this aspect. For this work thirty targets were created in the CIELAB color space. They increase in one dimension while keeping the others constant. The CIE ΔE^*_{94} color difference between consecutive steps is the same for all steps within each target.[9] There are four types of targets: lightness, hue, chroma, and cross axis. Figure 2, gives an example of a target and the metric used to evaluate the smoothness aspect. The 2^{nd} derivative metric was used to evaluate this aspect.[12] The final value returned was a sum of the number of points that exceeded the original CIE ΔE^*_{94} color difference that was used to create each target. The spike in the 2^{nd} derivative, may indicate the edge of the profile gamut.

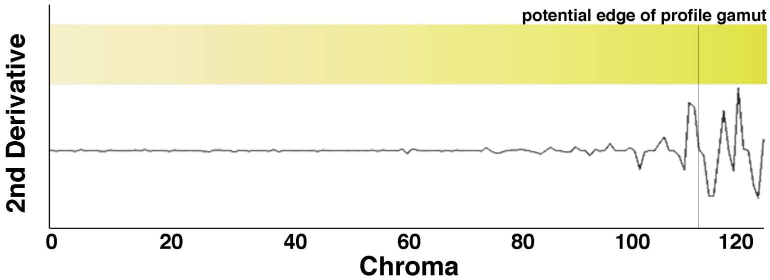

Figure 2. The smoothness aspect being assessed by the 2^{nd} derivative metric. The original target increases only in chroma, while keeping lightness and hue constant. The point where the 2^{nd} derivative spikes dramatically (the higher chroma values), may indicate the boundary of the profile.

A summary of the *z*-scores results of the targets are listed in Table 5. Profile 3 had the smallest or least frequent change in color differences between consecutive points for most targets. Profile 1 was the next smoothest. Profile 2 had larger 2^{nd} derivative values, the values exceeded the threshold more frequently and with higher values than both profiles 1 and 3.

3.5 Grayscale Reproduction

The grayscale reproduction aspect assesses how a system reproduces a set of neutral colors, which are relevant in both grayscale and color reproductions.[21] Inaccurate neutrals may cause a color cast on the overall impression of the document.[22] This aspect was tested by creating a CIELAB target that increases in lightness from 0 to 100, in steps of 1 ΔL^*. The profiles were applied to the target. The new CIELAB values are plotted in Figure 3. The metric used to assess the neutral axis considers only chroma and hue. The metric returns the product of the CIE $C^*\times\Delta h^*$, the hue difference is between consecutive values on the target. The metric results are summarized in Table 3.

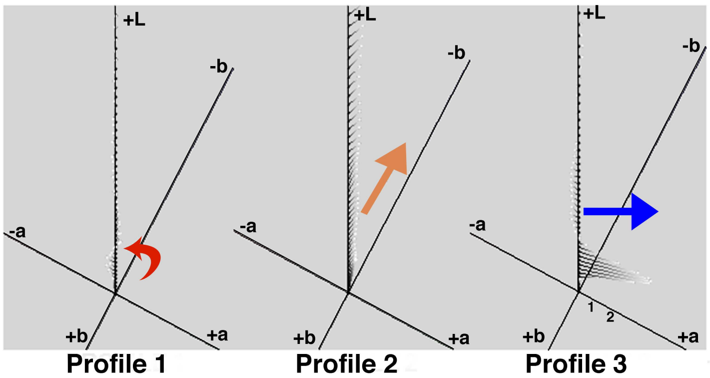

Figure 3. Grayscale reproduction: the lightness target increases in increments of 1 ΔL^*. Profile 1 has the smallest chroma deviation from the neutral axis and a large fluctuation in hue values. Profile 2, has a larger average ΔC^*, 0.5 and a consistent hue angle at approximately 220. Profile 3 has ΔC^* of 3, in the lower L^* values only.

Table 3. Grayscale Reproduction: neutral axis - the deviation from the neutral axis, CIE $C^*\times\ \Delta h^*$

Profile	Mean CIE C^*	CIE C^*_{STDV}	Mean CIE $\Delta\ h^*$	CIE $\Delta\ h^*_{STDV}$	CIE $C^*\times\ \Delta h^*$
1	0.10	0.07	40.60	62.86	0.00
2	0.49	0.10	4.37	4.30	0.96
3	0.56	0.71	11.18	32.07	0.76

Figure 3 illustrates the different ways a profile may reproduce the neutrals. Profile 1 has a very small average

ΔC^* but the hue angle fluctuates around the the neutral axis. The hue angle with profile 2 is much more constant, approximately 220, but the average ΔC^* is larger, approximately 0.5. This may cause the neutrals in profile 2 to appear cooler. The deviations from the neutral axis for profile 3 are small, except for the low L^* values. At the low L^* values, the ΔC^* increases to 3 with a h^* average of 330.

3.6 Black Point Compensation

Black Point Compensation (BPC), is often applied to address large gamut size differences between the source and destination. It is used to preserve details by maintaining tonal variations in the shadow regions, which may impact the document's perceived contrast. The method of applying BPC is different between vendors.[23] The BPC aspect is reported twice in the profile fact sheet, Table 5. The calculations represent the preservation of shadow details and overall contrast. The metric for shadow details is derived from the local differences for the lightness values that are less than 20. The contrast is calculated by the slope of the lightness curve, they are both using the CIE L^* target. The following plot, Figure 4, looks at the shadow details of the target.

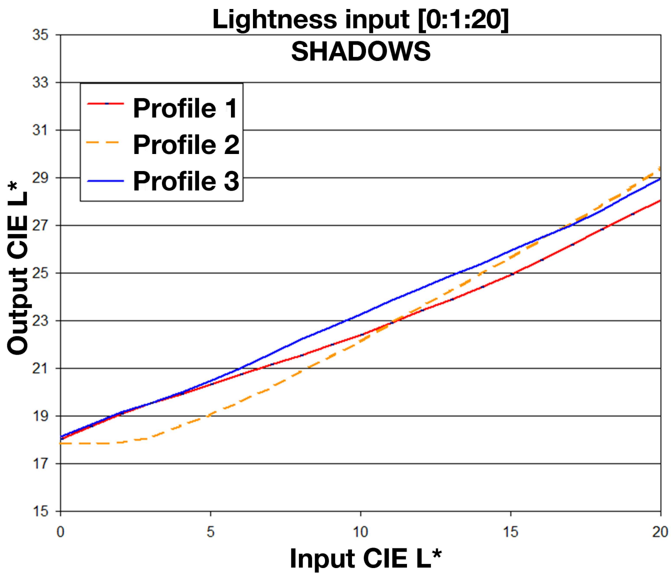

Figure 4. A plot comparing the curves of the low CIE L^* values of the reproductions. Profile 2, maps the very lowest values to the same point, and then has the most aggressive slope. Profile 1, has the most gradual incline, which will improve the shadow details at the expense of contrast.

Profile 1 has the most gradual slope, that will result in more shadow details at the expense of contrast. Profile 2 maps the lowest CIE L^* values to the same value and then has the most aggressive slope. This will result in some loss of shadow details but an increase in overall contrast. All three profiles have BPC applied, but the type and intensity varies. Profile 2 has some clipping at the very lowest levels, this is an indication that the BPC being applied is not as aggressive as the other profiles. The results reported in the fact sheet are scaled to sum to zero.

3.7 Gamut Mapping

Gamut mapping is the transformation of colors from an original to a reproduction. Different algorithms used to map colors between gamuts of different sizes are often evaluated by comparing them to each other.[24] The metric evaluates this aspect with ten targets, designed to test how out of gamut colors are mapped into the profile gamut. The targets are created in the CIELAB color space and consist of two color ramps. The targets have two columns, the difference between the vectors is dependent on which channel the target is testing.[9] Figure 5 illustrates an example of a lightness target. The target has two vectors with a difference of exactly 1 ΔL^*, the ramp increases in equal h^* steps, and the chroma is the same throughout the target. The top vector has a L^* value of 70 and the bottom is 69. After the profiles were applied, the color differences between the two vectors,

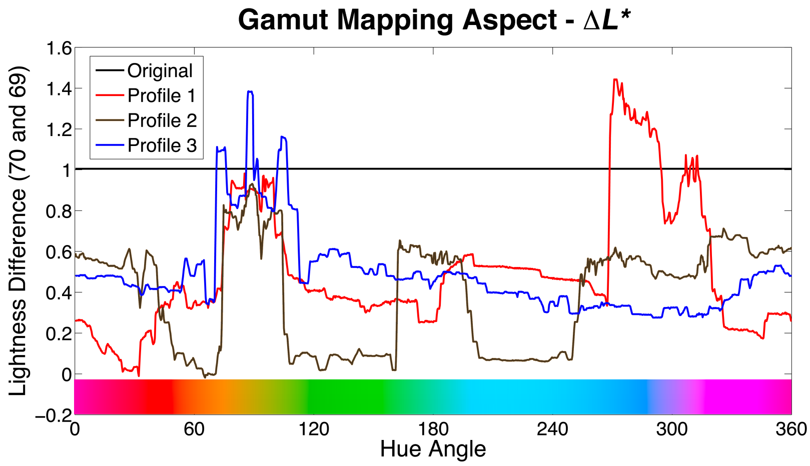

Figure 5. Example of the gamut mapping aspect, with the metric difference vectors plotted. The target used is at the bottom of the plot, a hue ramp target with two different lightness vectors, L^* 70 (top) and L^* 69 (bottom). A reproduction resulting in a straight line at one would be ideal, the original. Profile 3 had the smallest STDV, followed by profile 2. Profile 3 compresses the lightnesses differences without eliminating them. Profile 2 also compresses the differences, however at a few hue values, the differences go to zero. Profile 1 has the largest variation throughout the hue range.

for each target, were calculated. It is ideal for there to be a consistent difference between the vectors, to maintain details of the out of gamut colors. The metric returns the standard deviation, STDV, of the color difference. The final value reported in Table 5 is the average of the STDV for all the targets.

Overall profile 3 had the smallest average ΔL^*_{STDV}, indicating the differences between vectors were consistent. For this particular aspect, looking at the vectors plotted gives additional information about the way the colors are mapped. For example, a profile may perform well overall, but struggle to maintain the differences in specific regions of the gamut. Figure 5 is a plot comparing the vectors of the three profiles to the original of the first target. An ideal reproduction results in a straight line at one, the original ΔL^*. For this example, profile 2 had the smallest STDV followed by profile 3. There are two hue ranges, approximately 120 and 240, where profile 2 has nearly no difference. If the objective is to maintain details and the document is dominated by these color ranges, it may be best to use one of the other profiles. Profile 1 had the largest STDV, but only one hue range with no differences maintained.

3.8 Gamut Volume

The gamut volume describes the limits for a set of colors,[21] the outer boundary of possible colors that can be reproduced for a given profile. The profiles were opened in ICC3D and the gamut volumes were calculated.[13] The size of the profile gamut may indicate that the profile is able to reproduce colors with higher saturation but does not describe the shape of the gamut which could effect the smoothness or transitions within the gamut. The gamut volume values can be found in the fact sheet, Table 5. Profile 1 has the largest volume followed by profile 2.

3.9 Invertibility

The invertibility of an ICC printer profile is defined by an A2Bn LUT, the device color to the CIELAB color space. This is particularly relevant for soft-proofing applications. This aspect was tested with a software only round-trip test, called profileQA.[14] The script is used to test the colorimetric accuracy of a profile. The average color differences for the 1120 patch target for both rounds are shown in Table 4.

Profile 1 has a very small color difference average for this software only comparison. Unlike the colorimetric accuracy aspect, round 1 and 2 of the invertibility test showed profile 1 to have the smallest average color difference. For the round-trip test, when all colors were in the profile's gamut, the profile 1 A2B0 LUT was the most accurate. Profile 2 had the highest average color difference. The average color differences are reported in Table 5.

Table 4. Invertibility profileQA software only round-trip results. Profile 1 has the smallest average color difference for both round 1 and 2. Once all colors were in gamut, the average difference was very small for all three profiles.

Stats	**Round 1**			**Round 2**		
	Profile 1	Profile 2	Profile 3	Profile 1	Profile 2	Profile 3
Mean	0.47	1.35	2.4	0.29	1.14	0.92
STDV	0.57	0.61	2.44	0.23	0.48	0.83
max	3.94	4.28	10.7	1.37	2.07	3.76
min	0	0.06	0.02	0	0.07	0.014

3.10 Objective results, profile fact sheet

The objective metric results are summarized in the profile fact sheet, Table 5. They are also illustrated in Figure 6. Many of the objective metrics have been scaled to fit on a single figure. The further the profile is from the center the 'stronger' their performance is, in that particular aspect. So the profile with the largest area, was the strongest performer overall for the objective metrics. Keep in mind that more of an aspect does not always relate directly to more perceived IQ.

Table 5. The objective profile assessment - fact sheet. Listed are the aspects that have been objectively tested, with the corresponding metric and the final results.

Profile Fact Sheet				
Aspect	Metric	Profile 1	Profile 2	Profile 3
In gamut re-rendering	Colorimetric Accuracy (round-trip, ΔE^*_{94})	No (0.30)	Yes (3.84)	Yes (3.98)
6 primaries Lightness	Mean absolute ΔL^*	4.74	5.87	5.36
6 primaries Chroma	Mean absolute ΔC^*	31.92	33.02	39.85
6 primaries hue	Mean absolute Δh^*	5.43	4.32	4.02
Colorfulness	Cui 625 patch color target (scaled)	-0.63	0.61	0.02
Smoothness	2^{nd} derivative (z-scores)	0.34	-1.21	0.87
Grayscale Reproduction	CIE $C^* \times \Delta h^*$	0.00	0.96	0.76
BPC shadow details	ΔL^*_{STDV} (≤ 20)	0.08	0.23	0.06
BPC contrast	CIE ΔL^* (≤ 20)	9.41	10.78	10.18
Gamut Mapping	Mean ΔLCh^*_{STDV}	0.37	0.19	0.20
Gamut Volume	ICC 3D	470,574	460,803	448,334
Invertibility	profileQA ΔE^*_{94}	1.16	2.65	2.19

Profile 1 performs less re-rendering of the colors once they are in gamut, for both the round-trip and invertibility aspects. It also preserves shadow details at the expense of contrast in the low L^* values. Profile 2 is strongest in BPC contrast and colorfulness. It performed the worst for smoothness and BPC shadow details. It also reproduced the grayscale with a consistent blue cast. Profile 3 performed the strongest in smoothness, gamut mapping details and the preservation of the hue for the primaries. The grayscale reproduction was close to neutral, except for very low L^* values.

4. SUBJECTIVE ASPECTS - PERCEPTUAL IMAGE QUALITY

The motivation behind this section is to thoroughly investigate the perceptual IQ aspect. Perceptual differences and quality preferences are often very observer and document dependent, making it difficult to automate this process. Similar to the previous section, metrics are used to assess each aspect of the profile's performance. We investigate the use of more advanced IQMs (that simulate the human visual system) to assess observer quality preference of individual subjective aspects.

The choice of sub-aspects came from Pedersen et al.[10] Since the printer platform remained the same for the three reproductions and smoothness is addressed with the objective aspect list, we have eliminated artifacts and

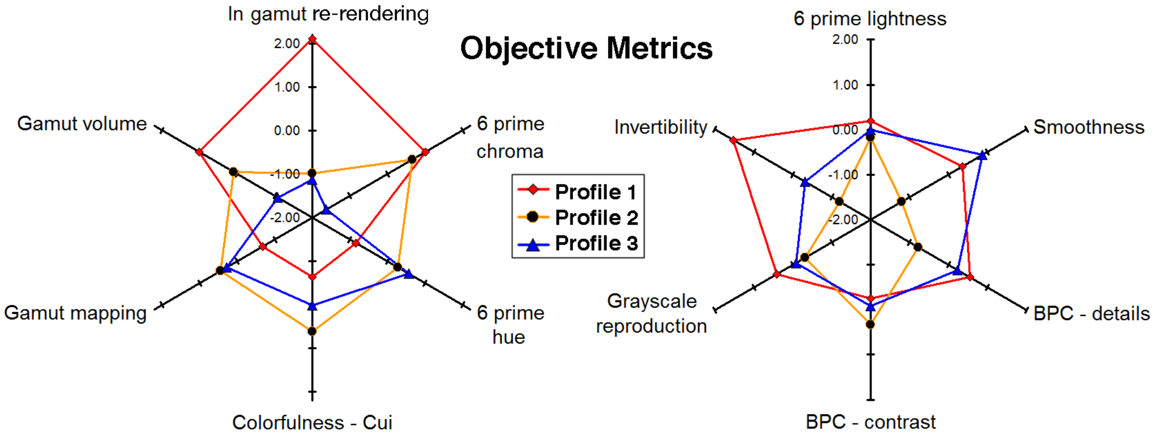

Figure 6. The objective metrics have been scaled, split into two groups, and plotted. The results are relatively compared by aspect. The profile with the largest area has the 'strongest' performance, for a specific aspect. Profile 1 has performed strongest for most aspects.

physical. Our list includes: overall preference, color preference, contrast preference, lightness preference, and sharpness preference.

A psychometric evaluation was conducted to test observer's preferences between the three profiles for the set of sub-aspects. Correlation statistics are calculated between the observer and IQM results to test the performance of the IQMs chosen. A psychometric test was not necessary for the objective metrics because they were not evaluating preference. The objective metrics test whether a profile reproduced more or less of each aspect while the subjective testing asks if a reproduction is more or less preferred based on a specific sub-aspect.

4.1 Image quality metrics

Numerous IQMs have been considered for this work, many of which came from the proposed metrics in the literature.[25] The IQMs are summarized in Table 6, with whether they are reference or no-reference, intended use, and which subjective aspects they evaluated. The selection of metrics was based on the original intended use, the goal of our evaluations, the popularity, and the authors' prior knowledge of the metrics. Only a subset of the IQMs considered are reported on.

Table 6. IQM summary: metrics are listed, whether they are reference based or no-reference, the creator's intended use, and which sub-aspect the metric is assessing.

Metric	Ref	Intended Use	Which subjective aspect(s)
busyness[26]	No	Quantifies the amount of details in a scene using a Sobel filter and morphological functions	Overall, contrast, lightness, sharpness
S-CIELAB[27]	Yes	Color difference with consideration to the human visual system	Overall, color, contrast
SSIM[8]	Yes	Uses luminance and contrast algorithms to compare local and global structural information	Overall, contrast, lightness, sharpness
VIF[28]	Yes	Assesses information in the original, quantifies how much of the original can be extracted from the reproduction	Overall, contrast, lightness, sharpness

4.2 Document Suite

A variety of document types were chosen for this evaluation. The following document characteristics were included in the suite: high, medium and low key, grayscale, natural, artificial, large areas of skintone, fine detail, clear gradients, primarily one color, memory colors, highly saturated colors, and documents that were mostly out of gamut. The documents were all 16 bits per pixel and in the CIELAB colorspace. Thumbnail examples of the documents can be found in Figure 7.

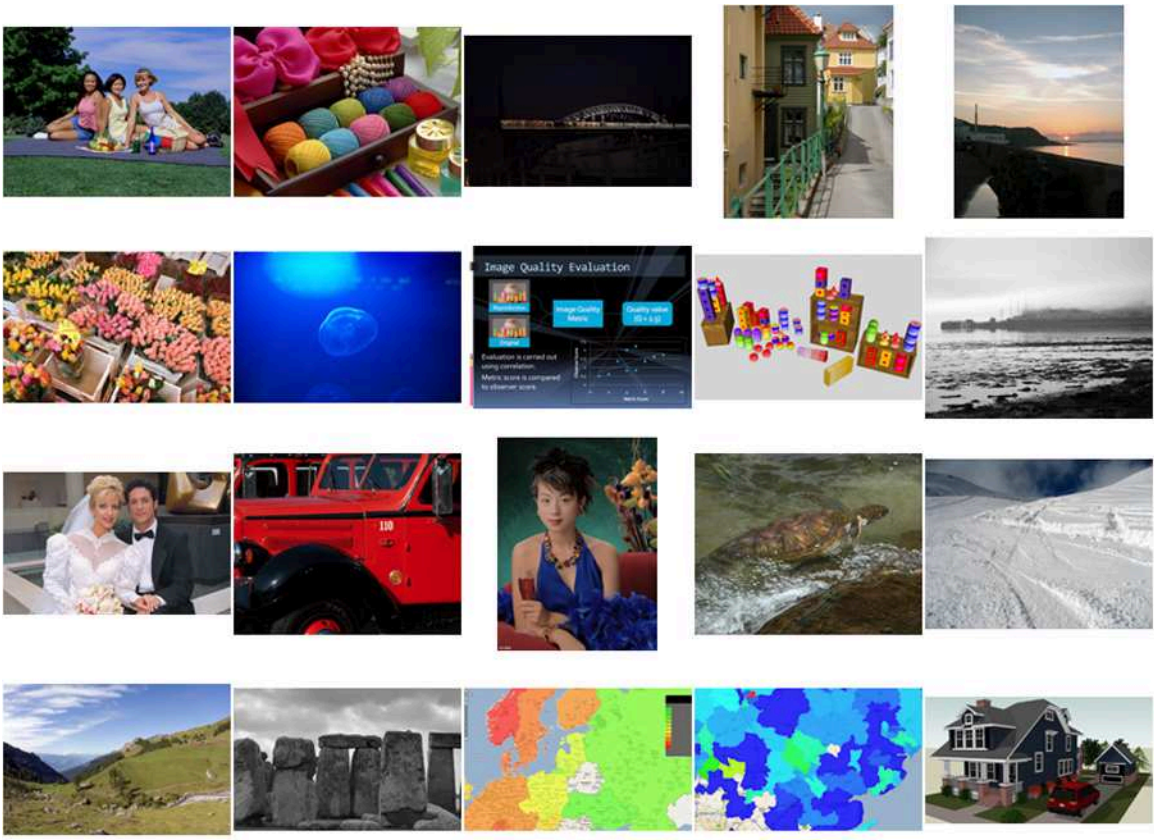

Figure 7. Thumbnail examples of the 20 documents used to test the perceptual IQ sub-aspects.

4.3 Psychometric evaluation details

The three reproductions were viewed simultaneously, at a viewing distance of approximately 24 inches. The documents were presented using the digram balanced latin square, to insure a unique viewing order for each observer.[29] The observers were asked to rate the three reproductions of each of the 20 documents based on a specific aspect. The ratings were on a scale from 1 to 5, 1 was most preferred and 5 was least preferred. Descriptions and synonyms were provided to clarify the meaning of each aspect. The color temperature of the room was approximately 5100K and the illumination was approximately 560 lux. The 15 observers ranged in age, experience, ethnicity and gender. 5 of the 15 observers were considered experts in print quality assessment. The synonyms associated with each aspect are listed below.

- Color preference - hue, saturation, colorfulness, and neutrality
- Contrast preference - visually meaningful differences (color and/or lightness), global and local
- Lightness preference - light, dark, and dynamic range
- Sharpness preference - details and edges

4.4 Psychometric evaluation results

A summary of the z-score results for the individual tests are illustrated in Figure 8. With most documents and most tests, the observers significantly preferred profile 1 over profiles 2 and 3. There was not a significant preference between profiles 2 and 3 for the overall, contrast and sharpness tests. There was a preference for profile 2 over 3 for the color test. There was no significant preference between the profiles for the lightess test. The preference differences were small for all the sub-aspects. The z-score range was less than 0.5, indicating that the observer preference between profiles was slight or that the observers disagreed with each other. With the

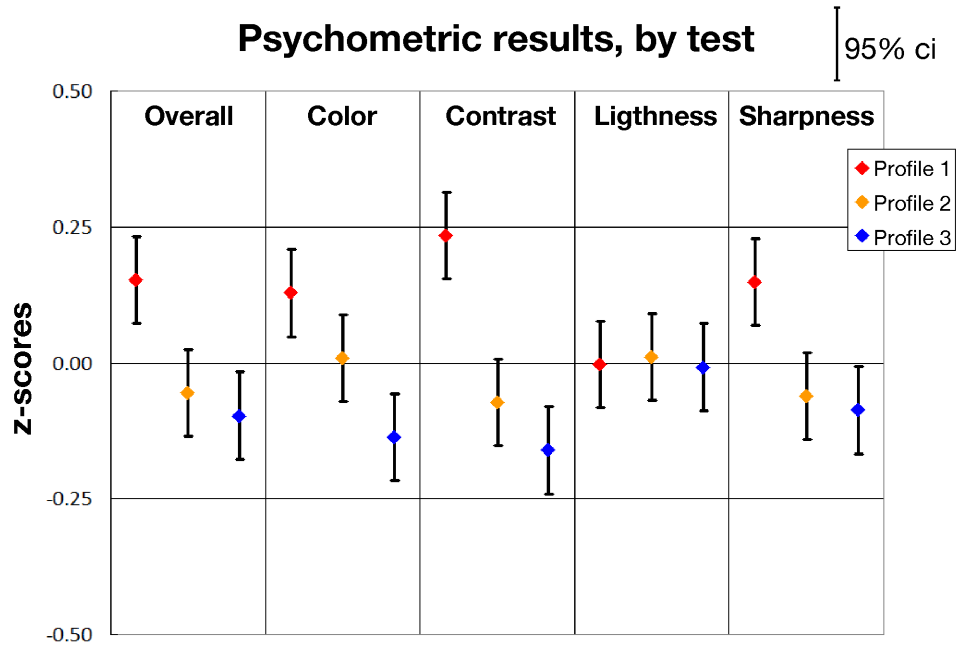

Figure 8. The category judgment z-score results of 15 observers rating 3 reproductions of 20 documents on their preference of a specific sub-aspect: overall, color, contrast, lightness, and sharpness.

overall preference test, profile 2 was most preferred for ten documents and profile 1 was most prefered for seven documents. However, profile 2 had five documents that were least preferred, suggesting that profile preference is document dependent.

4.4.1 Looking at the reproductions

Looking at the prints side-by-side, for most documents, the reproductions are similar with very subtle differences. Profile 1, maintained the contrast without loosing important document details. Profile 2 had slightly more contrast with most documents and the colors were more saturated. The higher contrast and loss of shadow detail may be caused by a different method of BPC, see Section 3.6. With many documents, profile 2 benefited from applying less BPC, ie. the roses document. For contrast and sharpness it was the most preferred, for the other tests it was preferred equally to profile 1. For many of the documents, profile 3 had a slightly softer focus or lower contrast than profiles 1 and 2. The red truck document was ranked differently depending on the sub-aspect. Figure 9 gives an example of the document with the key areas of interest cropped. Profile 1 was significantly preferred followed by profile 3 for the overall preference test. Profile 2 may have been preferred least overall, because of the loss of shadow details in the tire region. The observers preferred profile 2 over profile 3 when asked to judge only on their color preference.

4.5 Subjective metrics validated

The purpose of the IQMs, for this work, is to simulate the observer's preferences. The performance of the IQMs is measured by their correlation to the observer tests. The metric choices are evaluated on a per document and overall basis, by calculating the correlation to the observer results.[9] Pearson's correlation coefficient, $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \times \sigma_Y}$, was used to describe the correlation, per document. The coefficient indicates the linear relationship between two variables on a scale of +/- 1, the closer the values are to +1 the better the correlation was between the metric and visual results. For each document, the metric values and the z-scores of the observer results were used to find the correlation. The mean of these correlations are reported. The percentage of correlations above 0.6 are also listed, indicating the documents which have a higher than average correlation to the observer preferences.[30]

Figure 9. Example of the red truck document. The cropped images on the left compare the BPC contrast versus shadow details. On the left is the color differences of the red for the three profiles. The observers significantly preferred profile 1 for the overall preference test, followed by a significant preference of profile 3 over 2. The lost shadow details in the tire region with profile 2 may be the reason. When the observers were asked to judge color preference only, the observer preference between profiles 2 and 3 switched. The red of the truck is a lighter and more yellow with profile 3, the other profiles have a stronger clockwise hue shift with the red primary, based on the primaries objective metric.

A second, overall, correlation technique was used to evaluate the performance of the metrics, the rank order correlation.[31] The metric results of all the documents are ranked and the ranks are used to obtain metric z-scores. These are compared to the z-scores of the observers. The correlation values we reported with the p-values. Additionally, we indicate (based on the z-scores), if the metric and observer results report the same 'best' profile and if this profile was significantly preferred over the next preferred.

4.6 IQMs reviewed and correlations reported

The average observer preference between the profiles was very similar for all sub-aspects, with the exception of the lightness preference. Therefor the rank order correlations will be very similar across the aspects for a given metric. The rank order correlations of VIF IQM illustrates these similarities. This may or may not be true for the per document correlation. A summary of the correlation statistics are listed in Table 7.

Overall, contrast, and sharpness: The VIF and busyness IQMs both have strong correlations to these aspects. The VIF IQM ranked the profiles in the same order as the observer. Busyness reversed profiles 2 and 3. For these sub-aspects, busyness has a higher correlation per document, especially with contrast, while VIF has a higher rank order correlation.

Color: S-CIELAB had a strong correlation to the color aspect with 45% of the documents at a correlation of 0.6 or higher. The original documents were of good quality, so a color difference IQM may be accurate for assessing preference, even when a digital reference is not provided. The correlation with the artificial documents was much lower. Also observers preferred the grayscale documents with a slight color cast, over the ones with a lower average S-CIELAB value to the original.

Lightness: This aspect was difficult to match because the observer z-scores were very close together. SSIM had the best correlation on average with the same 'best' profile, however the average correlation per document was very low and only 25% of the documents had a correlation of above 0.6.

Many of the other IQMs investigated had strong correlations to the different aspects, however the average rank placed profile 2 above 1. This can be seen in Table 7, with SSIM, which ranked profile 2 above 1.

5. PROFILES COMPARED RELATIVELY

The performance of the profiles is summarized with the four radar plots. The first comparison, the objective metrics, is in Section 3.9, Figure 10. The second comparison, the the observer z-score and IQM results, is in Figure 10.

Table 7. The correlation performance of the IQMs used to assess the sub-aspects. The mean correlation uses Pearson's coefficient per document. The average is reported with the percentage of correlations above 0.6. The rank order correlation between the z-scores of the ranked metric data and of the observer results is also included. The p-values of the rank order correlation are listed next. The final column indicates whether or not the metric picked the same 'best' profile as the observers and if it was significantly preferred.

IQM	Aspect	Mean Correlation	Above 0.6	Rank order		
				Correlation	p-value	Yes/No
Busyness	Overall	0.18	35%	0.91	0.28	Yes (Yes)
S-CIELAB	Overall	0.19	35%	0.93	0.23	Yes (Yes)
SSIM	Overall	-0.18	20%	0.44	0.71	No
VIF	Overall	0.30	55%	1.00	0.06	Yes (Yes)
S-CIELAB	Color	0.35	45%	1.00	0.04	Yes (Yes)
Busyness	Contrast	0.50	55%	0.88	0.31	Yes (Yes)
S-CIELAB	Contrast	0.19	45%	0.91	0.27	Yes (Yes)
SSIM	Contrast	-0.17	30%	0.49	0.68	No
VIF	Contrast	0.41	50%	1.00	0.02	Yes (Yes)
Busyness	Lightness	0.14	35%	-0.54	0.64	No
SSIM	Lightness	0.01	25%	0.84	0.37	Yes (Yes)
VIF	Lightness	0.14	35%	0.04	0.98	No
Busyness	Sharpness	0.41	55%	0.93	0.24	Yes (Yes)
SSIM	Sharpness	-0.21	30%	0.38	0.75	No
VIF	Sharpness	0.40	60%	0.99	0.10	Yes (Yes)

Profile 1 was the most color accurate reproduction with: the least amount of in gamut re-rendering (smallest average color difference for the round-trip test), small color differences for the invertibility aspect, small differences with lightness and chroma for the primaries, and very small grayscale differences. It also maintained shadow details. All IQMs ranked profile 1 first, with the exception of SSIM. The observers significantly preferred profile 1 over profiles 2 and 3 for all tests but lightness.

Profile 2 is the most colorful and has the most contrast. It maintained out of gamut color differences in the gamut mapping aspect but did not maintain shadow details with the BPC aspect. It was not significantly preferred for any of the observer studies. It was significantly preferred over profile 3 for the color preference observer test. SSIM was the only IQM to rank profile 2 first, it was ranked second with VIF and S-CIELAB.

Profile 3 performed best with the smoothness metric and maintaining the hue values for the primaries. However, it lost the most chroma with the primaries aspect and has the smallest gamut volume of the three profiles. In the observer study and the IQMs, for all aspects, profile 3 was often ranked third and was never ranked first.

6. CONCLUSIONS AND FUTURE WORK

In this work we compiled a list of quality metrics used to evaluate the perceptual rendering intent. Each quality aspect, used to characterize the profiles, was evaluated with a metric and summarized in the profile fact sheet. The aspects that are preference based were evaluated with more complex IQMs that simulate the human visual system. An additional validation step was taken for the IQMs, where the results were compared to observer preference tests. A relative comparison of the profiles' performance is summarized by aspect in the radar plots.

Metrics help us in deciding the behavior of profiles, but we can not use them to determine an ultimate profile for all circumstances. For example, the red truck, profile 2 is least preferred overall, but has a stronger performance when the observers were asked to judge only on color. With the profile assessment tool, a user is enabled to make a much more informed decision between profile options, especially with consideration to their profiling objectives and the content of their documents. The final optimal profile is document and aspect dependent.

There are a number of items that could and should be considered for future work. An aspect that should be added is the monotonicity of the profiles. Additionally, a pooling strategy could be considered, to reduce

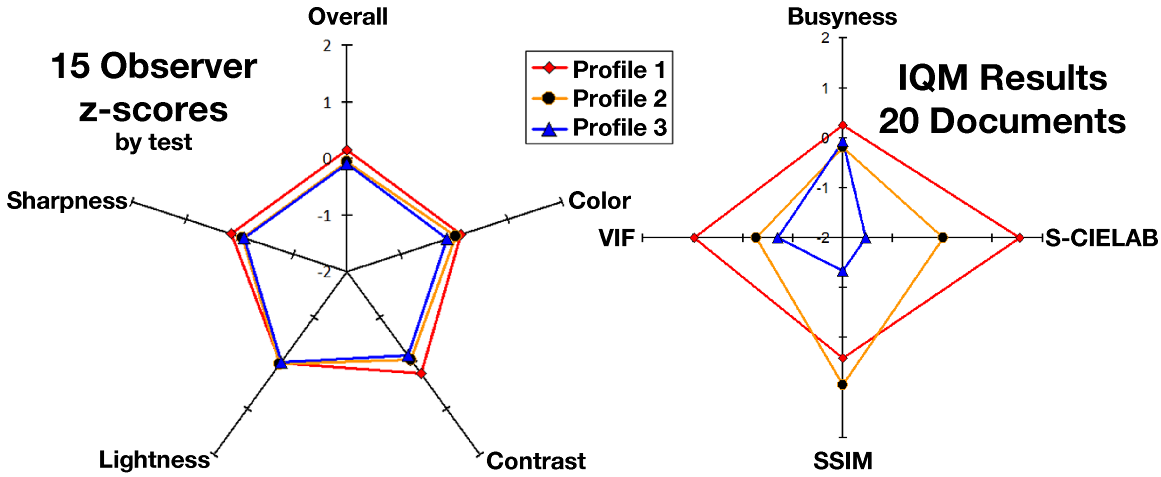

Figure 10. The observer z-scores and the IQM results are plotted. The profile with the largest area has the best performance, for a particular aspect. The observer preferences between profiles for all aspects are small. There is no significant preference between profiles 1 and 2 for the observer results, however the metrics all distinguish a preference between the two.

the amount of information given by the metrics and to weigh the importance of the metrics based on which rendering intent is being assessed. Ideally the weights would be sensitive to document content. Also using CIECAM02 to evaluate hue shifts rather than CIELAB. Finally, since the perceptual rendering intent aims for a best reproduction and a reference may not be provided, perhaps the reference IQMs should compare the different reproductions rather than against the digital original.

REFERENCES

[1] ICC.1:2004-10, "Image technology colour management - architecture, profile format, and data structure," (2004).

[2] Consortium, I. C., "ICC White Paper 1: Fundamentals of the Version 4 Perceptual Rendering Intent." 23/03/10: www.color.org (Dec 2004).

[3] Consortium, I. C., "ICC White Paper 4: Color Management - Conceptual Overview, Evolution, Structure & Color Rendering Options." 23/03/10: www.color.org (2004).

[4] Consortium, I. C., "ICC White Paper 9: Common Color Management Workflows and Rendering Intent Usage." 23/03/10: www.color.org (Mar 2005).

[5] Viggiano, J. A. S. and Moroney, N. M., "Color reproduction algorithms and intent," in [*3rd IS&T/SID Color Imaging Conference*], 152–154 (November 1995).

[6] Consortium, I. C., "ICC White Paper 2: Perceptual Rendering Intent Use Case Issues." 23/03/10: www.color.org (Jan 2005).

[7] Sharma, A., "Methodology for evaluating the quality of ICC profiles – scanner, monitor, and printer," *Journal of Imaging Science and Technology* **50**(5), 469–480 (2006).

[8] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

[9] Falkenstern, K., Bonnier, N., Brettel, H., Pedersen, M., and Viénot, F., "Using image quality metrics to evaluate an icc printer profile," in [*Color and Imaging Conference*], 244–249, IS&T and SID, San Antonio, TX (Nov 2010).

[10] Pedersen, M., Bonnier, N., Hardeberg, J. Y., and Albregtsen, F., "Image quality metrics for the evaluation of print quality," in [*Image Qualtiy and System Performance*], Gaykema, F. and Farnand, S., eds., *Proceedings of SPIE*, SPIE, San Francisco, CA (Jan 2011).

[11] Cui, C. and Weed, S. F., "Colorfulness? in search of a better metric for gamut limit colors," in [*PICS*], **3**, 183–187 (March 2000).

[12] Green, P., [*Color Management: Understanding and Using ICC Profiles*], Wiley-IS&T Series in Imaging Science & Technology (Feb 2010).

[13] Bakke, A. M., "ICC3D - A color management application." 23/07/10: http://colorlab.no/research/icc3d (Jun 2007).

[14] Boll, H., "Matlab script: ProfileQA." 23/03/10: http://www.mathworks.com/matlabcentral/fileexchange/3674-profileqa (July 2003).

[15] Sharma, A., "Measuring the quality of ICC profiles and color management software," *Seybold Report 4* **20**, 10–16 (2005).

[16] Green, P. and MacDonald, L., [*Color Engineering: Achieving Device Independent Colour*], John Wiley and Sons, Inc., New York (2002).

[17] Green, P., "A smoothness metric for colour transforms," in [*Color Imaging XIII: Processing, Hardcopy, and Applications*], Eschbach, Reiner; Marcu, G. G. T. S., ed., **6807**, 68070I–68070I–5, SPIE-IS&T (Jan 2008).

[18] Lindberg, S. and Green, P. J., "Methods of quality assessment for large sample sets," *Journal of Imaging Science and Technology* **49**, 442–449 (2005).

[19] Bang, Y., Pizlo, Z., and Allebach, J. P., "Discrimination based banding assessment," in [*NIP19: International Conference on Digital Printing Technologies*], 745–750 (Sep 2003).

[20] Zolliker, P. and Simon, K., "Continuity of gamut mapping algorithms," *Journal of Imaging Science and Technology* **15**, 220–233 (2006).

[21] Giorgianni, E. J. and Madden, T. E., [*Digital Color Management: Encoding Solutions*], John Wiley and Sons, Inc. (2009).

[22] Enoksson, E., "Digital test form for ICC-profiles," in [*TAGA Conference Proceedings*], 454–473 (Mar 2006).

[23] Consortium, I. C., "ICC White Paper 40: Black-Point Compensation: Theory and Application." 23/03/10: www.color.org (2010).

[24] Morovič, J., *To Develop a Universal Gamut Mapping Algorithm*, PhD thesis, University of Derby (1998).

[25] Pedersen, M., Bonnier, N., Hardeberg, J. Y., and Albregtsen, F., "Attributes of image quality for color prints," *Journal of Electronic Imaging* **19**, 011016–1 – 011016–13 (Jan 2010).

[26] Orfanidou, M., Triantaphillidou, S., and Allen, E., "Predicting image quality using a modular image difference model," in [*Image Quality and System Performance V*], Farnand, S. P. and Gaykema, F., eds., **6808**, 68080F, SPIE, San Jose, CA, USA (Jan 2008).

[27] Zhang, X. and Wandell, B. A., "A spatial extension of cielab for digital color-image reproduction," *Journal of the Society for Information Display* **5**, 61–63 (March 1997).

[28] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," in [*IEEE Transactions on Image Processing*], **15**(2), 430:444 (2006).

[29] Keppel, G. and Wickens, T., [*Design and Analysis. A Researcher's Handbook*], Pearson education, Upper Saddle River, New Jersey, USA (2004).

[30] Pedersen, M., Bonnier, N., Hardeberg, J. Y., and Albregtsen, F., "Estimating print quality attributes by image quality metrics," in [*Color and Imaging Conference*], 68–73, IS&T and SID, San Antonio, TX (Nov 2010).

[31] Pedersen, M. and Hardeberg, J. Y., "Rank order and image difference metrics," in [*CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision*], 120–125, IS&T, Terrassa, Spain (Jun 2008).