Improvement of CMC upon CIEDE2000 for a New Experimental Dataset

Received 15 October 2005; accepted 7 November 2005


Key words: color difference; CMC; CIE2000DE; CIE94

A recent article has reported a new experimental dataset of 96 textile color pairs (four color centers, each at two lightness levels), with color differences in the range 0.3–10.1 CIELAB units. These color pairs were visually assessed (pass/fail acceptability) by a panel of 21 inexperienced observers with nondefective color vision, using a commercial color assessment cabinet with daylight. Superior performance of CMC(2:1) upon CIEDE2000(2:1:1) was concluded on the basis of the results provided by a $\chi^2$ test. It must be acknowledged that the authors of Ref. 1 provided full data on the color coordinates and average visual differences of their color pairs for the use of other researchers. The authors of Ref. 1 qualify their work as “preliminary” and in their Conclusions they make some suggestions for expanding their research.

Improvement of CMC upon CIEDE2000 for a specific experimental dataset like the one reported in Ref. 1 may be accepted as a matter of fact. However, further analyses on the statistical significance of this improvement and their potential causes seem necessary. Note that for the 3657 color pairs from four independent laboratories that constitute the combined dataset used for the development of CIEDE2000, just the opposite conclusion has been reported (that is, CIEDE2000 significantly improves CMC), which also agrees with the results of the works carried out by CIE TC1-47 during the development of CIEDE2000. There is a noteworthy difference between the two last CIE-recommended formulas (CIE94 and CIEDE2000): while CIE94 was not significantly superior to CMC for the combined dataset, CIEDE2000 is. Other recently published papers also support that CIEDE2000 outperforms CMC for new visual datasets.

Responding to Mangine et al.’s invitation to complement their research, I have computed PF/3 values, widely employed in recent color-difference research, as well as minimum percentages of “wrong decisions” (WD%), usually used as a figure of merit in acceptability experiments. PF/3 and WD% are two additional appropriate measurements to test the merit of the color-difference formulas employed in Ref. 1. The lower the PF/3 or WD% values the better the performance of a given color-difference formula. In addition, two statistical analyses (Wilcoxon and $F$ tests) on the significance of the differences found between CMC(2:1) and CIEDE2000(2:1:1) have also been performed. All these computations were based on the data provided in Table AI (a minor mistake was corrected in this table: the $b^*$ coordinate of sample 56 was assumed to be $-31.70$ and Table I (the reported $\Delta V$ values were adopted) of Ref. 1. Some inconsistencies were found among the CIELAB, CMC, CIE94(2:1:1), and CIEDE2000(2:1:1) color differences reported in Table I and the ones achievable from the data provided in Table AI. Thus, for the forthcoming analyses, I used the values of the color differences computed by myself from Table AI (in the case of the CIEDE2000 color-difference formula, the software provided by Sharma et al. was employed).

Table I shows the PF/3 values and percentages of wrong decisions (WD%) for the whole dataset employed in Ref. 1 and three subsets, using the four color-difference formulas employed by Mangine et al. Specifically, these three subsets were used because it was suspected that color-difference pairs above 5.0 CIELAB units (the upper limit recommended for the use of CIEDE2000) and pass ratios of 1 and 0 (i.e., all the observers agreed on the pass/fail judgment of a color pair) may distort the conclusions concerning the merit of the different color-different formulas. For example, color pairs 17 and 46 have the same null pass ratio, but, from our current computations, their CIELAB color differences differ by more than a factor of 4, in such a way that these two pairs will distort the analyses on the relationship between perceived and computed color differences. In usual practice, the 100% accept or reject data must be discarded in the computations of the logit function providing $\Delta V$ values (Eq. [1] of Ref. 1).

PF/3 values shown in Table I indicate that for the whole dataset ($n = 96$) in Ref. 1, CIE94(2:1:1) is the best formula, followed by CMC(2:1), CIEDE2000(2:1:1), and CIELAB. However, the PF/3 values for these four formulas were in a narrow interval (5.8 units) and were also unexpectedly high (usual PF/3 values for CIELAB-based formulas are around 20–50 PF/3 units). PF/3 values for the three subsets are considerably lower than for the whole dataset, confirming that removal of pairs with color differences above 5.0
CIELAB units and pass ratios of 0 and 1 leads to great improvement of the predictions made by any of the four color-difference formulas. In any case, for the most reduced subset (n = 52) PF/3 values are also high (perhaps as a result of the use of nonexperienced observers), and the improvement of CMC(2:1) (now the best of the four formulas) upon CIEDE2000(2:1:1) is very slight (1.6 PF/3 units). With respect to the WD%, Table I shows that for the four datasets CIEDE2000(2:1:1) improves CMC(2:1), which also improves CIE94(2:1) or CIELAB. However, once again all these WD% improvements are very slight (at most two WD differences within a set of at least 52 color pairs). In summary, the results in Table I do not clearly support the contention that CMC(2:1) is the best formula, as claimed by Mangine et al.¹

Two statistical tests were performed in order to establish whether CMC(2:1) and CIEDE2000 (2:1:1) are significantly different. The results are shown in Table II. First, the nonparametric Wilcoxon test was used, computing for each color pair the parameter $V_{AB,i}$, as made in a previous paper²:

$$V_{AB,i} = \left[ \frac{\Delta E_i - F \Delta V_i}{\Delta E_i \Delta V_i} \right]^{1/2} \text{ with } F = \frac{\sum \Delta E_i}{\sum \Delta V_i},$$

(1)

Second, a modified $F$ test⁷ (originally proposed by Alman during the development of CIEDE2000) was used, computing the ratio of $V_M$ values for the two color-difference formulas and comparing this ratio with the critical value ($F_C$), which depends on the number of color pairs ($N$) and the assumed significance level ($P = 0.05$ in our case):

$$V_M = \sum (\Delta V_i - a_M \Delta E_i)/(N - 1)$$

$$\text{with } a_M = \sum \Delta E_i \Delta V_i / (\sum \Delta E_i)^2.$$

(2)

The results from the Wilcoxon test (Table II) indicate that, except for the whole dataset ($P = 0.015$), at 0.05 significance level, there are no significant differences between CMC(2:1) and CIEDE2000(2:1:1). With respect to the computed $F$ values (Table II), they are slightly greater than 1.0, indicating that CMC(2:1) is slightly superior to CIEDE2000(2:1:1) and far below the critical values $F_C$, which indicates that these two formulas do not significantly differ for the whole dataset or any of their three subsets. However, the results from the Wilcoxon and $F$ tests applied to the whole dataset ($n = 96$) are contradictory. This problem can be solved by analyzing the hypothesis of normality (Kolmogorov–Smirnov test), required for the application of the $F$ test. In our case, this hypothesis was always fulfilled, except for the visual differences in the whole dataset ($n = 96$), in such a way that the $F$ test cannot be properly applied to this particular dataset. Consequently, the $F$ value of 1.015 shown in Table II must be disregarded, and we must accept the conclusion given by the Wilcoxon test for the whole dataset (that is, CMC(2:1) and CIEDE2000(2:1:1) are significantly different for this dataset).

In summary, from the previous results it can be concluded that the improvement of CMC(2:1) upon CIEDE2000(2:1:1) for the dataset developed by Mangine et al.¹ is very slight, and in general it cannot be considered statistically significant. Only for the whole dataset¹ does CMC(2:1) significantly improve CIEDE2000(2:1:1), but it should also...

### Table I. PF/3 values and minimum “wrong decisions” percentages (WD%) for four color-difference formulas, from the experimental data reported in ref. 1.

<table>
<thead>
<tr>
<th></th>
<th>CIELAB</th>
<th>CIE94(2:1:1)</th>
<th>CMC(2:1)</th>
<th>CIEDE2000(2:1:1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pairs (n = 96)</td>
<td>PF/3 = 175.2</td>
<td>PF/3 = 173.1</td>
<td>PF/3 = 173.7</td>
<td>PF/3 = 178.9</td>
</tr>
<tr>
<td>WD% = 13.5 (1.8)</td>
<td>WD% = 13.5 (1.1)</td>
<td>WD% = 12.5 (0.9)</td>
<td>WD% = 10.4 (0.8)</td>
<td></td>
</tr>
<tr>
<td>Pairs with $\Delta E_{ab} &lt; 5.0$ (n = 71)</td>
<td>PF/3 = 135.3</td>
<td>PF/3 = 143.4</td>
<td>PF/3 = 140.7</td>
<td>PF/3 = 146.9</td>
</tr>
<tr>
<td>WD% = 18.3 (1.8)</td>
<td>WD% = 18.3 (1.1)</td>
<td>WD% = 16.9 (0.9)</td>
<td>WD% = 14.1 (0.8)</td>
<td></td>
</tr>
<tr>
<td>Pairs with pass ratio other than 1 or 0 (n = 58)</td>
<td>PF/3 = 101.7</td>
<td>PF/3 = 97.4</td>
<td>PF/3 = 97.4</td>
<td>PF/3 = 99.5</td>
</tr>
<tr>
<td>WD% = 22.4 (1.8)</td>
<td>WD% = 22.4 (0.7)</td>
<td>WD% = 20.7 (0.9)</td>
<td>WD% = 17.2 (0.8)</td>
<td></td>
</tr>
<tr>
<td>Pairs with $\Delta E_{ab} &lt; 5.0$ and pass ratio other than 1 or 0 (n = 52)</td>
<td>PF/3 = 78.7</td>
<td>PF/3 = 79.8</td>
<td>PF/3 = 77.0</td>
<td>PF/3 = 78.6</td>
</tr>
<tr>
<td>WD% = 25.0 (1.8)</td>
<td>WD% = 25.0 (0.7)</td>
<td>WD% = 23.1 (0.9)</td>
<td>WD% = 19.2 (0.8)</td>
<td></td>
</tr>
</tbody>
</table>

*Note. The numbers in parentheses are the values of the color-difference tolerances for the computed WD%.

### Table II. Significance of the differences between CMC(2:1) and CIEDE2000(2:1:1) formulas from two statistical tests.

<table>
<thead>
<tr>
<th></th>
<th>Wilcoxon test (P value)</th>
<th>Pairs with $\Delta E_{ab} &lt; 5.0$ (n = 71)</th>
<th>Pairs with pass ratio other than 1 or 0 (n = 58)</th>
<th>Pairs with $\Delta E_{ab} &lt; 5.0$ and pass ratio other than 1 or 0 (n = 52)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pairs (n = 96)</td>
<td>0.015</td>
<td>0.063</td>
<td>0.634</td>
<td>0.806</td>
</tr>
<tr>
<td>Pairs with $\Delta E_{ab} &lt; 5.0$ (n = 71)</td>
<td>1.015 (1.404)</td>
<td>1.043 (1.486)</td>
<td>1.030 (1.552)</td>
<td>1.033 (1.592)</td>
</tr>
</tbody>
</table>

*Note. Four experimental datasets (the whole data reported in Ref. 1 and three subsets) are distinguished.*
be remembered (Table I) that for this dataset the best color-difference formula was CIE94(2:1:1) and not CMC(2:1). In my opinion, the relatively poor performance of CIEDE2000 for Mangine et al.’s dataset might be explained by bearing in mind that CIEDE2000 was developed from only perceptibility data and homogeneous color pairs with color differences below 5.0 CIELAB units.

Of course, we must not consider CIEDE2000 as the final step in industrial color-difference evaluation. Although the CMC formula has been widely accepted by industries, in particular within the textile sector, perhaps the official recommendation of CMC as the standard color-difference formula for ISO and AATCC may be reconsidered in the near future. Note that CIEDE2000 (but not CIE94) proved significantly superior to CMC for a broad dataset with 3657 color pairs from four independent laboratories. In any event, new experimental datasets on color differences such as the one analyzed here are highly valuable and should be encouraged, in particular if they follow CIE guidelines, and current works carried out by CIE Technical Committees 1-55 (Uniform Colour Space for Industrial Colour-Difference Evaluation) and 1-63 (Validity of the Range of CIEDE2000).


BOOK REVIEWS


Gary Field has an international reputation as a teacher and authority on the reproduction of colour in print, based on a long career in the graphic arts and his position as a Professor at the California Polytechnic State University (CalPoly) in San Luis Obispo. His great contributions to the printing industry have been widely recognised through awards such as the Gold Medal from the Institute of Printing and the Honors Award from TAGA.

This is the third edition of *Color and Its Reproduction*, a work that has become firmly established, since its first edition in 1988, as one of the primary reference texts in the field. It has been very substantially revised, restructured and reformatted since the second edition, to reflect the shifting reality of how colour is evaluated within modern workflows. The content can now be considered as a fine distillation of the author’s forty years of practical experience in the industry. He reasons that the printing industry has been the driving force behind so many advances in colour reproduction because printing was the world’s first mass-production industry. Progress in colour reproduction has therefore always been driven by the twin needs for high quality and low costs.

What becomes immediately apparent when first handling the book is that this is indeed a high-quality publication. The paper, layout, typesetting, diagrams and format all make it a pleasure to handle and read. The colour images and illustrations are clean, sharp and attractive, and demonstrate most compellingly the principles expounded.

The structure may conveniently be divided into three parts. The introductory Chapters 1–4 might be regarded as a primer for a junior class, covering the fundamentals of colour perception, colour measurement and colour reproduction. These chapters dip into their subjects, engaging the interest of the reader without attempting to be comprehensive, yet they cover such diverse topics as colour vision variability, chromatic adaptation, gloss and fluorescence.

The second part, Chapters 5–11, covers all significant innovations. Many historical references endow the treatment with a context and authenticity so often lacking in lesser books on the ephemera of desktop publishing. An
example is the reproduction of a beautiful colour print from hand-engraved plates by Jacob Christoph Le Blon in the early 1700s. The discussion of colour separations, halftones, ink and paper is thorough and covers all relevant aspects without ever descending into tedium or trade jargon. One of the most pleasing chapters is on colour reproduction objectives and strategies. This condenses Hunt’s six different categories of colour reproduction into two—exact and optimum—and then subdivisions the latter into corrective, compromise and preferred. The author goes on to show how this classification applies in practice to reproducing artwork, merchandise samples, fine art and photographs.

The third part, Chapters 12–15, deals with the craft of printing, and all the elusive factors that go into making a pleasing picture that meets the needs of the customer. This treatment goes beyond almost every other book on the subject, and places Professor Field in a class of his own. He builds on the insights of earlier work by Evans and Yule to present guidelines for colour correction, combining objective calibration with subjective assessment of image definition, tone reproduction and colour balance. He considers the many issues related to proofs, both hard (printed) and soft (displayed). He shows how different methods of colour specification can be integrated within the general process of colour communication, which requires not only data communication between the source and destination media, but also person-to-person communication between the customer and the designer and the printer.

Throughout the book there is much emphasis on the assessment of image quality, including the choice of correct illumination and surround when viewing prints, the correct use of measuring instruments, and the adherence to industry standards. Colour management is introduced as an evolutionary development of systems for the control of colour reproduction, which have been employed in the pre-press and printing industries since the 1930s. At every stage of the workflow there is the need for evaluation of the image attributes for consistency, and the use of targets, metrics and heuristics is placed within the broader framework of achieving an optimum result and satisfying the requirements of both client and end-consumer.

The Appendices include excellent lists of symbols and abbreviations, standards, sources of information, a glossary, and nearly 200 references.

In summary, this is a book to be recommended to every student and practitioner of the graphic arts. It should also find a place in the library of every professional involved with the processing of colour images in print or digital media, alongside Hunt’s The Reproduction of Colour. All imaging scientists and engineers developing digital imaging systems would do well to read it carefully and absorb its wisdom on how to achieve optimum picture reproduction.

LINDSAY MACDONALD

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/col.20213


This is an excellent Handbook that should be read and available as a reference to everyone interested in exterior lighting. Its contents are very comprehensive by presenting a philosophy of night and the environment, and the purpose for lighting, addressing all aspects of light pollution and making recommendations. By starting with the grounds and concepts expressing fundamental beliefs (philosophy), this Handbook presents light pollution as a universal issue and not a concern by a few extremists.

The Handbook also contains information about vision and visibility, visual task performance, behavioral functions and lighting technique technology. This material starts with the seeing process and color concerns as well as discusses topics such as contrast, dazzle and glare. It further addresses the effects of exterior lighting on the public and the environment, details photometric concerns and the need to educate the public about outdoor lighting as well as a need for acceptable laws and regulations.

The authors have spent their careers in Japan and the Netherlands. The majority of the information in the Handbook therefore has a European and Japanese flavor. There is, however, information from the IESNA (North America) and IDA (International) about their recommendations related to proper outdoor lighting. There are extensive references provided throughout the Handbook. The references result in my concern that too many are listed as Anon. (anonymous) rather than many being able to be credited to the publication of an organization or company such as a Handbook by the IES of Japan or a report by Hewlett Packard.

Another incorrect reference notation is found in Section 11.2.9 where “(CIE, 1998)” is listed but not found in this or any chapter’s Reference list and the reference attributed list of 12 outdoor work areas is also inaccurate. The correct reference is “CIE S 015:2005, Lighting for outdoor work places” which lists 15 outdoor work areas whereas the authors may have incorrectly used material from an early draft with fewer areas. Similar incorrect use of draft CIE and CEN material is also found throughout the Handbook. Also the heading of Section 12 is incorrect. The present heading is “Effects of outdoor lighting on society and on the environment.” The correct heading would replace “outdoor” with “roadway” since only roadway lighting is covered in this section.

Despite the predominance of European and Japanese material, my concern about references and the incorrect heading of Section 12, I recommend this Handbook for your reading and use as a reference. Everyone that has an interest in outdoor lighting will greatly benefit from it.

THOMAS M. LEMONS

Published online in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/col.20214

COLOR research and application
Publications Briefly Mentioned


During the life of a lighting installation, the light available for the task progressively decreases due to accumulation of dirt on surface and aging of equipment. The rate of reduction is influenced by the equipment choice and the environmental and operating conditions. In lighting scheme design we must take account of this fall by the use of a maintenance factor and plan suitable maintenance schedules to limit the decay. Lighting standard “ISO 8995/CIE S 008-2001 Lighting of Indoor Workplaces” in Section 4.8, recommends a minimum maintenance factor. It states that “The lighting scheme should be designed with overall maintenance factor calculated for the selected lighting equipment, space environment and specified maintenance schedule.” A high maintenance factor together with an effective maintenance program promotes energy efficient design of lighting schemes and limits the installed lighting power requirements.

This revision of the guide describes the parameters influencing the depreciation process and develops the procedure for estimating the maintenance factor for indoor electric lighting systems. It provides information on the selection of equipment and the estimation of economic maintenance cycles and gives advice on servicing techniques. It shows some examples of data but for accurate data it recommends that data should be obtained from the manufacturers. This guide replaces CIE 97-1992 “Maintenance of indoor electric lighting systems.”

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/col.20215