Comparison Study of the Surface Colour Measurement Data Correlation

X. Li, W. Ji, C. Li, G. Cui and M. R. Luo
Department of Colour and Polymer Chemistry, University of Leeds, Leeds LS2 9JT, UK
Corresponding author: Changjun Li (C.LI@LEEDS.AC.UK)

ABSTRACT

Many methods were developed to correlate colour measurement results between different instruments. All methods were intended to use single or few reference materials (samples) to train the model. This may lead to the conclusion that the correction models are material independent. This study investigated the models for correcting measurements from different instruments and different materials. The results showed that the correction models studied are highly material dependent.

1. INTRODUCTION

Accurate surface colour measurements play important role in many industrial applications such as colour and colour imaging reproduction industries. In the former, it is quite often that one site requires another remote site to reproduce the colours as specified. In order to reduce the lead-time the colorimetric data rather than physical samples are sent to the remote site. One of the main factors for the exact colour reproductions is that the measurement instruments used in both sites give the same data if the same sample is measured. In the latter, the imaging devices are needed to be characterised based on the colour measurements of the targets. Once again, the colour fidelity can be achieved by accurate colour measurements. Hence it is essential to achieve acurate colorimetry for the demanding open environment applications.

However, different makes and types of colour measurement instruments may be used in two sites; this could lead to large disagreements between the measurements. A study from the NPL[1] showed that the typical discrepancies of the measurement data from 20 research Labs are in the range of 0.1 to 3.0 CIELAB ΔE unites. And a separate study by the NPL[2] found that only 50% of the measurement data from 4 national Labs agreed to $0.5\Delta E$ unites. Note that each instrument manufacturer trace their measurement results to one of the national Labs' standards and the disagreements between the national Labs are included in the instrument disagreements. Therefore, data correlation between different measurement instruments is highly required.

The earlier work about the instrumental corrections was given by Berns and Petersen[1]. They, based on the work of Robertson [4], derived a model using a single Ceramic Tile with cyan colour to diagnose and correct the systematic errors. The model is described below:

The Berns and Petersen Model:

$$R_T(\lambda) = R_m(\lambda) + B_0 X_0(\lambda) + B_1 X_1(\lambda) + \cdots + B_6 X_6(\lambda) \tag{1}$$

where $X_0(\lambda) = 1$, $X_1(\lambda) = R_m(\lambda)$, $X_2(\lambda) = [100 - R_m(\lambda)]R_m(\lambda)$, $X_3(\lambda) = dR_m / d\lambda$, $X_4(\lambda) = w_1(\lambda) dR_m / d\lambda$, $X_5(\lambda) = w_2(\lambda) dR_m / d\lambda$, and $X_6(\lambda) = d^2 R_m / d\lambda^2$, and $R_T(\lambda)$ and $R_m(\lambda)$ are the error-free and measured reflectances of the single standard reference material (SSRM) (Cyan, Ceramic Tile) respectively. The least squares method can be used for determining the coefficients B_0 to B_6 based on $R_T(\lambda)$ and $R_m(\lambda)$ of the SSRM. Once the coefficients are found, the correction can be made using eq. 2:

$$R_P(\lambda) = R_m(\lambda) + B_0 X_0(\lambda) + B_1 X_1(\lambda) + \cdots + B_6 X_6(\lambda) \tag{2}$$

Another correction method was given by NPL [1]. The method includes three correction steps using three Ceramic tiles: Black, White and Grey.

NPL Correction Method:

Step 1: Zero point error correction

$R_{c,1}(\lambda) = R_m(\lambda) + c_1(\lambda)$, with $c_1(\lambda) = R_T(\lambda,b) - R_m(\lambda,b)$

Here, $R(\lambda,b)$ means the reflectance of the Black tile.

Step 2: Scale error correction

$R_{c,2}(\lambda) = R_{c,1}(\lambda)(1 + c_2(\lambda))$, with $c_2(\lambda) = [R_T(\lambda,w) - R_{c,1}(\lambda,w)]/R_{c,1}(\lambda,w)$

Here, $R(\lambda,w)$ means the reflectance of the White tile.

Step 3: Linear error correction

$R_{c,3}(\lambda) = c_3(\lambda)R_{c,2}(\lambda) + c_4(\lambda)R_{c,2}^2(\lambda)$

with $c_4(\lambda) = [R_{c,2}(\lambda,g) - R_T(\lambda,g)]/[R_{c,2}(\lambda,g)(100 - R_{c,2}(\lambda,g))]$,

$c_3(\lambda) = 1 - 100c_4(\lambda)$. Here, $R(\lambda,g)$ means the reflectance of the grey tile.

The common idea of the above two correction models uses less reference materials as training set to correct the instrument measurements. This assumes that the correction between different measurements is material independent. This was confirmed by the work of Morovic et al. [6], where a modified Berns & Petersen model was used with the coefficients B_0 to B_6 varying with wavelength. The training data set consisted of the 12 BCRA-NPL tiles. The OSA paper samples were used for testing. They found an improvement of 75% for the training set and 45% for the testing set.

The aim of this study is to further investigate the material independent property of the correction models mentioned above.

2. EXPERIMENTAL SET UP

Four spectrophotometers were used for the measurements, which are named S_1, S_2, S_3 and S_4 respectively. 95 samples were measured using each of the four spectrophotometers. They were chosen from six different sample sets (according to material or property). The six sets are the BCRA-NPL series II ceramic matt tiles (12), the BCRA-NPL series II ceramic glossy tiles (12), thread samples (20), wool fabric samples (20), BAM/EBU standard samples (11), OSA paper samples (20).

All the samples, together with equipments, were placed in a conditioning laboratory for 24 hours before measurement. The conditionings were controlled at temperature 21 ± 2ºC and humidity 60% to 65%.

3. TESTING THE CORRELATION MODELS

The instrument performance was first analysed in terms of repeatability, accuracy and inter-instrument disagreement. It was found that three of the instruments performed well in terms of repeatability. Four ceramic tiles were measured 30 times continuously using each instrument. The maximum and average differences are 0.09 and 0.01 ΔE_{00} for three of the four instruments and one instrument had a maximum difference of 0.42 ΔE_{00} unites.

The accuracy for each instrument was measured using colour difference between the NPL standard (12 Glossy tiles) and the measurement results. It was found that the averaged accuracy error is between 0.3 and 0.4 ΔE_{00} unites for all the instruments. However, the maximum errors were ranged from 0.7 to 1.2 ΔE_{00} units.

For the disagreement of the measurements of the four instruments, it was found that S1 measurements agree well with the other three. Furthermore, according to its good repeatability and high accuracy, S1 measurement results were taken as standard in this study. The averaged and maximum measurement colour differences between S1 and others are listed in Tables 1 and 2.

Table 1: Averaged colour differences between the measurements of S1 and each of S2, S3 and S4 among each data set

	Matt	Glossy	Thread	Fabric	BAM/EBU	OSA
S1/S2	0.329	0.176	0.527	0.314	0.192	0.196
S1/S3	0.175	0.231	0.226	0.165	0.135	0.235
S1/S4	0.475	5.338	1.287	0.468	0.828	3.552

Table 2: Maximum colour differences between the measurements of S1 and each of S2, S3 and S4 among each data set

	Matt	Glossy	Thread	Fabric	BAM/EBU	OSA
S1/S2	0.444	0.233	3.980	0.579	0.297	0.307
S1/S3	0.224	0.288	0.404	0.315	0.206	0.332
S1/S4	0.918	13.35	1.819	1.368	1.654	9.493

It can be seen from the above tables, S1 agreed the best with S3, followed by S2. S1 and S4 had the largest difference amongst all. According to the maximum differences, the difference between S1 and S2 can be as high as 4 ΔE_{00} units and the difference between S1 and S4 can be as high as 13 ΔE_{00} units. This clearly demonstrates that the correction is necessary.

The Berns & Petersen, NPL and modified Berns & Petersen models were used for the corrections using the S1 measurements as standard. The averaged disagreement results from the three models' predictions against the standard are listed in Tables 3, 4, and 5 respectively.

Table 3: The colour differences between the results of the standard (S1) and the corrections of S2, S3, and S4 using Berns & Petersen method.

	Matt	Glossy	Thread	Fabric	BAM/EBU	OSA
S1/S2	**0.250**	0.364	0.752	0.507	0.407	0.476
S1/S3	0.244	0.270	0.364	0.382	0.253	0.287
S1/S4	0.581	**4.367**	**1.182**	**1.249**	0.561	**2.970**

Table 4: The colour differences between the results of the standard (S1) and the corrections of S2, S3, and S4 using **NPL** method.

	Matt	Glossy	Thread	Fabric	BAM/EBU	OSA
S1/S2	**0.271**	0.344	0.643	0.401	0.328	0.416
S1/S3	**0.111**	**0.132**	**0.191**	0.174	**0.110**	**0.127**
S1/S4	**0.202**	**4.727**	**1.105**	0.600	**0.353**	**3.085**

Table 5: The colour differences between the results of the standard (S1) and the corrections of S2, S3, and S4 using the **Modified Berns & Petersen** method.

	Matt	Glossy	Thread	Fabric	BAM/EBU	OSA
S1/S2	**0.171**	0.302	0.702	0.417	0.352	0.406
S1/S3	**0.069**	**0.125**	0.254	0.214	**0.124**	**0.175**
S1/S4	**0.211**	**4.542**	**1.186**	0.660	**0.334**	**2.989**

Comparing the results from Table 1 with Table 3, by applying the Berns & Petersen model, for correcting the S2 instrument, the improvement was achieved over the Matt data set, but there is no improvement for the other data sets. As for correcting the S3 measurements there is no improvement across all the data sets. However, four data sets were improved when correcting S4. Note only cyan tile was used for training the Berns & Petersen model from the Matt set.

Comparing the results from Table 1 with Table 4, after applying the NPL method to S2, S3, and S4 measurement results the corrections resulted in an improved accuracy over the Matt data set for all S2, S3 and S4 instruments, but had no improvements over the Fabric data set. In fact for the S3 and S4, an improvement occurred after the correction for all the data sets except the fabric dada set. Note that three neutral tiles chosen from the Matt set were used for training the NPL model. Therefore, the Matt set had the highest accuracy after the correction. The results proved that the correction model is material dependent.

Comparing the results from Table 1 with Table 5, once again after applying the modified Berns & Petersen method to S2, S3, and S4 measurement results the corrections resulted in an accuracy improvement over the Matt data set for all S2, S3 and S4 instruments, but had no improvements over the Fabric data set. In fact the model training data set used the whole Matt set. Hence it is expected to largely improve in this set for all the instruments. However, it was found that even the model was trained using 12 samples from the Matt, it cannot work for the Fabric data set. Besides, each of the other data sets was also used for training the modified Berns & Petersen model, and similar results were obtained. This once again demonstrates that the correction model is material dependent. This work disagreed with that found by Morovic, Xu and Luo [6] where the modified Burns & Petersen model gave the improvements for both the training set (ceramic tiles) and testing set(OSA).

4. CONCLUSIONS

The three correction models were tested for correcting the measurement data from three instruments over 95 samples from six different physical sample sets. The findings are summarised below:

- The NPL model performed better than the Berns & Petersen model with the the former using three tiles for training and the latter using one tile for the training.
- The modified Berns & Petersen model is the most accurate compared with NPL and Berns & Petersen models over the training data set.
- Correlation models are material dependent. This is contrary to the idea of Berns & Petersen and NPL models to use one or a few reference materials to training the model and contrary to the earlier work [6] in Derby.

References

1. FJ Verril, PJ Clarke, J O'Halloran and PC Knee, Inter comparison of colour measurements, NPL Report QU113, 1995.
2. FJ Verril, Synthesis Report, European project on Inter comparison of colour measurement, EUR14982EN, 1993.
3. RS Berns and KH Petersen, Empirical modelling of systematic spectrophotometric errors, Color Research and Application 13 (1988), 243-256.
4. AR Robertson, Diagnostic performance evaluation of spectrophotometers, presented on Advances in Standards and Methodology in Spectrophotometry, Oxford, England, 1986.
5. NPL Report, the Minutes of the FIG, Sept 12, 1998.
6. P Morovic, H Xu and M R Luo, Inter-comparison of color measuring instruments, Proceedings of International Colour Management Forum, 24-25 March, 1999, pp 43-49.