Regression based characterization of color measurement instruments in printing applications

Peter Nussbaum [a], Jon Y. Hardeberg [a] and Fritz Albregtsen [b]

[a] Gjøvik University College, P. O. Box 191, N-2815 Gjøvik, Norway
[b] Department of Informatics, University of Oslo, P. O. Box 1080 Blindern, N-0316 Oslo, Norway

ABSTRACT

In the context of print quality and process control colorimetric parameters and tolerance values are clearly defined. Calibration procedures are well defined for color measurement instruments in printing workflows. Still, using more than one color measurement instrument measuring the same color wedge can produce clearly different results due to random and systematic errors of the instruments. In certain situations where one instrument gives values which are just inside the given tolerances and another measurement instrument produces values which exceed the predefined tolerance parameters, the question arises whether the print or proof is approved or not accepted with regards to the standard parameters. The aim of this paper was to determine an appropriate model to characterize color measurement instruments for printing applications in order to improve the colorimetric performance and hence the inter-instrument agreement. The method proposed is derived from color image acquisition device characterization methods which have been applied by performing polynomial regression with a least square technique. Six commercial color measurement instruments were used for measuring color patches of a control color wedge on three different types of paper substrates. The characterization functions were derived using least square polynomial regression, based on the training set of 14 BCRA tiles colorimetric reference values and the corresponding colorimetric measurements obtained by the measurement instruments. The derived functions were then used to correct the colorimetric values of test sets of 46 measurements of the color control wedge patches. The corrected measurement results obtained from the applied regression model was then used as the starting point with which the corrected measurements from other instruments were compared to find the most appropriate polynomial, which results in the least color difference. The obtained results demonstrate that the proposed regression method works remarkably well with a range of different color measurement instruments used on three types of substrates. Finally, by extending the training set from 14 samples to 38 samples the obtained results clearly indicate that the model is robust.

Keywords: Color measurement, color measurement instrument characterization, ISO standards, print quality, process control, polynomial fitting technique, measurement uncertainties, Inter-instrument agreement.

1. INTRODUCTION

In general, process control is the basic requirement for ensuring satisfactory print and proof quality in the graphic art industry. To preserve the standardization concept colorimetric parameters and tolerance values for print and proof productions are defined in ISO 12647-2 [14] and ISO 12647-7 [15] respectively. Therefore, to ensure the quality control colorimetric values have to be obtained using color measurement instruments. Currently, there are many different models of color measurement instruments used in the printing industry, and this has been found to have significant consequences on print and proof quality [22].

In a modern color managed and standardized printing workflow, most of the printing houses use more than one color measurement instrument, typically one instrument in each department (pre-press, press, and post-press). Moreover, in the context of a Process Standard Offset (PSO) [21] certification process often the same color control wedge of a print or proof is measured first by the instrument of the company, that is to be certified, and secondly with the instrument of the certification body, to determine and confirm whether the colorimetric values are within the defined ISO tolerances.

However, measuring a control wedge with two different color measurement instruments will obviously result in different colorimetric data sets due to the nature of the instrument's uncertainties [22]. In a certification context assuming that both instruments give values that are within the given color difference tolerances according to the ISO standard, both measurements will be approved and the print or proof will be accepted. On the other hand, if one of the instruments gives values that exceed the tolerances, the question arises which of the measurement values are correct and which one has failed, even though both measurement instruments are certified. Depending on the applications and the customer's requirements the predefined ISO standard tolerances have been defined narrower to increase the print quality. Consequently, the color measurements performed with more than one instrument are even more critical in terms of the instrument uncertainty.

In the past, a number of studies have addressed the issues of color measurement instrument accuracy and uncertainties. For more details on assessment of color measuring instruments in general and inter-instrument reproducibility in particular see the works of Billmeyer [7], Briggs *et al.* [9], Billmeyer and Alessi [8], Rodgers *et al.* [25] and Wyble and Rich [29]. In a study by Rich *et al.* [24] the authors have observed that the differences between pairs of instruments can be quite significant, with maximum differences of up to ΔE^*_{ab} of 4.0. In a previous work by Nussbaum *et al*. [22] the authors conclude that in order to reduce the measurement errors in a color managed printing workflow the use of only one instrument product family (instruments of the same model from the same manufacturer using equal parameters) is recommended. However, due to a number of different reasons this advice seems to be rather difficult to implement in the daily printing production environment. A further technique to reduce the color differences obtained by measuring the same sample using more than one measurement instrument is applying a correction method to the obtained color measurements.

Therefore, the aim of the present work is to propose a method to reduce the variations in color measurement performed with more than one instrument measuring the same color target. In particular, the main contribution of this study is in characterizing measurement instruments using a colorimetric regression technique. Finally, the appropriate correction model applied to the measurement data sets will reduce the color errors between the measurements obtained by a master instrument and the measurements performed by a second instrument used (Figure 1). Consequently, the model will improve the colorimetric performance and inter-instrument and inter-model agreement.

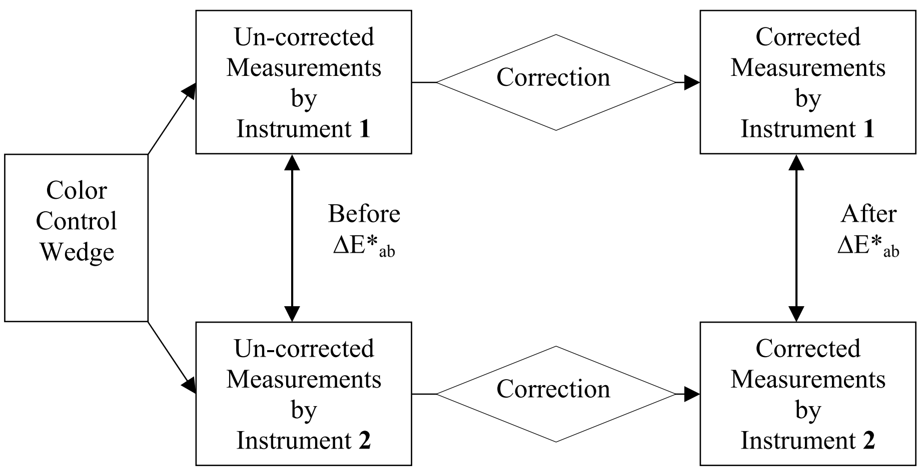

Figure 1. A schematic diagram of the color measurement workflow using two measurement instruments measuring a color control wedge, and applying a correction model to reduce the errors between instrument 1 and instrument 2.

In order to determine the performance of measurement instruments there are a number of parameters to consider. According to ASTM E2214 [3] the most important specification is the *repeatability* which defines how well an instrument repeats its reading of the same target over a certain period of time. *Reproducibility* is a form of repeatability in which one or more of the measurement parameters have been systematically changed, such that the target is being different, the time frame of measurements are being very long or the operator has being changed. *Inter-instrument agreement* describes the reproducibility of two or more instruments of the same design and *inter-model agreement* describes the reproducibility of two or more instruments of different design. Finally, *accuracy* describes the conformance of a series of readings to the accepted or true value. The measurement variations between instruments can be divided into systematic and random errors. According to Berns [6], repeatability is affected by random errors including drift, electronic noise and sample presentation. Random variations are difficult to avoid. On the other hand the accuracy is affected by systematic errors, which among other characteristics may be due to different measurement geometry, or detector linearity errors resulting from a change of wavelength. In the past several attempts have been made to reduce the systematic errors by characterizing color measurement instruments. The study by Berns [6] proposes the correction of various systematic errors applying to the spectral measurements data using multiple linear regression based on modeling the results to improve the colorimetric performance. In this work, however, the aim is to correct the instrument's systematic errors by applying a regression technique directly to the measured CIELAB data, and hence improve the inter-instrument and inter-model agreement.

Following this brief introduction including the definition of the aim in this work and discussing central references, we provide some more methodology information in Section 2, by illustrating the regression model, defining key concepts, and the experimental procedure including data collection. Then, in Section 3 we present and discuss our results, before concluding in Section 4.

2. METHODOLOGY

The method we propose in this work is based on color image acquisition device characterization, which has been applied by implementing polynomial regression with a least square technique [12]. The purpose of the characterization model of a color measurement instrument is to predict color measurement data from a given set of reference data (training set). Essentially, the derived model according to polynomial fitting technique describes the colorimetric relationship between a given sample set of reference data and the corresponding measurements taken by an instrument. Consequently, the derived model is applied to another set of measurements (test set) obtained by the same instrument. The obtained new data set is a corrected version according to the used regression model, as depicted in Figure 2. In this work, for each instrument a separate model has been derived and consequently applied to the test set to correct the measurement data set. It is assumed that by modeling the systematic errors of the measurement instruments the results will improve the colorimetric performance and hence the inter-instrument and inter-model agreement. In other words, the color difference between two corrected measurement data sets will be reduced.

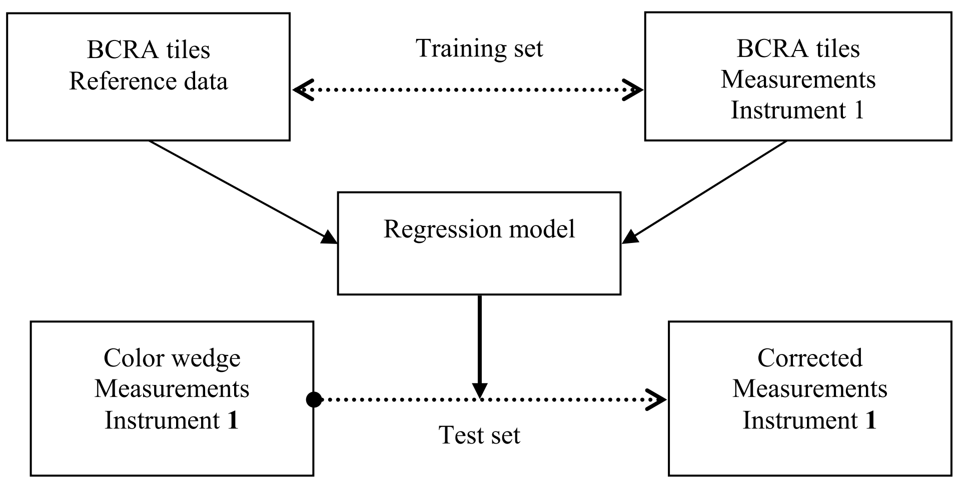

Figure 2. Schematic diagram of the regression method using a training set and a test set.

A set of 14 British Ceramic Research Association (BCRA) Ceramic Color Standards Series II (CCS II) ceramic gloss tiles including one Black and one White BCRA tile and printed substrates were measured using six spectrophotometers, according to the procedures outlined by ISO 13655 [16]. The instruments used are commercial industrial-oriented spectrophotometers typical utilized for daily production control in prepress and press applications. A spectrophotometer measures the ratio of reflected to incident light (the reflectance) from a sample at many points across the visible spectrum [5]. Table 1 presents the instruments employed including the corresponding specifications. Some of the instruments are typically from the same model and some of them represent different models from different manufacturers. According to Nussbaum *et al.* [22] the instruments used in this study show an acceptable performance in terms of repeatability and reproducibility.

Table 1. Overview of the six instruments used in this work and the corresponding specifications.

	Measuring without replacement	**Aperture**	**Measuring geometry**	**Manufacturer's claimed precision**	**Spectral range and interval**	**Manufacturer's claimed short term repeatability**

Master instrument	Yes	4mm	45°:0°	Mean ΔE^*_{ab} 0.3 Max ΔE^*_{ab} 0.8 on 12 BCRA tiles	380nm to 730nm at 10nm	ΔE^*_{ab} 0.02 (Standard shift from 10 measurements at 10 sec. interval on white)
*Secondary **E***						
*Secondary **A***	No	2mm	45°:0°	Mean $\Delta E^*_{94} < 1.0$ on 12 BCRA tiles	380nm to 780nm at 10nm	$\Delta E^*_{94} < 0.2$
*Secondary **D***	Yes	4.5mm	45°:0°	Mean ΔE^*_{ab} 0.3 on 12 BCRA tiles	380nm to 780nm at 10nm	ΔE^*_{ab} 0.02 (Standard shift from 10 measurements at 10 sec. interval on white)
*Secondary **B***	Yes	4.5mm	45°:0°	Mean ΔE^*_{94} 0.4 Max ΔE^*_{94} 1.0 on 12 BCRA tiles	380nm to 780nm at 10nm	$\Delta E^*_{94} < 0.1$ (From 10 measurements at 3 sec. interval on white)
*Secondary **C***						

One particular measurement instrument has been defined as a reference and is called the ‘master instrument’. It is worth pointing out that the chosen reference instrument is not meant to represent the best or ideal instrument but instead to be a state to which we compare the measurements obtained by other instruments. The corrected measurements obtained from the regression model were then used as the starting point with which the corrected measurements from other instruments were compared to find the polynomial which gives the least color difference. The other five instruments are called ‘secondary’s’. Note that because one of the manufacturers is requesting not to publish their name the devices were anonymized by identifying the instruments as ‘master instrument’, ‘secondary A’, ‘secondary B’ …’secondary E’.

2.1 Experimental procedure

To create the characterization model a training set is required. This should consist of a reference data set and the corresponding measurements performed by a measurement instrument. As seen above, ASTM [3] defines accuracy as the conformance of a series of measurements to the accepted value for a given sample. In other words how closely an instrument can conform to a certain reference. In this study the reference values have been provided by the manufacturer of the BCRA tiles. All measurement instruments have performed first a normal calibration procedure according to the manufacturer’s recommendations before measuring the tiles 15 times in a sequence.

Note that not only the precision between the instruments has to be improved but also the accuracy in terms of the appropriate colorimetric values. Therefore, the BCRA tile reference values, which are traceable, have been used to establish the model. The idea of using the 14 BCRA tiles for the training set is to make the appropriate adjustments in terms of the accuracy. Eventually the derived model has been tested using color measurements of 46 patches of the UGRA/FOGRA Media Wedge CMYK [26] on three different types of printed substrates (Figure 3).

The first paper substrate was a hard-copy digital proof print, printed according to the ISO 12647-7 graphic art standards for paper type 1 simulation by a commercial printing house. The second paper substrate was paper type 1 printed by the same commercial printing house aiming at the ISO 12647-2 graphic art standards. And the third paper substrate was paper type 5 Altona Test Suite reference print [23]. The CIELAB target values of the UGRA/FOGRA media Wedge CMYK are based on print conditions as stated in ISO 12647-2 to and the appropriate characterization tables for different paper types are provided by Fogra [19].

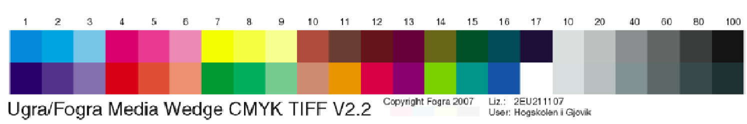

Figure 3. UGRA/FOGRA Media Wedge CMYK.

2.2 Polynomial regression

Essentially, polynomial regression is needed whenever you have two data sets, which are related and you are aiming to predict one of them if you know the other, typically device dependent colors from device in-dependent color and vice versa. Polynomial device characterization technique with least squares fitting for different application has been adequately explained and applied in a number of studies by Kang [18], Sharma [27], Hong et al. [13] and Johnson [17]. However, a brief description is given to demonstrate this technique applied to color measurement instruments.
The first data set, on which the regression model is based, is referred to as the training set and the second data set, which was not involved in deriving the model is used for evaluating it, and is referred to as the test set. As depicted in Figure 2 certain BCRA tiles sample colors are selected and defined as the reference values and the measurement instruments used

were measuring the corresponding color specifications. For simplification purposes let's call the BCRA reference values 'REF' and the corresponding measurement values from the instrument 'INS'.

Assume that the reference target has N samples. For each color sample the corresponding reference values R, E and F are represented by a 1 x 3 vector p_i ($i = 1 \ldots N$) and their corresponding I, N and S color measurement values obtained by the measurement instrument are represented by a 1 x 3 vector x_i ($i = 1 \ldots N$). Suppose that only I, N and S values are used in p, the transformation between INS and REF is a simple linear transform. However, the reason for using polynomials is that vector p_i can be extended by adding more terms such as I^2, N^2, S^2 etc. which may improve the accuracy of the model in terms of reducing the color differences over all the color samples [13]. While higher order polynomials will give a perfect fit to the data of the training set, this may result in over fitting as well as causing oscillations between the points, Runge's phenomenon [28]. Hence, in this work we applied only second order polynomial even for a 3 x 11 matrix.

The polynomials applied and analyzed in this work have the following form:

1. p_i = [I N S]
2. p_i = [I N S 1]
3. p_i = [I N S INS 1]
4. p_i = [I N S IN IS NS]
5. p_i = [I N S IN IS NS INS 1]
6. p_i = [I N S IN IS NS I^2 N^2 S^2]
7. p_i = [I N S IN IS NS I^2 N^2 S^2 INS 1]

Suppose **R** denotes an *3 x N* matrix of vectors p_i and **X** the predicted matrix of vector x_i. The mapping from INS to REF can be expressed by

$$X = M*R \tag{1}$$

M is the unknown transformation matrix that determines the accuracy of the model, which means minimizing the color differences over all color samples. The differences between Y' and Y can also be expressed as the Sum of the Squares of the Differences (SSD):

$$SSD = \sum_{i=1}^{n} \left(Y_i' - Y_i\right)^2 \tag{2}$$

where $Y' = MX$

$$SSD = \sum_{i=1}^{n} \left(MX_i - Y_i\right)^2 \tag{3}$$

Depending on the polynomial being solved the size of the matrix **M** in this work varies from 3 x 3 up to 3 x 11. On the following 1st order sample it is shown how the model can be derived.
Forward model: [R E F] = [I N S IN IS NS INS 1]***M**

In this case **M** is an 8x3 transformation matrix that contains the model parameter calculated from the training set by the equation:

$$M = (R^T*R)^{-1}*R^T*X \tag{4}$$

where $\mathbf{R}^T$ denotes the transpose of **R**, and $\mathbf{R}^{-1}$ denotes the inverse. In this example **R** is an n x 8 matrix which contains values of the I N S samples as well as corresponding IN, IS, NS, INS and 1 values calculated from them for each sample. **X** is an n x 3 matrix which contains the number of samples n used in the training set and the columns accommodate R, E and F values of all the samples.

The regression model is based on the training set containing 14 BCRA reference values and the corresponding measurements obtained by the measurement instruments from the BCRA tiles. The performance and accuracy of the characterization model has to be evaluated using an independent data test set, which in this work is represented by measurements of the 46 patches (UGRA/FOGRA Media Wedge) on different substrates. The best results with the least color differences are obtained by experimentation.

Finally, the most appropriate transformation matrix **M** is the one that results in the least color difference between the corrected measurements of the 'master instrument' and the corrected measurements of the 'secondary instruments'. Note that in this study CIELAB values are directly used in the characterization and evaluation procedure because CIELAB values have been reported from the spectral reflectance data initially measured by the instruments. Furthermore, the ISO tolerances given in the standards are communicated in CIELAB color space as well. Moreover and most important, Euclidian distance in CIELAB color space is corresponding quite well to the perceptual color differences [12].

2.3 Data collections

All instruments used in this study measured spectral reflectance factor values from 380nm to 730nm with 10nm intervals. Spectral measurements were converted to CIEXYZ tristimulus values according to the CIE 1931 2° observer and the CIE Standard illuminant D50 using the method proposed by ASTM 308, Table 1 [1]. Furthermore, to use a visually meaningful color space CIELAB (D50 as the reference white) values were calculated according to CIE 15 [10] specifications. Consequently, CIELAB data have been used for the regression model and colorimetric difference ΔE^*_{ab} values were computed between the master measurement instrument and the secondary instruments. Furthermore, the obtained results will be compared with the ISO tolerances. Because the colorimetric production control tolerances in the ISO standard 12647-2 and ISO standard 12647-7 are defined with ΔE^*_{ab} only, no further color difference metrics are used in this work.

3. RESULTS AND DISCUSSIONS

As mentioned previously the aim is to find a method to reduce the color difference between instruments measuring the same color patches. Furthermore the applied model shall improve the colorimetric performance and inter-instrument and inter-model agreement on three different types of substrates.

Figure 4 shows the color difference results between the BCRA tiles reference values and the ‘master instrument’. Moreover, the color difference results between BCRA tiles reference values and the ‘secondary instrument A’ and between the ‘master instrument’ and the ‘secondary instrument A’. Notice, that the ‘master instrument’ and the ‘secondary instrument A’ are not from the same instrument family (which in a practical application very often can be the case). We see that the color difference between the BCRA tiles reference values and the ‘master instrument’ has the highest values in the red (ΔE^*_{ab} 2.8 units), orange (ΔE^*_{ab} 4.2 units) and bright yellow (ΔE^*_{ab} 2.4 units) tiles. Comparing the BCRA tiles reference values with the ‘secondary instrument A’ only the bright yellow tile shows a rather high color difference value (ΔE^*_{ab} 2.4 units). On the other hand, comparing the measurement results between the ‘master instrument’ and ‘the secondary instrument A’ the results on the red, orange and bright yellow tiles again show very large color differences. Moreover, although the ‘master instrument’ and the ‘secondary instrument A’ show almost identical color difference compared to the BCRA tiles reference values on the bright yellow tile (approximately ΔE^*_{ab} 2.4 units), the direct comparison shows the largest color difference of ΔE^*_{ab} 4.5 units. This indicates that the accuracy of both measurement instrument, ‘master instrument’ and the ‘secondary instrument A’ on the bright yellow tile can be considered as very similar, However, the color difference between the ‘master instrument’ and the BCRA tiles reference values and between the ‘secondary instrument A’ and the BCRA tiles reference values points in different directions.

It is important to consider the inherent physical properties of the BCRA tiles. Fairchild and Grum [11] stated that the BCRA tiles red, orange and yellow can exhibit appreciable thermochromism due to sharp changes in their spectral reflectance curves. Based on this finding Berns [6] argued against using the tiles red, orange and yellow unless the temperature of the tiles at the time of calibration was known and this temperature was maintained both at the location where the tiles would be used and during their measurements. However, according to the results shown in Figure 4 there is no clear evidence of thermochromism for the ‘secondary instrument A’ except for the yellow tile. In contrast, the ‘master instrument’ demonstrates larger color differences due to possibly generating significant heat in the measuring process. According to Fairchild and Grum [11], it is important to make sure that the temperature of calibration standards remains constant during their use. On the other hand, no significant color changes have been observed with small temperature changes around room temperature. Furthermore, all measurements in this study have been conducted in the same location and the same room temperature conditions.

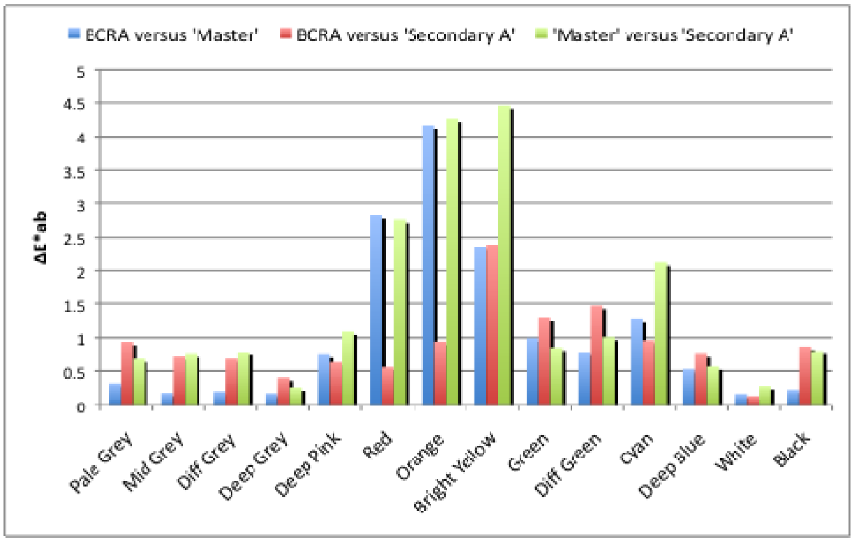

Figure 4. Color difference between the 14 BCRA tiles and the 'master instrument', BCRA tiles and the 'secondary instrument A' and between the 'master instrument' and the 'secondary instrument A'.

The impact of such a systematic error due to the nature of different instrument properties is illustrated in Table 4 which shows the color difference results between the measurements performed by the 'master instrument' and the 'secondary instrument A'. The measurements were conducted by measuring the control media wedge on proofing substrate. Additionally, the table indicates the defined ISO tolerances according to ISO 12647-7. It can be seen that the 'master instrument' gives values which qualify the proof as approved. On the other hand, the 'secondary instrument A' gives values which are outside the given tolerance (orange marked values) and therefore the proof might be not accepted. Moreover, the maximum color difference is clearly seen in the yellow color, which exceed the color difference tolerance of ΔE^*_{ab} 5 units. This is perhaps not unexpected due to the large difference in terms of the accuracy performance between the two measurement instruments on the bright yellow BCRA tile, seen previously. Nevertheless, the question may arise, whether the proof is really generated very close to the standard values, or it is just outside the colorimetric tolerance obtained by the two measurement instruments, which varies in terms of accuracy conformance.

In a real practical application this situation can be considered as very inappropriate, where due to systematic errors of the measurement instruments the one instrument results in approved and the other one not. Therefore in a first attempt the measurement instrument accuracy has to be ensured by modeling the relationship between the BCRA tile reference values and the actual measurements performed by the measurement instruments.

Table 2. Results from the regression using CIELAB values according to 14 BCRA tiles references values and the corresponding values of the 'master instrument' and the 'secondary instrument A'.

Trainings data set 14 BCRA tiles	**'Master instrument' versus 'Secondary instrument A'**		**BCRA (reference) versus 'Master instrument'**		**BCRA (reference) versus 'Secondary instrument A'**	
	Average ΔE^*_{ab}	Max ΔE^*_{ab}	Average ΔE^*_{ab}	Max ΔE^*_{ab}	Average ΔE^*_{ab}	Max ΔE^*_{ab}
Real ΔE^*_{ab}	1.47	4.45	1.07	4.17	0.91	2.39
Matrices						
3 x 3	0.44	1.20	0.43	1.02	0.29	0.63
3 x 4	0.45	1.10	0.44	1.01	0.28	0.64
3 x 5	0.43	1.07	0.34	0.84	0.21	0.41
3 x 6	0.29	0.49	0.18	0.25	0.16	0.25
3 x 8	0.15	0.30	0.12	0.25	0.09	0.20
3 x 9	0.11	0.39	0.10	0.21	0.06	0.21
3 x 11	0.05	0.12	0.06	0.18	0.03	0.10

Table 2 shows the results of the training set with 14 samples using polynomial regression minimizing the difference between reference values and corresponding measurements. As expected, the higher the degree of the polynomial the more reduction in the difference in terms of average ΔE^*_{ab} and maximum ΔE^*_{ab}. The regression is a function of CIELAB reducing differences in systematic errors to extremely low levels.

The derived training functions are used to correct the colorimetric values of the UGRA/FOGRA Media Wedge measured by the 'master instrument' and the 'secondary instrument A' on three different type of substrates. To determine the appropriate function in terms of the least color difference between the corrected measurement data conducted by the 'master instrument' and the 'secondary instrument A', all calculated training functions have been applied to the test data of the 46 color patches of the UGRA/FOGRA Media Wedge. Figure 5 shows the results (Mean and Max ΔE^*_{ab}) on proof substrate using different correction matrices on a 3-D surface. The horizontal axis and the depth axis indicate the polynomials used for the correction of the 'master instrument' and the 'secondary instrument A'. The vertical axis point out the corresponding corrected color differences. It is important to note, that the uncorrected color difference between the 'master instrument' and the 'secondary instrument A' on proofing paper results in Mean ΔE^*_{ab} 2.59 units and Max ΔE^*_{ab} 6.92 units respectively.

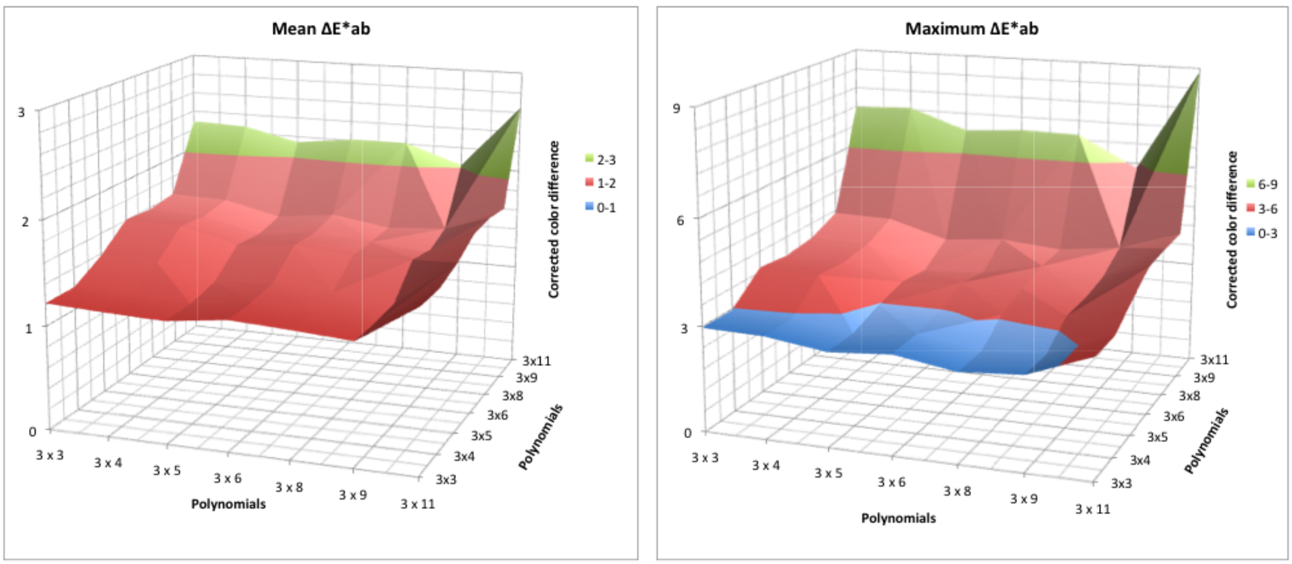

Figure 5. Results from the regression applying to the test set of the 'master instrument' and the 'secondary instrument A' on proofing substrate.

Further, it can be seen from Figure 5, that the systematic error of the 'master instrument' and the 'secondary A' can be corrected and the color differences reduced by applying almost all functions to the uncorrected measurement data. The goal is to find a connection between modeling the least color difference and the number of terms in the matrices. As can be seen there are a number of functions, from simple *3 x 3* to *3 x 9* polynomials, which reduce the color differences with more than half compare to the uncorrected data. Although, as seen in Table 2, higher polynomial functions perform excellent in the training set, testing the functions on the test set, the results from the test set indicate that using the higher polynomials (e.g. *3 x 11*) on the training set has resulted in over fitting. Therefore, it is generally recommended to use the smallest number of the polynomial terms which adequately fits the function while still smoothing out the noise [4]. Though, there is no single function performing significant best the *3 x 4* or *3 x 5* polynomials can be considered as most appropriate for both the 'master instrument' and the 'secondary instrument A'.

To find the relationship of the color differences in terms of different color attributes between the 'master instrument' and the 'secondary instrument A' of the uncorrected and corrected data set, ΔE^*_{ab} versus lightness (L^*), ΔE^*_{ab} versus chroma (C^*_{ab}), and ΔE^*_{ab} versus hue-angle (h^*_{ab}) are plotted in Figure 6.

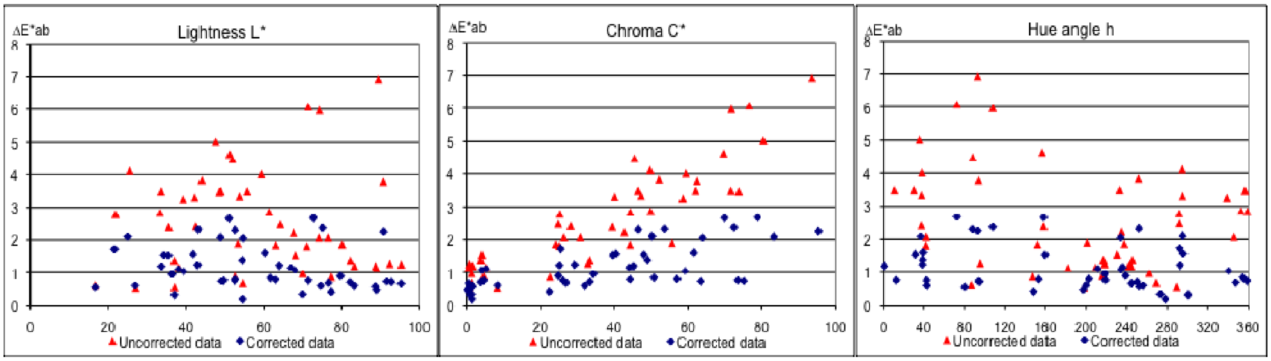

Figure 6. Color difference distribution between the 'master instrument' and the 'secondary instrument A' of the uncorrected and corrected data set: plot of lightness vs. ΔE^*_{ab}; plot of chroma vs. and ΔE^*_{ab} ; plot of hue angles vs. ΔE^*_{ab}.

The results of the corrected data set were obtained when a *3x5* matrix was used. Figure 6 illustrates the color difference distribution between the uncorrected and the corrected data set, which gradually affect the color attributes lightness, chroma and hue angle. It can be seen that larger color differences, especially in highly saturated colors, have been reduced significant by the correction model. As expected, the largest color difference in the uncorrected data set is in the yellow color (ΔE^*_{ab} 6.92 units). After the correction is applied to the measurements of the 'master instrument' and the 'secondary instrument A' the color difference is reduced to ΔE^*_{ab} 2.24 units.

To further evaluate the performance of the proposed model the same procedure has been applied for substrate paper type 1 and paper type 5. Moreover, the proposed technique has been used and tested for all the 'secondary instruments'. Table 3 presents the results from the regression and the corresponding color difference between the 'master instrument' and the 'secondary instruments (A-E)' on different types of paper substrates. In addition, the color differences between the 'master instrument' and the 'secondary instruments (A-E)' of the uncorrected measurement data are presented. Except for 'secondary instrument A' the color differences of the uncorrected measurement data are smaller between the other 'secondary instruments (B-E)' and the 'master instrument'. This is due to the different instrument product family. However, in addition to the proofing substrate the model is performing very well in terms of reducing the color differences on the two other substrates. As discussed previously, there is no polynomial which performs significantly best. Very small differences in the results among the functions can be observed. Similar performance patterns in terms of the size of the function can be found from the other 'secondary instruments (B-E)'. Note, that the polynomials for both the 'master instrument' and the 'secondary instrument (A-E)' have almost the same size for all three substrates. For the proofing substrates, the applied correction has reduced the color difference both in terms of mean and maximum for all the 'secondary instruments (A-E)'. The same is true for the results from the substrate paper type 5. For substrate paper type 1 and the 'secondary instrument B', 'secondary instrument D' and 'secondary instrument E' the correction model is not performing as good (slightly higher maximum values).

Table 3. Results from regression and the corresponding color difference between master instrument and the secondary instruments.

Master instrument versus	**Type Substrate**	**ΔE^*_{ab} Un-corrected**		**Polynomial**		**ΔE^*_{ab} Corrected**	
		Max	**Mean**	**Master**	**Secondary**	**Max**	**Mean**
*Secondary **A***	Proofing	6.92	2.59	3 x 4	3 x 5	2.67	1.17
	Paper type 1	6.14	2.36	3 x 4	3 x 4	2.09	0.98
	Paper type 5	4.38	1.72	3 x 5	3 x 5	2.34	1.18
*Secondary **B***	Proofing	1.21	0.47	3 x 5	3 x 5	0.70	0.34
	Paper type 1	1.28	0.47	3 x 5	3 x 5	1.59	0.43
	Paper type 5	1.25	0.51	3 x 5	3 x 5	1.09	0.41
*Secondary **C***	Proofing	4.67	2.48	3 x 4	3 x 6	4.49	2.26
	Paper type 1	4.67	2.18	3 x 4	3 x 5	4.48	1.89

	Paper type 5	1.56	0.82	3 x 8	3 x 8	1.42	0.71
*Secondary **D***	Proofing	1.41	0.48	3 x 5	3 x 5	1.07	0.41
	Paper type 1	1.08	0.56	3 x 6	3 x 6	1.59	0.75
	Paper type 5	1.82	0.77	3 x 5	3 x 4	1.71	0.94
*Secondary **E***	Proofing	0.62	0.33	3 x 4	3 x 3	0.61	0.25
	Paper type 1	0.66	0.32	3 x 5	3 x 5	0.83	0.28
	Paper type 5	0.67	0.34	3 x 6	3 x 6	0.60	0.32

As indicated previously instruments can be divided into product families which are instruments of the same model from the same manufacturer using equal specifications. Hence, the least inter-instrument color differences can be expected within a product family. In this work, the 'secondary instrument E' is the same model as the 'master instrument'. Regardless of the very small color differences of the uncorrected measurement data, applying the regression model has further minimized the color differences. It is noticeable that, taking the inter-instrument agreement from Table 1 into account which specifies the equipment accordance with mean ΔE^*_{ab} 0.3 units and maximum ΔE^*_{ab} 0.8 units on 12 BCRA tiles ceramics the proposed method is performing reasonable. On the other hand, the model is not able to handle the instrument's repeatability issues such as drift over time.

The 'secondary instrument B' and 'secondary instrument C' are considered as the same instrument model too. However, the 'secondary instrument C' shows rather large color differences, in particular the values given for proofing substrate and substrate paper type 1. The reason is an UV cut filter attached to the instrument which causes the color differences due to the concentration of optical brighteners to affect the CIE b^* values in the measurements. Such variations are not considered as systematic errors. Therefore, the applied model is not performing as expected, in terms of reducing the color differences.

In the context of quality control using more than one measurement instrument in the workflow, the proposed method can improve the inter-instrument and inter-model agreement significant. Table 4 shows the color differences on proofing substrate according to the uncorrected measurement data of the master instrument and the secondary instruments A-E. Furthermore, the orange marked numbers demonstrate the values, which are outside the tolerances defined by ISO 12647-7. According to the results presented in Table 4 the measurements from the 'secondary instrument A' and the measurements from the 'secondary instrument C' the proof would not be qualified as approved. On the other hand, measurements conducted with the 'master instrument', the secondary instruments B, D and E the proof would be qualified as approved. Again, the question may arise which of the instrument gives the appropriate results? Note, that the instruments random errors including repeatability performance has been tested in a previous study and concluded as acceptable [22].

Table 4. Color difference results on proofing substrate obtained by six instruments with respect to CIELAB ΔE^*_{ab} tolerances according to ISO 12647-7 (Orange marked values are outside the ISO tolerance).

Un corrected measurements	Substrate	Mean	Max	Primaries							Composed grey
	ΔE^*_{ab} 3	ΔE^*_{ab} 3	ΔE^*_{ab} 6	ΔE^*_{ab} 5				ΔH^*_{ab} 2,5			ΔH^*_{ab} 1,5
				C	M	Y	K	C	M	Y	Average
Master instrument	1.69	1.28	3.00	0.90	1.51	0.66	1.2	0.48	1.36	0.04	1.08
*Secondary **A***	1.40	2.54	7.5	2.96	3.03	7.5	1.36	2.56	0.05	0.71	0.55
*Secondary **B***	1.40	1.12	2.67	0.66	1.07	1.48	1.10	0.31	0.92	0.06	0.71
*Secondary **C***	6.34	3.04	6.34	3.27	2.36	2.49	1.68	3.14	2.19	0.33	3.47
*Secondary **D***	0.92	1.26	2.46	0.87	1.17	2.05	1.04	0.26	0.86	0.15	0.71

Secondary ***E***	1.52	1.4	3.12	1.43	1.70	0.9	1.38	1.00	1.59	0.45	0.93

Table 5 presents the results of all instruments after applying the regression model to the uncorrected data set. Although the 'secondary instrument A' results in values which now qualify the proof as approved it is important to emphasize that it is not the intention of the proposed method to get the values as close as possible to the ISO standard values but to reduce the color difference between the instruments.
Furthermore, it can be seen that the 'secondary instrument C' with the UV cut filter still results in values which qualifies the proof far from approved. The variations between the 'secondary instrument C' and the other instruments are large, especially the results obtained on the substrate and the composed grey. As stated previously, this effect of variation is considered as a systematic error and therefore the regression method is not handling this.

Table 5. Corrected measurement data set (based on 14 training samples) of seven instruments with respect to CIELAB ΔE^*_{ab} tolerances according to ISO 12647-7.

14 training samples	Substrate	Mean	Max	Primaries							Composed grey
Corrected Measurement data set	ΔE^*_{ab} 3	ΔE^*_{ab} 3	ΔE^*_{ab} 6	ΔE^*_{ab} 5				ΔH^*_{ab} 2,5			ΔH^*_{ab} 1,5
				C	M	Y	K	C	M	Y	Average
Master (3x4)	1.56.	1.55	3.22	1.96	1.81	2.87	1.25	0.57	0.98	0.19	0.78
Secondary ***A*** *(3x5)*	0.92	1.83	4.42	2.12	1.85	4.56	1.17	2.00	0.97	0.13	0.41
Secondary ***B*** *(3x5)*	1.40	1.33	3.39	0.90	0.94	2.62.	1.17	0.65	0.31	0.57	0.59
Secondary ***C*** *(3x6)*	6.04	3.01	6.04	1.90	2.90	3.89	1.64	1.68	2.82	0.35	3.20
Secondary ***D*** *(3x5)*	0.98	1.22	2.97	0.91	0.76	2.36	1.00	0.74	0.22	0.51	0.64
Secondary ***E*** *(3x3)*	1.45	1.60	3.40	1.80	2.05	2.75	1.4	0.28	1.56	0.27	0.68

So far the training set for building the model was limited to 14 samples (14 BCRA tiles). To test if the model will improve the performance in terms of reducing the color difference between the 'master instrument' and the 'secondary instrument (A-E)' the sample number of the training set has been increased with 24 patches from the ColorChecker, which is a color rendition chart including traceable reference values [20]. Consequently, the regression method has been applied again for all the measurement on all three substrates. Although there is no significant improvement in terms of reducing the mean color difference, the maximum ΔE^*_{ab} could be reduced substantially in all instrument combinations and all three substrates. Moreover, also functions with second order polynomial (such as *3x11*) give reasonable results reducing the color difference significantly, in particular the maximum color difference. This indicates clearly that the model with 38 sample points is more robust.

Table 6. Corrected measurement data set (based on 38 training samples) of seven instruments with respect to CIELAB ΔE^*_{ab} tolerances according to ISO 12647-7.

38 training samples	Substrate	Mean	Max	Primaries							Composed grey
Corrected Measurement data set	ΔE^*_{ab} 3	ΔE^*_{ab} 3	ΔE^*_{ab} 6	ΔE^*_{ab} 5				ΔH^*_{ab} 2,5			ΔH^*_{ab} 1,5
				C	M	Y	K	C	M	Y	Average
Master (3x4)	0.60	1.24	3.05	2.16	1.19	2.38	1.37	1.42	0.09	0.92	0.56
Secondary ***A*** *(3x3)*	1.00	2.02	4.29	3.00	1.94	4.02	1.68	2.69	1.02	0.30	0.39

*Secondary **B** (3x4)*	0.46	1.30	3.16	2.01	1.62	2.75	1.35	1.42	0.03	1.12	0.39
*Secondary **C** (3x5)*	4.89	2.39	4.89	1.42	1.90	3.39	1.78	1.36	1.77	1.06	2.49
*Secondary **D** (3x3)*	0.80	1.35	3.57	2.48	1.06	2.87	0.86	1.81	0.24	1.36	0.35
*Secondary **E** (3x6)*	0.52	1.38	3.62	1.08	0.84	3.17	1.28	0.89	0.41	1.51	0.31

Table 6 shows the results using the training set with 38 samples. It can bee seen that, except for the 'secondary instrument C' on composed grey tolerance, all corrected measurement values obtained will qualify the proof as approved. In other words, increasing the number of training samples in the model and applying the function to the test set (in our case to the measurements of the proofing substrate) will correct the measurements and reduce further the maximum color differences. It is important to note that by increasing the number in the training set, the model behaves more robustly in terms of using higher polynomials in correcting the test data set.

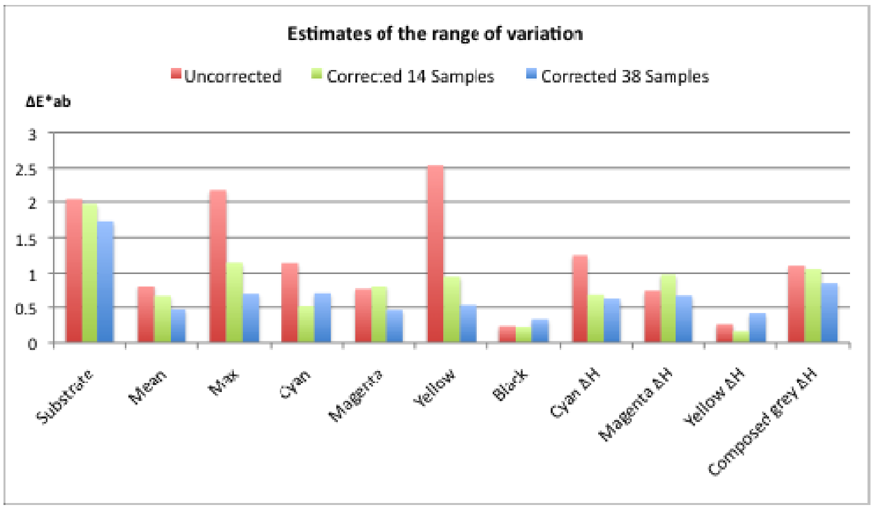

Figure 7. Range of variations with respect to uncorrected and corrected measurement data sets from all six instruments on proofing substrates.

Another way to describe the range of variation between the uncorrected and the corrected measurements is standard deviation σ. As mentioned previously the aim of the presented work is to reduce the variations related to color differences between measurement instruments in measuring the same color patches. Figure 7 illustrates the range of variations between the uncorrected and corrected measurement data sets (including 14 training samples and 38 training samples) on proofing substrates in terms of the standard deviation (according to the results given in Table 4, Table 5 and Table 6). It can be seen that the applied regression model is reducing the range of color difference variations among the six instruments compared to the uncorrected data sets. In particular the maximum values could be reduced significantly and therefore the mean results have been affected too. Looking at color difference results for the substrate it can be seen that the range of variations between the uncorrected data set and the corrected data set remains almost the same. This is again due to the 'secondary instrument C' with the UV cut filter, as explained above. Except for cyan and black, the correction model based on the training set with 38 samples is performing better compared to the training set with 14 samples.

To further verify the proposed method, two other commercial measurement instruments from different manufactures representing different models have been used. The first is considered as a spectrophotometer and the second as a spectrocolorimeter. ASTM E1347 defines a spectrocolorimeter as a spectrometer that provides colorimetric data, but not the underlying spectral data [2]. For the training set, 38 samples (14 BCRA tile and 24 ColorChecker) including the reference values and the corresponding measurements have been used to derive the model. The test data set has been extended from 46 patches (UGRA/FOGRA Media Wedge CMYK, version 2.2) to the new 72 patches (UGRA/FOGRA Media Wedge CMYK, version 3.0) and the measurements have been conducted on proofing substrates. Table 7 presents the results correcting the measurements of both instruments according to the training set for each device. As can be seen from the results, the color difference between the two instruments has been reduced by approximately 30% in both the mean and maximum respectively. The least difference has been obtained by using a *3x5* polynomial for both instruments and indicating that the applied model performs reasonable.

Table 7. Color difference results from the uncorrected and corrected measurements between two different instruments on proofing substrate.

Spectrophotometer versus Spectrocolorimeter	**Type Substrate**	**ΔE^*_{ab} Un-corrected**		**Polynomial**		**ΔE^*_{ab} Corrected**	
		Max	**Mean**	**Manu A**	**Manu B**	**Max**	**Mean**
	Proofing	3.64	1.73	3 x 5	3 x 5	2.32	1.21

It has to be mentioned that only one single measurement with each instrument on the training samples (14 BCRA tiles and 24 ColorChecker) has been conducted. Presumably, averaging multiple measurements per sample will reduce the noise, increase the performance of the model, and further reduce the color differences between the two corrected test data sets from each instrument.

4. CONCLUSIONS

It is known that the accuracy and inter-instrument and inter-model agreement of measurement instruments are limited. In this work we have described a method to correct the instrument's systematic errors by applying a regression technique directly onto the measurement output values in the CIELAB color space to improve the colorimetric performance and hence the inter-instrument and inter-model agreement.

The study compares different terms of polynomials derived using least-squares regression to determine the appropriate correction for six different measurement instrument's measured on three different types of substrates. One of the measurement instrument used has been defined as the reference instrument. Reference data from 14 BCRA tiles and the corresponding obtained measurements from each instrument has been used to derive a model. The model has been applied to a test set containing 46 measurements from the UGRA/FOGRA Media Wedge on three different substrates. To determine the most appropriate polynomial color differences have been calculated between the corrected measurements of the 'master instrument' the corrected measurement of the 'secondary instruments'. We conclude that first order polynomials (more precise *3x5* polynomial) in most cases produce the best results in terms of reducing the color differences between the instruments on different substrates.

Although there is no significant difference in the performance of the model on the three different types of substrates, the proofing substrate results in the least color differences. Moreover, with instruments from different product families the inter-model agreement can be significantly improved by applying the characterization model, reducing the color differences between the measurement instruments by more than 50%. Increasing the size of the training set from 14 to 38 samples is slightly reducing the maximum color differences, but much more important, the model's behavior is more robust in terms of different applied polynomials. To justify whether thermochromism affected the model, the BCRA tiles red, orange and yellow could be left out in the training sample.

As seen, the proposed regression method works remarkably well with a range of instruments used on the three types of substrates. However, for future work, the proposed method could by further investigated using different paper substrates (e.g. glossy paper, newspaper) and material (e.g. plastic, textile, aluminum, glass). To improve the performance of the model further extension of the sample number including different types of sample surfaces could be considered to derive the model. Furthermore, the method can be extended and tested on emission measurements using different models of spectrophotometers, spectrocolorimeters and colorimeters. Perhaps, a combination of Bern's [6] proposed method correcting various systematic errors in the spectral domain and the presented technique adjusting the output CIELAB data set may further improve the colorimetric performance and the inter-instrument and inter-model agreement. Finally, considering a relevant application, the proposed model could be implemented into a measurement software system where the correction model is directly applied to the obtained measured values from the instruments.

REFERENCES

[1] ASTM E308-08, *Standard Practice for Computing the Colors of Objects by Using the CIE-System*,American Society for Testing and Materials, West Conshohocken, PA, 2008

[2] ASTM E1347-06, *Standard Test Method for Color and Color-Difference Measurement by Tristimulus Colorimetry*,American Society for the Testing of Materials, West Conshohocken, PA, 2006

[3] ASTM E2214-08, *Standard Practice for Specifying and Verifying the Performance of Color Measuring Instruments*,American Society for the Testing of Materials, West Conshohocken, PA, 2008

[4] R. Bala, "Device characterization," in *Digital color imaging handbook*. vol. CRC Press LLC, S. Gaurav. 2003.

[5] R. Berns, *Billmeyer and Saltzman's principles of color technology*: Wiley New York, 2000.

[6] R. Berns and K. Petersen, "Empirical modeling of systematic spectrophotometric errors," *Color Research & Application,* vol. 13, pp. 243-256, 1988.

[7] F. Billmeyer Jr, "Comparative performance of color-measuring instruments," *Applied Optics,* vol. 8, pp. 775-783, 1969.

[8] F. Billmeyer Jr and P. Alessi, "Assessment of color-measuring instruments," *Color Research & Application,* vol. 6, pp. 195-202, 2007.

[9] J. Briggs, D. Forrest, M. Tse, and I. QEA, "Reliability Issues for Color Measurement in Quality Control Applications," in *IS&Ts NIP* 1998, 14 pp. 595-602.

[10] CIE15, *Colorimetry*,CIE Central Bureau, Vienna, 2004

[11] M. Fairchild and F. Grum, "Thermochromism of ceramic reference tile," *Applied Optics,* vol. 24, pp. 3432-3, 1985.

[12] J. Hardeberg, *Acquisition and Reproduction of Color Images, Colorimetric and Multispectral Approaches* Dissertation.com, 2001. ISBN 1-58112-135-0

[13] G. Hong, M. Luo, and P. Rhodes, "A study of digital camera colorimetric characterization based on polynomial modeling," *Color Research & Application,* vol. 26, pp. 76-84, 2001.

[14] ISO12647-2, *Graphic technology – Process control for the production of half-tone colour separations, proof and production prints – Part 2: Offset printing processes*,ISO, 2004

[15] ISO12647-7, *Graphic technology – Process control for the production of half-tone color separations, proof and production prints – Part 7: Proofing processes working directly from digital data*,ISO, 2007

[16] ISO13655, *Graphic technology – Spectral measurement and colorimetric computation for graphic arts images* ISO, 1996

[17] T. Johnson, "Methods for characterizing colour scanners and digital cameras," *Displays,* vol. 16, pp. 183-191, 1996.

[18] H. Kang, *Color technology for electronic imaging devices*: SPIE-International Society for Optical Engineering, 1997.

[19] A. Kraushaar. Characterization data for standardized printing conditions 2008. Available: http://www.fogra.de/

[20] C. McCamy, H. Marcus, and J. Davidson, "A color-rendition chart," *J. App. Photog. Eng ; now called "The Society for Imaging Science and Technology (IS&T)"),* vol. 2, pp. 95-99, 1976.

[21] K.-G. Nickel. (2010, November 10th). *Process Standard Offset, German Printing and Media Industries Federation (bvdm) and Graphic Technology Research Association.* Available: http://www.pso-insider.de/index.php?id=48&L=1

[22] P. Nussbaum, A. Sole, and J. Y. Hardeberg, "Consequences of using a number of different color measurement instruments in a color managed printing workflow," in *TAGA*, 2009.

[23] Print & Media Forum AG, *Altona Test Suite Anwendungspaket*,Bundesverband Druck und Medien, 2004

[24] D. C. Rich, Y. Okumura, and V. Lovell, "The Effect of Spectrocolorimeter Reproducibility on a Fully Color-Managed Print Production Workflow," in *4th European Conference on Color in Graphics, Imaging, and Vision (CGIV)*, Barcelona, Spain 2008, pp. 77-79.

[25] J. Rodgers, K. Wolf, N. Willis, D. Hamilton, R. Ledbetter, and C. Stewart, "A comparative study of color measurement lntstrumentation," *Color Research & Application,* vol. 19, pp. 322-331, 2007.

[26] U. Schmitt and F. Dolezalek, *Ugra/FOGRA Media Wedge CMYK V 2.0*,Munich, 2004

[27] G. Sharma, *Digital color imaging handbook*: CRC, Press, New York, 2003.

[28] Wikipedia. (2010, retrieved 12. November 2010). *Runge's phenomenon*. Available: http://en.wikipedia.org/wiki/Runge%27s_phenomenon

[29] D. Wyble and D. Rich, "Evaluation of methods for verifying the performance of color-measuring instruments. Part I: Repeatability," *Color Research & Application,* vol. 32, pp. 166-175, 2007.