Minimal-Effort Characterization of Color Printers for Additional Substrates

Mark Shaw, Gaurav Sharma+, Raja Bala+, and Edul N. Dalal+*
**Xerox Engineering Systems, Santa Clara, California*
+Xerox Innovation Group, Webster, New York

Abstract

The use of multiple substrates in color printers requires color characterization for each of the individual media. A full re-characterization for each substrate is measurement and labor intensive. In this paper, a variety of methods are proposed and evaluated for determining the color characterization for a new substrate based on a complete characterization on a reference substrate and a small number of additional measurements for the new substrate. This saves significant time and effort in comparison to the traditional method of repeating the color characterization for each new substrate. The methods developed and tested include model-based approaches based on Beer's law, Kubelka-Munk theory, and Neugebauer equations; and an empirical technique based on principal component analysis. Results indicate that the model based techniques offer only a small improvement over direct use of the reference characterization, whereas, the empirical technique offers a more significant improvement with as few as 16-26 measurements on the new substrate.

Introduction

Color calibration of a printer is typically a two-step process. In the first step, the printer response is characterized by printing a number of color patches with known device control values, measuring the colors obtained, and generating a characterization function that maps device control values, such as CMYK, to corresponding colors specified in a device independent color space, such as CIELAB. In the second step, the characterization function is inverted to determine the device control values required to produce a color specified in device independent color space. The final color correction that inverts the characterization function is often implemented as a 3D look-up table that maps from a device independent color space (e.g. CIELAB) to the device control values (e.g. CMYK).

One approach is to ignore the change in substrate and use the original characterization for the new substrate. This approach is far from satisfactory and will in some cases yield large color errors. The other straightforward alternative is to repeat the entire characterization procedure for each new substrate. This approach is both measurement and labor intensive and can be prohibitively costly in a system supporting many different print substrate. The intermediate approach explored in this paper is to "correct" an available reference substrate characterization for additional substrates by using a small number of additional measurements.

Color Characterization for Additional Substrates

Several approaches were developed to re-characterize the printer with a small effort when the substrate is changed. Each of these methods is based on techniques that attempt to predict the spectral reflectance of a patch with specified device control values on the test substrate from the known reflectance of the patch on the reference substrate with the same device control values and characteristics of the test substrate. These techniques include model-based approaches such as Beer's law, Kubelka-Munk theory and Neugebauer equations, and an empirical regression approach based on Principal Component Analysis.

Beer's Law

The first method investigated uses principles motivated by Beer's law.[1] The method assumes that the colorant layers have no scattering, uniform thickness, and that multiple colorant layers combine additively in spectral density. For individual colors and solid overprints, under these assumptions, the reflectance of a calibration patch under the test substrate can be obtained by multiplying the corresponding reflectance for the reference substrate with the ratio of test substrate to the reference substrate reflectance. This is shown in the equation below. The reflectance for the new substrate ($R_{2(\lambda)}$), can be obtained from the reflectance spectrum ($R_{1(\lambda)}$) of that colorant mixture printed on a known substrate with reflectance ($Rp_{1(\lambda)}$), and the reflectance of the new substrate ($Rp_{2(\lambda)}$). The method generalizes to halftone tints under the assumption of a spectral Neugebauer model (with or without an exponential Yule-Nielsen correction).

$$R_{2(\lambda)} = \left(\frac{R_{1(\lambda)}}{Rp_{1(\lambda)}} \right) \cdot Rp_{2(\lambda)}$$

Kubelka-Munk Model

When a colorant is not completely transparent, the Beer's law assumption of zero scattering does not hold and a Kubelka-Munk (KM) model that accounts for a colorant's absorption and scattering properties is more suitable. According to the KM model,[2] the reflectance of a color sample is determined by the absorption and scattering ($K_{(\lambda)}$ and $S_{(\lambda)}$) coefficients of the colorant material, the thickness X of the colorant layer and the reflectance of the substrate $Rp_{(\lambda)}$:

$$R_{(\lambda)} = \frac{\left(\frac{R_{p(\lambda)} - R_{\infty(\lambda)}}{R_{\infty(\lambda)}}\right) - R_{\infty(\lambda)}\left(R_{p(\lambda)} - \frac{1}{R_{\infty(\lambda)}}\right)\exp\left[S_{(\lambda)}X\left(\frac{1}{R_{\infty(\lambda)}} - R_{\infty(\lambda)}\right)\right]}{R_{p(\lambda)} - R_{\infty(\lambda)} - \left(R_{p(\lambda)} - \frac{1}{R_{\infty(\lambda)}}\right)\exp\left[S_{(\lambda)}X\left(\frac{1}{R_{\infty(\lambda)}} - R_{\infty(\lambda)}\right)\right]}$$

where $(R_{\infty(\lambda)})$ is the reflectance of a colorant sample of an infinitely thick sample of the colorant given by

$$R_{\infty(\lambda)} = 1 + \frac{K_{(\lambda)}}{S_{(\lambda)}} - \sqrt{\left(\frac{K_{(\lambda)}}{S_{(\lambda)}}\right)^2 + 2\left(\frac{K_{(\lambda)}}{S_{(\lambda)}}\right)}$$

The KM model works for single colorants. For solid overprints of multiple colorants we use an extension of this model derived by Hoffman[3], which predicts the reflectance of a print composed of multiple colorant layers by successively treating each predicted layer reflectance as the substrate reflectance for the subsequent layer. Even with the extension the KM model is inapplicable for halftone tints. We therefore use it with a Neugebauer[4,5] model in order to predict the spectral response of halftone color prints.

Although the use of Kubelka-Munk is relatively straightforward, it requires a-priori knowledge of many of the parameters for the model to work. This work assumes that a minimal amount of data is available and that some parameters do not change from substrate to substrate.

Neugebauer Model

Both the "Beer's law" and the KM approaches utilize only the test substrate reflectance in the prediction process and therefore do not require the measurement of any printed patches on the test substrate. A small number of measurements on the test substrate may be used to assist the model-based schemes. The Neugebauer model has been widely used to predict the colorimetric response of halftone color printers. The original model is essentially an extension of the Murray-Davies equation[6]. The reflectance of a print is predicted as the weighted average of the reflectance for the Neugebauer primaries that correspond to the prints and solid overprints for the colorants, where the weights correspond to the fractional area coverages of the Neugebauer primaries. For CMYK printers, there are 16 Neugebauer primaries, which include the unprinted substrate and all possible combinations of the four-color mixtures and the Neugebauer model is given by

$$R_{CMYK\,(\lambda)} = \left[\sum_{i=1}^{16} w_i \cdot P_{i(\lambda)}^{\;1/n}\right]^n$$

where w_i represent the fractional area of primary (i), P_i is the reflectance of the (i)th solid color, R_{CMYK} is the predicted patch reflectance, and the exponent n is an empirical Yule-Nielsen[7] correction factor included to account for light scattering within the substrate. The weights w_i are calculated from the individual colorant area coverages using the Demichel or dot-on-dot model equations.[4,5]

In the technique investigated, we explored two options: one in which the 16 Neugebauer primaries were printed and measured on the test substrate to be modeled and the second in which they were derived based on the KM model.

PCA-Based Empirical Regression

Instead of using model-based approaches, an alternative is to use an empirical technique that treats the problem of predicting the color/reflectance on the new substrate as a data-fitting problem using a limited number of measurements of corresponding patches on both substrates. The dimensionality of the problem can be reduced to tractable levels using Principal Component Analysis (PCA). The spectral reflectance data is approximated using a small number of PCA basis vectors. Each reflectance is represented as the corresponding set of weighting factors for the PCA representation. A spectral reflectance $R_{(\lambda)}$ is represented in its sampled form as a vector $\mathbf{r}$, which is approximated using a set of basis vectors obtained through PCA as

$$\hat{\mathbf{r}} = \mathbf{P}\mathbf{x}_{\mathbf{r}}$$

where $\mathbf{P}$ is the matrix with the orthonormal PCA bases as its columns, $\mathbf{P}=[\mathbf{p}_1\ \mathbf{p}_2\ \ldots\mathbf{p}_r]$ and $\mathbf{x}_\mathbf{r}=\mathbf{P}^\mathbf{T}\mathbf{r}$ is the $r \times 1$vector of PCA basis weights representing $\mathbf{r}$. In our analysis, we used $r=10$ PCA basis vectors, which significantly reduces the dimensionality from the typically 31-36 samples in $\mathbf{r}$ while still ensuring that the PCA approximation is accurate.

Based on a small set of training samples corresponding to identical CMYK values printed on the reference and test substrates, a linear transformation is estimated from the PCA representation on the reference substrate to the PCA representation on the test substrate. The estimated transformation is then applied to other reflectance measurements available on the reference substrate to estimate corresponding reflectance spectra on the test substrate. Specifically, if x is the $r \times 1$ vector of PCA basis weights for a patch printed on the reference substrate with a given set of device control values, the corresponding $r \times 1$ vector y of PCA basis weights for a patch printed on the test substrate with the same set of device control values is estimated as

$$\hat{\mathbf{y}} = \mathbf{T}\mathbf{x}$$

where $\mathbf{T}$ is an $r \times r$ matrix. If the vectors of PCA basis weights for the set of training patches printed on the reference substrate are $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, … $\mathbf{x}_K$ and corresponding vectors of PCA basis weights for the test substrate are $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$, … $\mathbf{y}_K$, respectively, the least-squares optimal linear transformation $\mathbf{T}$ is given by:

$$\mathbf{T} = \mathbf{Y}\mathbf{X}^{\mathbf{T}}\left(\mathbf{X}\mathbf{X}^{\mathbf{T}}\right)^{-1}$$

where

$$\mathbf{X} = [\,\mathbf{x}_1 \mid \mathbf{x}_2 \mid \mathbf{x}_3 \mid \ldots \mathbf{x}_K\,]$$

$$\mathbf{Y} = [\,\mathbf{y}_1 \mid \mathbf{y}_2 \mid \mathbf{y}_3 \mid \ldots \mathbf{y}_K\,]$$

and $\mathbf{X}^{\mathbf{T}}(\mathbf{X}\mathbf{X}^{\mathbf{T}})^{-1}$ represents the pseudo-inverse of the $r \times K$ matrix $\mathbf{X}$.

Experimental Results and Analysis

Twelve different substrates varying in coating characteristics, weight, and paper color were chosen for the experimental investigation. Due to spare constraints, only the results for a subset of 4 substrates are presented here. Complete results and more details can be found in a companion journal paper.[8] Table 1 lists the substrates grouping them by coated and uncoated types and indicating their specified weight in grams per square meter (gsm).

Experiments were conducted on a four color CMYK Xerographic printer. For each of the substrates included in the study, a complete printer characterization target having 289 patches was printed and measured. For each of the techniques investigated, the spectral reflectance values corresponding to the characterization target on the test substrate were estimated from the spectral reflectance values on a reference target. The color error in these estimates in comparison to the measured values was used as a figure of merit for comparing the different schemes. An exhaustive test was conducted in which each of the substrates was designated the reference substrate in turn and used to predict the measurements corresponding to the characterization target for each of the remaining 11 substrates. As indicated earlier, only a small subset of representative results are presented here to illustrate the observed trends.

The various methods utilize varying degrees of information about the colorants and the test substrate, and roughly in order of increasing effort are a) no re-characterization, b) adjustment based on Beer's law using measured test substrate reflectance, c) Adjustment based on a Kubelka-Munk (KM) Model for predicting the Neugebauer primaries and a Neugebauer model for estimating the spectra of halftone tints, d) Adjustment based on the Neugebauer model for halftones with actual measuments of the 16 Neugebauer primaries, e) Adjustment based on the PCA based empirical regression technique, where the transformation from the PCA representation on the reference substrate to the PCA representation on the test substrate was determined using a subset of the printer characterization target consisting of 26 corresponding patches on the respective substrates. There is also a sixth case, which involves complete recharacterization by measuring 289 patches on the test substrate. For the chosen figure of merit, the estimation error in this case is zero.

Figure 1a depicts a plot of the average color error of these compensation techniques for four substrates. The absicssa of the plot represents the test substrate for which reflectance values are estimated using each of the other substrates as reference substrates in turn. The ordinate represents the average color error in the estimates in CIE ΔE_{94}^* units across all the patches in the target and across the three choices of the reference substrate. The five individual curves in the plot represent the different adjustment techniques and are identified in the legend. Figure 1b shows the 95th percentile values of the color errors along the ordinate axis and has other parameters identical to Fig. 6a. From the curve corresponding to no adjustment (none), we can see that there are significant differences among the different substrates with average color difference around 3.5 ΔE_{94}^* units and a 95-percentile value for the color difference around 6.0 ΔE_{94}^* units. The curves for the minimal measurement model based approaches (Beer's Law and KM) on these plots are quite close to the case of no adjustment, indicating that these approaches offer only small improvements. The approach based on the Neugebauer model incorporating actual measurements of the 16 Neugebauer primaries offers reasonable improvement for substrates 2, 4, and 9 but only a minor improvement for substrate 6. The empirical regression technique based on PCA consistently provides significantly lower error than the no adjustment case and among the techniques investigated, it is the most successful at predicting the colors for the characterization target on the test substrate.

Figure 2 summarizes the performance of each of the techniques in a bar graph. The abscissa of the bar graph is a rough depiction of the effort required in terms of additional measurements on the test substrate and the height of the bars indicate the average and 95-percentile values of the ΔE_{94}^* color error in the prediction. At the lowest end of the effort scale is the case of using the measurements on the reference substrate directly with no adjustment and at the highest end is the case of complete re-characterization on the test substrate by re-measuring the entire characterization target. Each data point is the average across all patches. The figure shows that as the number of measurements increases, the accuracy of the spectral prediction increases.

Of the four prediction techniques implemented, the best results were obtained with the empirical regression technique using PCA. The performance of the model based Beer's law and Kubelka-Munk techniques is rather poor and offers only a small improvement over direct use of the reference substrate data. Several modeling assumptions and uncontrollable variables contribute to the poor performance of the model-based schemes. Beer's law assumes that the toners are transparent, exhibit zero scattering and have constant thickness across the different substrates. All of these are unrealistic assumptions. Kubelka-Munk, although trying to account for some of Beer's law's deficiencies by compensating for absorption and scattering coefficients separately, also assumes a planar non-interspersed configuration for the colorant layers and thicknesses for the layers that are invariant under change in substrate and constant over the spatial extent of a patch. These assumptions are unrealistic but cannot be readily improved

upon without additional measurement data for the test substrate.

Both the 'Neugebauer with measured primaries', and 'principal component analysis' techniques improve results significantly. While these techniques do require additional measurements, the number of measurements (16 and 26) is much smaller than that of the complete characterization target and can significantly reduce the effort for re-characterizing the printer in response to drift when a large number of substrates are involved.

For the PCA based regression technique, the distribution of color errors is depicted in Figure 3, where histograms of the color errors over the printer characterization target are presented for the case of no compensation and compensation based on the PCA regression technique. From the histograms, it is clear that the use of an incorrect substrate calibration will result in significant errors. The PCA substrate compensation model significantly reduces both the mean color difference and the standard deviation.

The distribution of errors in color space is further illustrated in Figs. 4 and 5, where the errors for the individual patches on the printer characterization target for a sample reference and test substrate pair are shown in CIELAB space for the case of no compensation and for the PCA based regression technique, respectively. In both figures the errors are depicted in plots of a* vs L*, b* vs L*, and a* vs b* so that the three dimensional distribution of the errors may be visualized. In the plots of Figure 4 one can see that the dominant error caused by the direct use of the reference substrate characterization on the test substrate is a decrease in lightness. One can also see that the change in substrate causes a change in the black point. If not accounted for correctly, this will result in a reduction of the substrate's characterized dynamic range. In addition to the change in lightness, systematic trends in the color errors can also be seen in the a* vs b* plot. By using the PCA based empirical regression technique to compensate for the change in substrate, it is possible to improve the substrate characterization significantly. This can be seen in Figure 5, where the errors are significantly smaller than those in Fig 4 and only relatively minor trends can be seen. The algorithm compensates for the change in black point of the substrate, and significantly reduced the magnitude of the errors in hue and chroma.

Conclusion

In this paper we evaluated several techniques for estimating the printer characterization for a new substrate based on the available characterization for a reference substrate and a small number of measurements on the new substrate. The research is motivated by the requirement for supporting a large number of substrates in a typical printing environment and providing color management capability for the multiple substrates in the presence of printer drift without excessive effort. The proposed techniques include three model-based approaches: Beer's Law, Kubelka-Munk theory, and Neugebauer model; and an empirical regression technique that exploits principal components analysis (PCA) for reducing the dimensionality of the problem to a tractable level. The techniques are experimentally evaluated using a CMYK xerographic printer. While the only knowledge required for the Beer's law and (planar) Kubelka-Munk methods is the reflectance of the new substrate, these methods provide only small improvement over the direct use of the reference substrate characterization on the test substrate. The unrealistic modeling assumptions required in the absence of additional measurement data on the new substrate are the primary reason for the poor performance of these techniques. The Neugebauer model and the empirical regression technique based on PCA each utilize more measurements, 16 Neugebauer primaries in the former case and 26 in the latter. Both provide a significant improvement over the direct use of the reference substrate characterization with the empirical regression technique providing a more consistent and better improvement than the Neugebauer model. While the techniques do require measurement of additional patches on the test substrate, the effort in measuring the additional patches is still significantly smaller than the effort required for measuring an entire printer characterization target on the test substrate.

References

1. F. W. Billmeyer and M. Saltzman, "Principles of Color Technology," Second Edition, John Wiley & Sons, New York (1981)
2. E. Allen, "Optical Radiation Measurements – Colorant Formulation and Shading," Vol. 2, Academic Press(1980)
3. K. Hoffman, "Applications of the Kubelka-Munk Color Model to Xerographic Images," BS Thesis, Rochester Institute of Technology, Center for Imaging Science, 1998, available online at http://www.cis.rit.edu/research/theses.shtml
4. R. Balasubramanian, "Optimization of the Spectral Neugebauer Model for Printer Characterization," *Journal of Electronic Imaging*, vol. **8**, no. 2, April 1999, pp. 156- 166.
5. M. Xia, E. Saber, G. Sharma, and A. M. Tekalp, "End-to-end Color Printer Calibration by Total Least Squares Regression, *IEEE Transactions on Image Processing*, vol. **8**, no. 5, May 1999, pp. 700-716.
6. J. A. C. Yule, "Principles of Color Reproduction," John Wiley & Sons, New York (1967)
7. J. A. C. Yule and W. J. Nielsen, "The penetration of light into paper and its effect on halftone reproduction," *Proc. TAGA*, pp. 65-76 (1951)
8. M. Shaw, G. Sharma, R. Bala and E, Dalal, "Paper Characterization Adjustment for Different Substrates", submitted to *Journal of Color Research and Application*.

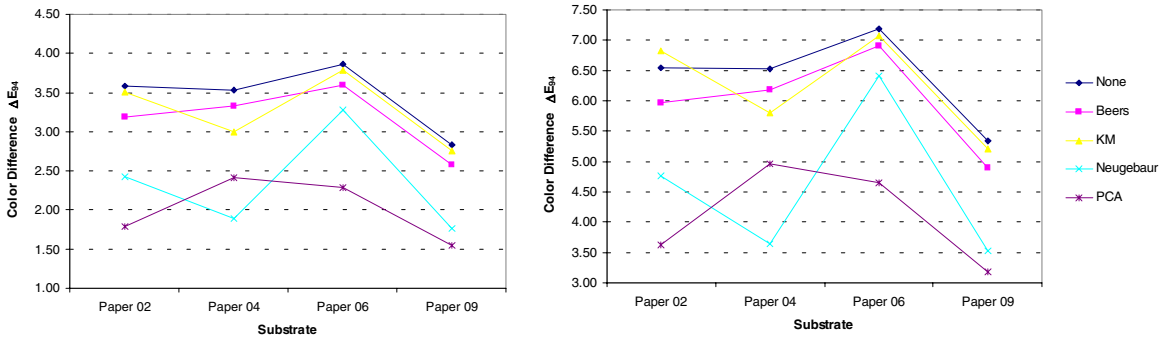
Biography

Mark Shaw received his M.S. degree in Color Science from the Munsell Color Science Laboratory, RIT and his B.S. degree in Graphic Arts Printing Technology from the

Univeristy of Hertfordshire, UK. He is employed as a Color and Imaging Scientist by Xerox Engineering Systems, Santa Clara, CA. He is a member of TAGA, ISCC, and IS&T. His interests include applied colorimetry, image rendering, color modeling and color management. In 2000 he was awarded the Grum Scholarship from the Munsell Color Science Laboratory, RIT. In has also been awarded the Varns Excellence in Printing Award, the AGFA Printing Award, and the Institute of Printing Bronze Medal.

Table 1. Names, Coating Characteristics and Weights for Four Substrates Used in the Investigation

Substrate		***Coated***	***Weight***
Paper 02	3R3874	No	90 gsm
Paper 04	Potlach Vintage Velvet Crème	Yes	80 gsm
Paper 06	Xerox Ultraspec Gloss	Yes	80 gsm
Paper 09	Consolidated Centura Gloss	Yes	80 gsm



(a) *(b)*

Figure 1. (a) Average and (b) 95 percentile color error (ΔE_{94}^{*}) for predicted test substrate reflectances.

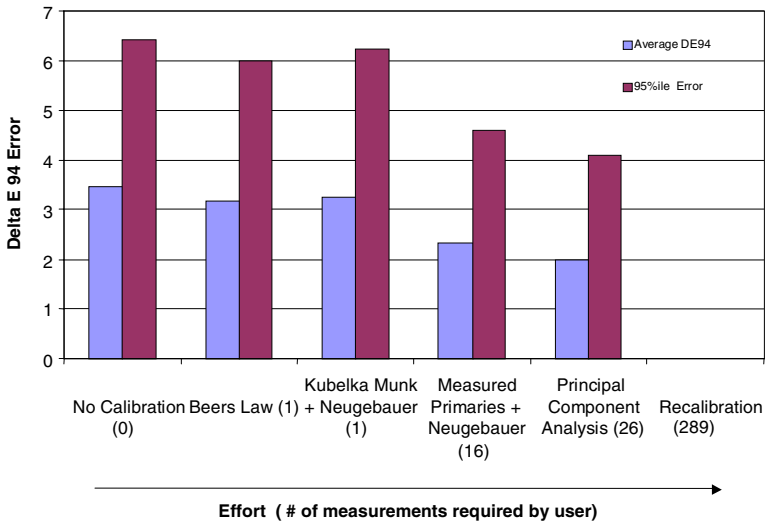


Figure 2. Effort vs. accuracy for prediction techniques.

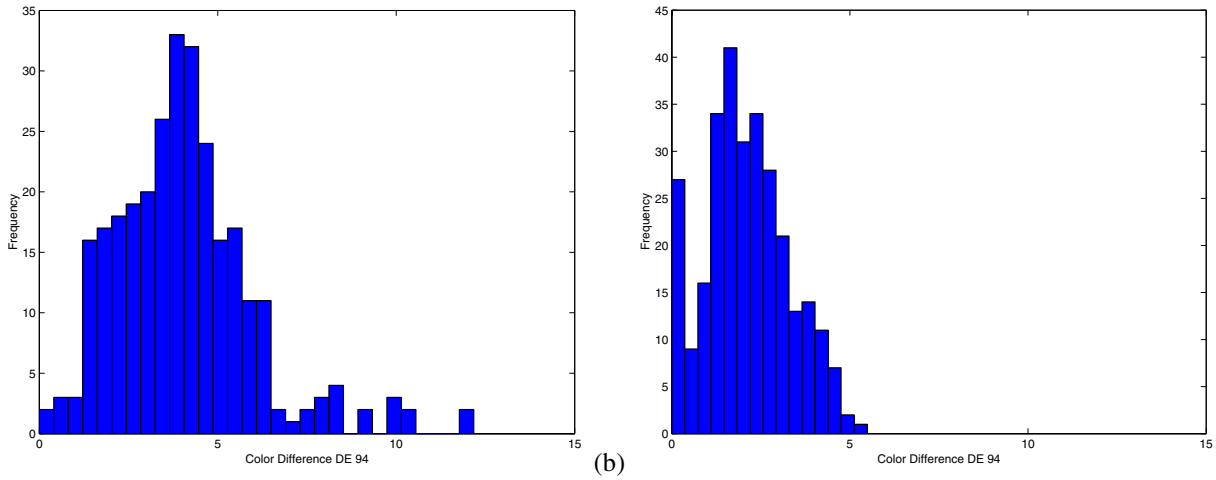


*Figure 3. (a) Histogram of ΔE*94 errors between CIELAB values measured on the reference (substrate 1) and test (substrate 4); (b) Histogram of ΔE*94 errors between the CIELAB values predicted by PCA and values measured on the test substrate.*

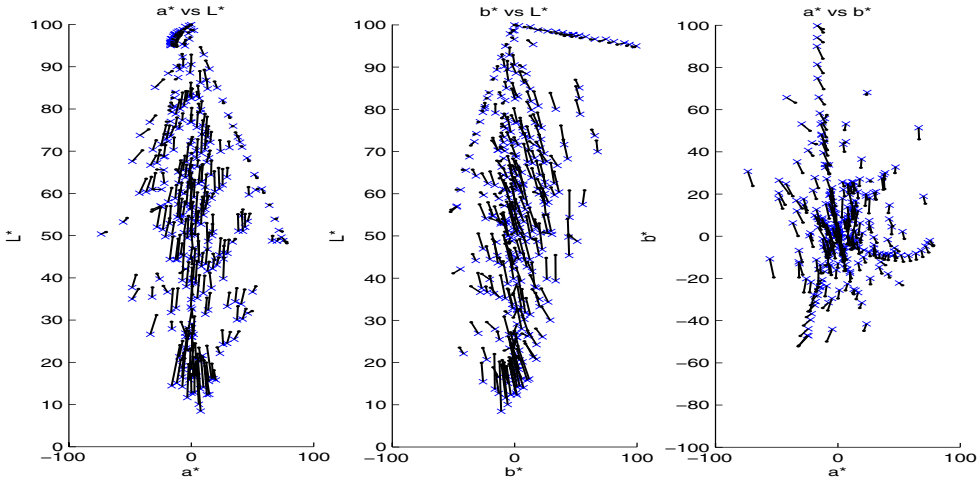


Figure 4. CIELAB values measured on a reference substrate (x) when compared to the CIELAB values (•) measured on the test substrate.

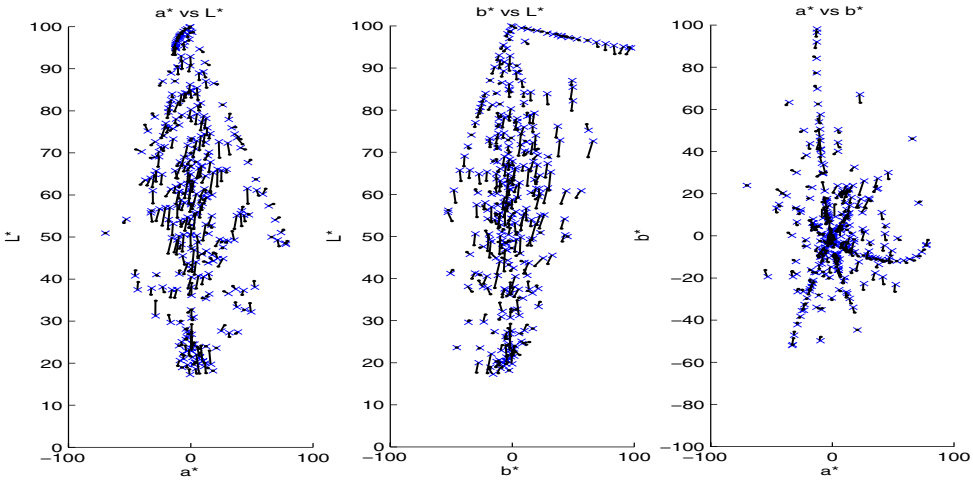


Figure 5. CIELAB values measured on a reference substrate (x) when compared to the reconstruction CIELAB values (•) predicted using the model.