Upgrading color-difference formulas

Ingmar Lissner and Philipp Urban*
Institute of Printing Science and Technology, Technische Universität Darmstadt, Magdalenenstrasse 2, 64289 Darmstadt, Germany
*Corresponding author: urban@idd.tu-darmstadt.de

Received January 26, 2010; revised April 15, 2010; accepted April 19, 2010; posted April 21, 2010 [Doc. ID 123347]; published June 11, 2010

We propose a method to improve the prediction performance of existing color-difference formulas with additional visual data. The formula is treated as the mean function of a Gaussian process, which is trained with experimentally determined color-discrimination data. Color-difference predictions are calculated using Gaussian process regression (GPR) considering the uncertainty of the visual data. The prediction accuracy of the CIE94 formula is significantly improved with the GPR approach for the Leeds and the Witt datasets. By upgrading CIE94 with GPR we achieve a significantly lower STRESS value of 26.58 compared with that for CIEDE2000 (27.49) on a combined dataset. The method could serve to improve the prediction performance of existing color-difference equations around particular color centers without changing the equations themselves.

1. INTRODUCTION
A measure of perceived difference between colors is required in many color-related applications, such as color imaging and color reproduction, and for defining and testing color tolerances. For this purpose, color-difference formulas were designed based on the results of various color-discrimination experiments. To allow the exchange of color-difference information, several formulas were standardized, such as CIE76, CMC [1], CIE94 [2], and the recent CIEDE2000 [3] formula, which are mainly used in industrial applications. The formulas were created by fitting predefined functions to the results of color-discrimination experiments [4–6] on the CIELAB color space.

A main problem in designing color-difference equations is the variability of the underlying experimental data. Investigations show that the inter-observer and intra-observer uncertainties are high and that the composition of the observer panel has a great influence on the results. Color-difference formulas are typically adjusted to the mean or median results of visual experiments. However, even if a formula can be considered accurate for the average observer, it predicts the judgments of only about one third of actual observers within ±20% of the calculated color difference [7], and the standardized formulas listed above already show a large disagreement with the average observer [8]. A method to evaluate the prediction performance of color-difference formulas considering visual uncertainties was recently proposed [9].

Another problem of the formulas mentioned above is the composition of the underlying data. Color-discrimination experiments were conducted in different laboratories, at partially different color centers, using various materials for the samples and different psychophysical methods. The CIE94 color-difference formula was fitted to the RIT-DuPont [10] and the BFD-P [11] datasets, whereas the CIEDE2000 formula was fitted to a combined dataset composed of the RIT-DuPont, BFD-P, Leeds [12], and Witt [13] datasets. Investigations show that the combination of visual data from differently designed experiments introduces additional uncertainty that needs to be considered when fitting color-difference formulas [8,14].

Because the uncertainty of the experimental visual data is high, an equation that predicts the data extremely accurately may overfit it. In this case the formula fits noise instead of the actual visual data. This is why the prediction performance of a color-difference formula must be evaluated on test data that differs from the training data used to fit the formula.

Furthermore, color centers from discrimination experiments cover the underlying color space only sparsely. The prediction accuracy of color-difference formulas is likely to decrease for color pairs located far from any training data. This effect is amplified by the high uncertainty mentioned above.

In this paper we propose a method that enhances existing color-difference formulas using additional visual data. It improves the prediction accuracy especially for color pairs located far from the formula’s training data. We previously described a related method to upgrade color-difference formulas with additional visual data [15]. To our knowledge, the method proposed in this paper is the first such approach that includes a detailed justification of the model and an automatic adjustment of its parameters. The key features of our approach can be summarized as follows:

1. The prediction performance of the color-difference formula is improved by adding a correction term. The formula itself remains unchanged.
2. Intrinsic uncertainties, such as inter- and intra-observer variability and different experimental conditions, are considered to avoid overfitting.
3. The parameters of the method are automatically ad-
justed based on a statistical model.

We believe that this method could be used to improve standardized color-difference equations around particular color centers (a company’s corporate colors, important reference colors, etc.) to allow, for instance, a more reliable selection of tolerance bounds. Furthermore, standardized color-difference formulas, which are designed to predict small color differences of up to five CIELAB units [16], could be upgraded with data for medium and large color differences.

2. MODELING COLOR DIFFERENCES

The basic idea of our approach is to use a Gaussian process to model color differences. In this section we explain the concept of a Gaussian process and how to employ it for the prediction of perceived color differences. We then apply our Gaussian process model to a combination of existing color-difference formulas and experimentally determined color differences.

A. Gaussian Processes for Predicting Color Differences

The relationship between two colors and their perceived difference can be written as

\[ x \rightarrow v(x), \quad (1) \]

where \( x = (x_1, x_2) \) is a color pair with \( x_1, x_2 \in \text{CIELAB} \), and \( v(x) \) is the perceived difference between \( x_1 \) and \( x_2 \). However, there is no definitive color difference that can be assigned to a color pair \( x \), because the perceived difference of two colors varies between observers [7]. This is why we model the perceived color difference \( v(x) \) as a normally distributed random variable. In addition, we assume that any collection of color differences \( v = (v(x_1), \ldots, v(x_n)) \) of a finite set of color pairs \( X = \{x_1, \ldots, x_n\} \) follows a joint Gaussian distribution in accordance with the definition of a Gaussian process [17]:

Definition 1: A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

A Gaussian process is fully specified by a mean function \( m(x) \) and a covariance function \( k(x, \hat{x}) \). In Subsection 2.B we will propose specific functions adjusted to the color-difference prediction problem. Using these functions we can express the joint Gaussian distribution of a set of color differences \( v \) as

\[ p(v|X) = \mathcal{N}(m, K), \quad (2) \]

where the vector \( m = (m(x_1), \ldots, m(x_n))^T \) contains the values of the mean function \( m(x) \) of the Gaussian process for the color pairs \( X = \{x_1, \ldots, x_n\} \), and the covariance matrix \( K \) is defined as

\[ K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}, \quad (3) \]

where the covariance function \( k(x, \hat{x}) \) of the Gaussian process models the correlation between two color pairs \( x \) and \( \hat{x} \). If we extend Eq. (2) to an additional color difference \( v_* \) of a color pair \( x_* \), our joint Gaussian distribution [17] changes to

\[ p(v_*, v|X) = \mathcal{N}\left( \begin{bmatrix} m_* \\ K_* \end{bmatrix}, \begin{bmatrix} K & k_* \\ k_* & k_*^T \end{bmatrix} \right), \quad (4) \]

where \( m_* = m(x_*) \) is the value of the mean function for \( x_* \), \( k_* = (k(x_*, x_1), \ldots, k(x_*, x_n))^T \) is the vector of covariances between \( x_* \) and the color pairs in \( X \), and \( k_* = k(x_*, x_*) \) is the autocovariance of the additional color pair \( x_* \). From Eq. (4) we can derive the conditional probability of the color difference \( v_* \) (see Rasmussen and Williams [17]):

\[ p(v_*|v, x_*, X) = \mathcal{N}(E(v_*), \text{var}(v_*)), \quad (5) \]

\[ E(v_*) = m_* + k_*^T K_*^{-1}(v - m), \quad (6) \]

\[ \text{var}(v_*) = k_* - k_*^T K_*^{-1} k_* \quad (7) \]

Given a set of color pairs \( X \) and corresponding color differences \( v \), Eq. (6) allows us to predict the unknown difference \( v_* \) of a color pair \( x_* \). We thus employ experimentally determined color differences to predict unknown color differences. Unfortunately, the results of visual experiments are affected by measurement noise \( \epsilon \), which has to be incorporated into our model. Instead of \( v(x) \), we measure

\[ v'(x) = v(x) + \epsilon. \quad (8) \]

In this paper we assume uncorrelated zero-mean Gaussian noise \( \epsilon \sim \mathcal{N}(0, \sigma^2 I) \) with variance \( \sigma^2 \). Please note that more complex noise models, which describe the measurement uncertainties more accurately, can be used with the Gaussian process framework without changing the concept. Using our simplified noise model, the color-difference prediction [Eqs. (6) and (7)] becomes

\[ E(v'_*) = m_* + k_*^T (K + \sigma^2 I)^{-1}(v' - m), \quad (9) \]

\[ \text{var}(v'_*) = k_* - k_*^T (K + \sigma^2 I)^{-1} k_* \quad (10) \]

Equations (9) and (10) are referred to as Gaussian process regression (GPR).

It should be noted that the GPR method can be interpreted as a Bayesian approach for functions. In our case, the function maps color pairs to their perceived color differences. In Bayesian inference notation, GPR gives us the posterior distribution of an unknown color difference given our Gaussian process model and noisy, experimentally determined visual data. For an in-depth description of GPR and its relation to Bayesian approaches, see Rasmussen and Williams [17].

It is worth mentioning that the concept of modeling small color differences by normally distributed random variables can be derived from the results of color-discrimination and color-matching experiments. Investigations have shown that color matchings around particular color centers can be well modeled as normal distributions within an intensity-linear color space (e.g., CIEXYZ) [18–20]. Since the color space transformation from CIEXYZ into CIELAB can be well approximated locally by a linear transformation, the deviation from normality of the corresponding distribution in CIELAB can
be assumed to be small. In this regard a Gaussian process approach may model the case of small color differences accurately. However, we have not found any analysis on the distribution of larger color differences. Such an analysis would be required to validate our assumption for arbitrary color pairs but would go beyond the scope of this paper.

**B. Specifying the Gaussian Process**

As already mentioned, a Gaussian process is specified by its mean and covariance functions. In our application, the color-difference formula that we want to improve serves as the mean function of the Gaussian process. To improve CIE94, the mean function is

\[ m(x) = \Delta E_{\alpha\beta}(x), \]

where \( x = (x^1, x^2) \) and \( x^1, x^2 \) ∈ CIELAB. Without additional visual data, the GPR predictions equal the predictions of the mean function for any color pair \( x_\alpha \), i.e., \( v_\alpha = m(x_\alpha), \forall x_\alpha \). This means that in CIELAB regions where no new experimental data are available, we can force our GPR-based color-difference predictions to fall back on the mean function. Thus, divergence from the mean function is possible only close to visual data.

GPR enables us to combine the color-difference predictions of the mean function with experimentally determined color differences. An assumption of the correlation between the underlying data is included in the form of the covariance function \( k(x, \hat{x}) \), whose selection is a crucial step in GPR. This covariance function has to fulfill certain properties, namely,

1. Symmetry, i.e., \( k(x, \hat{x}) = k(\hat{x}, x) \).
2. Positive semidefiniteness, i.e., for a measure \( \mu \) and any \( L^2 \)-integrable function \( f(L^2; \text{Hilbert space}) \) \( k(x, \hat{x})f(x)f(\hat{x})d\mu(x)d\mu(\hat{x}) \geq 0 \). This causes the corresponding Gram matrix \( K_{\alpha j} = \{k(x_\alpha, x_j)\} \) [see Eq. (3)] to be positive semidefinite [17].

Following an analysis of experimentally determined color-discrimination data, we chose three key factors affecting the correlation of two color pairs \( x = (x^1, x^2) \) and \( \hat{x} = (\hat{x}^1, \hat{x}^2) \) with their corresponding color-difference vectors \( x_d = (x^2 - x^1) \) and \( \hat{x}_d = (\hat{x}^2 - \hat{x}^1) \):

1. The Euclidean distance between color pairs \( x \) and \( \hat{x} \):

\[ \phi_{x, \hat{x}} = \left( \frac{x^1 + \hat{x}^1}{2} - \frac{x^2 + \hat{x}^2}{2} \right)^2. \]

2. The angle between color-difference vectors \( x_d \) and \( \hat{x}_d \):

\[ \alpha_{x, \hat{x}} = \frac{x_d^T \cdot \hat{x}_d}{\|x_d\|_2 \cdot \|\hat{x}_d\|_2} = \cos(\hat{x}_d, x_d). \]

3. The length difference of \( x_d \) and \( \hat{x}_d \):

\[ \rho_{x, \hat{x}} = \|x_d\|_2 - \|\hat{x}_d\|_2. \]

We evaluated the first two influence factors \( \phi_{x, \hat{x}} \) and \( \alpha_{x, \hat{x}} \) on the RIT-DuPont dataset [10] using the length difference \( \rho_{x, \hat{x}} \) as a simple measure of disagreement. This disagreement is meaningful because all color-difference vectors of the RIT-DuPont dataset correspond to the same perceived distance \( \Delta V = 1.02 \). Figure 1 shows how \( \rho_{x, \hat{x}} \) changes with increasing distance \( \phi_{x, \hat{x}} \).

Even though the disagreement does not increase monotonically, there is an obvious trend toward increasing disagreement with increasing distance, i.e., close color pairs are more correlated than distant pairs.

In Fig. 2 the mean length difference \( \rho_{x, \hat{x}} \) is related to the angles between the color-difference vectors. On average, the length difference increases for angles \( \hat{x}_d(x_d, \hat{x}_d) \in [0^\circ, 90^\circ] \) and decreases for \( \hat{x}_d(x_d, \hat{x}_d) \in [90^\circ, 180^\circ] \). Therefore, we have modeled the angle-dependent correlation by the term \( \alpha_{x, \hat{x}}^2 \) in our covariance function, which is small for angles corresponding to large length differences and vice versa. In Fig. 2 we illustrate the model performance by juxtaposing the average length-difference bars and \( 1-\alpha_{x, \hat{x}}^2 \) (normalized to the maximum bar height).

To justify the length-difference component [Eq. (14)], imagine two color-difference vectors that differ only in their Euclidean lengths, i.e., the distance and angle terms are zero. Without a length-difference component, these vectors would be strongly correlated. In addition, this length-difference factor allows us to combine small, medium, and large perceived distances at the same color center.

While the distance component \( \phi_{x, \hat{x}} \) is a global correlation factor governing the covariance of color pairs at distant color centers, the angle and the length-difference components \( \alpha_{x, \hat{x}} \) and \( \rho_{x, \hat{x}} \) describe local correlations at one particular color center. For instance, two color-difference vectors at distant color centers may have the same direction and Euclidean length—they should, however, not be strongly correlated, which is why we need the distance component. For a suitable covariance function, all three components are required.

Note that our model is based on experimental data that are sparsely distributed across the color space. This means that it cannot be generalized to arbitrary color differences, and visual data from new experiments may require changes to the model.

On the basis of our data analysis we introduce the following covariance function:

**Fig. 1.** (Color online) Relation between color pair distance \( \phi_{x, \hat{x}} \) and Euclidean length difference \( \rho_{x, \hat{x}} \) evaluated on the RIT-DuPont dataset. One bar represents the mean length difference for all color pair distances within the corresponding distance range.
From the properties of $k(x, \hat{x})$ it follows that $K$ is symmetric and positive semidefinite (normalized) it follows that $K$ is symmetric and positive semidefinite.

Summarizing our findings, we can predict the expectation value and the variance of an unknown color difference $v'$, of a color pair $x$, according to Eqs. (9) and (10):
A common concept of adjusting the hyperparameters to the problem is maximizing the (log) marginal likelihood with respect to \( \theta \) \([17,21,22]\):

\[
\log p(v'|X, \theta) = \log \mathcal{N}(m, K + \sigma^2 I)
\]

\[
= -\frac{1}{2} (v' - m)^T (K + \sigma^2 I)^{-1} (v' - m) - \frac{1}{2} \log |K + \sigma^2 I| - \frac{n}{2} \log 2\pi.
\]

(20)

The marginal likelihood is the probability of the experimentally determined visual data \( v' \), given the model that depends on the corresponding color pairs \( X = \{x_1, \ldots, x_n\} \) and the hyperparameters \( \theta \) \([17]\). The higher the marginal likelihood is, the more probable are the experimental data \( v' \) given a particular set of hyperparameters \( \theta \). In a practical application, local optima of the marginal likelihood are likely. This does not necessarily indicate an unsuitable model, as different local optima can be considered as “particular interpretations of the data” \([17]\).

Standard numerical optimization methods can be applied to maximize the marginal likelihood (e.g., conjugate gradients). To avoid local maxima with a low marginal likelihood, it is recommended to perform multiple optimizations with randomized initial values and to select the hyperparameters corresponding to the highest marginal likelihood (lowest negative log marginal likelihood; see Fig. 7).

Note that there are other methods for hyperparameter optimization, such as cross-validation as described by Rasmussen and Williams \([17]\). Because of its well-defined role in the concept of Gaussian process regression and because a reliable MATLAB implementation is available by

Fig. 4. (Color online) Damping function \( f(\delta, m) \). For \(-m < \delta < m\) the damping function has no effect on the correction term \( \delta \). For \( \delta \to \infty \) and \( \delta \to -\infty \), the damped correction term converges against \( m \), and \(-m\), respectively.

Fig. 5. Damping function \( f(\delta, m) \) for varying \( m\). The allowed range of values of the correction term \( \delta \) increases with increasing mean function value \( m\). For \( \delta \to \infty \) and \( \delta \to -\infty \) the damped correction term converges against \( m \), and \(-m\), respectively.

Fig. 6. (Color online) CIE Gray color center at \((L^*, a^*, b^*) = (59.3, -0.78, 1.05)\) with selected RIT-DuPont ±T50 color-difference vectors \((\Delta V = 1.02)\), CIE94 iso-distance contour (nearly circular ellipse, blue online) with \( \Delta E^*_{94} = 1 \), GPR iso-distance contour (ellipse, red online) with \( \Delta GPR = 1 \), and 95% confidence interval \((\pm 2 \cdot \text{var}(\tilde{v}^*))\) of the predictions (shaded gray ellipse) computed according to Eq. (17). Note that due to our damping function \( f \) this variance is only an approximation. The Gaussian process uses CIE94 as its mean function and was trained with the RIT-DuPont dataset.

Fig. 7. (Color online) Negative log marginal likelihood under varying noise. There is a distinct minimum at \( \sigma_e^2 = 0.0080 \) \((\sigma_e = 0.0894)\) with a negative log marginal likelihood of \(-233.4\). The remaining hyperparameters \( l_1, l_2, \) and \( l_3 \) are set to their optimal values determined by hyperparameter optimization (Section 3). The Gaussian process uses CIE94 as its mean function and was trained with the Witt dataset.
4. RESULTS AND DISCUSSION

To evaluate our Gaussian process approach for color-difference prediction, we chose CIE94 as the Gaussian process’s mean function. The method will therefore be referred to as GPR$_{\text{CIE94}}$ in the following. No color-discrimination experiments were conducted for this paper: to make our results reproducible, we used visual data already employed to fit the CIEDE2000 formula, namely, the RIT-DuPont [10], BFD-P [11], Leeds [12], and Witt [13] datasets (see Table 1).

In the case where the mean function already overfits its training data, adjusting the Gaussian process to the same data would amplify this behavior. Therefore, data used to fit the mean function were not considered in training the Gaussian process. Since the RIT-DuPont dataset and parts of the BFD-P dataset were used in the development of the CIE94 formula [4], we trained our Gaussian process with the Leeds and the Witt data. The hyperparameters $l_1, l_2, l_3,$ and $\sigma^2$ were optimized using marginal likelihood maximization as described in Section 3. The resulting hyperparameters are listed in Table 2.

To evaluate the generalization ability of the method, we also used test data that differed from the training data and were not used to adjust the mean function (CIE94). This was necessary to assess the prediction accuracy of the GPR approach on completely unknown data. A combined dataset (COM) was generated for the performance evaluation equivalently to the dataset used by Melgosa et al. [8].

The prediction performance of all color-difference measures on the visual data was assessed using the PF/3 measure [24] and the recently introduced STRESS (standardized residual sum of squares) index [8,25]. The STRESS index allows a statistical judgment regarding whether the predictions of two color-difference formulas differ significantly on a given confidence level (see García et al. [25] for details). A 95% confidence level was used for all tests of statistical significance. The results are shown in Tables 3 (STRESS index) and 4 (PF/3 measure). Our key findings can be summarized as follows:

- GPR$_{\text{CIE94}}$ achieves high prediction accuracy on its training data. Because the data are assumed to be noisy, the STRESS values on these data are not zero.
- If the Witt data are used as training set, the prediction performance decreases significantly on the Leeds data. If trained with the Leeds dataset, the prediction performance improves nonsignificantly on the Witt dataset.
- The GPR$_{\text{CIE94}}$ approach achieves significantly better results on the COM dataset than CIE94 for all evaluated combinations of training data.
- If trained with a combination of the Leeds and the Witt data, GPR$_{\text{CIE94}}$ performs significantly better on the COM data than CIEDE2000.

Low STRESS values are achieved with the GPR-based method on its own training data (underlined values in Tables 3 and 4). This is illustrated in Fig. 8, where the disagreement between visual and computed distances is shown for CIE94 and its upgraded GPR$_{\text{CIE94}}$ counterpart. Although these STRESS values are quite low, they do not vanish, due to our assumption of noisy visual data (Eq. (8)). The corresponding noise variances are shown in Table 2. A zero-noise assumption would force these STRESS values to zero, i.e., the test data would be perfectly predicted. In this case, however, the fact that the visual data are affected by intra- and inter-observer noise would be neglected. In addition, a good generalization capability is desirable rather than extremely accurate predictions on the training data.

The STRESS values for the COM dataset indicate that Gaussian process regression enhances the color-difference prediction accuracy: The results are significantly better for GPR with a CIE94 mean function (GPR$_{\text{CIE94}}$) than for plain CIE94 on all test sets. If the Leeds and Witt data are combined to train the Gaussian process, the predictions on the COM data are even more accurate than the CIEDE2000 predictions. An F-test on the corresponding STRESS values (26.58 and 27.49) shows that they differ significantly at a 95% confidence level.

Since the GPR predictions are based on correlations between test and training color pairs (according to the model in Subsection 2.B), the results implicitly show how well visual datasets agree with each other. Interestingly, if the Leeds data are used for training, the prediction per-

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Color Pairs</th>
<th>$\Delta E_{ab}^*$ Range</th>
<th>Sample Material</th>
<th>Psychophysical Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIT-DuPont</td>
<td>312</td>
<td>0.78–4.41</td>
<td>Glossy paint</td>
<td>Constant stimuli</td>
</tr>
<tr>
<td>BFD-P</td>
<td>2776</td>
<td>0.04–18.21</td>
<td>Various materials, relative scales of individual sets adjusted using textile samples</td>
<td>Grayscale</td>
</tr>
<tr>
<td>Leeds GS</td>
<td>203</td>
<td>0.40–4.74</td>
<td>Glossy paint</td>
<td>Grayscale</td>
</tr>
<tr>
<td>Leeds PC</td>
<td>104</td>
<td>0.53–4.02</td>
<td>Glossy paint</td>
<td>Constant stimuli</td>
</tr>
<tr>
<td>Witt</td>
<td>418</td>
<td>0.12–10.63</td>
<td>Glossy paint</td>
<td>Grayscale</td>
</tr>
<tr>
<td>COM</td>
<td>11273</td>
<td>0.04–18.21</td>
<td>All above</td>
<td></td>
</tr>
</tbody>
</table>

*The Leeds dataset contains two subsets: Leeds GS is based on the grayscale method; Leeds PC is based on the method of constant stimuli. The COM dataset is a weighted combination of the other four datasets as described by Melgosa et al. [8]. Table content based on Table 1 in Luo et al. [5] and Table 5.1 in Shen [26].
formance improves on the RIT-DuPont dataset. The RIT-DuPont data and parts of the Leeds data were collected with the method of constant stimuli. In contrast, if the method is trained with the grayscale-based Witt data, the prediction performance on the RIT-DuPont and the Leeds data decreases.

As the Leeds dataset is composed of both grayscale-based and constant-stimuli-based data, it enables a more detailed investigation of contradicting test and training data and their effect on the GPR predictions. If the method is trained with the grayscale-based Witt data, applying GPR to the grayscale-based Leeds data yields a STRESS value of 33.45, which indicates that the predictions do not significantly differ from the CIE94 predictions (29.50) at a 95% confidence level. In contrast, the Witt dataset is rather consistent, as indicated by a small noise variance of $\sigma^2 = 0.0080$. The effect of the noise level on the GPR predictions can be stated as follows: The larger the noise variance $\sigma^2$ is, the closer the GPR predictions are to the mean function.

Another important issue is the intrinsic noise of the visual data. This noise is considered in a simplified manner using a statistically determined hyperparameter $l_2$. This parameter also indicates how consistent a dataset is. As already mentioned, the Leeds data are a combination of two heterogeneous datasets, collected with either the method of constant stimuli or the grayscale method. The noise variance for the combined Leeds data is comparably high ($\sigma^2 = 0.0202$), reflecting inconsistencies between the subsets. In contrast, the Witt dataset is rather consistent, as indicated by a small noise variance of $\sigma^2 = 0.0080$. The effect of the noise level on the GPR predictions can be stated as follows: The larger the noise variance $\sigma^2$ is, the closer the GPR predictions are to the mean function.

Possibly contradictory color-difference information in the training data is also indicated by a small hyperparameter $l_1$. For the combined Leeds and Witt training set $l_1 = 1.18$ (see Table 2), which means that the correction term $\delta$ does not have significant influence unless we are predicting close to new visual data. As a result, the color-difference formula used as the Gaussian process's mean

| Training Set | $l_1$ | $l_2$ | $l_3$ | $\sigma^2$ | log $p(\psi | X, \theta)$ |
|--------------|-------|-------|-------|------------|--------------------------|
| Leeds        | 26.99 | 1.95  | 6.05  | 0.0202     | 65.95                   |
| Witt         | 7.95  | 2.93  | 3.48  | 0.0080     | 233.41                  |
| Leeds, Witt  | 1.18  | 2.05  | 6.35  | 0.0109     | 143.92                  |

The corresponding log marginal likelihood is shown in the last column.

Table 3. STRESS Values of Color-Difference Measures Applied to Five Sets of Visual Data

<table>
<thead>
<tr>
<th>Color-Difference Measure</th>
<th>RIT-DuPont</th>
<th>BFD-P</th>
<th>Leeds</th>
<th>Witt</th>
<th>COM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIELAB</td>
<td>33.42</td>
<td>42.46</td>
<td>40.09</td>
<td>51.71</td>
<td>43.93</td>
</tr>
<tr>
<td>CMC</td>
<td>27.44</td>
<td>30.61</td>
<td>24.90</td>
<td>35.04</td>
<td>30.48</td>
</tr>
<tr>
<td>CIE94</td>
<td>20.31</td>
<td>33.88</td>
<td>30.57</td>
<td>31.94</td>
<td>32.07</td>
</tr>
<tr>
<td>GPR CIELAB</td>
<td>17.85</td>
<td>33.96</td>
<td>11.31</td>
<td>30.49</td>
<td>29.59*</td>
</tr>
<tr>
<td>GPR CIELAB</td>
<td>22.94</td>
<td>33.62</td>
<td>35.86</td>
<td>06.79*</td>
<td>30.43*</td>
</tr>
<tr>
<td>GPR CIELAB</td>
<td>20.19</td>
<td>33.73</td>
<td>08.42</td>
<td>08.42*</td>
<td>26.58*</td>
</tr>
<tr>
<td>CIEDE2000</td>
<td>19.47</td>
<td>29.55</td>
<td>19.25</td>
<td>30.22</td>
<td>27.49</td>
</tr>
</tbody>
</table>

GPR CIELAB symbolizes GPR with CIELAB as a mean function. An underlined STRESS value means that the Gaussian process was trained with the corresponding dataset. An asterisk indicates that the GPR CIELAB predictions are significantly different from the corresponding CIELAB predictions as determined by an F-test with a 95% confidence level on the STRESS values.

Table 4. PF/3 Values of Color-Difference Measures Applied to Five Sets of Visual Data

<table>
<thead>
<tr>
<th>Color-Difference Measure</th>
<th>RIT-DuPont</th>
<th>BFD-P</th>
<th>Leeds</th>
<th>Witt</th>
<th>COM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIELAB</td>
<td>33.79</td>
<td>56.32</td>
<td>47.23</td>
<td>70.94</td>
<td>55.71</td>
</tr>
<tr>
<td>CMC</td>
<td>28.26</td>
<td>39.76</td>
<td>27.51</td>
<td>46.81</td>
<td>37.45</td>
</tr>
<tr>
<td>CIE94</td>
<td>20.98</td>
<td>42.95</td>
<td>33.35</td>
<td>42.33</td>
<td>37.72</td>
</tr>
<tr>
<td>GPR CIELAB</td>
<td>18.97</td>
<td>45.21</td>
<td>12.47</td>
<td>49.24</td>
<td>37.17</td>
</tr>
<tr>
<td>GPR CIELAB</td>
<td>23.73</td>
<td>42.99</td>
<td>35.86</td>
<td>06.79*</td>
<td>33.38</td>
</tr>
<tr>
<td>GPR CIELAB</td>
<td>20.88</td>
<td>44.47</td>
<td>08.87</td>
<td>08.87*</td>
<td>27.43</td>
</tr>
<tr>
<td>CIEDE2000</td>
<td>19.56</td>
<td>37.31</td>
<td>22.02</td>
<td>38.78</td>
<td>32.11</td>
</tr>
</tbody>
</table>

GPR CIELAB symbolizes GPR with CIELAB as a mean function. An underlined PF/3 value means that the Gaussian process was trained with the corresponding dataset.
function is improved only in regions where additional visual data are available. This is not necessarily undesired, as we strive for a locally enhanced prediction accuracy around particular color centers. Figure 9 illustrates the location dependence of the correction term.

5. CONCLUSIONS

We proposed a method to upgrade existing color-difference formulas using additional visual data. The method uses a Gaussian process model. It predicts color-differences via Gaussian process regression (GPR), treating the existing color-difference formula as its mean function and the visual data as observations. Measurement noise in the visual data is considered using a simplified noise model.

Upgrading the CIE94 color-difference equation improves the prediction accuracy on a combined dataset when the method is trained on the Leeds data, the Witt data, or a combination of the two. This improvement is statistically significant at a 95% confidence level according to an F-test on the resulting STRESS values. Trained on both the Leeds and the Witt data, the achieved STRESS value on the combined visual data was significantly lower than for CIEDE2000. The prediction accuracy on the RIT-DuPont and the BFD-P data, which were not used to train the Gaussian process, was not significantly different from that of CIE94.

Our investigations show that particular care should be taken in the selection of the visual data, which should be collected under the same experimental conditions as the data used to fit the mean function.

The proposed Gaussian process approach could be em-

![Fig. 8.](image-url)
ployed to upgrade existing color-difference formulas with new visual data around especially important color centers. Since the method is based on adding a correction term, modifications to the color-difference formula are not required.

Merging visual data for medium and large color differences with formulas designed for small color differences was not tested in this paper. We believe that the GPR approach is generally capable of upgrading small-difference-based formulas to enable the prediction of medium or large distances around particular color centers. Further investigations are required to validate this assumption.

ACKNOWLEDGMENTS
This work would not have been possible without the financial support of the Deutsche Forschungsgemeinschaft (German Research Foundation).

REFERENCES
2. CIE Publication No. 116, “Industrial colour-difference