Research on Computer Color Matching in Textile Dyeing by the Method of Regression Analysis

Xiaojie Liu, Bingsen Zhang, Anling Zhang

1, First Author Qingdao University, qdlxj@qdu.edu.cn
*2, Corresponding Author Qingdao University, bszhang@qdu.edu.cn
3 Qingdao University, anling_zhang@126.com

Abstract
In the study of color matching in textile dyeing, regression algorithm is used to analyze the nonlinear relationship between the dye concentration and the color data in CIE-L*a*b* color space of dyeing sample, thereby providing theory basis for establishing mathematical model for textile dyeing. Statistical analysis is conducted on a number of experimental data and the model is tested by actual data in this paper. From the comparison, stepwise regression analysis of coloring creates the relationship between the color information and dye concentration effectively, and at the same time avoids the model to be a very complicated one which is helpful to the recipe prediction.

Keywords: Color Matching, Textile Dyeing, Stepwise Regression, CIE-L*a*b* color space

1. Introduction
During recent years, Computer Color Matching (CCM) has received more and more attention and has been widely used in various fields related to color, such as printing, coating and dyeing, due to its convenience, accuracy and acceptability. This paper focuses on CCM in textile dyeing.

Computer color matching systems based on data query have a wide range of applications both at home and abroad, such as Gretag Macbeth, DataColor, domestic Winsrici computer's measuring and matching color system. But these color matching systems are complicated, expensive, and not suitable for domestic dyeing enterprises with small scale and unstable dyes series conditions.

Data query color matching method gradually turns to the research on dyeing modeling in recent years. Genetic algorithm is applied to the research of color recipe prediction by M.A mani Tehran and S.Gorji Kandi in 2006[1]. Nayatanir Sakai Hideki Yoshinobu proposed the improving method of color matching model In-CAM[2]. Maozu Guo etc. put forward the BP neural network color matching method in the year 2000[3]. Huifeng Wang etc. raised the neural network color identification method in the application of color matching system in wool in 2006[4]. Tao Yang developed the neural network computer color matching system of paints in 2006[5]. Yáñez Cornelio etc. presented an automatic color matching system which introduced Alpha-Beta associative memories in 2006[6]. Some research applied principal component analysis in computer color matching[7][8].

But there are some challenges about the neural network color-matching method which is applied to the research of textile dyeing color matching: one is because the artificial neural network color model is not visible; we can't see the real textile color matching mathematical model. So it is relatively difficult to grasp and explain the color-matching inner laws and characteristics. Then it is instability and without reproducibility characteristics of the artificial neural network. They are the practical difficulties of applying the color-matching method. Therefore, this paper tries to use numerical analysis method to establish color-matching models in order to solve the problem of textile dyeing color-matching.

As the complex inherent rule of dyeing and the limited of human understanding, we can not identify the internal causal connections and set up the mathematical model according to dyeing mechanism completely. Through the statistical analysis of the data, the best data fitting model is hoped to be found out. Regression model is a kind of most-used model established by the analysis methods[9][10].
2. Experiments and Numerical Analysis

Obviously the color of dyeing sample is closely related to the dye concentration. But the corresponding relationship is complicated and dye has different effect on each color component. We hope to select the independent variables which influence color more significant in order to simplify the mathematical model of the problem. Stepwise regression is a method to do effective selection from many variables [11]. To optimize the color matching model, this paper attempts to do regression analysis with each component of dyeing sample CIE-la*b* color information and dyeing concentration.

The core issue of the paper is to establish the dyeing mathematical model and to forecast the recipe in a given dyeing sample and color-matching precision. Hundreds of different sets dyes series, different proportion of the dyeing processing lists are collected in the experiments.

67 groups of mixed dyeing samples of a neutral reactive dye color series (part of them in Table1 are analyzed in this paper) active cyan FBN, activity yellow 3RS and active magenta 3BS, are gotten from a dyeing factory. The three components of the CIE-La*b* color space: the lightness factor L, two chromaticity factors a* and b*, are independent of each other. The three components are treated respectively as the three dependent variables which are explained by the concentration variables in this paper. Concentration units (g/10 grams water) is showed in Table 1 below.

### Table 1. dyeing data with same concentrations

<table>
<thead>
<tr>
<th>FB</th>
<th>3BS</th>
<th>3RS</th>
<th>CIE-L</th>
<th>a*</th>
<th>b*</th>
<th>FBN</th>
<th>3BS</th>
<th>3RS</th>
<th>CIE-L</th>
<th>a*</th>
<th>b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>0.05</td>
<td>0.05</td>
<td>64.77</td>
<td>5.92</td>
<td>-4.95</td>
<td>0.80</td>
<td>0.20</td>
<td>0.40</td>
<td>33.87</td>
<td>5.94</td>
<td>-7.16</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>57.48</td>
<td>7.53</td>
<td>-4.50</td>
<td>0.80</td>
<td>0.20</td>
<td>0.80</td>
<td>33.90</td>
<td>3.30</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.15</td>
<td>50.57</td>
<td>8.29</td>
<td>-5.00</td>
<td>1.00</td>
<td>0.20</td>
<td>0.40</td>
<td>30.69</td>
<td>7.41</td>
<td>-9.46</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.20</td>
<td>47.84</td>
<td>5.91</td>
<td>3.41</td>
<td>1.00</td>
<td>0.20</td>
<td>0.80</td>
<td>31.16</td>
<td>4.84</td>
<td>-2.96</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.20</td>
<td>47.75</td>
<td>2.65</td>
<td>14.57</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>44.29</td>
<td>9.09</td>
<td>-5.51</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.20</td>
<td>40.66</td>
<td>5.62</td>
<td>-3.11</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>41.54</td>
<td>9.59</td>
<td>-5.40</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.20</td>
<td>40.58</td>
<td>2.35</td>
<td>6.61</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>40.27</td>
<td>9.43</td>
<td>-4.86</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.20</td>
<td>36.86</td>
<td>5.79</td>
<td>-5.76</td>
<td>0.20</td>
<td>0.40</td>
<td>0.40</td>
<td>43.76</td>
<td>12.90</td>
<td>-0.75</td>
</tr>
</tbody>
</table>

2.1. Suppose

The linear relationship between L and the concentrations of FBN, 3BS, 3RS with L as dependent variable, the concentration of FBN as independent variable $x_1$, the concentration of 3BS as $x_2$, the concentration of 3RS as $x_3$.

Then we can establish the basic model:

$$ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon $$  \hspace{1cm} \text{Formula 1}

$\beta_0, \beta_1, \beta_2$ and $\beta_3$ are regression coefficients, $\epsilon$ is random error (normal distribution variable and the mean is zero). The function regress(y, x) in MatLab is used to sketch the linear relation and the fitted function is formula 2.

$$ y = 52.9360 -18.1208 x_1 - 9.8441 x_2 - 2.9082 x_3 $$  \hspace{1cm} \text{Formula 2}

Error analysis of MLR(Multiple Linear Regression):

$$ y = \beta x + \epsilon : $$

The error $\epsilon$ is the total sum of square $SST$ which represents the error estimating y with the mean y (i.e. $\beta=0$). It can be disintegrated: $SST=SSE+SSR$.

$SSE$ is residual sum of squares and $SSR$ is regression sum of squares.

An ideal model is SSE as small as possible, SSR as large as possible. $R^2$ is $SSR/SST$, which indicates the degree of the model explaining the information. $R^2$ is closer to 1 showing significant of the model.
The significance tests for the model: when statistics $F$ is large ($P$ value is small corresponding), the model is significant.

The statistics in regression (1): $R^2=0.8716$, $F=142.5274$, $p=0.0000$ $\sigma^2=7.18650$. From $R^2=0.8716$, it can be known that the 87.16% of $y$ is determined by the model. The value is far from the critical value. $P$ is far below $\alpha=0.05$. Overall, the model can be valid, but the model is not very significant.

### 2.2. Four kinds of regression model are compared

Comparative tests with the function rstool(X,Y) in MatLab is shown in Figure 1.

![Figure 1. Four kinds of regression model](image)

Only linear terms in regression function (We also got this in the previous part in Formula 1):

$$y = 52.9360 - 18.1208x_1 - 9.8441x_2 - 2.9082x_3 + \varepsilon$$

Pure quadratic terms:

$$y = 65.3353 - 34.7252x_1 - 14.8548x_2 - 39.9971x_3 + 15.1829x_1^2 + 5.5215x_2^2 + 33.5993x_3^2 + \varepsilon$$

Merely interactions:

$$y = 60.7494 - 29.5947x_1 - 22.1642x_2 - 11.4795x_3 + 12.6855x_1x_2 + 8.2840x_1x_3 + 9.6943x_2x_3 + \varepsilon$$

Full quadratic polynomial form:

$$y = -40.0903 - 40.0903x_1 - 21.0661x_2 - 27.0624x_3 + 8.2960x_1^2 - 0.4951x_2^2 + 0.9152x_3^2 + 15.7529x_1x_2 + 6.0914x_1x_3 + 22.6092x_2x_3 + \varepsilon$$

By comparing the output data of the four figures, it can be seen that the full quadratic fitting has the least residual standard deviation, which the RMSE of it is the least. So the dyeing model is most probably a full quadratic model. But if the model has so many coefficients, there will be large amount of calculation in the recipe prediction to make the problem complicated a lot. So we use the method of stepwise regression analysis and select important factors, remove insignificant factors of items for the dyeing mathematic model to make the model more reasonable.
2.3. Stepwise regression analysis of the regression function

The function stepwise(x,y) in MatLab is an interactive tool for creating a regression model to predict the vector y, using a subset of the predictors given by columns of the matrix X [12].

In this paper, X is columns of all terms of full quadratic polynomial forms.

Initially, no predictors are included in the model, as is shown in Figure 2. For each term on the y-axis, the plot shows the regression (least squares) coefficient as a dot with horizontal bars indicating confidence intervals. Blue dots represent terms that are in the model, while red dots indicate terms that are not currently in the model.

Figure 2. stepwise regression

Coeff is the value of the regression coefficient for that term, t-stat is its t-statistic and p-val is p-value. The coefficient for a term that is not in the model is the coefficient that would result from adding that term to the current model. When the t-stat of a term is larger and p-val is less than other terms, which can be chosen to move into the model.

Several diagnostic statistics appear below the plot:
- Intercept - the estimated value of the constant term
- RMSE - the root mean squared error of the current model
- R-square - the amount of response variability explained by the model
- Adjusted R-square - the R-square statistic adjusted for the residual degrees of freedom
- F - the overall F statistic for the regression
- P - the associated significance probability

When X1 is moved in the model, the stepwise GUI displays the next term to add or remove. When there are no more recommended steps, the GUI displays "Move no terms." The process is move X1, X2, X7, X4, X8 in model step by step. The RMSE changed from 7.30875 to 3.9765, 2.72916, 1.8623, 1.37741 and 1.18825 at last.

The finally output is showed in Table 2:

| Coefficients: | -48.8833 | -29.0129 | 0 | 13.0145 | 0 | 0 | 20.1509 | 10.4895 | 0 |
| Confidence interval: | -54.1982 | -34.3278 | 0 | 9.4845 | 0 | 0 | 15.7103 | 6.0489 | 0 |

The regression function (Formula 3):

\[ \hat{y} = 62.7768 - 48.8833 x_1 - 29.0129 x_2 + 13.0145 x_1^2 + 20.1509 x_1 x_2 + 10.4895 x_1 x_3 \]

From the statistics (as is shown in table 3), we can see the model is reasonable.
### 3. Results and Discussion

As we see in Formula 3, it responses the influence of the three dyes concentration. $x_1$ represents the concentration of the active cyan FBN, with the linear coefficient negative showing that the increase of the concentration of FBN will cause the luminance dropping quickly. While the change of concentration for 3BS the luminance will change a little smaller. The influence of the active yellow is positive. All these results from Formula 3 are useful in modeling.

A new method is employed in color matching analysis for textile in this paper by using a ternary mixture of three dyes and regression analysis technique.

All colored samples are divided into two parts, each of which contains 100 samples. The first data set was used to evaluate the formula coefficients of the three dyes mixed, where it can be seen that the active yellow has the highest coefficients and others are lower.

The other data set is used to check the formula and the accuracy was determined by the root mean square (RMS) error of the CIELAB color fitting.

### 4. Conclusion and Outlooks

In this paper, data analysis method is used in the problems of mathematics modeling in textile dyeing, the processes of which are described in detail. On the basis of the data analysis of dyeing samples, the rules between concentration of three mixed colorants and CIE-$L^*a^*b^*$ values are detected. Experiment illustrates that the mathematical models of color information based on the different concentration of three mixed dyes are effective. A compromised color matching model in dyeing using aforesaid formula is further studied instead of conventional color matching method that using spectrum analysis.

The study in the paper provides new reference for color matching in textile dyeing and also has certain value for theory study and application research. In a study of color matching method, the accuracy of the mathematics modeling for the match prediction indicates that the regression function can reflect the actual environment characteristics of dyeing mixed. The colorimetric accuracy depends on the coefficients of different dye’s concentration and about which, the regression performs well, with sufficient accuracy in comparison with the normal method. This would allow a more wide-ranging and effective system where we need not concern the characteristics of the textile dyes and the performance of dyeing equipment, the noise effects can also been removed all or some depending on the application. In addition, with this method, regression analysis could be used to compress and reduce the size of database, so it will have a good prospect in computer color matching in textile dyeing.

### 5. References


