IFSCC 2025 full paper (IFSCC2025-1512)

"From Color to Decision: A Python-Based Approach to Predicting Delta E Ranges"

Franciane de Oliveira Cortez[1*], Carlos Magno Molinaro Fonseca[1], Vitoria Luisa Cavalheiro[1]; Rebecca Lucila Lemes Graciolli[1]; Juliana Cibi Amorim[1]; Fernanda Gabriela Henning[1]; Desirée Cigaran Schuck[1].

[1] Grupo Boticário, São José dos Pinhais, Paraná, Brazil.

1. Introduction

Color is no longer perceived by the consumer as a mere attribute, but as an important influence in the purchase decision of a product, the perception of a brand, the attractivity of the marketing campaign, the intention of making a political statement, or even the identity expressivity [5, 7]. In the cosmetic industry, color evaluation traditionally limits itself fot batch analysis, being insufficient to guarantee the customer loyalty. The act of rebuying a product, fundamental to the business longevity, relies directly on the industry capacity of delivering a product that is chromatically faithful to the marketed one, becoming a central challenge [3,7].

Methodologies based exclusively in color scales and instrumental data such as spectrophotometers, colorimeters and digital cameras, although crucial, have limitations. The complexity of color formulation and the precise combining of pigments and dyes, affects not only the visual appeal, but also texture, stability and perceived performance of the product [1,4]. Despite advances in instrumental measurement, sensorial evaluation remains crucial to determine the border between quality control requirements and perception and acceptance from the final clients. Due to the vast portfolio of lipstick colors available and desired by its users, the determination of an acceptance range to each individual shade is unfeasible. Therefore, the need to escalate its evaluation methods and the pursuit for a technique that can both translate the subjective consumers' perception and use quantitative parameters become imperative [3].

To access this challenge, we propose an innovative data-drive solution, using programming language Python. The choice of Python is justified by its clear and concise syntax, facilitating code development and maintenance, a crucial feature for building robust and scalable tools for complex data analysis. Its vast library of scientific modules, such as NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn and Plotly provide powerful tools for data manipulation, statistical analysis, and visualization, optimizing the development and analysis process. In addition, frameworks such as scikit-learn allow the efficient implementation of machine learning algorithms, essential for building accurate predictive models. Python's large user community and extensive available documentation ensure support and accessible resources, contributing to the efficiency and reliability of tool development. The large user community and extensive online documentation provide accessible support and resources for programmers of all levels, contributing to a gentle learning curve and a user-friendly development environment, ideal for both beginners and experts [10].

In this work, we explore this programming language potential to develop a tool capable of predicting acceptance of the DeltaE (ΔE) ranges of bullet shaped lipsticks. This tool integrates sensorial data collected from a consumers panel and instrumental analysis using the CIEDE2000 system, allowing us to objectively evaluate production batches and colors under development, explicitly considering our target audience perception, facilitating the process of making informed and efficient decisions in the process of developing and launching new products.

2. Materials and Methods

2.1 Data bases

2.1.1 Instrumental database

This study employed two different data bases. The first comprised instrumental data from the application of 40 shades of lipstick from a product catalog, obtained with colorimeter and expressed in the colorspace CIELab and CIELCh. Then, the data was clusterized using resources from the Scikit-learn library within four main groups (reds, pinks, browns and purples).

2.1.2 Sensorial database

Originated from a sensory panel of 600 lipstick users (men, women, and diverse skin tones) who evaluated 9 to 11 variations of nine representative shades of each subgroup from the instrumental database previously described. A discriminative sensorial analysis (difference-from-control test) was conducted using a 5-point categorical scale with complete randomization in controlled lighting and temperature dry booths. Participation was voluntary, with informed consent obtained from all participants, adhering to ethical principles, approved by an ethics committee.

The samples from both databases were measured in triplicate using the DigiEye® colorimeter (with DigiProduction v3.2.4.6 software) and the CIEDE2000 color space to quantify color differences.

2.2 Stabilishing ΔE tolerance

Each subgroup went through extensive graphical and numerical analysis, such as boxplots, scatterplots, distribution analysis and descriptive statistical analysis, all performed using the NumPy, Pandas and Matplotlib libraries in the Google Collaboratory platform.To determine ΔE thresholds, samples which had all of its variations evaluated approved by the consumers panel, had the acceptable ΔE limit defined by the third quartile of the distribution of their ΔE values. Although we have maximum ΔE limits for each color cluster (4 for browns/pinks; 3-5.6 for reds; 6.8 for purples), for these samples, the average ΔE between approved variations generates the data presented in Table I.

Table 1. ΔE thresholds defined from the mean of each main group. (Cortez et al., 2025).

Chromatic group	ΔE threshold (group mean)	max ΔE acceptance found
brown	2.83	4
pink	2.5	4
purple	4.8	6,8
red	3.55	5.6

2.3 Color space definition

The definition of the color space was performed using graphical libraries available within the Google Collaboration platform, such as Matplotlib, Seaborn and Plotly.

1. Graphical representation: Three bidimensional graphical scatter plots were created to each one of the four groups, containing the plotted data from the accepted variations provenient from the sensorial analysis. In each one of the three projections of the color space (a* x L*; a* x b* and b* x L*), an ellipse was drowned around each data set, to better fit the points.
2. Scaling the ellipse: The dimension of each ellipse was determined by the maximum difference of the coordinates x and y of the points within each data set, with a scale factor added to create a safety margin.
3. Ellipse rotation angle: The ellipse's rotation angle was calculated using the arctangent of the gradient from a linear regression on the points. A correction factor of ±90 degrees was applied to ensure the angle was within the correct 180-degree range for plotting, addressing the periodic nature of the arctangent function.
4. Creating an ellipsoid: From the superposition of all the ellipses created, consumer acceptability is defined within the CIELAB color space for each of the four subgroups. The standard of each subgroup is used to position this ellipsoid in the color sphere.

2.4 Tool development

Based on the ellipsoids defined for each cluster, a computational tool was developed in Python-3 to predict ΔE limits for new products and for batch approval/rejection in quality control. Although the source code is confidential, its logic and operation are described below:

- New Product Development: To define the specification ranges for a new product, the tool requires the Lab* color coordinates of the standard and the first production, so it can accurately positionate the tolerance space (the ellipsoid centroid). By consulting the database containing the instrumental color information (CIELab) of the 40 lipstick colors in the catalog, the tool uses Euclidean distance to determine the groups and subgroups closest to the new sample in color space. Based on this proximity analysis, the tool generates the most appropriate color specification, considering the similarity with the colors already cataloged.
- Quality Control: For products already in the catalog, the tool receives the color data (CIELab) recorded by the colorimeter during the evaluation of a production batch. The tool compares this data with the catalog's reference color specification and, using the ΔE (Delta E) calculation, identifies whether the sample is within the predefined tolerance limits. If the sample is outside the specification, the tool indicates which color coordinates (L*, a*, b*) need to be adjusted to achieve compliance with the standard and meet the established ΔE tolerance. This information assists in the process of correcting the production process to ensure product quality and consistency.

3. Results

3.1 Color space definition

For each of the four subgroups (reds, pinks, browns, and purples), the acceptable color space was defined by an ellipsoid obtained by fitting ellipses to their two-dimensional projections (a* x L*, a* x b*, b* x L*). Figures 1 to 4 show the ellipses corresponding to each cluster.

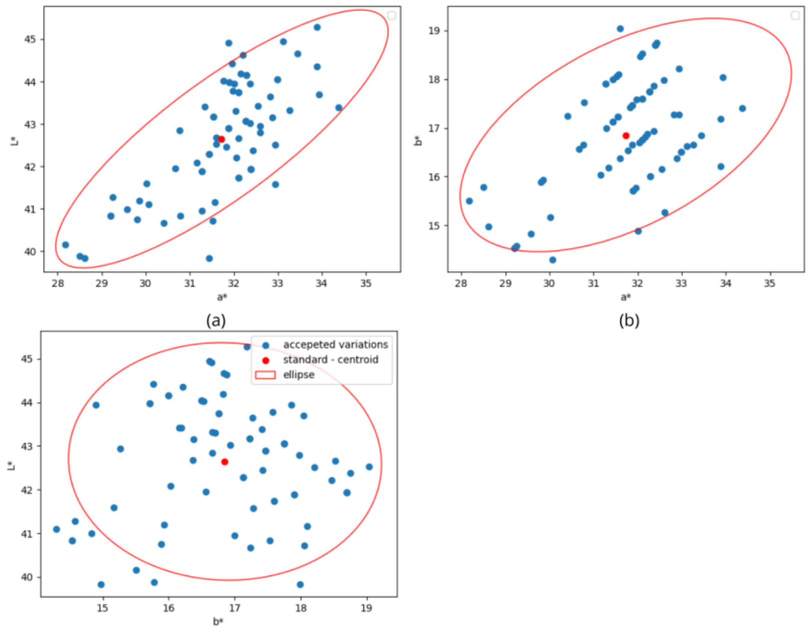

Figure 1. Ellipses representing the acceptable color space for the red cluster, in the two-dimensional projections a* x L* (a), a* x b* (b) and b* x L* (c).

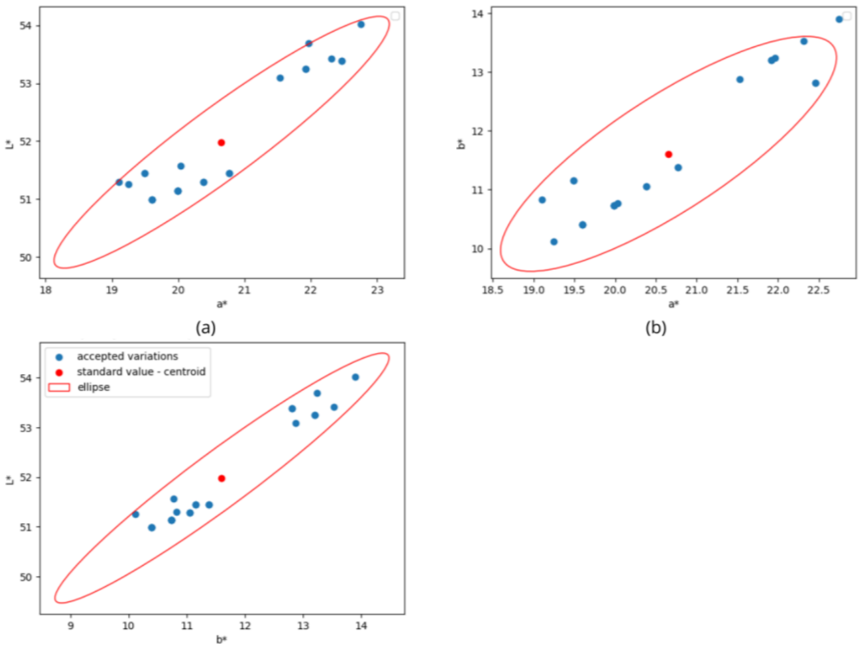

Figure 2. Ellipses representing the acceptable color space for the pink cluster, in the two-dimensional projections a* x L* (a), a* x b* (b) and b* x L* (c).

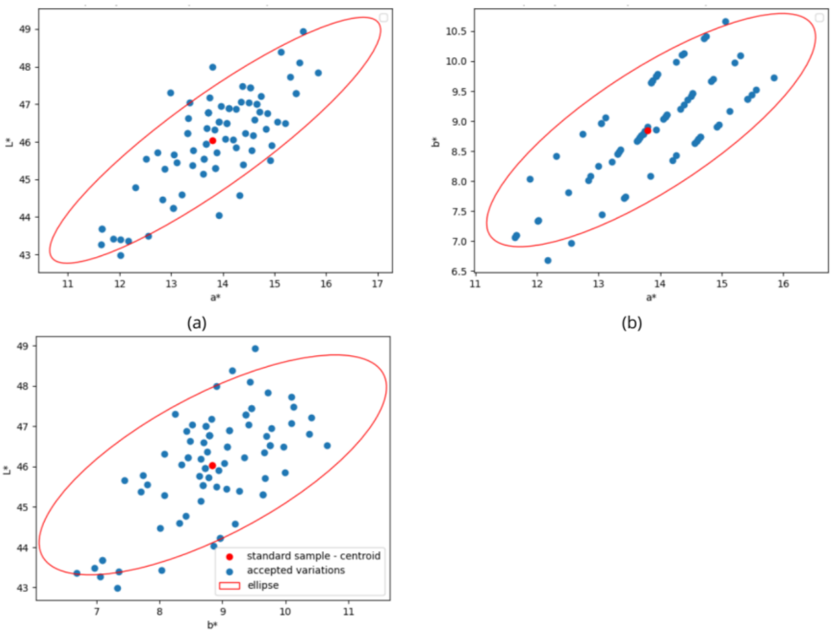

Figure 3. Ellipses representing the acceptable color space for the brown cluster, in the two-dimensional projections a* x L* (a), a* x b* (b) and b* x L* (c).

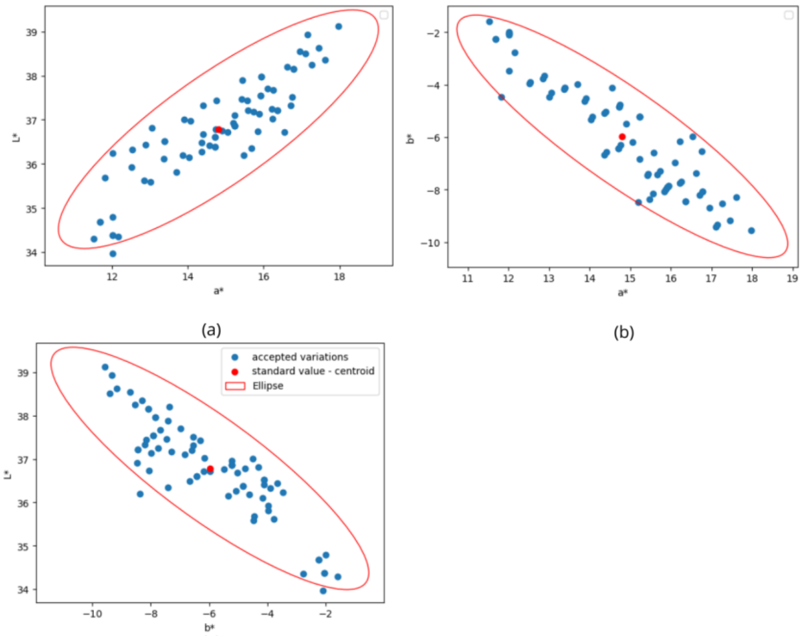

Figure 4. Ellipses representing the acceptable color space for the purple cluster, in the two-dimensional projections a* x L* (a), a* x b* (b) and b* x L* (c).

Using the ellipsoids defined for each cluster, a computational tool in Python 3 was developed to predict the ΔE thresholds, allowing the evaluation of new products and the approval/rejection of batches in quality control (Figure 5).

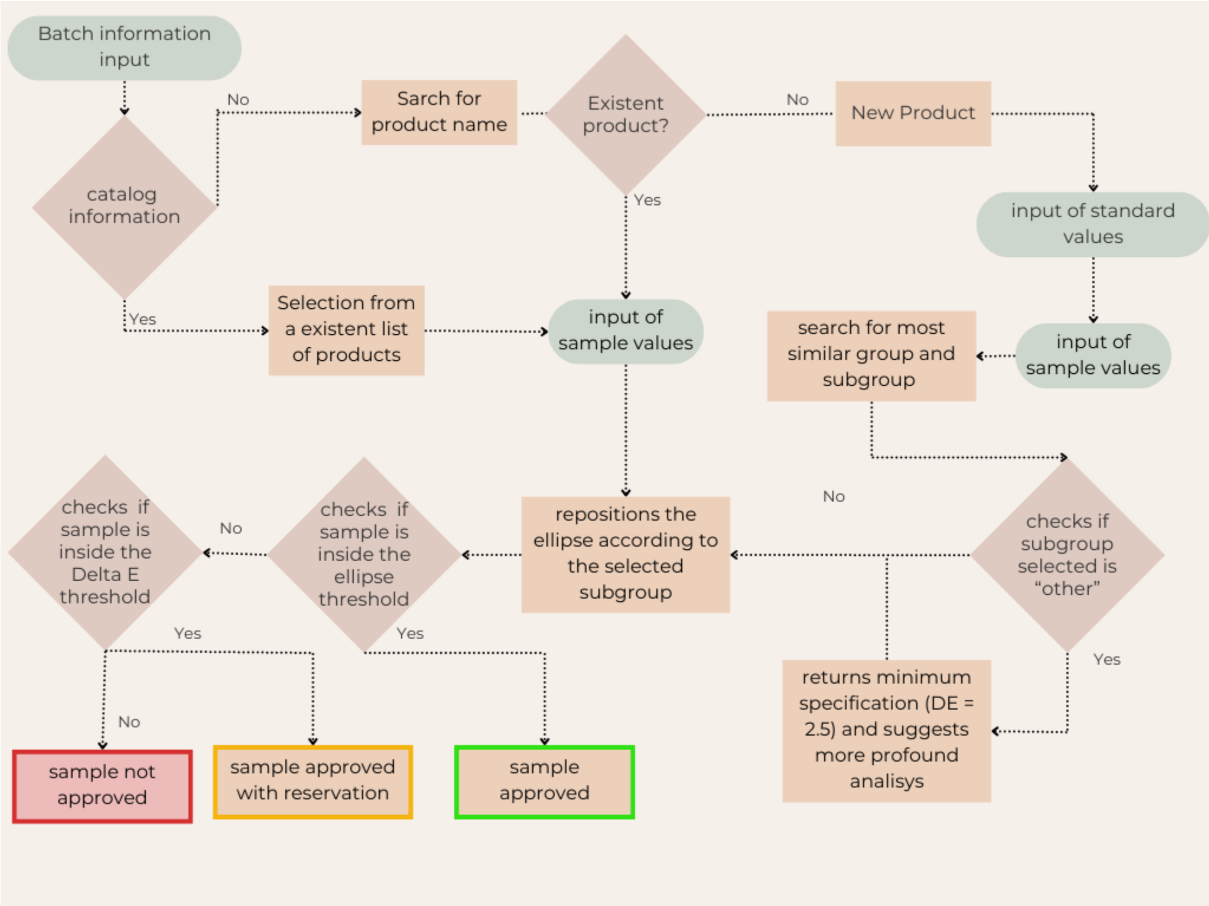

Figure 5. Tool function exemplification.

Figure 6 illustrates the tool screen interface for evaluating new products and for quality control analysis.

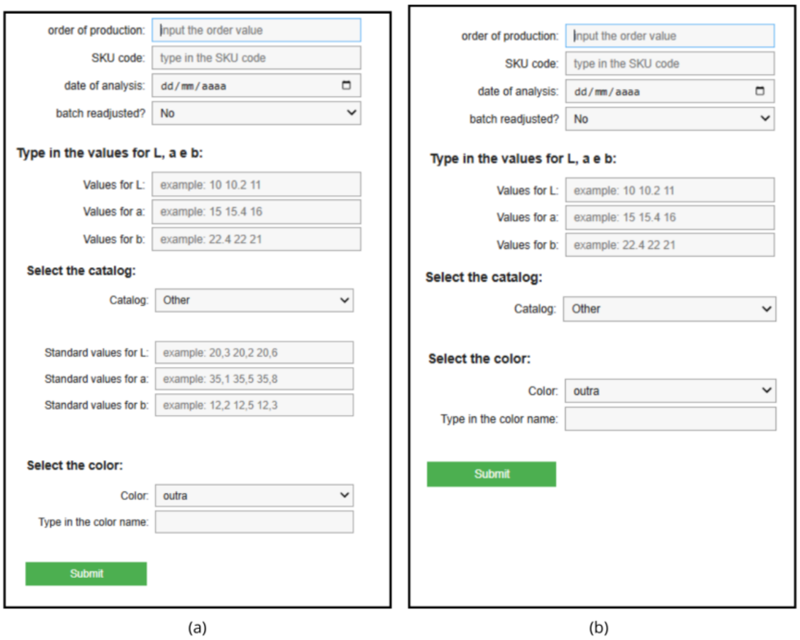

Figure 6. Tool interface for evaluating new products (a) and for evaluating products in quality control (b).

Figure 7 illustrates the type of output for the batch analysis with an approved sample, and Figure 8 points to an approved sample for the development of a new product.

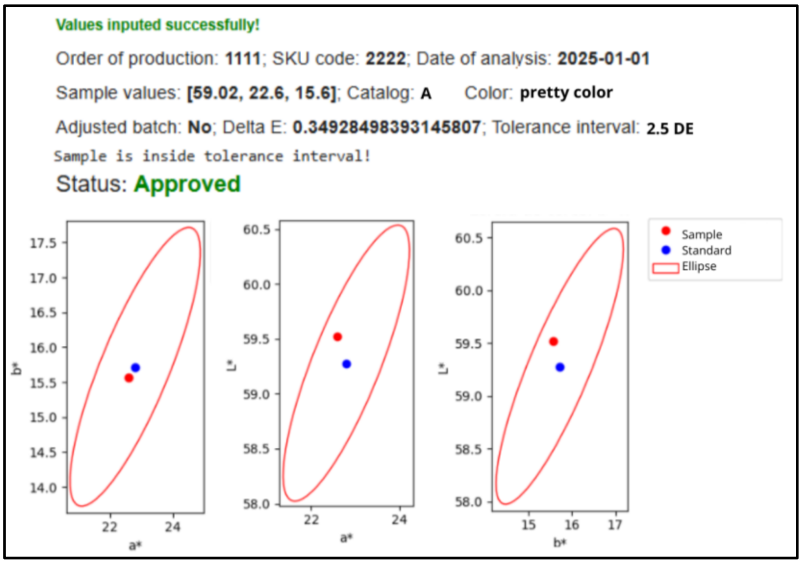

Figure 7. Quality control analysis output for an approved sample exemplification

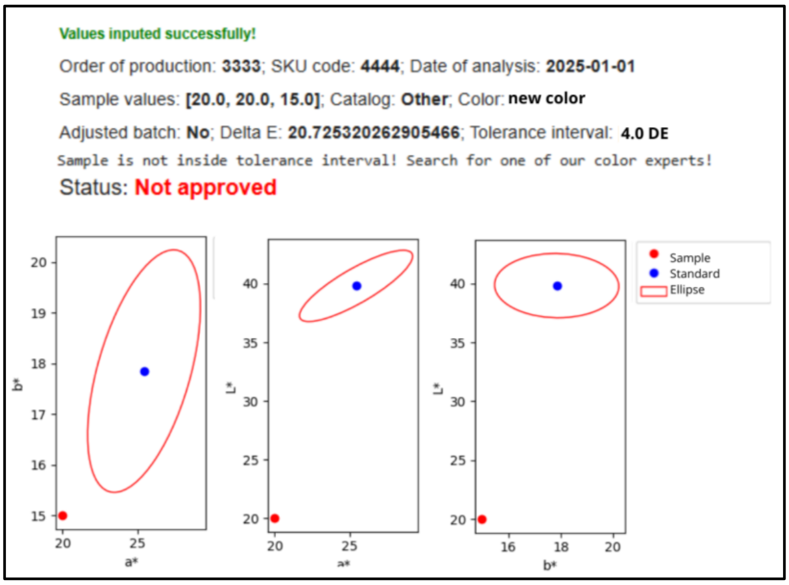

Figure 8. New product DeltaE specification output for an rejected sample exemplification.

4. Discussion

This work takes on the challenge of developing a data driven solution using Python-3 to accurately evaluate color, overcoming the traditional metrics of color differentiation and its inherent variability. By choosing this programming language, we were able to guarantee a concise and legible code, with a vast options of libraries to choose from - such as NumPy for statistical analysis, Pandas for data manipulation, and Seaborn, Plotly and Matplotlib for graphical representation - and its extensive online community, robust support community, facilitating development, maintenance and troubleshooting. The tool proposes a new methodology for validating results and adjusting colors, offering greater accuracy and reliability [10].

When considering specificities of each color and chromatic group, the thresholds defined in Table 1 corroborate with evidence previously described by Cabezas (2018), who establishes values between 2 and 4 [3], and a narrower range of 2 to 3.5 proposed by Mokrzycki et al. (2011) [15] as being non-distinguishable by common observer when comparing colors. However, one could also argue the need of embracing not only ΔE values when establishing limits of human perception, as this measurement, as effective as it is, indicates distance but not direction.

Instead of solely relying on the DeltaE (ΔE) values, the tool employs ellipsoids as a suggestion to determine human perception tolerance regarding the color differences. This approach is based on the MacAdam model, recognizing the non-uniformity in color perception by the human eye [2, 6, 8]. To each proposed group, we generate a two-dimensional ellipses representing an ellipsoid, considering its three-dimensionality (a* x L*, a* x b*, and b* x L*). The ellipses, resulting from a mathematical adjustment process, are calculated from experimental color tolerance data.

The variability of shape and size of these ellipsoids reflect the sensitivity of the human visual tolerance in different regions of the color space, more faithfully representing the perception of color difference than a simple circle or sphere of radius ΔE (Figures 1 to 4). The use of ellipsoids, instead of spheres, considers the anisotropy of color perception, where sensitivity varies in different directions from a reference point [6,8].

The tool offers two main functionalities, illustrated in Figure 5: Product Development (Figure 6A) and Quality Control (Figure 6B). For the Product Development functionality, the required input are the Lab* coordinates of the standard sample, so it can reposition the given ellipses correctly in the color space, and colorimetric data from the first product batch, comparing them with the predefined groups and subgroups. The tool then determines whether the colors of the batch fall within the acceptable range defined by the corresponding subgroup, providing a specification for the next batches of this product.

For the Quality Control function, the user inputs the Lab* coordinates of the sample and selects the target color or product for comparison. The tool compares the sample data with the registered standard, indicating its proximity within the created ellipses and whether the sample passes or fails based on its position in the ellipsoid and the predetermined DeltaE ranges (Table 1). Figures 7 and 8 show example output for passing and failing samples, including the calculated ΔE value and corresponding specification range. The tool defines color difference tolerance as the magnitude of color difference between two pre-established judgment categories (e.g., “exact match” vs. “no match”), allowing identification of the specific coordinate requiring adjustment [6, 9].

The tool is crucial in industrial colorimetry, where agreement between subjective visual color assessment and objective measurements performed with instruments is determinant. Quality control, including shade grading and evaluation of color fastness tests, is a critical application area, where high correlation between visual assessment and calculated results is essential. Reliable results require not only a robust color difference formula, but also an accurate chromatic adjustment method [2, 3, 6, 8].

In tests with lipsticks, the tool achieved a 60% approval rate (considering approvals and approvals with reservations), demonstrating its effectiveness in objectively defining color acceptance criteria. Its applicability can be extended to other cosmetic products (foundations, liquid lipsticks, hair dyes) and to various situations that require precise color control, such as the replacement of formulas or ingredients and the updating of color catalogs. By offering objective and precise criteria, the tool contributes to more efficient color management and reduces the subjectivity often associated with visual evaluation.

5. Conclusion

This paper presents a computational tool in Python that significantly improves color evaluation in industry, especially in the cosmetic sector. By integrating sensory and instrumental data, and using MacAdam’s ellipsoid modeling to represent human color perception, the tool overcomes the limitations of traditional metrics based solely on ΔE. Its ability to assist both in the development of new products and in quality control demonstrates its versatility and practical applicability. The 60% approval rate in lipstick tests indicates its effectiveness, although future research is needed to optimize the model and expand its validation to different products and contexts.

The choice of using Python as a programming language, with its robust libraries and concise syntax, has facilitated the development and maintenance of a tool that promises to increase the accuracy, efficiency and objectivity of color management, reducing the subjectivity inherent in visual evaluation and contributing to a more efficient and reliable production process. The innovative approach proposed in this work paves the way for more accurate and customized solutions in the color quality control industry.

References

[1] Chu SJ, Trushkowsky RD, Paravina RD (2010) Dental Color Matching Instruments And Systems . Review Of Clinical And Research Aspects. 2–16.

[2] Budak, V. P., Zheltov, V. S., Meshkova, T. V., & Notfullin, R. Sh. (2017). Evaluation of illumination quality based on spatial-angular luminance distribution. Light & Engineering Svetotekhnika, 25(4), 24-31.

[3] David H. - Performance Testing of Color-Difference Metrics Using a ColorTolerance Datase

[4] Hofer H, Carroll J, Neitz J, NeitzM, Williams DR (2005) Organization Of The Human Trichromatic Cone Mosaic. Journal Neuroscience.

[5] Kim A, Kim H, Park S (2011) Measuring Of The Perceptibility And Acceptability In Various Color Quality Measures. Journal Of The Optical Society Of Korea, 310-317.

[6] Klaus Witt - Three-Dimensional T hreshold ofColor-Difference Perceptibility inPainted Samples: Variability ofObservers in Four CIEColor Regions*

[7] Mohebbi B (2014) The Art Of Packaging: An Investigation Into The Role Of Color In Packaging, Marketing, And Branding. International Journal Of Organizational Leadership, 92-102.

[8] M. R. Luo B. Rigg Chromaticity=DiscriminationEllipses for Surface Colours

[9] Rolf G. - Industrial Color Difference:Progress and Problems

[10] Wrolstad RE, Smith DE (2017) Color Analysis. Department Of Food Science And Technology, 545-554.