Symbolic Regression: An Alternative Method to Model the Optical Response of Photonic Biological and Bio-inspired Structures

JULIAN SIERRA-VELEZ[1], MARINA INCHAUSSANDAGUE[2,3], DIANA SKIGIN[2,3], ALEXANDRE VIAL[1], HENDRIK HÖLSCHER[4], AND DEMETRIO MACIAS[1,*]

[1] *Laboratory Light, nanomaterials & nanotechnologies - L2n, University of Technology of Troyes & CNRS EMR 7004, 12 Rue Marie Curie, 10004 Troyes, France*
[2] *Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Física, Buenos Aires, Argentina*
[3] *CONICET - Universidad de Buenos Aires, Instituto de Física de Buenos Aires (IFIBA), Buenos Aires, Argentina*
[4] *Institute for Microstructure Technology, Karlsruhe Institute of Technology (KIT), H.-v.-Helmholtz Platz 1, 76344 Eggenstein-Leopoldshafen, Germany*
[*] *demetrio.macias@utt.fr*

Compiled December 17, 2025

In this contribution we assess the performance of Symbolic Regression (SR) when used to model the optical response of biological and bio-inspired structures. To this end, we search for analytical closed-form expressions that model the reflectance spectra related to the *Tersina viridis* bird's plumage and the porous structure inspired on the *Cyphochilus insulanus* beetle. Our numerical results demonstrate the high prediction accuracy of the employed Symbolic Regression scheme. The retrieved models not only capture the dependency of the optical response on various relevant geometrical and illumination parameters, they are are dimensionally homogeneous.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

The advent of increasingly powerful computers, together with great advances in computer science over the past decades, have made the use of Artificial Intelligence (AI) a common activity in physical sciences [1]. Notwithstanding this progress, the readability and interpretability of Deep Learning models are still under discussion, and significant efforts are currently being conducted in that direction [2]. On the other hand, Symbolic Regression (SR) is an AI method that explores, unlike traditional regression techniques which adjust data to predefined models, a wide space of functions to identify an appropriate closed-form expression that describes a fundamental problem [3].

Recently, some of us have successfully employed SR to characterize the optical properties of dielectric and biological materials [4, 5]. Those works illustrated the capability of SR to find readable dispersion models from far-field spectral information, without the need for any preliminary hypothesis concerning the algebraic form of the expression to be retrieved. Furthermore, the results obtained also suggested the possibility of using SR as a meta-model to compute the optical response of a given structure. This means that, at the end of the SR process, there is a closed-form expression that provides results equivalent to those obtained with a computationally expensive electromagnetic method. This expression is an explicit relationship between the optical response and different geometrical parameters of the studied structure. The approach based on SR presents an important advantage over a neural network-based meta-model [6], since the latter does not provide a closed-form expression but only a numerical model.

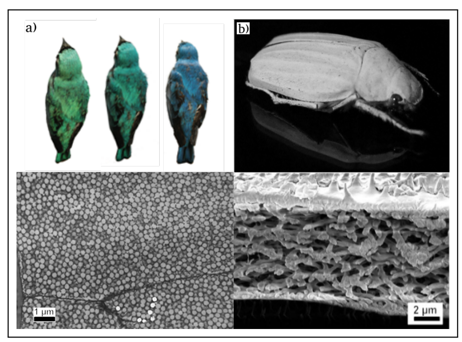

Fig. 1. Two examples of photonic biological structures found in nature, and the structural morphology responsible for their coloration. a) Male *Tersina viridis* bird, with a TEM image of a transversal cut section of a barb from a back feather. b) White beetle *Cyphochilus insulanus*, with a sectional view of an SEM image of the white scales that cover its body (adapted from Ref. [7] Licensed under CC BY 4.0. [No changes were made to the original image]).

In this letter, we make use of SR to model the optical responses of the *Tersina viridis* bird's plumage [8] and of a scattering polymeric porous film, inspired by the scale of the *Cyphochilus insulanus* beetle's wing [7], both shown in Fig. 1. These are interesting illustrations of the structural coloration mechanism, a consequence of the interaction between light and the multiple scattering centers present in their complex morphology [9, 10]. The novelty and importance of this work lies in its potential to establish, from experimentally measured or numerically generated data, readable closed-form meta-models that characterize the optical response of the biological structure studied. This can be particularly useful in situations where the existing models are not suitable for the problem at hand or they are computationally expensive. SR also provides an alternative approach to settle the basis for the characterization of the optical response and design of photonic bio-inspired structures.

Throughout this work we use the state-of-the-art open-source Python library, recently proposed by Tenachi *et al.*, known as *Physical Symbolic Optimization* (PhySO) [11]. This novel framework enhances SR's capabilities by incorporating dimensional analysis into the optimization process. Traditional SR methods frequently overlook the physical units of the data, often leading to solutions without a physical meaning. PhySO addresses this issue by narrowing the search space to physically plausible solutions. This ensures that the expressions found adhere to unit constraints. PhySO's workflow begins with the data collection and preprocessing, followed by the generation of symbolic expressions using a combination of deep reinforcement learning and Recurrent Neural Networks (RNNs). These expressions are then iteratively refined through optimization techniques, balancing model complexity and performance to produce readable and accurate models. This iterative refinement ensures that the resulting symbolic expressions are dimensionally homogeneous, making PhySO particularly useful for applications in physics and related disciplines. In a practical situation, PhySO requires to pre-define the dimensions of the target expression, the variables involved, and any constants that might be present in the final expression. It is noteworthy to mention that in this work we follow the rule of thumb established by the authors of Ref. [11]: we allow at least one free constant for each independent variable with its corresponding dimensions and at least one free constant with the dimensions corresponding to the expression sought. Furthermore, to ensure dimensional homogeneity PhySO enforces that all the operations take into account the units of its operands.

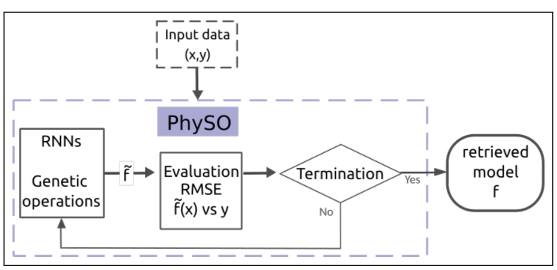

Fig. 2. Flux-diagram of the PhySO framework as "glass-box". It illustrates the generality of the SR implementation, where there is an input dataset (x, y), and a closed-form f expression as output.

Figure 2 illustrates the work-flow of PhySO, where for the sake of clarity we consider only one one-dimensional function f such that $y = f(x)$. For more details, we refer the interested reader to Ref. [11] describing PhySO's operating principles. The work-flow of PhySO begins with the generation, as described in the previous paragraph, of a dimensionally homogeneous closed-form expression $\tilde{f}$, which is evaluated on each of the input values of x to give $\tilde{y} = \tilde{f}(x)$. Then, as indicated in Fig. 2, $\tilde{y}$ is compared to the input data y using a metric, the Root Mean Square Error (RMSE) in this work. The RMSE quantifies the fit, with a perfect match at $RMSE = 0$. The numerical evidence in our work suggests that values $RMSE < 0.05$ are an acceptable match. If $\tilde{f}$ does not fit the precision criterion to reproduce the input data y, then it is redirected again to the first step to undergo genetic variations and optimization. On the other hand, if $\tilde{f}$ accurately reproduces the input data y, the regression loop ends and the expression f is given as a result.

As stated previously, to assess the performance of SR within the context of this contribution, we consider as our case studies the green-blue colored feather's structure depicted in Fig. 1(a) and the white scattering polymeric porous structure depicted in Fig. 1(b). To facilitate the visualization and discussion of our results, we use the following line styles for all our numerical experiments: the spectra that serve as input to the SR scheme are depicted with a dotted blue line while the spectra generated with the expression retrieved through SR are depicted with orange cross-like markers.

Case Study 1: Spectral reflectance of *Tersina viridis* bird's plumage. Some of us have previously studied the structural color mechanism of the *Tersina viridis* plumage [8] depicted in Fig. 1(a), whose hue changes remarkably from greens to blues as the angle between the illumination and observation directions increases. This color effect is a consequence of the microstructure present in the feathers' barbs, which consist of quasi-spherical air voids in a β-keratin matrix. The barbs' microstructure was modeled considering the geometrical model illustrated in Fig. 3. It consists of an N-layer system of air-filled spherical voids of radius r in a matrix. The voids are arranged in a hexagonal lattice with lattice parameter a. To simulate the disorder present in the natural prototype, the reflectance curves were obtained by averaging the responses corresponding to a set of identical structures that differ in their lattice constant.

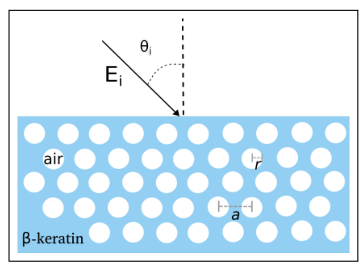

Fig. 3. Geometrical model considered for the KKR computations. The composite system is illuminated with an non-polarized plane wave of wavelength λ, at an incident angle θ_i with respect to the surface normal.

The spectral information that serves as input to PhySO could have been measured experimentally, or numerically generated with any well suited numerical method. However, for this case study, we made use of the KKR method and the averaging technique described in Ref. [8]. The results of our numerical simulations are depicted in Fig. 4, where the spectra were obtained assuming unpolarized illumination and six different angles of incidence. The closed-form expression retrieved through our SR scheme, from the input spectra in Fig. 4(a) is

$$R(\lambda, \theta_i) = \frac{A_1}{-A_2 + \lambda(B_1 + \sin(3B_1 + B_2 + \theta_i + \frac{2(A_1+\lambda)}{A_2}))} \quad , \tag{1}$$

where the values of the free constants are $A_1 = -0.063\ \mu m$, $A_2 = 0.227\ \mu m$, $B_1 = -0.331$, and $B_2 = -0.809$. Note that the variable θ_i in Eq. 1 should be considered in radians, but for clarity we use degrees in the text and figures.

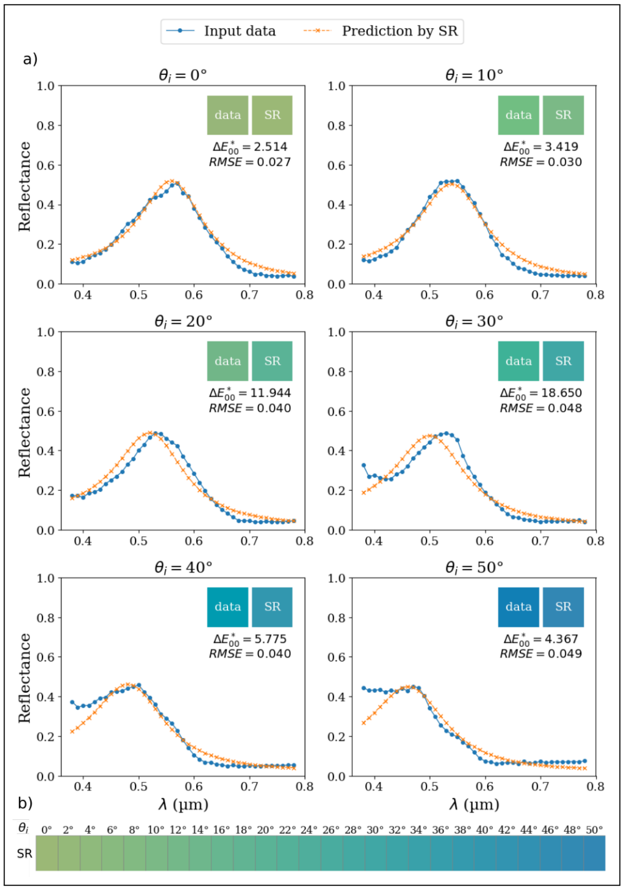

Fig. 4. (a) Reflectance spectra and corresponding sRGB colors for *Tersina viridis* feathers at different angles θ_i. The upper right corner shows the spectrum-related color and ΔE^*_{00} in CIE-Lab. RMSE between the input data and SR prediction is also shown. (b) Colors predicted by Eq. 1.

In Fig. 4(a) we present a visual comparison between the input spectra and the predictions of Eq. 1, and their corresponding RMSE metric values. We observe a good agreement for angles of incidence up to $\theta_i = 50°$. The SR model captures the essential features of all the input reflectance spectra, including the peak position and the overall shape, showing its effectiveness in approximating the bird's spectral response. Furthermore, the model accurately follows the trend of decreasing reflectance at higher wavelengths. As the incident angle increases, there is a noticeable shift and broadening of the reflectance peak. The SR model successfully captures these changes, illustrating its robustness in handling variations in the optical response due to different incident angles.

A complementary test, to assess the performance of our SR scheme, is to compare the colors related to Eq. 1 with those associated with the input spectra shown in the insets in Fig. 4(a). To make this comparison in an objective way we used the color difference ΔE^*_{00} metric from the CIE-Lab color space, which stipulates that two colors are perceptually indistinguishable if $\Delta E^*_{00} \leq 1$ [12].

The results presented in Fig. 4(a) indicate that, despite significant color differences ΔE^*_{00} in some cases, the spectra predicted by Eq. 1 closely replicate the expected coloration of the bird's plumage. Additionally, in Fig. 4(b), we showcase SR's potential to predict the structure's chromatic response at different angles of incidence not included in the input dataset. To do so, we used Eq. 1 to calculate $R(\lambda, \theta_i)$ for intermediate θ_i values in the training range [0°,50°], and then we used these spectra to compute their corresponding colors. These figures visually illustrate the high accuracy of the model's predictions, as they not only reproduce the coloration for the input θ_i angles but also show good agreement with the expected green-blue tones of the bird's plumage.

Case Study 2: Spectral reflectance of *Cyphochilus insulanus* beetle bio-inspired structure. For the second case study, the input to our SR scheme are the experimentally measured reflectance spectra originally reported by some of us in Ref. [7]. In that work, the structure of the scales of the *Cyphochilus insulanus* beetle, shown in Fig 1(b), were replicated using a foaming process by saturation with CO_2. As schematically represented in Fig. 5(a), a PMMA resist was spin-coated on glass, covered with a second glass slide, and clamped between two neodymium magnets to prevent film deformations during foaming. This assembly was placed in a custom-built high-pressure cell connected to a CO_2 source. As shown in Fig. 5(b), the application of suitable pressure up to 50 MPa, along with controlled temperature and saturation times, forms a nano-cellular foam during the final rapid depressurization step.

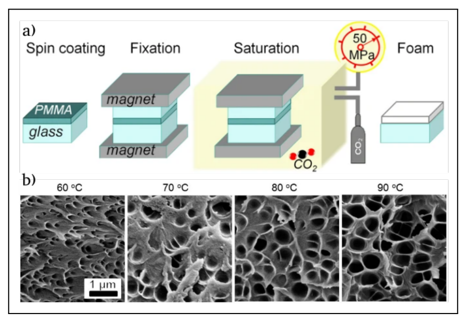

Fig. 5. a) Schematic of the foaming process by saturation with CO_2, b) SEM images of the porous films, showing that the geometrical parameters of the structure depend on the fabrication temperature varied between 60° and 90° (adapted from Ref. [7], Licensed under CC BY 4.0.

In this example we searched, from the reflectance spectra in Fig. 6, for a closed-form expression of R dependent on relevant geometrical parameters of the porous film such as, the thickness D and the average pore size d. The model retrieved by the SR scheme is of the form

$$R(\lambda, D, d) = -B_1 + B_2 + \frac{A_2 - \lambda}{A_3} - \frac{\lambda}{\lambda + A_2} - \frac{A_1 - A_2}{D} - \frac{A_1}{A_3 - d - D}, \tag{2}$$

with free constants values $A_1 = -0.003\ \mu m$, $A_2 = 3.866\ \mu m$, $A_3 = 10.274\ \mu m$, $B_1 = 0.643$, and $B_2 = 1.357$. Eq. 2 provides a compact and readable model for the reflectance of the porous PMMA structure. The third and fourth terms capture the decrease in reflectance with increasing wavelength observed in the experimental data. The last two terms indicate an inverse relationship of the reflectance with the geometrical parameters of the structure.

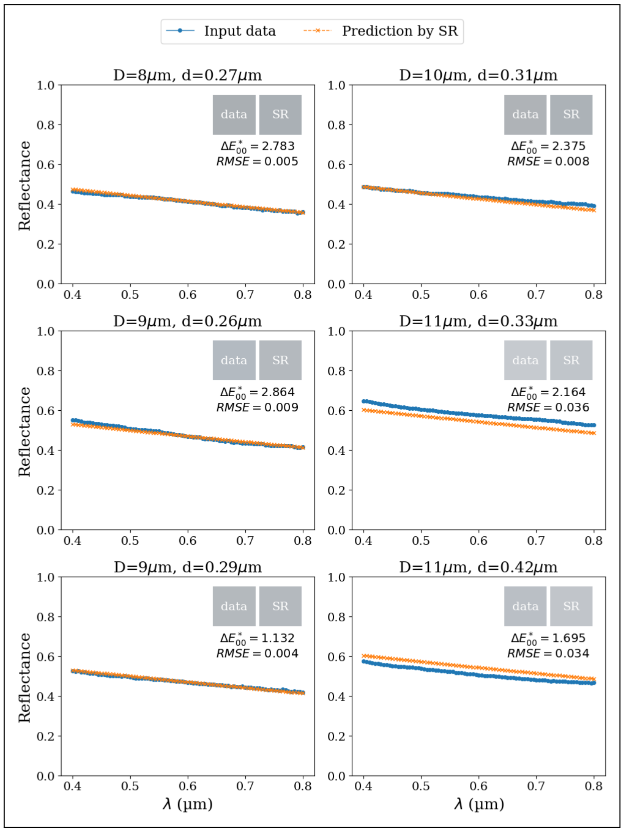

Fig. 6. Reflectance spectra and sRGB colors for the porous film in Fig. 5(b) for varying film thickness D and average pore size d. The upper right corner of each subplot shows the color and ΔE^*_{00} in CIE-Lab. RMSE values between the input data and SR prediction are also provided. The colors appear gray on the white background.

In Fig. 6 we visually compare the input experimental reflectance spectra, related to the polymeric porous film in Fig. 5, and the reflectance spectra predicted by Eq. 2. As in the previous case study, we also use the RMSE metric to make our comparisons quantitatively. We observe a good agreement between the experimental and predicted spectra. Moreover, each plot corresponds to a different set of parameters (λ, D, d), with $\lambda \in [0.4, 0.8]\ \mu m$. The SR model captures the trend of decreasing reflectance with increasing wavelength across all conditions. Although there are slight differences between the predicted and the experimental data, the SR model reliably approximates the overall behavior of the reflectance spectra. As we did in the first case study, in Fig. 6 we also show the colors generated by the experimental and the predicted reflectance spectra, together with their respective color differences ΔE^*_{00} in the CIE-Lab color space. The previous comparison suggests that the spectra generated through the SR's model reproduce the white scattering coloration of the *Cyphochilus insulanus* beetle-inspired structure.

In both of the case studies presented, the differences we observed between the target data and the SR's predictions can be attributed to two causes. The first is the random nature of the SR's operating principles, as the solution obtained depends on the initial state of the algorithm. The second cause is that the same closed-form expression that models the reflectance spectrum, should reproduce not only the input data but any other spectrum that also verifies the illumination conditions established for our numerical experiments. Despite this strong constraint, SR provides accurate and readable models.

A first conclusion of this work is that, contrary to Deep Learning-based approaches, SR is a glass-box that finds a closed-form expression, which establishes an explicit relationship between the optical response of complex biological or bio-inspired structures and the geometrical or illumination parameters of the problem. Furthermore, SR can be used as a meta-model to solve the direct problem, and in this way, avoid the use of computationally expensive methods.

The results presented in this contribution are encouraging. However, further work is still required to establish the validity limits of SR-based approaches. Furthermore, although the closed-form expressions found are readable and their dimensional homogeneity is verified, their physical meaning is still an open question.

Due to its generic nature, there are not visible restrictions to extend the predictive capability of SR to other applications requiring precise optical characterizations, such as in materials science and biomimetics. The ability to generate readable closed-form models, that capture dependencies on multiple parameters, makes SR a valuable tool for designing and understanding complex optical materials and systems.

Acknowledgments.

This work has been done within the framework of the Graduate School NANO-PHOT (École Universitaire de Recherche, PIA3, contract ANR-18-EURE-0013).

M.I. and D.S. acknowledge partial support from Universidad de Buenos Aires (UBACyT 20020190100108BA) and CONICET (PIP 11220170100633CO and PIP 11220210100299CO).

The images in Fig. 1(b) and Fig. 5 are adapted from Ref.[7].

Data availability. No data were generated on the present research.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

REFERENCES

1. D. Angelis, F. Sofos, and T. E. Karakasidis, Arch. Comput. Methods Eng. 2023 **1**, 1 (2023).
2. D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, Artif. Intell. Rev. pp. 1–66 (2022).
3. L. Billard and E. Diday, *Symbolic Regression Analysis* (Springer Berlin Heidelberg, 2002), pp. 281–288.
4. Q. Li, D. Macias, and A. Vial, Opt. Express **30**, 41862 (2022).
5. J. Sierra-Velez, D. Macias, A. Vial, and M. A. Giraldo, IFMBE Proc. **89**, 377 (2023).
6. V. Kalt, A. K. González-Alcalde, S. Es-Saidi, *et al.*, JOSA A **36**, 79 (2019).
7. J. Syurik, R. H. Siddique, A. Dollmann, *et al.*, Sci. Reports 2017 7:1 **7**, 1 (2017).
8. D. C. Skigin, M. E. Inchaussandague, C. D'Ambrosio, *et al.*, Optik **182**, 639 (2019).
9. K. Vynck, A. Pitelet, L. Bellando, and P. Lalanne, *Specular Reflection and Transmission of Electromagnetic Waves by Disordered Metasurfaces* (Springer International Publishing, Cham, 2023), pp. 389–417.
10. A. Garcia-Valenzuela, E. Gutierrez-Reyes, and R. G. Barrera, J. Opt. Soc. Am. A **29**, 1161 (2012).
11. W. Tenachi, R. Ibata, and F. I. Diakogiannis, The Astrophys. J. **959**, 99 (2023).
12. G. Sharma, W. Wu, and E. N. Dalal, Color. Res. & Appl. **30**, 21 (2005).

FULL REFERENCES

1. D. Angelis, F. Sofos, and T. E. Karakasidis, “Artificial intelligence in physical sciences: Symbolic regression trends and perspectives,” Arch. Comput. Methods Eng. 2023 **1**, 1–21 (2023).
2. D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, “Explainable artificial intelligence: a comprehensive review,” Artif. Intell. Rev. pp. 1–66 (2022).
3. L. Billard and E. Diday, *Symbolic Regression Analysis* (Springer Berlin Heidelberg, 2002), pp. 281–288.
4. Q. Li, D. Macias, and A. Vial, “Modeling the optical properties of transparent and absorbing dielectrics by means of symbolic regression,” Opt. Express **30**, 41862–41873 (2022).
5. J. Sierra-Velez, D. Macias, A. Vial, and M. A. Giraldo, “Retrieving the refractive index of a biological material via symbolic regression,” IFMBE Proc. **89**, 377–384 (2023).
6. V. Kalt, A. K. González-Alcalde, S. Es-Saidi, *et al.*, “Metamodeling of high-contrast-index gratings for color reproduction,” JOSA A **36**, 79–88 (2019).
7. J. Syurik, R. H. Siddique, A. Dollmann, *et al.*, “Bio-inspired, large scale, highly-scattering films for nanoparticle-alternative white surfaces,” Sci. Reports 2017 7:1 **7**, 1–11 (2017).
8. D. C. Skigin, M. E. Inchaussandague, C. D’Ambrosio, *et al.*, “How the observed color of the swallow tanager (tersina viridis) changes with viewing geometry,” Optik **182**, 639–646 (2019).
9. K. Vynck, A. Pitelet, L. Bellando, and P. Lalanne, *Specular Reflection and Transmission of Electromagnetic Waves by Disordered Metasurfaces* (Springer International Publishing, Cham, 2023), pp. 389–417.
10. A. Garcia-Valenzuela, E. Gutierrez-Reyes, and R. G. Barrera, “Multiple-scattering model for the coherent reflection and transmission of light from a disordered monolayer of particles,” J. Opt. Soc. Am. A **29**, 1161–1179 (2012).
11. W. Tenachi, R. Ibata, and F. I. Diakogiannis, “Deep symbolic regression for physics guided by units constraints: Toward the automated discovery of physical laws,” The Astrophys. J. **959**, 99 (2023).
12. G. Sharma, W. Wu, and E. N. Dalal, “The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations,” Color. Res. & Appl. **30**, 21–30 (2005).