Shifting paradigms in Ocean Color: Bayesian Inference for Uncertainty-Aware Chlorophyll Estimation

Erdem M. Karaköylü[1], **Susanne E. Craig**[2]

[1]Independant Researcher,
[2]NASA/UMBC,

Corresponding author: Erdem M. Karaköylü, `erdemk@protonmail.com`

Abstract

Placeholder

Plain Language Summary

Placeholder

1 Introduction

1.1 Historical Context of Chlorophyll Algorithms

Satellite ocean color observations have long been fundamental for monitoring marine ecosystems, as they enable global estimation of chlorophyll-a (Chl_a) — a key indicator of phytoplankton biomass and ocean productivity. Early empirical algorithms, notably developed by O'Reilly et al. (John E. O'Reilly et al., 1998; John E. O'Reilly et al., 2000), established the OCx family (where x denotes the number of bands used) of polynomial regression models. These models relate blue-to-green reflectance ratios (after log-transformation) to in situ Chl_a, employing either straight band ratios (BR) or maximum band ratios (MBR)—the latter selecting the highest available blue-to-green ratio for any given observation as input to a high-order polynomial. These formulations have served as the operational foundation for chlorophyll-a products across a broad range of satellite ocean color sensors—from the pioneering Coastal Zone Color Scanner (CZCS) through SeaWiFS, MODIS, and MERIS to more recent missions—offering a straightforward and robust approach for Case-1 waters. However, their performance is more limited in optically complex Case-2 waters and remains sensitive to atmospheric correction errors.

Subsequent refinements were introduced to address these deficiencies. For example, Hu et al. (Hu et al., 2012) proposed a Color Index (CI) formulation that employs a band-difference approach to reduce sensitivity to residual atmospheric errors and instrument noise, with further improvements enhancing inter-sensor agreement (Hu et al., 2019). The increasing availability of calibration data (e.g., (**Valente2015?**)) and ongoing algorithmic improvements have led to the development of additional variants of the OCx algorithms—specifically, the OC5 and OC6 formulations. O'Reilly and Werdell (John E. O'Reilly & Werdell, 2019) maintain that OC5 extends the spectral basis by incorporating the 412 nm band, thereby exploiting its strong signal in clear, oligotrophic waters, while OC6 replaces the traditional denominator with the mean of the 555 and 670 nm reflectances, with the aim of improving the dynamic range at low chlorophyll concentrations. In total, (John E. O'Reilly & Werdell, 2019) propose 65 versions of BR/MBR OCx type algorithms for 25 sensors—on average, two or more variants per sensor. With this arsenal, it is hoped, researchers are better equipped to address the wide array of bio-optical environments encountered in global ocean color applications.

1.2 Limitations of Existing Approaches

Regrettably, the development of traditional ocean color predictive models relies on a fundamental statistical error that plagues most of observational science today; that of conflating sampling probability, with inferential probability(Scheemaekere & Szafarz (2011).) Consider a dataset D that consists of input-output pairs - e.g. Remote sensing reflectance (Rrs) and chlorophyll-a concentration Chl_a - and a model M, such as OCx, hypothesized to describe the statistical association between them. The sampling probability $p(D|M)$ represents the probability of observing the data D if model M were "true". Standard model fitting maximizes this quantity (i.e. the likelihood) by tuning the parameters of M.

The unspoken assumption is that the model that best fits the data also best represents the underlying process. This constitutes an epistemic fallacy; substituting $p(D|M)$ for $p(M|D)$, in a violation of the rules of conditional probability inversion as framed by Bayes' theorem. While in well-behaved data-rich problems, the maxima of both expressions may coincide, this is an exception, not the rule. The consequences of this mistake are far-reaching as argued by (Clayton, 2022), with implications extending from scientific modeling to medicine, law, and public policy.

In science, this fallacy contributes to models fail to generalize well, an impetus to produce ad-hoc or post-hoc uncertainty quantification, and yield published results that resist replication(Cobey et al. (2024)). These limiations are not confined to traditional hypothesis testing; they are embedded in the training.; they are embedded in the training and deployment of modern machine learning models as well. In regression and classification tasks, likelihood maximization is treated as sufficient for inference, even though it results in a single point estimate and fails to account for parameter uncertainty or model plausibility.

This epistemic shortcut has already been explicitly critiqued in the machine learning community. Gal (Gal, 2016) and Ghahramani (**gahramani2015probabilistic?**) emphasize that most ML models discard uncertainty entirely and treat the output of optimization as inference. This leads to overconfident predctions and brittle generalization, echoing Clayton's concerncs. Bishop (Bishop, 2006) also distiguishes between the utility of models and the inferential core required to quantify uncertainty, underscoring that likelihood is not enough and that the Bernoulli fallacy pervades the very logic of applied machine learning.

1.3 Overcoming limitations

There have been attempts to address these issues. (Seegers et al., 2018) have proposed alternative metrics to circumvent the inadequate assumptions of the frequentist approach. Others have tried to go a step futher incorporate Bayesian concepts. E.g. (Frouin & Pelletier, 2013) have proposed a Bayesian inversion scheme for atmospheric correction. (**shi2015?**) proposed a probabilistic method to merge data from different sensors. (Craig & Karaköylü, 2019) have proposed a Bayesian neural network (BNN) approach using Hamiltonian Monte Carlo sampling to retrieve Inherent Optical Properties (IOP) from Top-of-the-atmosphere (TOA) radiance. (**werther2022?**) used Monte-Carlo dropout to approximate a deep BNN. (Erickson et al., 2023) have proposed using conjugate Gaussian prior and likelihood to frame the GIOP as Bayesian model to predict IOPs with uncertainty. (**hammout2024?**) have proposed a BNN approximation via Stochastic Variational Inference to predict Chl_a from ocean color observations. Yet most of these approaches retain variable levels of frequentism by applying only part of what is now commonly referred to as the Bayesian workflow(**bgelman2019?**).

1.4 In search of our Bayesian workflow

The Bayesian workflow is a recognizable and unifying way of approaching statistical modeling, initially championed by statisticians at Columnbia University (**gelman_bda?**) subsequently quick endorsed and further developed by researchers dealing with scant and noisy data (e.g. (**mcelreath_stat_rethink?**)). That is not to say that all steps are applicable to all fields of research and all intents. The workflow can and should be tailored to the researcher's field. (Wolkovich et al., 2024) has proposed a four-step process for Ecological modeling. However there are some core steps that should be followed. (1) Begin with one, or better yet, more than one conceptual models and code them up. Model building include prior formulation to encode existing knowledge. Priors are part of the model structure and make the researcher's assumptions transparent, providing a first avenue of critique and improvement. (2) Check assumptions by a simulation process known as *Prior Predictive Checks*. Bayesian models are generative, meaning they can run on empty, that is produce results without data. This allows a sanity check of the built-in assumptions. Non-sensical simulation results indicate assumptions must be revisited. (3) Collect data and conduct exploratory data analysis. Knowing the data will be critical to understanding potential problems during fitting; e..g multicollinearity of input features. (4) Diagnostics; the fitting process and the resulting posterior distribution provide rich constructs that can be mined for a great deal of information. Most modern packages (e.g. Stan, PyMC, Numpyro, etc) offer a great many tools to extract insights. (5) Model compasison and selection. Often a model will perform markedly better, which is useful in terms predictive performance, but more importantly in terms of generating insights. (6) Sequential update; a model's output, the posterior distribution becomes the new prior every time new data becomes available for a new sequence of model fitting.

In this paper we leverage past work and recast some of the pre-existing *OCx* models into their probabilistic version. We also propose alternative models, namely Bayesian Additive Regression Trees (BART, (Chipman et al., 2010)) as a robust flexible method that has been used for diverse application ((Linero, 2017).)

2 Methods

2.1 Prior Elicitation and Model Formulation

2.2 Data Preprocessing

3 Data Preprocessing

Data for this study were acquired from multiple satellite ocean color sensors and corresponding in situ chlorophyll-(a) measurements obtained from sources such as the NASA Bio-Optical Marine Algorithm Data set (NOMAD) and the compilation by Valente et al. (2015). To ensure consistency across sensors, the spectral reflectance data ((R_{rs})) were interpolated as needed to common wavelength centers.

For the empirical (OCx) formulation, blue-to-green band ratios were computed for each observation. In particular, the maximum band ratio (MBR) was determined by taking the highest value among the available blue-band ratios (e.g., (Rrs(443)/Rrs(555)), (Rrs(490)/Rrs(555)), and (Rrs(510)/Rrs(555))). This maximum value was then log-transformed:

$$\log R = \log_{10}\left(\frac{R_{rs}(\lambda_{\text{blue}})}{R_{rs}(555)}\right).$$

For the Color Index (CI) formulation of Hu et al. (2012), the CI was calculated as:

$$\text{CI} = R_{rs}(555) - \left[R_{rs}(443) + \frac{555-443}{670-443}\Big(R_{rs}(670) - R_{rs}(443)\Big)\right],$$

and the corresponding in situ chlorophyll-(a) concentrations were log-transformed:

$$\log \text{Chl} = \log_{10}(\text{Chl}).$$

These transformations standardize the data to a common scale, ensuring that variability is appropriately captured for subsequent regression and uncertainty quantification. Detailed descriptions of the interpolation methods and quality control procedures are provided in the Supplementary Material.

4 Statement of Contribution

In this study, we develop and demonstrate a new global chlorophyll retrieval model based on Bayesian Additive Regression Trees implemented in PyMC. We train the BART model on a large, standardized dataset of satellite remote-sensing reflectance (Rrs) spectra matched with in situ chlorophyll measurements, using log - transformed Chl-a as the response to stabilize variance. The resulting model is applied globally to produce chlorophyll-a estimates from multi-spectral satellite data, with associated uncertainty estimates for each prediction. We show that this Bayesian tree-based model can serve as a general-purpose ocean color algorithm that is sensor-agnostic (provided reflectances are harmonized to common wavebands), interpretable, and uncertainty-aware. Unlike conventional empirical algorithms, the

BART approach allows users to examine the inferred Rrs–Chl relationships and trust the model's performance across regimes, while also quantifying confidence in each retrieval. This work thus contributes a novel methodological advance to satellite ocean color science: a unified chlorophyll retrieval model that marries the strengths of empirical algorithms (global applicability and simplicity) with the benefits of modern Bayesian machine learning (flexibility, interpretability, and rigorous uncertainty quantification). Our introduction of BART for global chlorophyll prediction opens the door for more robust monitoring of ocean biogeochemistry and improved integration of ocean color data into scientific and management applications.

5 Acknowledgments

6 Open research

References

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266–298. https://doi.org/10.1214/09-AOAS285

Clayton, A. (2022). *Bernoulli's fallacy: Statistical illogic and the crisis of modern science*. Columbia University Press. Retrieved from https://books.google.com/books?id=BT4CzwEACAAJ

Cobey, K. D., Ebrahimzadeh, S., Page, M. J., Thibault, R. T., Nguyen, P.-Y., Abu-Dalfa, F., & Moher, D. (2024). Biomedical researchers' perspectives on the reproducibility of research. *PLoS Biology*, *22*(11), e3002870.

Craig, S. E., & Karaköylü, E. M. (2019). Bayesian models for deriving biogeochemical information from satellite ocean color. *EarthArXiv*. https://doi.org/10.31223/osf.io/shp6y

Erickson, Z. K., McKinna, L. I. W., Werdell, P. J., & Cetinić, I. (2023). Bayesian approach to a generalized inherent optical property model. *Optics Express*, *31*, 22790–22801. https://doi.org/10.1364/oe.486581

Frouin, R., & Pelletier, B. (2013). Bayesian methodology for ocean color remote sensing. Retrieved from https://hal.archives-ouvertes.fr/hal-00822032

Gal, Y. (2016). *Uncertainty in deep learning* (PhD thesis). University of Cambridge.

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research: Oceans*, *117*(C1). https://doi.org/https://doi.org/10.1029/2011JC007395

Hu, C., Feng, L., Lee, Z., Franz, B. A., Bailey, S. W., Werdell, P. J., & Proctor, C. W. (2019). Improving satellite global chlorophyll a data products through algorithm refinement and data recovery. *Journal of Geophysical Research: Oceans*, *124*(3), 1524–1543. https://doi.org/https://doi.org/10.1029/2019JC014941

Jaynes, E. T., & Bretthorst, G. L. (2003). *Probability theory: The logic of science*. Cambridge University Press. Retrieved from https://books.google.com/books?id=tTN4HuUNXjgC

Linero, A. R. (2017). A review of tree-based bayesian methods. *Communications for Statistical Applications and Methods*, *24*, 543–559. https://doi.org/10.29220/CSAM.2017.24.6.543

O'Reilly, John E., & Werdell, P. J. (2019). Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. *Remote Sensing of Environment*, *229*, 32–47. https://doi.org/https://doi.org/10.1016/j.rse.2019.04.021

O'Reilly, John E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., et al. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans*, *103*(C11), 24937–24953. https://doi.org/https://doi.org/10.1029/98JC02160

O'Reilly, John E., Maritorena, S., Siegel, D. A., O'Brien, M. C., Toole, D., Mitchell, B. G., et al. (2000). Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. *SeaWiFS Postlaunch Calibration and Validation Analyses, Part*, *3*, 9–23.

Scheemaekere, X. D., & Szafarz, A. (2011). The inference fallacy from bernoulli to kolmogorov. Retrieved from https://www.researchgate.net/publication/254450993_The_Inference_Fallacy_From_Bernoulli_to_Kolmogorov

Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., & Werdell, P. J. (2018). Performance metrics for the assessment of satellite data products: An ocean color case study. *Opt. Express*, *26*(6), 7404–7422. https://doi.org/10.1364/OE.26.007404

Wolkovich, E., Davies, T. J., Pearse, W. D., & Betancourt, M. (2024). A four-step bayesian workflow for improving ecological science. Retrieved from https://arxiv.org/abs/2408.02603