AGU ADVANCING EARTH AND SPACE SCIENCES

Earth and Space Science

RESEARCH ARTICLE
10.1029/2025EA004551

From Standard to Bayesian: Revisiting Ocean Color Model Evaluation

Abed Hammoud[1], **Robert J. W. Brewin**[2], **Susanne E. Craig**[3], **and Elie Bou-Zeid**[1]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [2]Centre for Geography and Environmental Science, Department of Earth and Environmental Sciences, Faculty of Environment, Science and Economy, University of Exeter, Cornwall, UK, [3]NASA Goddard Space Flight Center/University of Maryland Baltimore County GESTAR II, Greenbelt, MD, USA

Key Points:

- We recast classic OC algorithms (OC3–OC6, OCI) in a Bayesian framework, providing predictive uncertainty for $\log_{10}$[CHL-*a*], alongside point estimates
- Information criteria (Bayesian information criterion/Akaike information criterion) offer a principled, complexity-aware preference when conventional error metrics are tied
- For the data sets analyzed here, criteria favor simpler OC models; performance remains sensor-, region-, and data set-dependent

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Hammoud,
ah1389@princeton.edu

Citation:

Hammoud, A., Brewin, R. J. W., Craig, S. E., & Bou-Zeid, E. (2026). From standard to Bayesian: Revisiting ocean color model evaluation. *Earth and Space Science*, *13*, e2025EA004551. https://doi.org/10.1029/2025EA004551

Received 23 JUN 2025
Accepted 18 JAN 2026

Author Contributions:

Conceptualization: Abed Hammoud, Robert J. W. Brewin, Susanne E. Craig, Elie Bou-Zeid
Data curation: Abed Hammoud
Formal analysis: Abed Hammoud, Robert J. W. Brewin, Susanne E. Craig, Elie Bou-Zeid
Funding acquisition: Elie Bou-Zeid
Investigation: Abed Hammoud, Robert J. W. Brewin, Susanne E. Craig, Elie Bou-Zeid
Methodology: Abed Hammoud, Robert J. W. Brewin, Susanne E. Craig, Elie Bou-Zeid

Abstract

Ocean color models, facilitated by satellite-based remote sensing technologies, have transformed our ability to monitor chlorophyll-a concentrations at the ocean's surface. These models decode changes in the spectral properties of reflected sunlight to map phytoplankton biomass across extensive spatial and temporal scales. Such capabilities are crucial for understanding marine primary productivity, carbon cycling, and ecosystem reactions to environmental change. Additionally, improvements in atmospheric correction and in-water algorithms over the past decades have reduced the uncertainties associated with chlorophyll-a retrieval, rendering satellite-derived ocean color data more suitable for fisheries management, climate research, and coastal water quality monitoring. However, the most commonly used ocean color products (e.g., Sentinel data sets) do not provide uncertainties, which poses challenges in their application for analyzing events like algal blooms and for risk assessment. Moreover, these models are typically evaluated using deterministic error metrics that may not always capture the subtleties in model performance. This paper presents a probabilistic alternative to traditional ocean color models, including the maximum band and color index models, utilizing Bayesian regression and underpinned by the largest database of validated bio-optical match-ups. We further propose a method for comparing ocean color models using information criteria. When conventional errors are indistinguishable, information criteria (Bayesian and Akaike information criteria) provide a principled, complexity-aware preference among ocean color models for the data sets analyzed, with the OC3 generally favored when considering likelihood-based approaches, suggesting that additional parameters might not be warranted.

Plain Language Summary

Satellites help us monitor the health of our oceans by measuring their color, which tells us about the amount of tiny marine plants-like algae called phytoplankton. These algae are crucial as they form the base of the marine food web and play a big role in Earth's climate. Scientists use mathematical tools, or models, to turn satellite ocean color measurements into estimates of phytoplankton abundance (specifically, a pigment called chlorophyll-a). However, current standard models often do not tell us how confident we can be in their estimates, and it can be hard to choose the best-performing model among many options. In this study, we introduce a new way to approach these ocean color models using the so-called Bayesian statistical framework. This allows models to not only predict chlorophyll-a but also to provide a measure of uncertainty—telling us how reliable each prediction is. We also propose a more robust method for comparing different ocean color models, helping us identify which ones perform most reliably by balancing accuracy with simplicity. Our results show that the most commonly used model (known as OC4) is not always the top performer. Sometimes, a simpler model (like OC3) can provide better results. This work helps improve how we monitor phytoplankton, providing more reliable information for understanding ocean ecosystems, managing fisheries, and assessing climate change impacts.

1. Introduction

Phytoplankton, a diverse group of microscopic photosynthetic organisms, form the foundation of marine ecosystems and drive global biogeochemical cycles. As the primary producers in oceanic food webs, they underpin the survival of higher trophic levels, including fish and marine mammals (Harris, 2012; Lévy et al., 2012). Their ecological role extends beyond nutrient cycling; phytoplankton mediate the exchange of essential elements such as carbon, oxygen, and nitrogen between the ocean and atmosphere (Cetinić et al., 2024; Jones, 1998). Notably, their ability to absorb atmospheric carbon dioxide during photosynthesis positions them as key players in mitigating climate change, helping regulate the Earth's climate system (Basu & Mackey, 2018; Haeder et al., 2014).

Project administration: Susanne E. Craig, Elie Bou-Zeid
Resources: Elie Bou-Zeid
Software: Abed Hammoud
Supervision: Robert J. W. Brewin, Susanne E. Craig, Elie Bou-Zeid
Validation: Abed Hammoud, Robert J. W. Brewin, Susanne E. Craig
Visualization: Abed Hammoud
Writing – original draft: Abed Hammoud
Writing – review & editing: Abed Hammoud, Robert J. W. Brewin, Susanne E. Craig, Elie Bou-Zeid

Remote sensing has revolutionized our ability to study phytoplankton on a global scale, offering high-resolution data over vast spatial and temporal extents (Cael et al., 2023). This technology allows researchers to capture the dynamic processes that regulate phytoplankton ecosystems, from daily variability to long-term trends, and to detect ecological hotspots (Racault et al., 2014). Compared to traditional sampling methods, remote sensing provides synoptic ocean coverage, overcoming logistical and temporal limitations (Platt et al., 2009). Furthermore, remotely sensed data are instrumental in calibrating and validating biogeochemical models that predict marine ecosystem responses to climate change, including phenomena such as ocean warming, acidification, and eutrophication (Shu et al., 2022).

Central to remote sensing approaches are ocean color algorithms, which estimate concentrations of phytoplankton and associated water constituents by analyzing light reflected from the ocean's surface (Blondeau-Patissier et al., 2014; Brewin et al., 2011). These algorithms exploit the spectral properties of chlorophyll-a, the primary pigment in phytoplankton, which absorbs more blue and red light relative to green (Morel, 1980; Morel & Prieur, 1977). The scattering of light by phytoplankton is thought to be controlled by its size and structural complexity (Sun et al., 2025). For instance, by measuring the ratio of blue to green light, ocean color algorithms can derive chlorophyll-a concentrations ([CHL-a]) as an indicator of phytoplankton biomass (Bricaud et al., 1995). Advances in satellite technology and algorithm development have improved the accuracy of these estimates, though challenges remain in regions with complex optical properties, such as coastal waters and estuaries (Groom et al., 2019; Joint & Groom, 2000). In natural waters, absorption by chlorophyll-*a* is only one of several factors influencing apparent color; [CHL-a] often covaries with other optically active constituents such as colored dissolved organic matter, detrital material, and suspended sediments, all of which can exert an equal or greater influence on the spectral reflectance (Houskeeper & Hooker, 2025; Prochaska & Frouin, 2025). A plethora of ocean color models have been developed; however, standard error and correlation metrics do not do a good job of differentiating among them (O'Reilly & Werdell, 2019). This necessitates developing complementary approaches for assessing models to better distinguish their performance.

Standard algorithms, such as the band ratio (O'Reilly & Werdell, 2019) and band difference (Hu et al., 2012) among other ocean color models, are generally deterministic and lack awareness of the uncertain nature of the retrieval (IOCCG, 2019). A probabilistic framework, however, can capture uncertainties arising from factors such as atmospheric interference, sensor noise, and variable water constituents, which often challenge remote sensing applications (Brewin et al., 2012; Hammoud et al., 2025; Wang et al., 2024). Bayesian inference provides a robust framework for statistical analysis and decision-making by incorporating prior knowledge and observed data to update the probability distribution of uncertain parameters (Gelman et al., 2013; Najm, 2017). In the context of ocean color models, Bayesian methods extend classical approaches by treating the regression parameters and predictions as probabilistic quantities, governed by their predictive distributions, also called posterior distribution (Craig & Karaköylü, 2019; Frouin & Pelletier, 2015). This probabilistic perspective enables the quantification of uncertainty in parameter estimates and model predictions, making it particularly advantageous for complex and noisy systems to enhance our understanding of phytoplankton dynamics and develop tools in applications ranging from managing fisheries to understanding extreme algal bloom events.

Building on the principles of Bayesian inference, Bayesian model selection provides a systematic framework for comparing competing models by evaluating their posterior probabilities (Chipman et al., 2001). Unlike standard error and correlation measures, Bayesian model selection directly incorporates uncertainty and penalizes model complexity through the marginal likelihood, or evidence, which integrates over all possible parameter values (Hooten & Hobbs, 2015). This approach allows for the identification of models that not only fit the data well but are also parsimonious and robust to overfitting (MacKay, 2003). These approaches are often beneficial in fields where competing hypotheses must be rigorously evaluated or where uncertainty quantification is critical to decision-making.

This study investigates Bayesian counterparts of standard ocean color algorithms, namely, band ratio and difference algorithms. In doing so, an interpretable approach for probabilistic ocean color models is established, allowing for uncertainty quantification of [CHL-a] predictions. By following a more traditional approach for constructing a probabilistic version of standard ocean color models, our model is different from other recent works (e.g., Frouin and Pelletier (2015)) that focused on a Bayesian atmospheric correction, absorption coefficients (e.g., Prochaska and Frouin (2025) and Houskeeper and Hooker (2025)), and those that rely on a Gaussian mixture approach with reflectances as input, which might lack in interpretability (e.g., Saranathan

et al. (2023)). The performance of the probabilistic models is then compared with their deterministic counterparts, verifying that both perform similarly on standard statistical measures, building confidence in the Bayesian models. Bayesian model selection based on the Bayesian information criterion (BIC) is then adopted as an alternative approach for assessing the performance of various ocean color models when statistical measures are tied, enabling a more reliable metric for assessing model performance. The importance of which is further underscored for future developments of ocean color models based on hyperspectral data, taking advantage of a refined spectrum (Vandermeulen et al., 2020). Finally, we explore the application of the Bayesian ocean color model using a realistic Sentinel-3 image, as a case study example of these probabilistic models.

The manuscript is organized as follows. Section 2 presents the openly accessible data sets considered in this study. In Section 3, the standard ocean color models based on the band ratio and difference are outlined. Section 4 presents the Bayesian inference tools employed in this study, namely, Bayes' rule, Bayesian regression, and Bayesian model selection. Results are then analyzed in Section 5. Finally, the main takeaways are discussed and conclusions are summarized in Section 6.

2. Ocean Color Data

The present study relies primarily on a comprehensive repository of bio-optical match-ups of in situ measurements (Valente et al., 2022a, 2022b). The repository is a compilation of 27 data sets that were individually processed to ensure the highest level of quality control. Remotely sensed reflectances were retrieved from the European Space Agency's Medium Resolution Imaging Spectrometer (MERIS), ESA's Ocean and Land Color Instrument (OLCI), National Aeronautics and Space Administration's (NASA) Sea-viewing Wide Field-of-view Sensor (SeaWiFS), NASA's Visible Infrared Imaging Radiometer Suite (VIIRS), and NASA's MODerate resolution Imaging Spectro-radiometer Aqua (MODIS). The match-ups cover global waters, inclusive of coastal and open waters, making the repository attractive for developing and testing ocean color models.

The repository employs data sourced from multiple project archives accessible through open internet platforms and directly from data providers. Specifically, the match-ups were assembled from 27 in situ data sets acquired from a variety of sources, including SeaBASS (Werdell & Bailey, 2002), NOMAD (Werdell & Bailey, 2005), MERMAID (Barker et al., 2008), ICES (https://www.ices.dk/data/dataset-collections/Pages/Plankton.aspx), ARCSSPP (Matrai et al., 2013), BIOCHEM (Devine et al., 2015), BODC (https://www.bodc.ac.uk/data/bodc_database/), COASTCOLOUR (Nechad et al., 2015), MAREDAT (Peloquin et al., 2013), and SEADATANET (seadatanet.org). Additional data sets were obtained from projects such as MOBY (Clark et al., 2003), BOUSSOLE, AERONET-OC, HOT, GeP&CO, AMT, AWI, BARENTSSEA, BATS, CALCOFI, CCELTER, CIMT, ESTOC, IMOS, PALMER, TPSS, and TARA. Remotely sensed reflectances were retrieved from various instruments, including ESA's MERIS, ESA's OLCI, NASA's Sea-viewing Wide Field-of-view Sensor (SeaWiFS), NASA's VIIRS, and NASA's MODerate resolution Imaging Spectro-radiometer (MODIS). With its global distribution encompassing both coastal and open waters, this database serves as an excellent resource for the development and validation of ocean color models, as illustrated by O'Reilly and Werdell (2019).

The Valente et al. (2022a, 2022b) repository compiles a matchup of geographically-aligned bio-optical properties, including the remotely sensed reflectances (R_{rs}), in situ measured [CHL-a] that are measured by fluorometry $\left([\text{CHL-a}]_{fluor}\right)$ or high-performance liquid chromatography $\left([\text{CHL-a}]_{HPLC}\right)$, alongside other inherent optical properties (IOPs). The IOPs include the absorption coefficient of phytoplankton, the absorption coefficient of detrital matter, the colored dissolved organic matter, the backscattering coefficients of particles, the diffusive attenuation coefficient for downward irradiance, and the total suspended matter. The repository discards all missing observations, those indicated with an incorrect date and/or geographic coordinates, obtained using an incompatible measurement method, or exhibiting extreme values. Furthermore, the repository was constructed by applying a threshold, based on the coefficient of variation (CoV), to discard data samples with high spatial and/or temporal variability.

While this study heavily relies on one of the most comprehensive and openly available matchup data sets for ocean color algorithm development, it is important to acknowledge several limitations. The data set inevitably reflects regional and temporal sampling biases and may not capture the full diversity of optical conditions encountered globally. Nonetheless, it has been widely used in prior studies for the derivation and validation of

ocean color algorithms (e.g., O'Reilly and Werdell (2019)), which supports its continued use as a community-standard benchmark.

In addition to the Valente et al. (2022a, 2022b) repository, this study also incorporates data from the NASA SeaWiFS Bio-optical Archive and Storage System (SeaBASS) (Werdell et al., 2003). While partially overlapping, the SeaBASS data set was accessed independently and used to extract regionally-specific observations along the United States of America coasts to complement the global data set. The SeaBASS repository provides a comprehensive match-up data set of bio-optical and oceanographic field measurements, which adheres to the match-ups criteria proposed by Werdell and Bailey (2005)–spatial measurement and temporal measurements are within ±5 km and ±3 hours, respectively. The data set includes a diverse range of bio-optical measurements, such as absorption coefficients, scattering coefficients, [CHL-a], and remote sensing reflectance (R_{rs}). While the data set compiled field campaigns, shipboard observations, and laboratory analyses, encompassing various global oceanic regions and temporal spans, a subset covering the eastern and western coasts of the contiguous United States was employed to derive regionally-tuned models. The data set is accompanied by detailed metadata describing instrumentation, sampling methodologies, and calibration protocols, ensuring data reliability and reproducibility.

3. Ocean Color Models

The maximum band ratio (MBR) algorithm is an established, empirical ocean-color retrieval method used to estimate [CHL-a] concentrations in surface waters from satellite-derived reflectances (O'Reilly & Werdell, 2019). At its core, this approach leverages the spectral characteristics of phytoplankton, which preferentially absorb light in the blue region of the visible spectrum and reflect relatively more in the green. By examining the ratio of R_{rs} at selected blue and green wavelengths, the MBR algorithm capitalizes on the strong relationship between these spectral signatures and [CHL-a]. The MBR technique systematically evaluates multiple candidate blue-to-green ratios and selects the one that yields the maximum ratio value. This procedure allows the algorithm to maximize the signal and adapt to subtle differences in sensor-specific band positions and regional optical conditions, ultimately improving its robustness and generality across diverse oceanic environments.

Once the maximum ratio is identified, it is translated into a [CHL-a] through an empirically derived polynomial regression. These polynomial coefficients are calibrated against extensive in situ data sets, including direct chlorophyll measurements and radiative transfer simulations. This calibration ensures that the algorithm's parameterization accurately represents a wide range of oceanic conditions. Prominent MBR implementations are encapsulated in the family of "OCp" algorithms, where p is the degree of the polynomial relating the MBR to the log [CHL-a]. For example, the OC4 maximum band algorithm is a 4th order polynomial equation with an MBR defined as:

$$\mathrm{MBR} = \max[R_{rs}(443), R_{rs}(490), R_{rs}(510)]/R_{rs}(555). \quad (1)$$

The $\log_{10}$[CHL-a] is then estimated according to:

$$\log_{10}[\text{CHL-a}] = c_0 + c_1\log_{10}(\mathrm{MBR}) + c_2\log_{10}(\mathrm{MBR})^2 + c_3\log_{10}(\mathrm{MBR})^3 + c_4\log_{10}(\mathrm{MBR})^4, \quad (2)$$

where $c_0 = 0.310$, $c_1 = -2.621$, $c_2 = 2.910$, $c_3 = -3.238$ and $c_4 = 1.036$ for the deterministic case using the comprehensive Valente et al. (2022a,2022b) data set and a LASSO regularized regression approach of Tibshirani (1996). Note that the Bayesian counterpart of the OC algorithm treats each of the coefficients c_i as a random variable whose distribution is inferred. The MBR algorithms are thoroughly described in O'Reilly and Werdell (2019). While higher-order variants such as OC5 and OC6 can, in principle, reduce residuals within the calibration range, they also risk introducing non-physical behavior outside that range. In particular, high-degree polynomials fitted to $\log_{10}$(MBR) may exhibit unrealistic oscillations or behavior at very small or large ratios, leading to implausible [CHL-a] estimates. Such overfitting underscores the importance of evaluating model order not only by goodness-of-fit but also by its physical plausibility and extrapolation stability. This consideration further motivates the use of information-criterion approaches (Section 4.3) to objectively balance complexity and likelihood as a complementary approach for model comparison and selection.

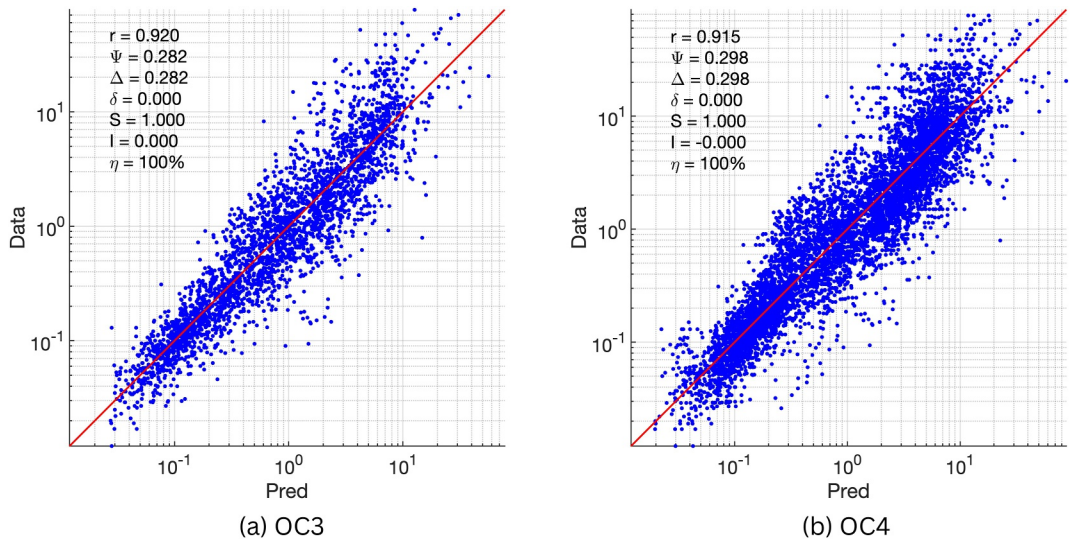

Figure 1. Scatter plots illustrating the in situ [CHL-a] against the predictions using (a) OC3 and (b) OC4. The plots are accompanied by the standard error and correlation metrics, which clearly indicate a similar performance, underscoring the importance of exploring alternative approaches to compare ocean color models.

Another commonly used ocean color algorithm is the Ocean Color Index (OCI) algorithm, often referred to as the band-difference algorithm (Hu et al., 2012). It has proven particularly effective for estimating phytoplankton concentrations below 0.25 mg m^{-3} (Hu et al., 2012). The OCI algorithm is given by:

$$\xi = R_{rs}(555) - 0.5[R_{rs}(443) + R_{rs}(670)]. \tag{3}$$

Here, ξ, color index, measures the discrepancy between the reflectance in the green part of the visible spectrum and the mean of the blue and red reflectances. This index is then used to estimate [CHL-a], according to:

$$[\text{CHL-a}] = 10^{A+B\xi}, \tag{4}$$

where the constants $A = -0.4451$ and $B = 217.145$ are derived via a LASSO regularized regression. Note that testing of the OCI algorithm is performed for [CHL-a] below 0.25 mg m^{-3} since it becomes unclear how to compute the model selection criterion for two models with different numbers of parameters that are trained on different data sets.

3.1. Model Evaluation

Model performance is typically assessed using the deviation and correlation metrics such as those outlined by Brewin et al. (2015). Error metrics generally involve the root mean squared error (Ψ), the unbiased root mean squared error (Δ). Correlation metrics encompass the Pearson correlation coefficient (r), mean bias between observations and model predictions (δ), the percentage of retrieval (η), and the parameters of a Type-II regression, namely the slope (S) and intercept (I). Detailed descriptions of these metrics and the associated computational formulas can be found in Section 4.1 of Brewin et al. (2015). Consistent with established practices within the scientific community, these metrics, applied in $\log_{10}$ space, fulfill the comparability criteria outlined by Seegers et al. (2018) for non-Gaussian matchup data, obviating the need for additional robust estimators.

Prediction uncertainty is fully quantified by a distribution described by an ensemble of [CHL-a] predictions, and reported in terms of the CoV, which quantifies prediction uncertainty as the ratio of the standard deviation to the mean of the predictions. In particular, for a probabilistic model, an ensemble of predictions is generated to construct the posterior distribution of [CHL-a]. The statistical metrics were subsequently computed on the maximum probability estimate of the [CHL-a] posterior probability distribution, known as the maximum a posteriori (MAP) estimate.

Figure 1 illustrates the scatter of the predicted [CHL-a] against its measured counterpart, using the OC3 and OC4 models. The conventional summary statistics (Ψ, Δ, r, S, I, δ) differ only in the second decimal place, leaving no statistically defensible basis for preferring one model over the other. In other words, the standard error and correlation metrics are too alike to resolve which algorithm delivers genuinely smaller residuals once sampling noise is taken into account. This motivates the need to investigate alternative model selection approaches for a more objective ranking, considering prediction uncertainties as well as the associated errors when evaluating model predictions. To this end, we choose to examine Bayesian model selection through the BIC and Akaike information criterion, which weighs each model's likelihood against its complexity and quantifies the relative evidence in favor of one algorithm versus the other.

4. Bayesian Statistics

4.1. Bayesian Inference

Bayesian inference is employed to estimate the probability distribution of the ocean color model parameters. We denote the inverse solution, vector of the model parameters, by $\boldsymbol{\theta}$. Each random sample from the posterior distribution of the model parameters, $\boldsymbol{\theta}$, produces one [CHL-a] estimate; repeating this process with a sufficiently large number of draws yields an ensemble that adequately populates the posterior distribution of [CHL-a]. Bayes' rule (Gelman et al., 2013) estimates the posterior distribution of model parameters as:

$$\pi(\boldsymbol{\theta}|\mathrm{D}) \propto \pi(\mathrm{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{5}$$

where, $\pi(\boldsymbol{\theta}|D)$ is the posterior distribution, $\pi(\mathrm{D}|\boldsymbol{\theta})$ is the likelihood and $\pi(\boldsymbol{\theta})$ is the prior distribution. The MAP prediction, which is the prediction with the highest probability, is typically adopted as the most probable estimate of the model parameters.

In this study, the prior distribution on θ is a uniform distribution, meaning no prior knowledge is available. Specifically, the prior over the polynomial coefficients spans the range $]-\infty, +\infty[$, whereas that for σ^2 is $]0, +\infty[$. The likelihood function is chosen to be a Gaussian distribution, which is a common choice in the literature (e.g. Schulz et al. (2018)), taking the form of:

$$\pi(\mathrm{D}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_o^2}} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma_o^2}\right) \tag{6}$$

where, y and $\hat{y}$ are the observed and predicted data; n is the number of observations; and σ_o^2 is the variance in the observed data. Note that the results for the Gaussian prior distribution, which require major computational simplifications in the sampling, are presented in the Supporting Information S1. For completeness, we note that Gaussian priors over the coefficients were centered over their deterministic solution with a unit standard deviation, whereas that for σ was a standard Gaussian prior.

The Bayesian regression problem for the band ratio ocean color model can be written as:

$$\hat{y} = \mathbf{X} \cdot \boldsymbol{\theta} + \epsilon \tag{7}$$

where, $\mathbf{X} = [1, \boldsymbol{\xi}, \ldots, \boldsymbol{\xi}^p]$ is the Vandermonde matrix of the independent variable, such as the band ratio, up to order p that is being evaluated; $\boldsymbol{\theta}$ are the coefficients of the expansion; and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise perturbing the predictions and having variance σ that is inverted for. Note, however, that this model structure inherently allows for heteroskedasticity in the predictions, enabling the predictive uncertainty to scale with the magnitude of the estimated value and with the density of observations.

This σ provides a means for regularization, preventing overfitting, and for an intuitive measure of the prediction uncertainty. In fact, σ is commonly introduced in Bayesian regression problems to explicitly represent aleatoric uncertainty (Carlin & Louis, 2008; Gelman et al., 2013). Finally, we remark that the present analysis addresses two primary sources of uncertainty: parametric uncertainty in the ocean color model parameters, and observational uncertainty in the available data sets. However, uncertainties in the matchups themselves—arising from

measurement errors in the remote-sensing reflectances or the in situ [CHL-a]—are typically not reported or are difficult to quantify. This limitation makes it challenging, at least for the data sets used here, to explicitly incorporate these additional uncertainty components into the probabilistic framework, necessitating in situ data providers to report uncertainties in the observations, ideally following standard protocols (Lin et al., 2022). Nevertheless, σ here represents an aggregate discrepancy, lumping unresolved variability and model-data mismatch.

4.2. Markov Chain Monte Carlo

In this study, we utilized Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distribution of model parameters given some data (Marzouk et al., 2007). MCMC is a powerful computational approach for Bayesian inference, particularly effective for exploring high-dimensional or complex parameter spaces. Specifically, we implemented the random walk Metropolis-Hastings (MH) algorithm (Gilks et al., 1995), which constructs a Markov chain by proposing candidate parameter sets from a specified proposal distribution and accepting or rejecting them based on an acceptance probability (Brown & Jones, 2024). This probability, derived from the ratio of posterior probabilities of the proposed and current states, ensures that the chain converges to the target posterior distribution (Marshall et al., 2004). The forward model was evaluated at each sampled parameter set, and the likelihood function was computed by comparing the model outputs to observed data. This algorithm has been well-described in the literature and is briefly described in what follows.

Algorithm 1

Algorithm 1. Random walk Metropolis-Hastings algorithm

Data: Target distribution $\pi(\boldsymbol{\theta}|D)$, proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$
Result: Samples $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(N)}\}$
Initialize: Choose an initial state $\boldsymbol{\theta}^{(0)}$;
for $i = 1$ *to* N **do**
  Propose a new state $\boldsymbol{\theta}'$ from $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(i-1)})$;
  Evaluate the forward model $F(\boldsymbol{\theta}')$ using the proposed parameters $\boldsymbol{\theta}'$;
  Compute the likelihood $\pi(D|\boldsymbol{\theta}')$ based on $F(\boldsymbol{\theta}')$ and the observed data;
  Compute the acceptance probability:

$$\alpha = \min\left(1, \frac{\pi(\boldsymbol{\theta}'|D)}{\pi(\boldsymbol{\theta}^{(i-1)}|D)}\right)$$

  Generate a uniform random variable $u \sim \mathcal{U}(0,1)$;
  if $u \leq \alpha$ **then**
    Accept the proposal: $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}'$;
  else
    Reject the proposal: $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$;
  end
end

In the present study, a standard procedure for Bayesian regression was followed to ensure reproducibility for the widest audience. In particular, an uninformative uniform prior distribution was associated with the model parameters, and a normal likelihood function was adopted for the error model. The MH algorithm generated 1 million samples of $\boldsymbol{\theta}$ across all experiments, after 100,000 burn-in samples. This is typically larger than most Bayesian inference cases, but was achievable because of the simplicity of standard ocean color models.

4.3. Bayesian Model Selection

Bayesian model selection provides a robust framework for comparing competing models based on their posterior probabilities, which systematically incorporates both the goodness-of-fit and model complexity (Hooten & Hobbs, 2015; Neath & Cavanaugh, 2012). The posterior probability of a model M_k is proportional to its marginal likelihood, also known as the model evidence, $\pi(D|M_k)$, combined with a prior $\pi(M_k)$ on the model itself. The

marginal likelihood is computed by integrating the product of the likelihood $\pi(D|\boldsymbol{\theta}_k, M_k)$ and the prior $\pi(\boldsymbol{\theta}_k|M_k)$ over the parameter space $\boldsymbol{\theta}_k$ of the model:

$$\pi(D|M_k) = \int \pi(D|\boldsymbol{\theta}_k, M_k)\pi(\boldsymbol{\theta}_k|M_k)\, d\boldsymbol{\theta}_k.$$

This integral inherently penalizes overly complex models by spreading the prior probability across a larger parameter space. Models with higher marginal likelihoods are more supported by the data, and the posterior model probabilities can be used to compare and select among multiple competing hypotheses.

In practical applications, computing the marginal likelihood directly can be computationally intensive. Alternatively, approximate criteria like the BIC and AIC are often used. The BIC is derived as an asymptotic approximation to the log marginal likelihood and is defined as (Wit et al., 2012):

$$\mathrm{BIC} = -2\log\hat{L} + k\log(n),$$

where $\hat{L}$ is the likelihood of the model, k is the number of model parameters, and n is the number of observations. The BIC penalizes model complexity by incorporating the term $k\log(n)$, which increases with the number of parameters and observations. BIC offers a convenient and computationally efficient approximation to full Bayesian model comparison, albeit more suitable in the case of large data sets and not overly complex models, which is the setting of a wide-class of ocean color models. Lower BIC values indicate models that achieve a better trade-off between fit and simplicity. Finally, we note that while the BIC provides a convenient and interpretable criterion for model selection, its reliance on asymptotic approximations makes it less suitable for small sample sizes or when strong prior information is available, where full Bayesian approaches may be more appropriate.

Unlike BIC, AIC targets expected out-of-sample predictive performance by estimating the (relative) Kullback–Leibler risk (Akaike, 1998):

$$\mathrm{AIC} = -2\log\hat{L} + 2k.$$

Because its penalty is $2k$ (independent of n), AIC will tend to prefer richer models than BIC when n is large and the maximized likelihoods are similar across candidates. This provides another interpretable measure that is more suited for the condition when the data set is small. Generally, when BIC and AIC agree, the evidence for the selected model is stronger; when they differ, we interpret the preference in light of the study objective (complexity-aware parsimony versus predictive performance) and report results alongside conventional error metrics as complements.

5. Results

The results of the present study are presented as follows. First, the deterministic and probabilistic OC models are contrasted to confirm that the Bayesian model performs comparably to its deterministic counterpart. Then, the standard validation metrics are presented for the Bayesian ocean color models for both OC*p* and OCI. The BIC for the ocean color models considered are presented for various sensors present in the repository described in Section 2. Finally, the most reliable probabilistic model is used to estimate the [CHL-a] field using a high-resolution satellite image obtained from ESA's Sentinel-3.

5.1. Bayesian Ocean Color Models

The MH algorithm is first employed to sample the posterior distribution of the parameters of the Bayesian ocean color models. In particular, samples from the joint posterior of the OCx coefficients (c_j) and the likelihood noise variance σ^2, all defined in the $\log_{10}[\mathrm{CHL}\text{-}a]$ domain, were drawn using the MH sampler. As an example, Figure 2 illustrates the probability density function of the parameters of the Bayesian OC4 model. Note that these plots are intended to illustrate the Bayesian counterpart of standard deterministic models, OC4 was selected because it is widely accepted as the standard operational model, and the present methodology is extendable to virtually any ocean color model. The plot also shows, in red, the coefficients of a deterministic OC4 model obtained through

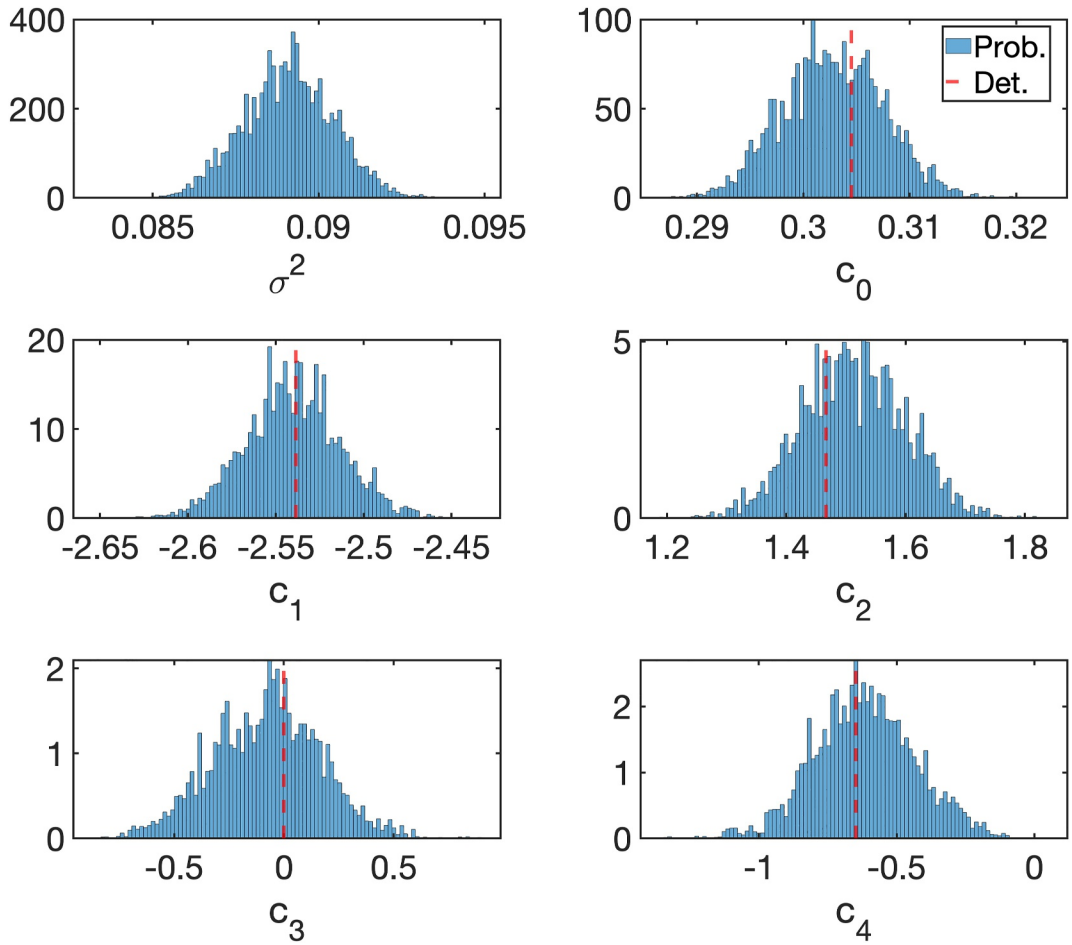

Figure 2. Histogram of the MH samples for the parameters of the OC4 ocean color model using an uninformative uniform prior. The histograms are plotted based on approximately a million samples. The deterministic model parameters based on standard regression are shown as vertical dashed red lines, which clearly indicate that a single deterministic model does not cover the whole data spread, providing a single realization/prediction.

LASSO regression, where the regularization parameter is tuned to minimize the mean squared error between the model predictions and the data. The figure shows that the coefficients of the deterministic model are close to the MAP of those from the probabilistic model, which boosts confidence in the models behaving similarly. The plots also indicate the uncertainty associated with each model parameter, which underscores the importance of probabilistic algorithms that capture the inherent uncertainty in the data, often missed with deterministic models. Note that the derived uncertainty follows a full Bayesian framework, where the posterior distribution naturally integrates both data and prior information, providing credible intervals that directly quantify the probability of parameter values. In contrast to frequentist confidence intervals, which rely on hypothetical repeated sampling, Bayesian uncertainty estimates offer a more intuitive and comprehensive characterization of uncertainty, reflecting the full range of plausible model behaviors given the observed data.

Further examining the qualitative behavior of the Bayesian models, [CHL-a] is plotted against MBR for different ocean color models in Figure 3. The plots indicate that the MAP [CHL-a] predictions obtained using the sampled coefficients follow the behavior of the in situ data closely. Furthermore, the uncertainty around the predictions, represented by the posterior predictive intervals at 50%, 68%, 90%, and 95% in the shadowed regions, covers the majority of the scatter of data points. Note that a posterior predictive interval is a Bayesian uncertainty measure that quantifies the range within which data collected in the future are expected to fall, accounting for both parameter uncertainty and data noise as inferred from the posterior distribution. These results indicate that the Bayesian ocean color models are suitable for predicting [CHL-a] with reliable estimates of the associated uncertainties.

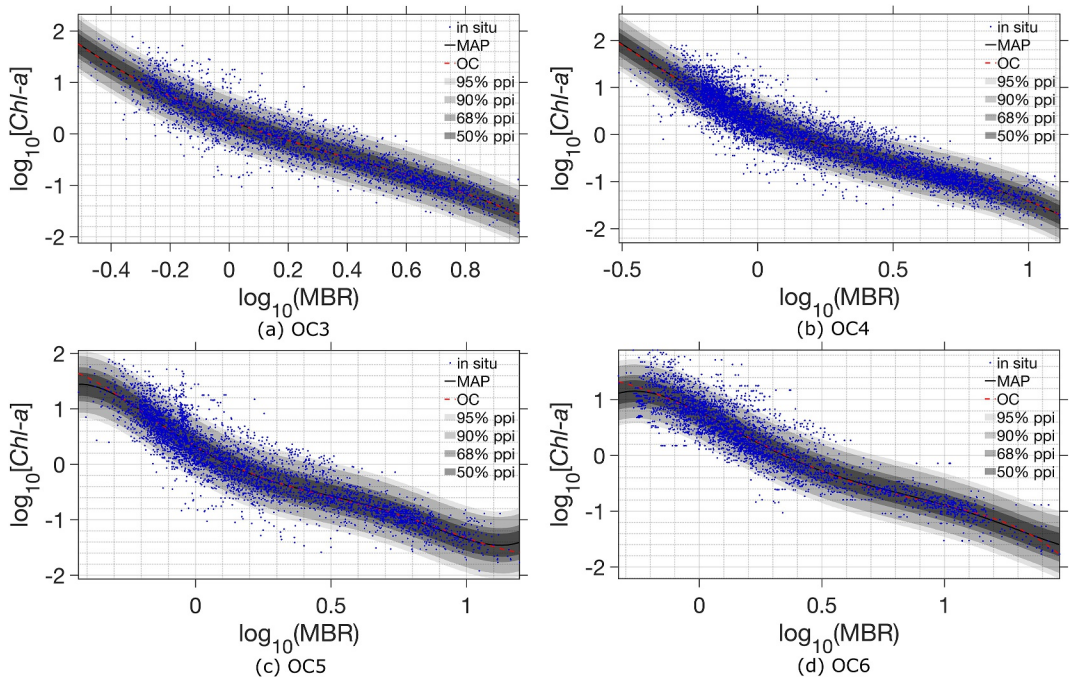

Figure 3. Scatter plots illustrating the [CHL-a] for the in situ (blue dots), Bayesian maximum a posteriori prediction (black), deterministic OC prediction (red) and uncertainties represented by the posterior predictive intervals at 50%, 68%, 90%, and 95% for different Bayesian ocean color models, indicated in the captions, regressed using the global match-ups repository.

Finally, we investigate the quantitative performance of the Bayesian ocean color models. Figure 4 presents scatter plots of the in situ data against the MAP predictions for all band ratio ocean color models. The plots also show the standard error metrics for all MAP predictions, which indicate that all models have high Pearson correlation coefficients between the data and predictions. The MAP predictions have similarly low errors with respect to the data, where the root mean squared error $\Psi < 0.3$ for all band ratio models, and within approximately 5% of each other. Similarly, the bias and unbiased root mean squared errors are comparably small for all band ratio models. Finally, the slope and intercept of a type-2 regression indicate that the correlation is high and bias is low between the data and model predictions, where S is within 2% of unity and δ is of the order of 10^{-3}.

These results underscore the need for probabilistic models because subtle differences on the order of 0.05–0.11 in $\log_{10}$[CHL-a] correspond to $\approx 12-29\%$ in linear space. Such shifts can advance or delay surpassing fixed bloom thresholds, alter the effective observation error used in biogeochemical data assimilation (and thus analysis increments), and bias long-term trend estimates if systematic across time or sensors. Note that the most-probable OC6 model derived by the MCMC algorithm exhibits an upper limit for [CHL-a], which is a byproduct of the data's distribution and the 6th order polynomial, indicating that the model with the lowest errors might not necessarily maintain physical consistency across the prediction space. This also suggests that higher-order models might be susceptible to overfitting, and that for the present ocean color data set, a low- or moderate-order model has reasonable expressivity. These results clearly underscore the importance of studying alternative measures that support model selection because the present results indicate that all four plots have similar performance metrics despite looking rather different.

5.2. Model Selection

This study examines an alternative complementary approach for model selection of ocean color models that relies on the probabilistic distribution of the outputs rather than standard error and correlation measures. Table 1 presents the BIC (lower BIC is better) for the band ratio and the different ocean color models categorized by different sensors, where the global match-ups are used for the analysis. The table shows that the OC3 model has the lowest BIC of 961, and AIC of 929 for the full merged Valente et al. (2022a, 2022b) data set, and BIC/AIC of 65.8/47.8 for the MODIS-SeaWiFS subset. The OC4 model, on the other hand, has the highest BIC/AIC of 3750/

10.1029/2025EA004551

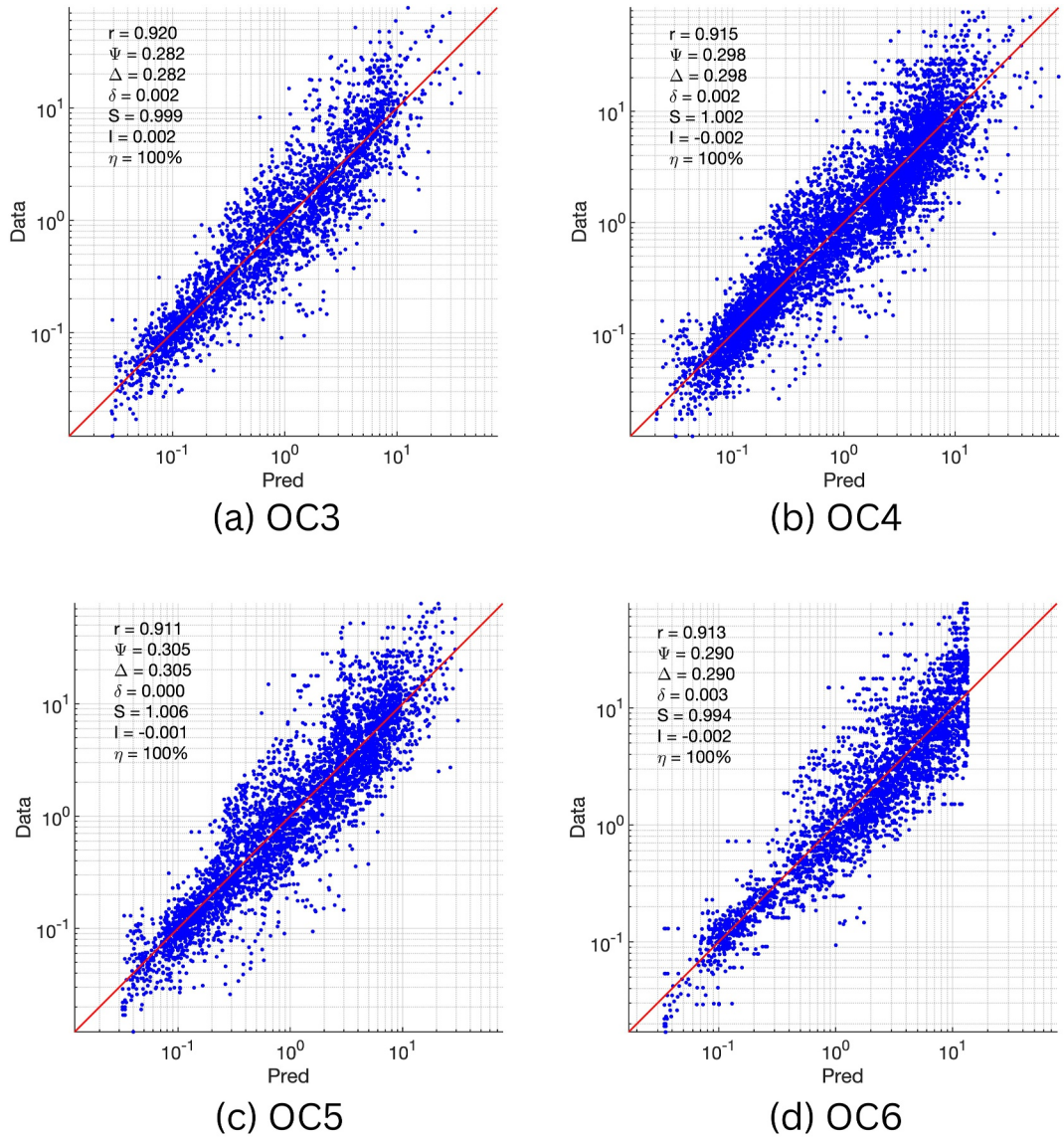

Figure 4. Scatter plots illustrating the in situ [CHL-a] against the maximum a posteriori prediction for the various ocean color models, indicated in the captions. The plots are accompanied by the standard error and correlation metrics, which clearly indicate a similar performance, illustrating the difficulty in delineating the differences across models.

3705 for the merged data set, while OC5 has the largest BIC/AIC of 1280/1236 on the MODIS-SeaWiFS subset. This suggests that for these products, the simpler OC3 model could achieve better likelihood-based scores, while performing at least as well on conventional statistical error metrics. This suggests that the additional parameters associated with OC4 or higher-order models might not be warranted. Note that the number of samples for each data set are presented in Tables S1 and S2 in Supporting Information S1, which further motivates the use of AIC for smaller subsets where asymptotic approximations may be less reliable.

The table also shows results for the match-up subsets corresponding to VIIRS, where the OC3 model slightly outperforms the OC4 model, where the BIC/AIC values for OC3 are 894.9/863.2, respectively, and 1,167/1,129 for OC4. On the other hand, the BIC and AIC computed using the match-ups from OLCI and MERIS is lowest with the OC6 model, indicating that the OC6 model achieves the best likelihood-based scores. Note, however, that for both MERIS and OLCI O'Reilly and Werdell (2019) have no suggested low-complexity model, further motivating the development of simple ocean color models tailored for these sensors. It is noteworthy to highlight the fact that both AIC and BIC values are in agreement, supporting the suitability of the selected models to be

Table 1
Bayesian Information Criterion (BIC) Values Corresponding to Various Ocean Color Band Ratio Models Regressed on a Global Repository of Bio-Optical Match-Ups

	Valente et al. (2022a, 2022b) repo.		Modis-SeaWiFS merged		VIIRS		MERIS		OLCI	
Criterion (lower is better)	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC
OC3	**961**	**929.0**	**65.8**	**47.8**	**894.9**	**863.2**	x	x	x	x
OC4	3,750	3,705	817.3	779.3	1,167	1,129	631.5	597.3	630.9	596.7
OC5	2,880	2,830	1,280	1,236	x	x	769.1	729.5	769.1	729.5
OC6	1,397	1,345	500.8	456.7	x	x	**409.2**	**368.4**	**480.8**	**437.3**
OCI ([CHL-a] < 0.25)	−1,248.3	−1,264.0	14.9	12.2	−57	−71.2	18.8	8.9	15.8	13

Note. The Bayesian models were derived using uninformative uniform prior distributions, and the lowest BICs are highlighted in bold font.

favored against others. Table 1 also shows the BIC/AIC values corresponding to different sensors present in the Valente et al. (2022a, 2022b), repository.

In addition, Table 2 shows the difference in BIC and AIC of a given model relative to the one with $\Delta BIC/AIC = 0$, which shows the relative improvement/deterioration of a given model relative to the lowest order model. The lowest value in the table indicates the most preferred model, and the relative preference in comparison to a baseline. By quantifying inter-model differences, the relative preference inferred from likelihood-based metrics is established, underscoring the strength of these metrics in the selection of ocean color models. While the comparison with the band ratio algorithms is not feasible because the OCI algorithm was constructed to predict [CHL-a] smaller than 0.25 mg/m^3, it remains valuable to compute the BIC for the OCI algorithm to show the flexibility of the Bayesian model selection methodology. The BIC/AIC values for the OCI model illustrate the trade-off between model complexity and prediction errors, where the BIC values are large, indicating that the model might be too simple to provide reliable predictions, nevertheless, more data is needed to provide conclusive remarks.

Table 3 outlines the BIC values for a matchup data set obtained from NASA's SeaBASS validation data set (Werdell et al., 2003). The match-ups are collected for the Atlantic and Pacific oceans in a region closest to the contiguous United States as a means to examine the model selection methodology for regionally-tuned ocean color models. The table shows that the OC3 model outperforms, from a likelihood-based perspective, all other band ratio models for both regions, with an appreciable advantage of the OC3 algorithm in the Pacific Ocean. This indicates that Bayesian model selection approaches provide additional, complementary insights and valuable

Table 2
ΔBIC Values Corresponding to Various Ocean Color Band Ratio Models Regressed on a Global Repository of Bio-Optical Match-Ups

	Valente et al. (2022a, 2022b) repo.		Modis-SeaWiFS merged		VIIRS		MERIS		OLCI	
Criterion (lower is better)	ΔBIC	ΔAIC	ΔBIC	ΔAIC	ΔBIC	ΔAIC	ΔBIC	ΔAIC	ΔBIC	ΔAIC
OC3	**0**	**0**	**0**	**0**	**0**	**0**	x	x	x	x
OC4	2,789	2,776	751.5	731.5	272.1	265.8	0	0	0	0
OC5	1,919	1,901	1,214.2	1,188.2	x	x	137.6	132.2	138.2	132.8
OC6	436	416	435	408.9	x	x	**−222.3**	**−228.9**	**−150.1**	**−159.4**

Note. The Bayesian models were derived using uninformative uniform prior distributions, and the lowest ΔBICs are highlighted in bold font.

Table 3
Information Criteria Values Corresponding to Various Ocean Color Band Ratio Models Regressed on Regional Data Sets of Bio-Optical Match-Ups for the Atlantic and Pacific Oceans in an Area Near the Contiguous United States

	SeaBASS Pacific USA				SeaBASS (Atlantic USA)			
	BIC	ΔBIC	AIC	ΔAIC	BIC	ΔBIC	AIC	ΔAIC
OC3	**17.69**	–	**7.12**	–	**279.1**	–	**251.3**	–
OC4	200.2	182.6	175.1	167.9	1,090.2	811.1	1,051.9	800.6
OC5	223.8	206.1	194.8	187.6	1,669.4	1,390.3	1,625.8	1,374.5
OC6	235.9	218.2	203.0	195.9	1,133.6	854.5	1,084.1	832.7

Note. The Bayesian models were derived using uninformative uniform prior distributions, and the lowest BICs and AICs are highlighted in bold font.

information to delineate the subtle differences in ocean color models, especially in the case when conventional statistical metrics are tied.

5.3. Application to Sentinel-3

The probabilistic ocean color models are employed in application to predicting [CHL-a] for a scene obtained using ESA's Sentinel-3 on the 22nd of October 2024. This section is intended to illustrate the application of the Bayesian models in a spatially explicit context—first, to visualize the associated predictive uncertainties, and second, to follow the common practice in ocean color studies of presenting algorithm outputs as maps, even though a direct reference or "ground truth" field is not available for validation (Brewin et al., 2015; O'Reilly & Werdell, 2019; Zoffoli et al., 2025). In particular, the Bayesian OC3 model was adopted to predict an ensemble of 50 [CHL-a] fields, from which the mean, MAP, and CoV are estimated, which are shown in Figure 5. To further assess representativeness, we computed the percentage of pixels in the satellite scene whose MBR values fall outside the range of the training data; in all cases less than 1% of the points lie beyond the min-max range of the training data set, highlighting the strong coverage and suitability of the training data for the observed conditions. The plot first shows that the MAP [CHL-a] field reasonably represents the concentration field, where the high concentrations follow plumes and eddy structures, with the highest concentration near the Gulf of the Farallones, in agreement with past studies (Legaard & Thomas, 2006). Figure 5b shows that the MAP and the mean offer similar predictions, suggesting that the posterior distribution of the predictions is almost symmetric. Notably, Figure 5c illustrates a larger difference, on log scale, between the MAP prediction and the OC4 model, highlighting the importance of using regionally-tuned models for reliable predictions. Finally, Figure 5d presents the CoV as a means to estimate prediction uncertainty by normalizing the standard deviation using the mean. The plot indicates

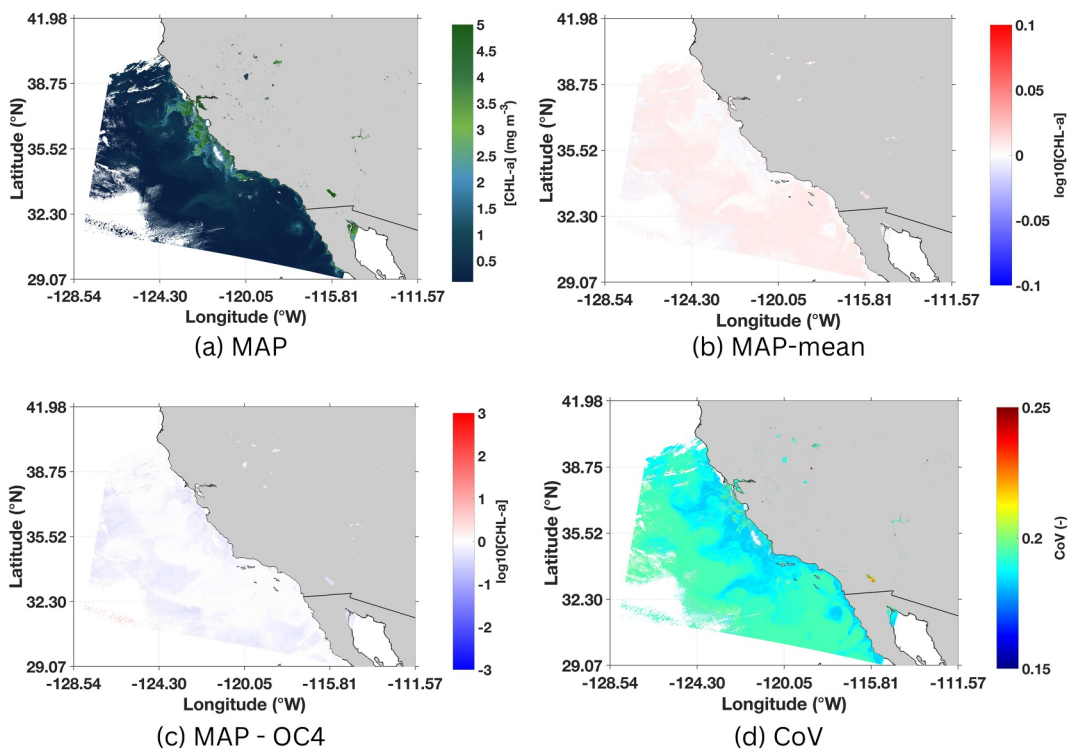

Figure 5. Spatial maps of (a) the maximum a posteriori (MAP) [CHL-a] obtained using the OC3 model (b) difference between the MAP and mean predictions from OC3, (c) difference between the OC3 MAP and Sentinel-3 OC4 prediction, and (d) coefficient of variation estimated using an ensemble of OC3 predictions. The plots are based on a scene retrieved from Sentinel-3 on 22nd of October 2024 for the Western coast of the contiguous United States.

10.1029/2025EA004551

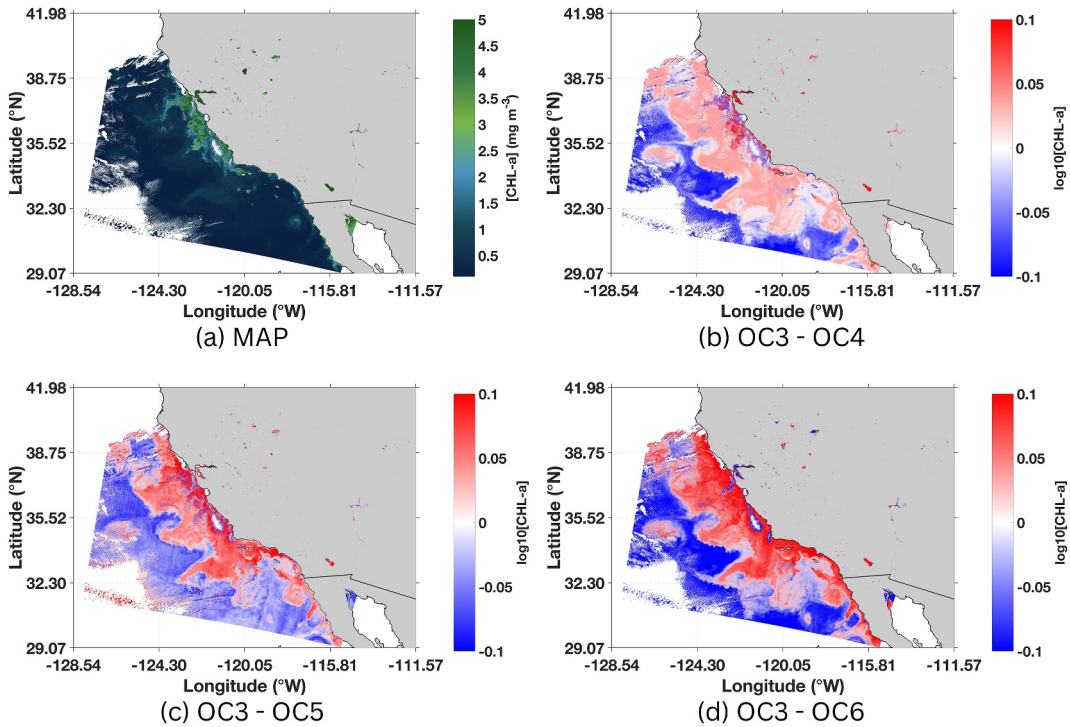

Figure 6. Spatial maps of (a) the maximum a posteriori (MAP) [CHL-a] obtained using the OC3 model, (b) difference between the MAP predictions from ocean color models OC3 and OC4, (c) OC3 and OC5, and (d) OC3 and OC6 models. The plots are based on a scene retrieved from Sentinel-3 on 22nd of October 2024 for the Western coast of the contiguous United States.

that the prediction standard deviation is approximately within 20% of the mean value, which is not extremely high but also not negligible. This means that while the data points vary from the mean, they do not do so extremely in either direction.

Figure 6 presents the OC3 MAP prediction alongside its difference with the MAP prediction from OC4, OC5 and OC6 on log scale for the same scene considered in Figure 5. The plot indicates that the OC3 and OC4 predictions are relatively similar to each other, such that predictions are within ± 0.05. However, this difference is almost doubled near the coast close to the Gulf of the Farallones, which is expected since band ratio models are not tailored toward coastal predictions. The differences between the OC3 MAP with the OC5 and OC6 MAP fields are more appreciable than those with OC4. In particular, the difference can reach up to ± 0.11, which is almost double the maximum difference between the OC3 and OC4 MAP fields. Furthermore, the large [CHL-a] values are predicted to be larger using OC5 and OC6 in comparison to those from OC3, similarly, the small values tend to be smaller with the OC5 and OC6 models. These results suggest that the OC3 and OC4 models might be more suitable for predicting [CHL-a], at least within the premises of the present study.

6. Conclusions

This study investigated established methodologies in Bayesian statistics, namely, Bayesian regression and Bayesian model selection in application to ocean color models. The study relies on the largest, and most commonly adopted, open-source bio-optical matchup repository that compiles 27 data sets of globally-sourced match-ups alongside the most commonly adopted validation data set SeaBASS. Bayesian regression was first employed to construct Bayesian ocean color models that are probabilistic counterparts of the conventional deterministic band ratio and band difference models. These probabilistic models could then be used to generate an ensemble of [CHL-a] predictions, which can ultimately be used for data assimilation, risk assessment, and uncertainty quantification.

The second contribution of the present study was adopting Bayesian model selection as a new approach for assessing ocean color models in addition to standard methods. This was motivated by the need to better delineate subtle performance differences across ocean color models, specifically where standard error and covariance metrics might not perform well. The Bayesian and Akaike information criteria (BIC/AIC) were computed for ocean color models based on band ratios and differences using global and regional data sets employing measurements from various sensors. Our results suggest that the likelihood-based measures generally favor the OC3 band ratio model, meaning that the additional terms associated with more complicated models might not be warranted from that standpoint. This is particularly interesting because standard remotely sensed ocean color products often provide predictions using the OC4 model, tuned using a global matchup data set. Our results recommend examining likelihood-based approaches for assessing the performance of ocean color models, especially when standard statistical measures are at a draw.

Diving further into the discussion, our information-criterion analysis, following a likelihood approach for model comparison, favors OC3 over OC4 for the data sets examined here (Tables 1–3). This does not contradict the community's widespread operational use of OC4. One objective of the present study is to underscore the importance of likelihood-based model selection criteria as an alternative to standard error metrics, which are typically similar in value over different models. In our analyses, and according to the likelihood-based performance metric used in our Bayesian framework, adding more parameters beyond those of OC3 did not significantly improve the model's ability to explain the data. This finding suggests that future algorithm developers might consider avoiding overly complex polynomial fits when simpler formulations capture the available information just as well.

The probabilistic ocean color models were then used to populate a posterior distribution of [CHL-a] predictions based on a Sentinel-3 image for the west coast of the contiguous United States. The prediction with the MAP and the mean of the ensemble of [CHL-a] predictions are comparable, suggesting that the posterior distribution is almost symmetric. The OC4 output provided with the Sentinel-3 product is far from the MAP prediction, which underscores the importance of regionally-tuned ocean color models. The results also highlight the magnitude of the uncertainty associated with the predictions, where predictions by the probabilistic model are within approximately 20% of the average prediction. These results illustrate the importance of probabilistic ocean color models for risk assessment studies related to ocean productivity, where posterior distributions enable a deeper understanding of the possible ocean state and potential scenarios. We conclude by noting that, without the loss of generalizability, these methods can be used to derive probabilistic models based on other ocean color algorithms.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The data on which this article is based are available in Valente et al. (2022a, 2022b). The codes will be made accessible on Github (https://github.com/mhammoud115/OCModelSelection) and Zenodo (https://doi.org/10.5281/zenodo.17474524) upon acceptance.

Acknowledgments
This research has been supported by the National Oceanic and Atmospheric Administration (US Department of Commerce Grant NA23OAR4320198) and Princeton University through the Cooperative Institute for Modeling the Earth System. RB is supported by a UKRI FLF grant (MR/V022792/1).

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0_15

Barker, K., Mazeran, C., Lerebourg, C., Bouvet, M., Antoine, D., Ondrusek, M., et al. (2008). Mermaid: The meris matchup in-situ database. In *Proceedings of the 2nd (a) atsr and MERIS workshop, Frascati, Italy* (pp. 22–26).

Basu, S., & Mackey, K. R. (2018). Phytoplankton as key mediators of the biological carbon pump: Their responses to a changing climate. *Sustainability*, *10*(3), 869. https://doi.org/10.3390/su10030869

Blondeau-Patissier, D., Gower, J. F., Dekker, A. G., Phinn, S. R., & Brando, V. E. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in Oceanography*, *123*, 123–144. https://doi.org/10.1016/j.pocean.2013.12.008

Brewin, R., Devred, E., Sathyendranath, S., Lavender, S. J., & Hardman-Mountford, N. J. (2011). Model of phytoplankton absorption based on three size classes. *Applied Optics*, *50*(22), 4535–4549. https://doi.org/10.1364/AO.50.004535

Brewin, R., Hirata, T., Hardman-Mountford, N. J., Lavender, S. J., Sathyendranath, S., & Barlow, R. (2012). The influence of the Indian Ocean dipole on interannual variations in phytoplankton size structure as revealed by Earth observation. *Deep Sea Research Part II: Topical Studies in Oceanography*, *77–80*, 117–127. https://doi.org/10.1016/j.dsr2.2012.04.009

Brewin, R., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E., et al. (2015). The ocean colour climate change initiative: III. A round-robin comparison on in-water bio-optical algorithms. *Remote Sensing of Environment*, *162*, 271–294. https://doi.org/10.1016/j.rse.2013.09.016

Bricaud, A., Babin, M., Morel, A., & Claustre, H. (1995). Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. *Journal of Geophysical Research*, *100*(C7), 13321–13332. https://doi.org/10.1029/95jc00463

Brown, A., & Jones, G. L. (2024). Convergence rates of Metropolis–Hastings algorithms. *WIREs Computational Statistics*, *16*(5), e70002. https://doi.org/10.1002/wics.70002

Cael, B. B., Bisson, K., Boss, E., Dutkiewicz, S., & Henson, S. (2023). Global climate-change trends detected in indicators of ocean ecology. *Nature*, *619*(7970), 551–554. https://doi.org/10.1038/s41586-023-06321-z

Carlin, B. P., & Louis, T. A. (2008) In *Bayesian methods for data analysis* (3rd ed.). Chapman and Hall/CRC.

Cetinić, I., Rousseaux, C. S., Carroll, I. T., Chase, A. P., Kramer, S. J., Werdell, J., et al. (2024). Phytoplankton composition from space: Requirements, opportunities, and challenges. *Remote Sensing of Environment*, *302*, 113964. https://doi.org/10.1016/j.rse.2023.113964

Chipman, H., George, E. I., & McCulloch, R. E. (2001). The practical implementation of Bayesian model selection. In *Institute of mathematical statistics lecture notes - Monograph series* (pp. 65–116). Institute of Mathematical Statistics.

Clark, D., Murphy, M., Yarbrough, M., Feinholz, M., Flora, S., Broenkow, W., et al. (2003). Moby, a radiometric buoy for performance monitoring and vicarious calibration of satellite ocean color sensors: Measurement and data analysis protocols (no. 6). *SeaWiFS Postlaunch Technical Report Series*.

Craig, S. E., & Karaköylü, E. M. (2019). Bayesian models for deriving biogeochemical information from satellite ocean color. *Down to Earth*. https://doi.org/10.31223/OSF.IO/SHP6Y

Devine, L., Galbraith, P. S., Joly, P., Plourde, S., Saint-Amand, Pierre, J. S., & Starr, M. (2015). Chemical and biological oceanographic conditions in the Estuary and Gulf of St. Lawrence during 2015. Fisheries and Oceans Canada. *Ecosystems and Oceans Science*.

Frouin, R., & Pelletier, B. (2015). Bayesian methodology for inverting satellite ocean-color data. *Remote Sensing of Environment*, *159*, 332–360. https://doi.org/10.1016/j.rse.2014.12.001

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (Eds.) (1995)., *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., et al. (2019). Satellite ocean colour: Current status and future perspective. *Frontiers in Marine Science*, *6*, 485. https://doi.org/10.3389/fmars.2019.00485

Haeder, D.-P., Villafane, V. E., & Helbling, E. W. (2014). Productivity of aquatic primary producers under global climate change. *Photochemical and Photobiological Sciences*, *13*(10), 1370–1392.

Hammoud, M. A. E. R., Papagiannopoulos, N., Krokos, G., Brewin, R., Raitsos, D. E., Knio, O., & Hoteit, I. (2025). On the potential of Bayesian neural networks for estimating chlorophyll-a concentration from satellite data. *Remote Sensing*, *17*(11), 1826. https://doi.org/10.3390/rs17111826

Harris, G. (2012). *Phytoplankton ecology: Structure, function and fluctuation*. Springer Science and Business Media.

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28. https://doi.org/10.1890/14-0661.1

Houskeeper, H. F., & Hooker, S. B. (2025). The primacy of dissolved organic matter to aquatic light variability. *Biogeosciences*, *22*(17), 4367–4385. https://doi.org/10.5194/bg-22-4367-2025

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll *a* algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research*, *117*(C1), C01011. https://doi.org/10.1029/2011JC007395

IOCCG (2019). Uncertainties in ocean colour remote sensing. *International Ocean Colour Coordinating Group*. https://doi.org/10.25607/OBP696

Joint, I., & Groom, S. B. (2000). Estimation of phytoplankton production from space: Current status and future potential of satellite remote sensing. *Journal of Experimental Marine Biology and Ecology*, *250*(1–2), 233–255. https://doi.org/10.1016/s0022-0981(00)00199-4

Jones, R. I. (1998). Phytoplankton, primary production and nutrient cycling. *Aquatic humic substances: ecology and biogeochemistry*, 145–175.

Legaard, K. R., & Thomas, A. C. (2006). Spatial patterns in seasonal and interannual variability of chlorophyll and sea surface temperature in the California current. *Journal of Geophysical Research*, *111*(C6), C06032. https://doi.org/10.1029/2005JC003282

Lévy, M., Iovino, D., Resplandy, L., Klein, P., Madec, G., Tréguier, A.-M., & Takahashi, K. (2012). Large-scale impacts of submesoscale dynamics on phytoplankton: Local and remote effects. *Ocean Modelling*, *43–44*, 77–93. https://doi.org/10.1016/j.ocemod.2011.12.003

Lin, J., Dall'Olmo, G., Tilstone, G. H., Brewin, R. J. W., Vabson, V., Ansko, I., et al. (2022). Derivation of uncertainty budgets for continuous above-water radiometric measurements along an Atlantic meridional transect. *Optics Express*, *30*(25), 45648–45675. https://doi.org/10.1364/OE.470994

MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.

Marshall, L., Nott, D., & Sharma, A. (2004). A comparative study of Markov Chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research*, *40*(2), W02501. https://doi.org/10.1029/2003WR002378

Marzouk, Y. M., Najm, H. N., & Rahn, L. A. (2007). Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, *224*(2), 560–586. https://doi.org/10.1016/j.jcp.2006.10.010

Matrai, P., Olson, E., Suttles, S., Hill, V., Codispoti, L., Light, B., & Steele, M. (2013). Synthesis of primary production in the Arctic Ocean: I. Surface waters, 1954–2007. *Progress in Oceanography*, *110*, 93–106. https://doi.org/10.1016/j.pocean.2012.11.004

Morel, A. (1980). In-water and remote measurements of ocean color. *Boundary-Layer Meteorology*, *18*(2), 177–201. https://doi.org/10.1007/BF00121323

Morel, A., & Prieur, L. (1977). Analysis of variations in ocean color1. *Limnology & Oceanography*, *22*(4), 709–722. https://doi.org/10.4319/lo.1977.22.4.0709

Najm, H. N. (2017). An introduction to statistical inverse problems and bayesian inference. Retrieved from https://www.osti.gov/biblio/1465087

Neath, A. A., & Cavanaugh, J. E. (2012). The bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics*, *4*(2), 199–203. https://doi.org/10.1002/wics.199

Nechad, B., Ruddick, K., Schroeder, T., Oubelkheir, K., Blondeau-Patissier, D., Cherukuru, N., et al. (2015). Coastcolour round robin data sets: A database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters. *Earth System Science Data*, *7*(2), 319–348. https://doi.org/10.5194/essd-7-319-2015

O'Reilly, J. E., & Werdell, J. (2019). Chlorophyll algorithms for ocean color sensors - oc4, oc5 and oc6. *Remote Sensing of Environment*, *229*, 32–47. https://doi.org/10.1016/j.rse.2019.04.021

Peloquin, J. M., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., et al. (2013). The MAREDAT global database of high performance liquid chromatography marine pigment measurements - Gridded data product (NetCDF) - Contribution to the MAREDAT world ocean atlas of plankton functional Types [dataset]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.793246

Platt, T., WhiteIII., G. N., Zhai, L., Sathyendranath, S., & Roy, S. (2009). The phenology of phytoplankton blooms: Ecosystem indicators from remote sensing. *Ecological Modelling*, *220*(21), 3057–3069. https://doi.org/10.1016/j.ecolmodel.2008.11.022

Prochaska, J. X., & Frouin, R. J. (2025). On the challenges of retrieving phytoplankton properties from remote-sensing observations. *Biogeosciences*, *22*(18), 4705–4728. https://doi.org/10.5194/bg-22-4705-2025

Racault, M.-F., Platt, T., Sathyendranath, S., Ağirbaş, E., Martinez Vicente, V., & Brewin, R. (2014). Plankton indicators and ocean observing systems: Support to the marine ecosystem state assessment. *Journal of Plankton Research*, *36*(3), 621–629. https://doi.org/10.1093/plankt/fbu016

Saranathan, A. M., Smith, B., & Pahlevan, N. (2023). Per-pixel uncertainty quantification and reporting for satellite-derived chlorophyll-a estimates via mixture density networks. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–18. https://doi.org/10.1109/TGRS.2023.3234465

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16. https://doi.org/10.1016/j.jmp.2018.03.001

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

Shu, C., Xiu, P., Xing, X., Qiu, G., Ma, W., Brewin, R., & Ciavatta, S. (2022). Biogeochemical model optimization by using satellite-derived phytoplankton functional type data and BGC-ARGO observations in the northern South China Sea. *Remote Sensing*, *14*(5), 1297. https://doi.org/10.3390/rs14051297

Sun, X., Brewin, R., Sathyendranath, S., Dall'Olmo, G., Antoine, D., Barlow, R., et al. (2025). Coupling ecological concepts with an ocean-colour model: Parameterisation and forward modelling. *Remote Sensing of Environment*, *316*, 114487. https://doi.org/10.1016/j.rse.2024.114487

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Jackson, T., et al. (2022b). A compilation of global bio-optical in situ data for ocean colour satellite applications – Version three. *Earth System Science Data*, *14*(12), 5737–5770. https://doi.org/10.5194/essd-14-5737-2022

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Jackson, T., et al. (2022a). A compilation of global bio-optical in situ data for ocean-colour satellite applications - Version 3 [dataset bundled publication]. *PANGAEA*. https://doi.org/10.1594/PANGAEA.941318

Vandermeulen, R. A., Mannino, A., Craig, S. E., & Werdell, J. (2020). 150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean. *Remote Sensing of Environment*, *247*, 111900. https://doi.org/10.1016/j.rse.2020.111900

Wang, Y., Song, Q., Wasif, D., Shahzad, M., Koller, C., Bamber, J., & Zhu, X. X. (2024). How certain are uncertainty estimates? Three novel Earth observation datasets for benchmarking uncertainty quantification in machine learning. Retrieved from https://arxiv.org/abs/2412.06451

Werdell, J., & Bailey, S. (2002). *The SeaWiFS bio-optical archive and storage system (seabass): Current architecture and implementation (technical memorandum no. 2002-211617)*. NASA Goddard Space Flight Center.

Werdell, J., Bailey, S., Fargion, G., Pietras, C., Knobelspiesse, K., Feldman, G., & McClain, C. (2003). Unique data repository facilitates ocean color satellite validation. *Eos, Transactions American Geophysical Union*, *84*(38), 377–387. https://doi.org/10.1029/2003EO380001

Werdell, J., & Bailey, S. W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, *98*(1), 122–140. https://doi.org/10.1016/j.rse.2005.07.001

Wit, E., Heuvel, E. v. d., & Romeijn, J.-W. (2012). 'All models are wrong…': An introduction to model uncertainty. *Statistica Neerlandica*, *66*(3), 217–236. https://doi.org/10.1111/j.1467-9574.2012.00530.x

Zoffoli, M. L., Brando, V., Volpe, G., González Vilas, L., Davies, B. F. R., Frouin, R., et al. (2025). Ciao: A machine-learning algorithm for mapping Arctic Ocean chlorophyll-a from space. *Science of Remote Sensing*, *11*, 100212. https://doi.org/10.1016/j.srs.2025.100212