Reviewer #2

Review of "Machine learning for estimating phytoplankton size structure from satellite ocean color imagery in optically complex Pacific Arctic waters"

This paper is interesting and timely. It tackles an important gap in Arctic Ocean remote sensing: how to retrieve phytoplankton size structure these optically complex waters. The authors explore machine learning to deal with the high uncertainty in going from Rrs to aph. The main finding, that a simple Rrs-based multivariable linear regression model performs best for satellite applications, is significant.

Overall, the study is well-motivated, the methodology is sound, and the results are clear. Revisions are needed, however, to improve the focus and the clarity.

> Thank you for the insightful comments that sharpened the paper's focus and transparency. A detailed, point-by-point response follows.

**Major comments:**

1. The paper is slightly unfocused on its main contribution. Is the novelty in the methodology (new CSD model) or in the application (Arctic η distribution)? The Abstract and Introduction should be revised to make it clear.

> We thank the reviewer and clarify that the primary novelty is methodological, not the application. We have revised the Abstract to state this explicitly. The paper develops and rigorously benchmarks retrievals of CSD slope directly from Rrs(λ), and evaluates ML and aph(λ)–based alternatives to identify viable non-PCA formulations. The maps of the CSD slope are included only as a brief demonstration of feasibility. The final paragraph of the Introduction already frames the study in this method-first way, so we have not made additional changes there.

Abstract: In this study, we have developed a method based on ML to use remote sensing reflectance ($R_{rs}(\lambda)$) for directly retrieving $\eta$, thus avoiding uncertainties due to the inversion of $a_{ph}(\lambda)$ from $R_{rs}(\lambda)$.

Introduction: The current study aims to (1) parameterize CSD models for the Pacific Arctic using spectral features of $R_{rs}(\lambda)$ and $a_{ph}(\lambda)$, (2) assess satellite algorithm performance using an *in situ* dataset, and (3) compare newly developed models with the previously developed PCA-based CSD model.

2. The paper discusses the trade-off between model accuracy and interpretability, in terms of machine learning methods. Since the Rrs-based linear regression model performed best for application, the authors should emphasize its transparency and robustness. I recommend expanding the Methods with more detail on this model.

> We appreciate the suggestion. To keep the Methods concise while ensuring full transparency, we added a pointer to the Supplement that provides the complete model parameters of the Rrs-based linear regression model in Table S4. Additionally, we will make the MATLAB codes of the developed CSD model available through GitHub, which help readers reproduce our model.

2. The strong performance of the ML model with in situ aph reflects a closer fundamental optical link between η and aph than between η and Rrs (Tables 6 and 7). Please clarify that the choice of the Rrs-based model is a practical solution to inversion limitations in optically complex waters, not an indication of a stronger fundamental relationship.

> We agree. The higher in-situ skill of the aph–based models reflects a closer optical linkage between η and aph than between η and Rrs (see Table 4). Our selection of the Rrs—based linear regression is therefore a practical choice for satellite application in optically complex waters, where inversions to aph are uncertain, and their errors would propagate to η. We have clarified this in the text and explicitly note that choosing the Rrs-based model does not imply a stronger fundamental relationship than the aph linkage; it reflects operational robustness (fewer assumptions, stable performance, and straightforward uncertainty quantification).

To state this clearly, we revised the Abstract and Section 4.4 as follows:

Abstract: Nevertheless, models using in-situ $a_{\mathrm{ph}}(\lambda)$ yielded better accuracy, reflecting a closer optical linkage between $\eta$ and $a_{\mathrm{ph}}(\lambda)$ than between $\eta$ and $R_{\mathrm{rs}}(\lambda)$. Our choice of an $R_{\mathrm{rs}}(\lambda)$-based model for satellite application is therefore practical, motivated by the limitations and uncertainty of $a_{\mathrm{ph}}(\lambda)$ inversions in optically complex waters.

The second paragraph in Section 4.4: The superior in-situ performance of aph(λ)-based models reflects a stronger physical coupling between η and aph(λ) (Tables 6 and 7). Our preference for the Rrs(λ)-based model is operational, as it avoids uncertain aph(λ) inversions in optically complex waters and yields stable retrievals at satellite scale; it should not be taken as evidence that η is more fundamentally linked to Rrs(λ) than to aph(λ).

3. The statement that the random selection was performed only once "in order to develop and compare different models using the consistent dataset" (Line 230) is a weakness for ML studies. Please comment on the feasibility of cross-validation or at least repeat the split several times and report mean and standard deviation of the metrics.

> Thank you for flagging this—our earlier wording was misleading. The 30% subset was a single external test set, held out once and used only for final validation (Figure 6). Model development and comparison were conducted on the remaining 70% using repeated 5-fold cross-validation. We have clarified this in the text and now report cross-validation performance as mean ± SD across repeats, alongside the external-test metrics. This

protocol cleanly separates model development/selection (70%) from final assessment (independent 30%).

In the revised version, we repeated the cross-validation for 10 times. Mean ± std across 10 repeats are shown in Table 4, as well as Tables S4 and S5 in the Supplement.

4. The paper is comprehensive but could be streamlined:

a. The Discussion section repeats quantitative findings already shown in the Results. Please trim or merge with Results.

> We agree and have removed duplicative sentences and folded essential interpretive points into the Results where appropriate. The revised Discussion now focuses on synthesis, limitations, and implications rather than re-reporting numerical results.

b. Some supporting material (e.g., phytoplankton community groups, Figure 4, Table 3) could be moved to the Supplementary.

> Implemented. We relocated the requested supporting material to the Supplement. The main text now contains 7 figures and 4 tables, improving readability while preserving full detail in the Supplement.

c. Figure 5 mainly illustrates optical complexity and could also be moved to the Supplementary.

> Implemented. Figure 5 now appears in the Supplement, with a pointer in the main text.

d. Since many models are compared, but not all are equally important, the main text should focus on the most effective model, with detailed comparison tables in the Supplementary.

> Implemented. The main text now emphasizes the most effective model for application, with concise comparative context. Comprehensive performance tables and model-by-model diagnostics are provided in the Supplement, ensuring transparency without overloading the narrative.

**Minor comments:**

1. Line 32: Remove “repeat”.

> Done.

2. Line 55. Define “AOPs” at first use.

> It's defined in the third paragraph in the Introduction: "The ocean color variables used in these spectral approaches are grouped into two categories: apparent optical properties (AOPs, e.g., Rrs(λ)) and inherent optical properties (IOPs, e.g., aph(λ))."

3. Line 61. Please add relevant references on ocean biogeochemical models that use the CSD slope.

> Direct assimilation of the CSD slope into ocean biogeochemical models is still emerging; to our knowledge it has not yet been widely implemented.

4. Line 84. There are many recent publications applying machine learning in ocean colour remote sensing, e.g., : https://doi.org/10.1016/j.rse.2023.113596 and https://doi.org/10.1016/j.rse.2023.113628, and etc.

> Added the two suggested Remote Sensing of Environment (2023) papers to the Introduction's ML paragraph.

5. I recommend adding a table listing all symbols and abbreviations used in the paper for clarity.

> Implemented. We added a list of definitions and units of symbols used in the manuscript as Table 1.