Dear Reviewer, we sincerely appreciate your valuable review and feedback on our manuscript. We agree that it is important to emphasize the key contributions and innovations of our study. In response to your comments, we have rewritten the abstract and discussion sections to clearly summarize our findings. Additionally, we have explicitly highlighted the main contributions of our work at the end of the introduction to enhance clarity and impact. A new version will be uploaded into the system when permitted. Some excerpts from the text have been copied into the responses for clarity.

The current manuscript builds upon existing research to explore the use of multi-output random forest models for retrieving backscattering coefficient (bbp) data at various depths using different input datasets. These inputs include enhancements in spatial resolution and a diversity of data types. However, the draft in its present form is somewhat rudimentary. The work presented is not effectively summarized in the abstract and discussion sections, and the highlights and innovations of the study are not prominently featured. I recommend that the authors address these issues by clearly outlining the study's contributions and innovations at the end of the introduction.

Below are specific suggestions for improvement:

1. The introduction initially mentions Particulate Organic Carbon (POC) using its abbreviation without first presenting its full name and explaining its significance within the study's context. This may confuse readers unfamiliar with the term.

   The full name and abbreviation were first introduced in the abstract but are now also included in the introduction of the new version.

   Additionally, the discussion on why profiling POC is challenging is insufficiently developed. A more detailed explanation of these measurement difficulties is necessary to establish the research problem's significance and to clearly justify the study's objectives. Providing a comprehensive background on POC and elaborating on the challenges in measuring it will better prepare readers for the research presented.

   Thanks for pointing this out. We have modified the introduction and added a more detailed description in the new version of the manuscript (see in question number 3).

2. The second paragraph of the introduction discusses Apparent Optical Properties (AOP), which does not appear to be directly related to the paper's main focus. This section may detract from the introduction's clarity and coherence by introducing a topic that is not central to the study's objectives. It is important to ensure that the introduction remains focused on the key themes and research questions. If AOP is not essential to the main argument, consider removing this section or significantly condensing it to maintain the introduction's focus and engage readers with the paper's central themes.

   One of the aims of the manuscript is to evaluate if the Inherent Optical Properties (IOP), which are provided as a derived product from the Sentinel-3 OLCI-C2RCC data, can improve the retrieval of bpp. This product is directly related to AOP and a comprehensive background is provided in the introduction for this reason. However, we have rewritten this paragraph to highlight why it is important in this study.

3. The logical flow between the introduction's first two paragraphs is somewhat disjointed, potentially hindering the reader's understanding of the paper's overall direction. Furthermore, the latter paragraphs lack a detailed analysis of the current research landscape. The discussion of existing studies is limited and does not clearly identify the knowledge gaps this paper aims to address. To enhance the introduction, revise the first two paragraphs to improve their logical structure and coherence. Additionally, include a more comprehensive review of the current research, highlighting specific gaps in the literature and the problems this study seeks to solve. Incorporating more examples of relevant previous research will strengthen the context and rationale for the study, providing a clearer foundation for the paper's contributions.

   Thank you for your comments. We have rewritten the introduction accordingly to clarify and highlight the objectives and novelties of this study.

   "*The ocean covers approximately 70\% of Earth's surface and plays a fundamental role in regulating climate dynamics. It redistributes energy and carbon through a variety of physical and biogeochemical processes. Among these processes, the biological carbon pump facilitates the transfer of CO${_2}$ from the atmosphere to the ocean floor by enabling the production and sinking of particulate organic carbon (POC), which becomes sequestered in deep-ocean sediments. POC originates from living organic carbon, primarily produced by photosynthetic organisms such as phytoplankton, which thrive in the sunlit upper ocean layers. These organisms require carbon compounds, along with light and nutrients, to survive and reproduce(Falkowski et al., 1998; Siegel et al., 2014).Their presence and abundance reflects the interplay of resources and losses in the environment(Behrenfeld et al., 2006) with populations maintaining daily division cycles even in regions where nutrients appear to be depleted beyond detection limits (Ribalet et al., 2015; Vaulot and Marie, 1999). The observed populations represent a balance where new biomass produced each day is matched by consumption through grazing and other loss processes (Landry and Hassett, 1982; Calbet and Landry, 2004) maintaining relatively stable populations despite continuous growth and turnover. Quantifying phytoplankton biomass and carbon content is crucial to understanding these ecosystem dynamics and their role in carbon cycling. Traditionally, chlorophyll-a (chl-a}) concentration has been used as a proxy for phytoplankton biomass, but its interpretation is complicated due to physiological photoacclimation, which affects intracellular pigment content without necessarily indicating changes in biomass. The Particulate Backscattering Coefficient (bbp) has been recognized as a stable optical proxy for phytoplankton biomass and carbon content as it is sensitive to the abundance, size distribution, and composition of suspended particles, rather than pigment concentration alone (Behrenfeld and Boss, 2006; Graff et al., 2015; Martinez-Vicente et al., 2013). Unlike chl-\textit{a}, which can underestimate biomass in stratified and oligotrophic waters, bbp remains relatively unaffected by photoacclimation effects, making it particularly useful for studying carbon fluxes across different oceanic regions and depth layers. The complex interaction between key variables (usually non-linear) and limited sampling resolution in dynamic environments, combined with the technical challenges of depth-resolved measurements, contribute to gaps in our understanding of specific marine processes, such as carbon sequestration, nutrient cycling, sedimentation and the ocean-atmosphere CO2 exchange (...)*"

4. The introduction should underscore the importance of bbp in POC measurement, as well as the deficiencies and areas for improvement in current bbp products. While the introduction currently highlights the significance of POC, it does not adequately stress the critical role of bbp. Clarify whether POC estimation relies solely on bbp and discuss its specific importance in this context. Additionally, expand upon the current state of bbp data by discussing the limitations of existing bbp products and the shortcomings of related algorithms. For instance, accurately deducing inherent optical properties (IOPs) from apparent optical properties (AOPs) is crucial for POC retrieval models based on IOPs, but this process can be challenging. Furthermore, the complex optical conditions in coastal areas can lead to significant spatial heterogeneity in POC distribution, introducing uncertainty in POC estimation even when using advanced methods. Addressing these points will provide a clearer context for the study's objectives and the need for improved bbp products.

   We have added a paragraph in the Introduction where we tackle the following points:

   - Remark the bbp use as a proxy of POC
   - How we can derive bbp: from in situ data using the BGC-Argo floats (sensors of backscatter at 700 nm); from satellite, where several algorithms have been developed like NASA's OBG group bbp_sat product.
   - From satellites Rrs is used and an inversion model is applied to derive IOPs. The retrieval of bbp from satellites requires of a forward model that shows the relation between Rrs and bbp (i.e. GIOP, QAA and others).

   It now reads:

   "The bbp parameter is an inherent optical property (IOP) of water, and it has been widely recognized as a robust bio-optical proxy for POC (Cetini´c et al., 2012; Sullivan et al., 2013). However, b_bp measured by floats can have an uncertainty of the order of 10–15% (Bisson et al., 2019). These uncertainties stem from the instrumental drift, the sensor calibration limitations, and the reliance on manufacturer calibration files rather than sensor-specific calibrations using dark counts. While autonomous platforms provide extensive spatial and temporal coverage, these factors must be considered when interpreting bio-optical datasets to ensure accuracy and reliability. IOPs are intrinsic characteristics of water, determined solely by its composition and are independent of the external light field or the geometrical angle conditions during observation. These properties include absorption, elastic scattering, inelastic processes (such as fluorescence and Raman scattering), and attenuation, which describe how light behaves and propagates through water. IOPs are essential in studying light interactions in aquatic environments, as they reflect the presence of dissolved organic matter, phytoplankton and suspended particles. The b_bp can be measured by autonomous platforms spread out across the ocean, such as the Biogeochemical-Argo (BGC-Argo) profiling floats(Claustre et al., 2020); or estimated from onboard satellite sensors, such as the Sentinel-3 Ocean and Land Colour Instrument (EUMETSAT, 2019; Jorge et al., 2021; Koestner et al., 2024). Designing observational strategies based on

combining the two approaches constitutes a fundamental tool for improving knowledge of ocean processes."

The calculation of inherent optical properties (IOPs) or concentrations of water constituents from reflectances always involves some degree of uncertainty (IOCCG, 2019). Two primary factors contribute to this: 1) The inverse relationship between IOPs and water reflectance spectra is an underdetermined system. This means that the information contained in reflectance spectra is significantly lower than the number of variables influencing those spectra. As a result, different combinations of IOPs can produce nearly identical reflectance spectra, leading to inherent ambiguity between IOPs and remote sensing reflectance (Rrs). 2) The natural variability of all components that determine reflectance spectra is extremely high. This includes the optical properties of the atmosphere and water, their vertical distribution, and the characteristics of the air-sea interface. Because of this complexity, any retrieval algorithm simplifies the natural system and is only effective within the scope defined by its underlying optical model assumptions.

- [ ] The bbp from Argo floats is explained in detail in section 2.2.
- [ ] Derivation of the satellite IOPs from C2RCC is later explained in section 2.3

5. It is crucial to provide specific details about the data collected from each dataset, including the exact variables used, the time range of data collection, website links for accessing the data, and the dates when the data were accessed. Currently, Table 1 lacks sufficient information, and the time frames for the BGC-Argo data and other datasets are not clearly stated. To improve clarity and completeness, ensure that all necessary details are included in the data section, allowing readers to understand the scope and sources of the data used in this study.

   We have added the following text in the new version of the manuscript and changed figure 1.

   "The temporal distribution of the match-ups shows a clear seasonal bias, with most data concentrated between May and September, particularly during 2017. This uneven distribution is primarily due to the limited availability of cloud-free satellite observations required to match with BGC-Argo profiles, especially during winter months when cloud cover and low solar angles reduce the quality of remote sensing products."

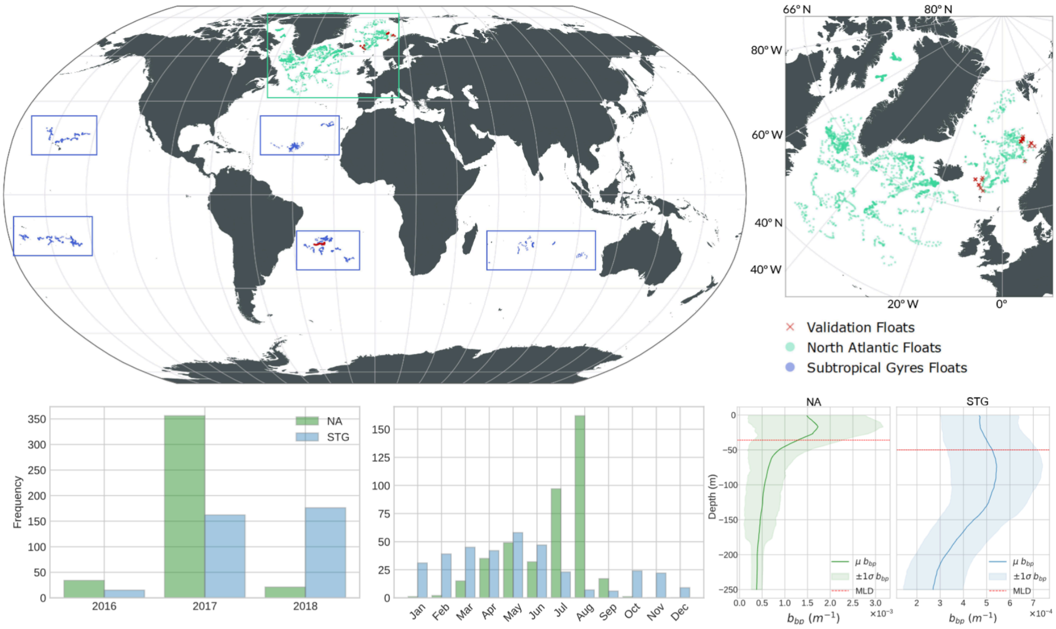

Figure 1. Global map showing the geographic locations of the BGC-Argo floats and satellite data matchups. [Botttom row] Temporal coverage of matchups by year (left) and month (middle) for the North Atlantic (NA, green) and Subtropical Gyres (STG, blue). Vertical profiles (right) of bbp from floats, where solid lines show mean values, shaded areas ±1 standard deviation, and the dashed red line the average Mixed Layer Depth (MLD).

6. In the methods section, the use of Principal Component Analysis (PCA) for dimensionality reduction of high-dimensional features is mentioned, stating that “After this feature reduction on the high-dimensional variables, the 250 m and 50 m measurements with 126 and 26 inputs are reduced to 5 components for each variable, resulting in a total of 20 features. This method still retains 99% of the information.” However, this section lacks supporting data and visualizations to illustrate the PCA results. To enhance clarity and effectiveness, include data tables or figures that demonstrate the specific components selected and their contributions to the overall variance. This will help readers better understand the impact of PCA on the feature set and validate the claim that 99% of the information is retained.

   The application of PCA for dimensionality reduction is a well-established and widely used technique for reducing the dimensionality of large datasets while retaining the most important information. Given its common application in the field, we did not initially include a graphical representation of the PCA results in the manuscript. However, in response to your suggestion, we could include as supplementary material the PCA plot in the revised manuscript, although we consider that it does not add much to the analysis. PCA graph for all the variables and depth are shown below:

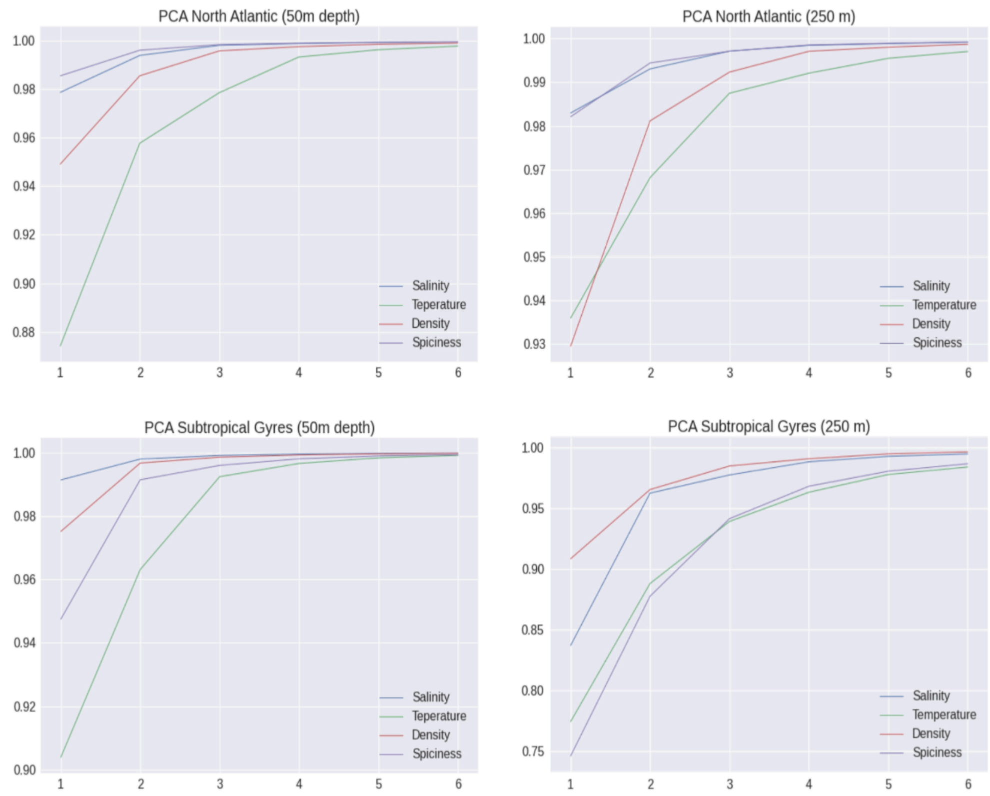

Figure. Cumulative explained variance of the first six PCA components.

7. In Section 2.5, the discussion is somewhat disorganized. The introduction of the Random Forest Regression model should precede the discussion of existing studies based on random forest models. Additionally, such content seems more appropriate for the introduction section, as it pertains to a review of existing research rather than the methods section. Moreover, the authors state, “All the previously mentioned algorithms, along with others such as Linear Regressor (LR), Ridge Linear Regressor (RLR), Random Forest Regressor (RFR), and Multi-Layer Perceptron (MLP), were tested for estimating bbp during the dataset preparation phase. Based on these results, the Random Forest Regressor (RFR) was selected as the most suitable algorithm for this multi-input/multi-output problem.” Comparative results should also be presented to illustrate the differences in inversion results and the stability of various models. This will help substantiate the choice of the Random Forest Regressor as the most suitable algorithm for the problem at hand.

We have restructured this section according to the comments. Thank you for pointing it out. Now it reads like this:

“There are two main approaches for dealing with multi-output regression problems. One way is to use univariate models, also known as problem transformation methods (Schmid et al., 2022; Borchani et al., 2015). These methods decompose the multi-output regression problem into multiple single-target problems, creating an independent model for each output. The predictions from these separate models are then combined. This approach ignores the relationships between the targets, which can adversely affect the prediction's overall accuracy. Alternatively, multivariate models are designed to capture dependencies and interactions between the outputs, potentially leading

to more accurate predictions (Borchani et al., 2015). When and how to apply these two approaches depends on the nature of the data and the correlation between the targets. In our preprocessing results, PCA decomposition indicates a high covariance among measurements at different depths in the water column. Since our regression models estimate bbp at different depths, it is logical to consider that nearby values in the water column are related to each other.

Random Forest Regressor (Breiman, 2001) has been widely applied in geosciences and marine environmental studies for classification and regression tasks (Cutler et al., 2007; Ruescas et al., 2018). Regression trees are at the model's core, which effectively handles complex data when there are non-linear dependencies between a numerical response variable and a diverse set of predictors, whether qualitative or quantitative (D'Ambrosio et al., 2017). RFR is an ensemble method that combines many weak decision tree learners, which are grown in parallel to reduce the bias and variance of the model simultaneously, enhancing the model's predictive performance. Furthermore, RFR provides insights into the importance of the training features, which reveals the variables that have the most significant impact on the predictions. This capability makes the model's mechanisms and results easier to interpret and explain.

Different algorithms have been tested in previous works (see Sauzède et al. (2016, 2020)) to estimate bbp at various depths. Both works are based on a multivariate model applied to all possible outputs. In SOCA16, a Multi-Layer Perceptron is developed, while in SOCA2020 a comparison between a linear model (Ridge) and an ensemble model (Random Forest) is done. The latter showed higher performance. The Multivariate Random Forest used in this study offers higher accuracy than the univariate Random Forest, especially when the outputs are highly correlated (Schmid et al., 2022) and when complex interactions demand structured inference to be effectively managed d (Xu et al., 2019). All the previously mentioned algorithms, including Linear Regressor (LR), Ridge Linear Regressor (RLR), Random Forest Regressor (RFR), and Multi-Layer Perceptron (MLP), were tested at both 50 and 250 m depth during the dataset preparation phase. Results for 250 m are shown in Figure 2. Based on these results, the Random Forest Regressor (RFR) was selected as the most suitable algorithm for this multi-input/multi-output problem."

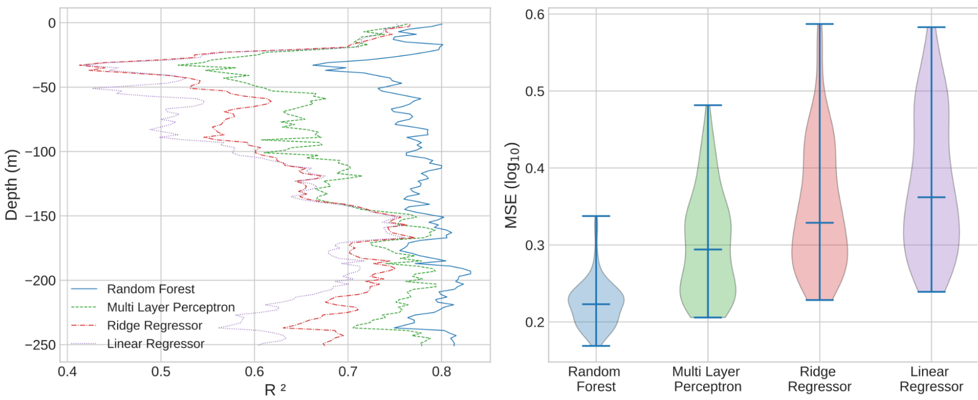

Figure 2. Comparison of different multi-output regression models for estimating vertical profiles of bbp up to 250 meters depth. Left: Depth-resolved R2 values for four regression models: Random Forest, Multi-Layer Perceptron, Ridge Regressor, and Linear Regressor. Right: Violin plots of the Mean Squared Error (MSE, log10-transformed) distributions for each model

8. In the initial paragraphs of Section 3, "Performance of the Random Forest Regressor," the authors refer to the content of Table 1, including the specific datasets corresponding to each abbreviation. However, this information should have been presented in the data introduction section. Instead, this section should provide details on the data volume obtained after feature engineering and data filtering, specifically how much data is used for training and how much for the independent validation set. This will give readers a clearer understanding of the data used in the study and its distribution between training and validation.

   We have moved this information to the data introduction section in the new version as suggested.

9. In the section "3 Performance of the Random Forest Regressor," the authors discuss the differential contribution of various features within the model. It would be beneficial to clarify the source of this feature importance data. Is it derived from the inherent parameters of the random forest model, or does it rely on additional algorithms? While the random forest, as an ensemble learning method, can assess feature importance through multiple decision trees, providing a measure of each feature's contribution to the predictive outcome, employing SHAP (SHapley Additive exPlanations) values could offer a more detailed and accurate attribution of feature importance. SHAP values provide a robust approach to explaining machine learning model outputs by assigning each feature an importance value for a particular prediction. Incorporating SHAP could enhance the transparency and depth of the analysis regarding each feature's influence on the model's performance.

   The feature importance (FI) for each model was obtained from the built-in feature importance algorithm. In this section we show the built-in FI of the different models in order to compare if the same variables are selected from the two satellite-derived products for the same training dataset (match-ups). A comparative of different feature importance algorithms, such as permutation importance and SHAP could be done, but we think that it is out of the scope of this manuscript. In the manuscript, we have clarified that it is calculated using the in-built feature importance of the random forest model.

10. In the same section, the authors depict the contribution of various features within the model. However, there are concerns regarding the clarity and utility of the presented feature importance data. Specifically, it should be clarified whether features with low contribution are consistently negligible across all depths. If these features do not significantly contribute to the model's performance at any depth, it might be beneficial to consider their removal to further reduce dimensionality and enhance the model's efficiency.

   We understand the concern regarding feature importance and dimensionality reduction. However, we have chosen to retain all features in our analysis to ensure consistency and comparability between both areas. The random forest model with fewer features is simpler, but usually similar results are yielded. Additionally, keeping all features allows for a more standardized evaluation and interpretation of results. We appreciate your suggestion and will consider discussing this point further in the manuscript to clarify our approach.

11. Additionally, some features are derived from PCA processing, and with the multitude of features used, it is challenging to distinguish between those

originating from different datasets or subjected to various treatments in the bar chart. To enhance the richness and readability of the visual information, it is suggested that the authors use distinct colors to represent bars corresponding to different types of features. This would allow for a clearer distinction between features from different datasets or processing methods, thereby providing a more informative and accessible visualization of the data. It is also worth noting that while random forest models can provide feature importances based on the model's internal assessment, these may not always reflect the true importance of features. The authors might also consider using alternative methods such as SHAP (SHapley Additive exPlanations) to calculate feature importances, which could offer a more nuanced understanding of each feature's contribution to the model's predictions.

Thank you for the suggestion. We have changed the figures using distinct colors for the different types of features in the new version.

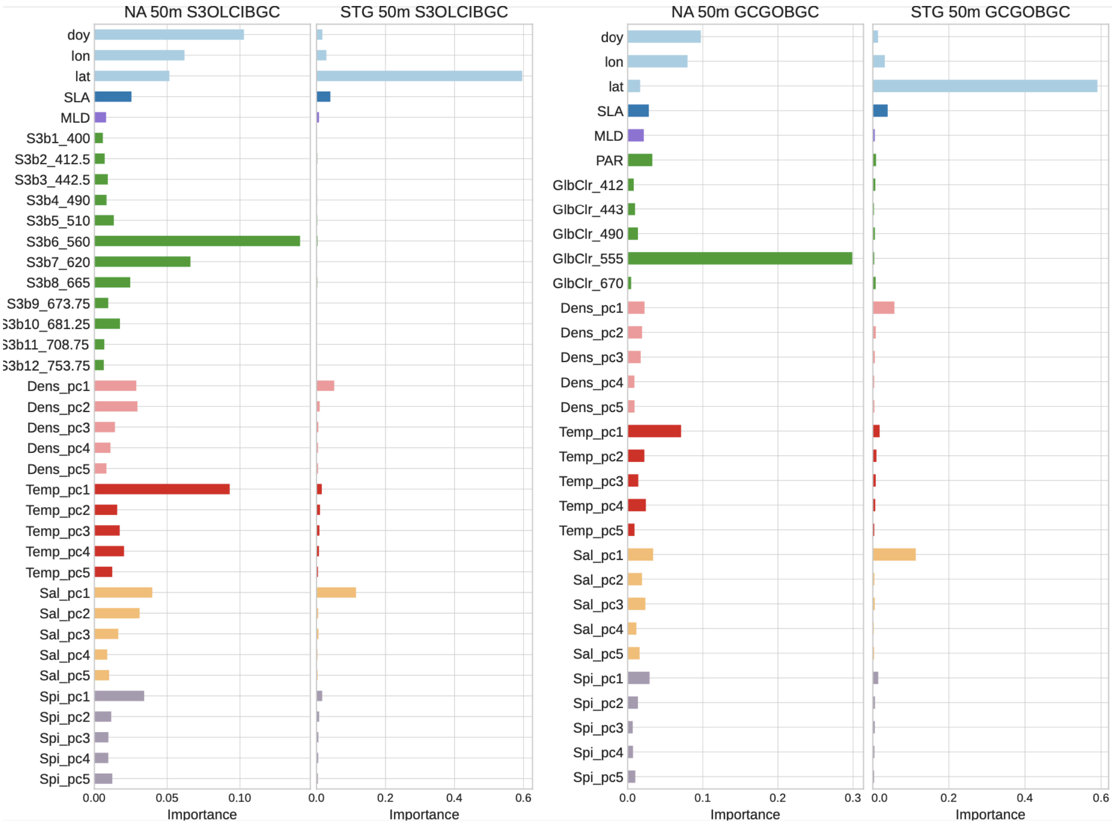

Figure 4. Feature importance for the models with S3OLCIBGC and GCGOBGC data for 50 m depth in the North Atlantic (NA) and in Subtropical Gyres (STG). Features are grouped by category according to color: (1) blue tones represent spatial and temporal descriptors, including day of year (doy), latitude, and longitude; (2) dark blue represents sea level anomaly (SLA); (3) purple indicates Mixed Layer Depth (MLD); (4) green corresponds to satellite reflectance bands from either Sentinel-3 OLCI or GlobColour with their central wavelengths; and (5) pink, red, orange, and light purple correspond to the first five principal components (PCs) derived from BGC-Argo profiles of density, temperature, salinity, and spiciness, respectively.

12. In the concluding part of the introduction, the authors outline the main content of the research, focusing on a detailed analysis of estimating bbp in the upper layers of the ocean surface using Sentinel-3 Ocean and Land Colour Instrument (S3OLCI) data. The study aims to enhance spatial resolution from the 4 km resolution of GlobColour level-3 merged products to the 300 m Full Resolution

(FR) of Sentinel-3 OLCI. Additionally, the research evaluates model performance after incorporating OLCI spectral wavelengths as features for bbp estimation and compares these results with those obtained using GlobColour. The study also explores whether the inclusion of Inherent Optical Properties (IOPs) derived from satellite data can improve the accuracy of bbp estimation compared to using reflectances alone. These IOPs, provided by the Sentinel-3 OLCI processor, are hypothesized to significantly enhance regression models. The comparison is made between BGC-Argo data and various satellite datasets for two depth layers: from the surface to either 50 m or 250 m. However, the abstract does not provide a comprehensive and concise summary of the work and its innovative aspects. After reading the abstract, it remains unclear what the specific contributions and novelties of this research are. I recommend that the authors revise the abstract to include a brief but complete overview of the study's objectives, methods, and key findings. The abstract should clearly communicate the innovative aspects of the research, such as the use of higher resolution data, the incorporation of IOPs, and the comparison of model performances, to give readers a clear understanding of the study's significance and contributions to the field.

We have changed the abstract to highlight the main contributions and innovations of this study.

"Abstract. As the second largest carbon reservoir on Earth, the ocean regulates the carbon balance through dissolved and particulate organic carbon forms. Monitoring carbon cycle processes is key to understanding climate system science. While most organic carbon in the ocean is dissolved, Particulate Organic Carbon (POC) plays a crucial role despite its smaller proportion, as it links surface biomass production, the deep ocean, and sedimentation. POC estimation is achieved by measuring proxies like the Particulate Backscattering Coefficient (bbp), obtained from satellite observations and in situ sensors, such as the BioGeoChemical-Argo (BGC-Argo) floats. Previous research has combined data from BGC-Argo floats and satellite sensors, demonstrating the potential of machine learning models to infer vertical bio-optical properties in the water column. By bridging the gap between surface optical properties and deep ocean processes, this approach enhances the estimation within the top 250 meters of the water column. This study focuses on such estimations with the inclusion of remote sensing data from the Sentinel-3 Ocean and Land Colour Instrument (OLCI) sensor at full resolution (300 m). The addition of optical information about absorption and scattering processes has improved the accuracy of the multi-output Random Forest models, which show promising results, especially within the first 50 meters in the Subtropical Gyres. However, in dynamic regions such as the North Atlantic, the results are less consistent, suggesting that further research is needed to understand how the complexity of the physical state of the water column modifies the bbp vertical fluxes."

End of the introduction section:

"Building on these results, this research proposes a more detailed analysis of estimating bbp in the upper layers of the ocean surface using the Sentinel-3 Ocean and Land Colour Instrument (S3OLCI). We enhance the spatial resolution from the 4 km resolution of GlobColour level-3 merged products (1/24° at the equator), used in previous studies, to the 300 m Full Resolution (FR) of Sentinel-3 OLCI. Additionally, we evaluate the model performance after incorporating OLCI spectral wavelengths as features for bbp estimation and

compare these results with those obtained using GlobColour. Another key aspect of this study is determining whether adding IOPs derived from satellite data (absorption and scattering) improves the accuracy of the bbp estimations compared to using reflectances alone. These IOPs, calculated from the OLCI processor, could significantly enhance regression models. Furthermore, bbp at different depths of the water column is estimated using multi-output models. These multi-output random forest models account for the high correlation between measurements at nearby depths. The comparison is conducted between BGC-Argo data and the satellite datasets for two depth layers: from the surface to either 50 m and 250 m."

13. The section "2.5 Multi-output Machine Learning Models" in the methods part of the paper should be clarified to determine whether it represents one of the study's innovative aspects. If this section indeed constitutes an innovation, it is essential to highlight it appropriately throughout the paper to ensure that readers recognize its significance. In the abstract, include a brief mention of the multi-output machine learning approach and its novelty to pique the interest of potential readers and set the stage for the detailed methodology presented later. In the introduction, provide a clear and concise explanation of what multi-output machine learning models are and how they are applied in this study. Emphasize the innovative nature of using these models, perhaps by comparing them to traditional single-output models or by discussing the advantages they offer in the research context. During the discussion, reflect on the implications of using multi-output machine learning models, including a comparison of their performance with other models, the benefits they provide in terms of accuracy or efficiency, and their potential applications in similar research endeavors. To ensure consistency and clarity, make sure that the term "multi-output" is consistently defined and used throughout the paper, and that its implications for the research are clearly articulated. If the multi-output approach is a key innovation, it should be a central theme in the narrative of the paper, guiding the reader through the methodology, results, and implications of the study.

    Multi-output models are commonly used in the machine learning field as they provide a better estimation of the output variables when they are related. In this case, we are estimating the bpp at different depths of the water column. Since measurements of bpp at nearby depths are highly correlated, using a multi-output model allows us to account for this correlation effectively. We have clarified and emphasized this in the new version of the manuscript. Thank you for the comment.

Overall, addressing these suggestions will significantly enhance the manuscript's clarity, coherence, and professionalism, thereby strengthening its contribution to the field of ocean physical remote sensing.

Citation: https://doi.org/10.5194/egusphere-2024-3942-RC1

Finally, we would like to thank the reviewer for their valuable comments and the time and effort dedicated to reviewing our work. Your valuable suggestions have contributed to enhancing the quality and clarity of our manuscript. We sincerely appreciate your effort and expertise.