Photosynthesis parameters play crucial roles in remote estimation of phytoplankton primary production. This study presents a new global dataset of photosynthesis parameters that derived from the in situ primary production profiles compiled by Mattei and Scardi (2021). The work addresses an important limitation in marine primary production modelling by improving the availability of photosynthesis parameters, which are difficult to obtain from remote sensing and remain sparsely sampled across the global oceans. The proposed inverse modelling framework is appropriate for the study and makes effective use of a large global observational archive by combining the shape of the normalized production profile with water column production. Overall, the manuscript is well organized, and the objectives, methods, results, and discussion are presented clearly. However, the following issues need to be further addressed before the manuscript is suitable for publication.

Abstract

Line 16 reports that 4160 photosynthesis irradiance parameters were recovered, whereas Lines 375-376 state that 3776 parameter pairs were obtained. These two numbers are confusing and should be explained more clearly. Please clarify what each number represents and use consistent terminology throughout the manuscript.

Revised as suggested.

Introduction

Line 32 “Global scale chlorophyll concentration …” — should this be revised to “chlorophyll-a concentration”? Please check and apply this correction consistently throughout the manuscript.

Revised as suggested.

Lines 34-36 The statement that direct measurements are the only way to estimate photosynthesis parameters may be misleading. Since this study estimates these parameters using an inverse modelling approach, it would be clearer to distinguish between directly measured parameters and model derived estimates.

Revised as suggested. This sentence was reformulated.

Lines 85-90, the authors mentioned the global database and codes for data access are made freely available. However, the Data Availability section appears to provide only the dataset location. The authors need to provide a dedicated Code Availability section describing the availability of the preprocessing, forward model, inverse optimization, and quality control code. If some parts cannot be made publicly available, this should be stated clearly.

Revised as suggested. The code is now made available on Zenodo repository via the following link: https://zenodo.org/records/22645060.

Inverse model

The terminology must be used consistently throughout the manuscript. For example, "watercolumn", "water column", and "water-column" should be standardized.

Revised as suggested.

It would also be helpful to include a table summarizing all symbols, their definitions, units, and whether they are measured or estimated.

Revised as suggested. A new Appendix was added with such a table.

Lines 105-110 use the light formulation of Platt et al. (1990) and Kovač et al. (2016a). Although this approach has been verified previously, the authors need to clarify why the approach is applicable for generating this global dataset of this study, particularly for profiles with reconstructed PAR and a single attenuation coefficient.

Revised as suggested. We do state in the manuscript (lines 112-114):

“Keeping in mind the limitations of the Mattei & Scardi (2021) dataset, namely a single value for the attenuation coefficient K and a single value for noon irradiance, we employ a relatively simple light penetration model:”.

Our model choice is guided by the data, given our approach uses a single attenuation coefficient, which is what is provided in the dataset. That is one of the benefits of our approach: it can achieve high accuracy despite using a single attenuation coefficient.

Lines 120-121, it is recommended that the authors integrate the Appendix A and Appendix B into the main text, as the inversion algorithm is one of the most crucial factors for determining the quality of the results in this study.

We have followed the journal submission instructions, available at: https://www.earth-system-science-data.net/submission.html#manuscriptcomposition, which state:

“Additional figures, tables, as well as technical and theoretical developments which are not critical to support the conclusion of the paper, but which provide extra detail and/or support useful for experts in the field and whose inclusion in the main text would disrupt the flow of descriptions or demonstrations may be presented as appendices.”

We consider the material presented in the Appendices as technical details, which can also be found in prior publications. For convenience they are gathered here in Appendices so that experts can replicate them easily.

Equation (10) combines the residuals at individual depths with one residual for normalized water column production. However, the manuscript does not explain whether these two terms are weighted or scaled before they are combined. Since they may differ in magnitude, their relative influence on the optimization should be clarified. Lines 337-343 state that the water column constraint improves convergence, but the authors should also explain how this term affects the estimated parameters.

This is an integral constraint on the optimization procedure and as such helps the optimization algorithm to narrow its search space. The watercolumn production term in the optimization function penalises the search algorithm if it deviates significantly. Upon testing the inverse model with this formulation, we have found that it helps the model to converge sooner and prevents blowup in parameter values. The inverse model without it also achieves high accuracy and the estimated parameter values do not differ significantly. However, this approach adds to the speed at which the optimization method converges, therefore helping to scale the inverse model to handle large datasets such as the Mattei and Scardi (2021) dataset used here.

Production profiles dataset

Lines 197-200 report 6084 profiles, including 5578 from the Northern Hemisphere and 478 from the Southern Hemisphere. Mattei and Scardi (2021) additionally classify 28 profiles as Equator, which completes the total of 6084. Please include this category explicitly.

Revised as suggested.

In addition, the manuscript reports 37722 observations in Line 198 but 37723 in Line 321. These numbers should also be verified and kept consistent throughout the manuscript.

Verified and revised as suggested.

Lines 201-209, some of the input data were filled using satellite products or interpolation rather than direct measurements. Since these data are used to estimate the photosynthesis parameters, it would be helpful to discuss how they may affect the final results. The authors are encouraged to compare the estimated parameters obtained from fully measured profiles with those containing gap filled data. In addition, including the data source and gap filling information for each published record would improve the usefulness of the dataset.

The gap filling was done already in the Mattei and Scardi (2021) paper and not in this paper. Mattei and Scardi (2021) provide details on the data sources and gap filling information in their section "2.1 Data merging and reconstruction". This is now stated in the manuscript. Also, a paragraph is now added to Section 2 (lines 158 - 169) on the topic of parameter estimation with, or without, information on surface irradiance and the attenuation coefficient.

Lines 246-254, the requirement of at least five measurement depths needs further justification. According to the manuscript, six depths are considered sufficient, while five depths represent the minimum requirement, but no quantitative evidence is provided to support this conclusion. I encourage the authors to include a simple sensitivity analysis showing how the estimated parameters change with different numbers of measurement depths.

The effect of the measurement depths on the estimation of parameter values was discussed at length in Kovač et al. (2016) ICES JMS. To estimate the assimilation number, incubations close to the surface are more informative, since production is mostly light saturated close to the surface. However, to estimate the initial slope, incubations at depth are needed, given production is light limited at depth. Also, in the optimal function we have added a term for watercolumn production to help constrain the parameter values, which helps the optimization method when the number of depths is small. We consider making sensitivity analysis to study the effect of the number of measurement depths on the estimated parameters beyond the scope of this paper.

Results

The Figure 5 caption states that the vertical axis represents the percentage of profiles, whereas the figure appears to show the number of profiles. The authors need to correct either the caption or the axis labels.

Corrected.

Lines 315-343 describe the results as model accuracy. However, the same production profiles and normalized water column production values are used to estimate the parameters, and water column production is also included in Equation (10). Therefore, Figure 6 represents goodness of fit or internal reconstruction performance rather than independent validation. The authors should conduct cross validations to verify the performances of the inverse model.

Revised as suggested. We have now written "goodness of fit" instead of "model accuracy". Regarding cross validations: we only employ the inverse model to extract the photosynthesis parameters for a given in situ measured production profile. By the very nature of the in situ production profile measurements, the estimated parameters are suitable for a given in situ profile.

In this context our forward model is not used in its predictive capacity, but simply to estimate the optimal parameter values for a given in situ production profile. Cross validation could be done to test how well an individual profile is modelled, say if data from some depths were to be removed, but here the problem lies with individual profiles often not having a sufficient number of measurements in the vertical to do cross validation experiments on individual profiles.

The reported R2 (=71.31%) (Line 336) is insufficient to characterize model performance. The authors need to report RMSE, MAE, bias, and sample size for both production at depth and water column production, as well.

Revised as suggested. We have now included all those measures following the work of Li et al. (2022) RSE added a paragraph in the Results (line 356 - 361) . We have also added Appendix D in which we state the equations for calculating these statistical measures explicitly.

The caption of Figure 6 should state the number of profiles included and the screening or quality flag criteria applied before plotting. In addition, given the substantial overlap among points, a density-based representation may be more informative than a conventional scatter plot.

Revised as suggested. All the profiles for which the parameters have been estimated are plotted, so in total there are 3776 points on the plot. We have tested a density based plot and find it visually saturated, therefore we opted to keep the figure as is.

Discussion

The manuscript discusses the differences between parameters derived from approximately 24-hour in situ incubations and those obtained from shorter controlled light experiments. (Lines 367-381, 441-453) These differences should also be reflected in the discussion of dataset complementarity and potential applications. It should be made clear that the two datasets are not directly comparable without considering the differences in the experimental methods.

Revised as suggested. A paragraph was added in the Discussion (lines 408 - 420) on this topic.

Lines 41-413, For your information, a previous study published on the journal of Remote Sensing of Environment (https://doi.org/10.1016/j.rse.2022.113027) has proposed a machine learning-based method to remotely estimate photosynthesis parameters.

We thank the reviewer for this reference. We have now used the same model accuracy assessment statistical measures in our work as were used in Li et al. (2022) RSE.

Conclusions

Lines 455-462, the Conclusions overstate the strength of the model evaluation. The model performance is described as "quite good" based on an of 0.7131, although the evaluation is based on the same data used for parameter estimation. Therefore, the results demonstrate internal reconstruction performance rather than independent predictive ability. This limitation should be acknowledged explicitly in the Conclusions.

Revised as suggested.

Citation: https://doi.org/10.5194/essd-2025-820-RC2