Response to reviewers

Yumeng Chen

June 2026

We thank the reviewer for their detailed comments for improving the manuscript. Below is a response letter for the reviewers' comments. The green texts show our responses to the comments. The blue texts show our proposed changes in the revised manuscript. Due to the procedure of the journal, the revised manuscript is separated from the response letter. The final revised manuscript may have adjustments compared to the proposed changes here to improve the coherence of the revised manuscript.

Reviewer 2

The paper is overall very well written and the methodology well explained. Even though it is not a breakthrough paper, as the novelty lies more in the technical implementation than in the idea itself, I consider it worth publishing with some clarifications that are hereby expressed, as it represents a solid contribution to the ocean biogeochemical data assimilation community.

The study investigates the impact of assimilating a novel satellite-derived phytoplankton carbon product from the ESA BICEP project into a global marine ecosystem model, using an ensemble data assimilation framework built on NEMO-MEDUSA coupled with the Parallel Data Assimilation Framework (PDAF). The authors compare several assimilation strategies using chlorophyll alone, carbon alone, or both simultaneously and evaluate their effects not only on phytoplankton biomass estimates but also on downstream ecosystem variables such as zooplankton, surface pCO_2, and oxygen. The main finding is that simultaneously assimilating both chlorophyll and carbon products yields more balanced and accurate estimates of phytoplankton biomass than assimilating either product alone, and produces the strongest response in ocean carbon and oxygen cycles.

The paper is within the scope of GMD. The source code and experiment setup are made available via Zenodo, which is commended and meets GMD reproducibility standards. Once the major concerns are addressed, the results are in principle sufficient to support the conclusions, with the caveat that the two-year experiment window and absence of independent validation currently limit the strength of those conclusions.

In order to facilitate the authors' revision, the concerns to be addressed before publication are reported in detail below.

Answer: We thank the reviewer for their efforts and for the positive and useful comments.

Major Concerns

1. Carbon observation-error model: not justified and not assessed.

   The carbon product does not provide its own error estimate, so the authors inflate the chlorophyll observation error by 10% as a proxy. The manuscript itself indicates this at line 518 as requiring further investigation but never returns to it. The choice of observation error directly controls the weight given to the carbon product in the DA update: an underestimated error leads to overconfident corrections; an overestimated error suppresses the carbon signal. Deferring this entirely to future work is not acceptable in a results paper. The authors should provide at minimum a sensitivity analysis or a qualitative bounding of how this assumption affects the conclusions. Additionally, the two products are not truly co-located, have different spatial resolutions, and share information from the same ocean-colour measurements, none of which is adequately accounted for the error model.

   Answer: We agree that the choice of observation errors is heuristic in this study. However, providing a sensitivity analysis or qualitative bounding of the impact on the conclusions will require multiple repeated runs of the DA system. This is computationally expensive and infeasible. Instead, we run a Desroziers diagnostic that supports our choice of observation error. The spatially and temporally averaged carbon observation error used in this study is 0.264 while the Desroziers diagnostic gives an observation error of 0.266. We will add the results from the Desroziers' diagnostic in Sec. 3.2. Because the Desroziers' diagnostic is based on linear statistical theory, it should account for the issue of co-location, and differing spatial resolutions. However, due to the limitation of the diagnostic as a global error statistic, further investigation of observation errors will be carried out in futuer studies.

   The choice of the observation error is supported by the statistics diagnostic in observation-space Desroziers et al. (2005), which is a common approach to tune observation errors based on the consistency of the linear statistical theory. The diagnostic shows that the spatial and temporal averaged carbon observation error used in this study is 0.264 while the diagnostic gives an observation error of 0.266. However, due to the limitation of the diagnostic as a global error statistic, further investigation of observation errors will be carried out in futuer studies.

2. Systematic carbon bias and the fixed C:N stoichiometry

   The negative mean misfit for carbon across all experiments indicates a systematic overestimation of phytoplankton carbon by the model relative to observations. This is a significant finding whose physical origin is not examined. In particular, MEDUSA assumes a fixed C:N stoichiometric ratio of 6.625:1, but phytoplankton C:N ratios vary considerably with nutrient availability, light conditions, and community composition. The manuscript mentions this fixed ratio only in the forward-looking sentence comparing MEDUSA with more complex models, never as a candidate explanation for the negative misfit already present in the results. The authors should explicitly discuss whether the fixed C:N assumption contributes to this systematic bias.

   Answer: The bias exists in both phytoplankton chlorophyll and carbon. The biases could also exist in both model forecast and observations. In the model forecast, the bias could be a result of inaccurate model parameters, the bulk representation of model processes and biased initial conditions, etc. (Mamnun et al., 2025). In MEDUSA, phytoplankton is modelled by nitrogen. The fixed ratio of C:N is only applied to the model output, which is a linear operation. As the fixed ratio can be viewed as a global average, biases occur when the distribution of the spatial and temporal varying C:N ratio is not symmetric (Tanioka et al., 2022). Nonetheless, a fixed ratio cannot be expected to result in a bias of fixed sign unless it is systematically either too large or too small.

   We will add this discussion in Sec. 5.1.1 before L238.

   The bias exists in both phytoplankton chlorophyll and carbon. The biases could exist in both model forecast and observations. In model forecast, the bias could be a result of inaccurate model parameters, bulk handling of model processes and biased initial conditions, etc. (Mamnun et al., 2025). In addition to model forecast, the application of fixed C:N ratio could affect the biases. The fixed ratio can be viewed as a global average, biases could occur when the distribution of the spatially and temporally varying C:N ratio is not symmetric (Tanioka et al., 2022).

3. Statistical significance of RMSD differences and the two-year experiment window

   RMSD reductions across experiments are reported without any assessment of statistical significance. The experiments cover only 2015 and 2016: a very short period for a global ocean biogeochemical study, and one that encompasses the tail of the strongest El Niño on record. This is never mentioned as a limitation despite the fact that anomalous phytoplankton dynamics in the eastern equatorial Pacific during this period are likely to influence several of the largest reported DA adjustments. The authors should provide confidence intervals or significance tests for the RMSD differences, or explicitly acknowledge the short experiment window and ENSO conditions as limitations.

   Answer: The two-year assimilation period is used for various other global ocean biogeochemical studies (Pradhan et al., 2019, 2020; Ford, 2021). It covers two full seasonal cycles and hence cannot be considered as ‘very short’, with the computational cost associated with running the ensemble methods prohibiting longer simulations. The limitation lies in the carbon product, which is only available monthly. We compute the 95% confidence interval of the monthly normalised RMSD over the experiment period as shown below. These results do not change our results qualitatively but we will add in Sect. 5.1.2 that Monthly C, Monthly C+ and Monthly C & Chl experiments have similar carbon RMSD reduction, and Monthly C & Chl shows lower chlorophyll RMSD reduction than Daily Chl and Monthly Chl+.

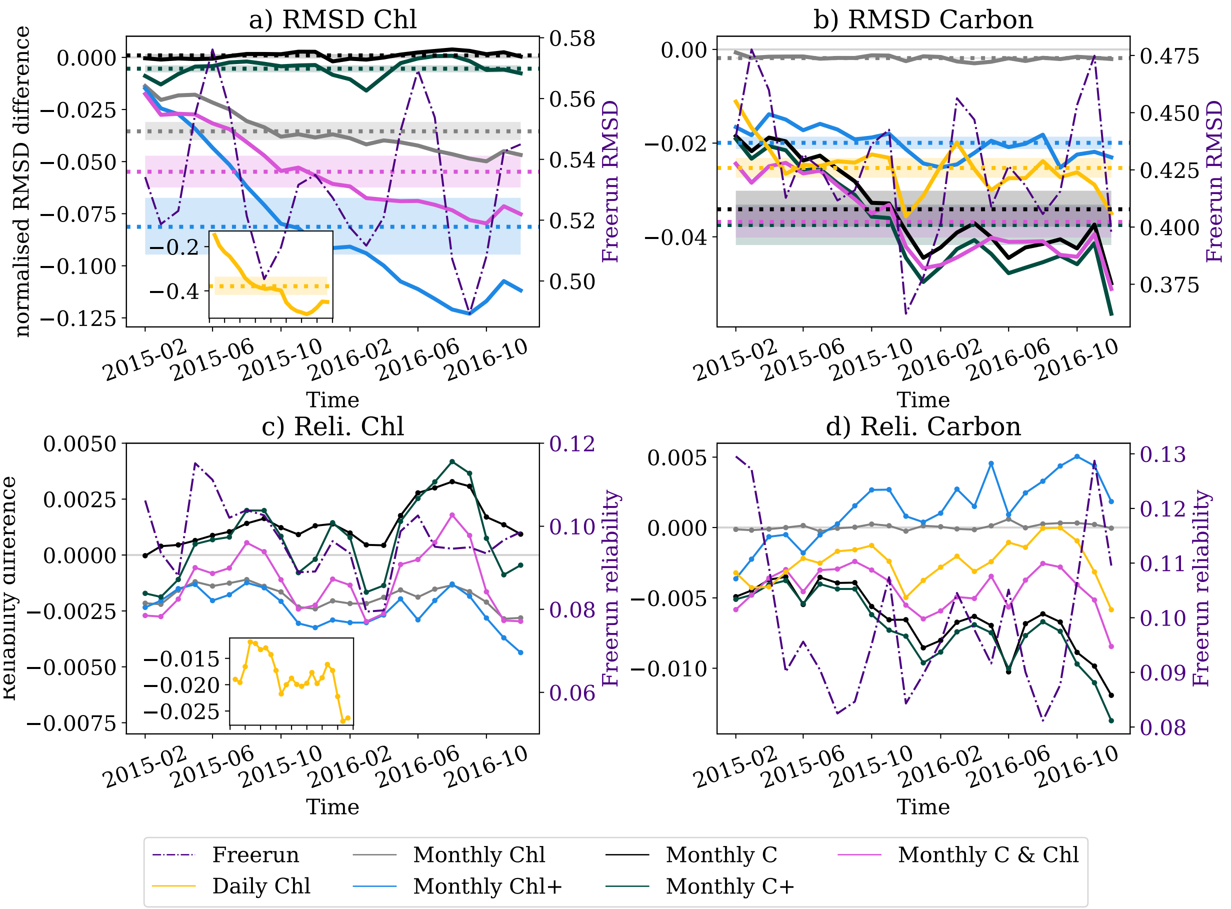

Figure 1: Upper row: time series of the RMSD of the ensemble mean of the logarithmic surface phytoplankton a) chlorophyll and b) carbon. The left axis shows the differences in RMSD between DA experiments and Freerun normalised by the RMSD of Freerun, and the right axis shows the RMSD of Freerun (red line). The light gray horizontal line is the zero line representing no global adjustments compared to Freerun. The inset in a) is the RMSD difference of chlorophyll in the "Daily Chl" experiment. The dotted horizontal lines are time-averaged normalised differences, and the shaded areas are the 95% confidence interval over the experiment period. Lower row: reliability score of the ensemble of surface phytoplankton a) chlorophyll and b) nitrogen compared with the assimilated observations derived from CRPS. The left axis is for the reliability differences between each DA experiment and Freerun, and the right axis is the reliability score of Freerun (red line). The inset in c) is the differences in the reliability of chlorophyll in the "Daily Chl" experiment.

The El Niño event is expected to lead to reduced phytoplankton in eastern tropical Pacific. The modelled results in the eastern tropical Pacific are discussed in Sect. 5.2.1. We will add a discussion that results in the eastern Pacific region may not be generalised to other years without El Ni no events.

The impact of the El Niño event is planned to be added in Sect.5.2.1:

It is worth noting that a strong El Niño occurred during our experiment period which had a strong impact on the eastern Equatorial Pacific. This could imply that results in the eastern Pacific region may not be generalised to other years without El Ni no events.

4. Unexplained seasonal discrepancy with Pradhan et al.

   The authors note that RMSD peaks in boreal autumn in their system, whereas Pradhan et al. (2019) found higher errors in boreal spring, but offer no explanation. Possible explanations such as differences in biogeochemical model formulation, grid resolution, ensemble size, or forcing data, should be explored, even if only qualitatively.

   Answer: Pradhan et al. (2019) uses the MITgcm-REcoM2 model at a variable resolution from 0.38° to 2° with 20 ensemble members forced by the Coordinated Ocean-Ice Reference Experiment (CORE). Our study uses NEMO-MEDUSA on a fixed ORCA1 grid with 30 ensemble members forced by JRA-55. Furthermore, REcoM2 is a quota model using dynamically varying ratios, while MEDUSA uses fixed ratios. Moreover, as discussed in point 3, our experiments coincide with a strong El Niño event. These could all lead to different responses to seasonality.

   We will add this discussion in L250.

   However, Pradhan et al. (2019) showed a higher RMS error in the boreal autumn than in spring, whereas our results show the opposite pattern, with larger RMSD in autumn. The differences can be explained by various factors including the model formulations, forcing, and ensemble size, which cannot be easily attributed. For example, Pradhan et al. (2019) uses the MITgcm-REcoM2 model on a variable resolution from 0.38° to 2° with 20 ensemble members forced by Coordinated Ocean-Ice Reference Experiment (CORE). Our study uses NEMO-MEDUSA on a fixed ORCA1 grid with 30 ensemble members forced by JRA-55. Apart from this, REcoM2 uses a varying stoichiometry, while MEDUSA uses fixed ratios. Moreover, as discussed in point 3, our experiments coincide with a strong El Niño event. These could all lead to different responses to seasonality.

5. Ensemble reliability and its implications for DA validity

   The reliability score of the Freerun is reported to be worse than in Popov et al. (2024), indicating that the global ensemble is considerably less reliable than regional configurations. The authors indicate this but do not discuss its consequences for the validity of the DA results. If the ensemble poorly represents true system uncertainty, LESTKF corrections may be suboptimal or misleading. The authors should discuss how ensemble unreliability affects the interpretation of their results and what steps could improve it. Santana-Falcón et al. 2020, (Ocean Science: `https://os.copernicus.org/articles/16/1297/2020/`) showed that DA improvements depend strongly on prior ensemble reliability and is directly relevant to this discussion.

   Answer: Reliability is not a normalised quantity, so that one cannot directly compare the values from different studies. To this end, we remove the comparison of the reliability score. Our experiments indicate that the ensemble spread is under-dispersive. This could lead to smaller adjustments than an optimal system. The ensemble reliability can benefit from perturbations in atmospheric forcing and spatially and temporally varying parameters. However, this will require further careful tuning and possible modifications to the model itself.

   We will add the following in Sec. 5.1.3.

   Fig. 3c)-d) focuses on the reliability score under the assumption that the climatology distributions are given by observations. The non-zero reliability score shows that the ensemble is not perfectly reliable. This highlights the challenge of achieving a reliable global ensemble system and underscores the need for further investigation into quantifying uncertainty in MEDUSA. The ensemble used in this study is in general under-dispersive. This could lead to smaller adjustments than an optimal system. The ensemble reliability can be benefited by perturbations in atmosphere forcing and spatially and temporally-dependent parameters. However, this will require further careful tuning and possible modifications to the model itself. Such a challenge is also investigated by Anugerahanti et al. (2018) and Mamnun et al. (2022).

6. Surface-only constraint: subsurface structure is unconstrained

   The DA system assimilates only surface satellite observations, leaving the subsurface phytoplankton structure entirely unconstrained. This is never explicitly justified or discussed as a limitation. Arteaga et al. 2022, (Global Biogeochemical Cycles: `https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022GB007389`) showed that surface-only retrievals systematically misrepresent subsurface productivity, particularly in the Southern Ocean. The manuscript should acknowledge surface-only assimilation as a known limitation.

   Answer: Most operational systems still only assimilate satellite chlorophyll. Hence, the surface-only constraint is a challenge faced by most operational systems and is not unique to this study (Skákala et al., 2025).

   To avoid updating purely the surface phytoplankton, our study follows the practice in some operational systems by updating the chlorophyll and carbon throughout the ocean mixed layer (Ford, 2021) as discussed in L144 in Sect. 3.1. This treatment could reflect the impact of assimilating carbon product on existing DA setup. Moreover, the ensemble system has sub-surface information from the model dynamics. For example, the deep chlorophyll maximum can be updated by related covariances from the ensemble (Pradhan et al., 2020), which is not used in this study.

   In L26 of introduction, we will add the reference.

   One limitation of satellite ocean colour is its inability to differentiate the vertical structure of the euphotic zone near the ocean surface (Arteaga et al., 2022).

   In L110 in Sect. 2.2, we will rephrase:

   In this study, two datasets are assimilated into NEMO-MEDUSA: a new monthly phytoplankton carbon product and a chlorophyll-*a* dataset for reference, which follows the common practice of operational DA systems (Skákala et al., 2025). We do not assimilate additional observations for improved vertical profiles.

   In conclusion, we will add:

   Furthermore, this study follows the common practice of operational DA systems by only assimilating satellite ocean colour (Skákala et al., 2025). This means that the update of subsurface structure is not as accurate as assimilating profiles of marine ecosystem variables such as BGC-Argo data. It is of interest to investigate the update by ensemble covariance in MEDUSA similar to Pradhan et al. (2020), or a DA system that assimilates both vertical profiles and ocean colour data.

7. Absence of independent validation

   All evaluation metrics in the paper are computed against the same products that were assimilated chlorophyll and carbon from satellite, which means the reported RMSD reductions measure how well the system fits its training data, not whether the model state has become more realistic. This limitation is acknowledged only in the final sentence of the manuscript. It should be

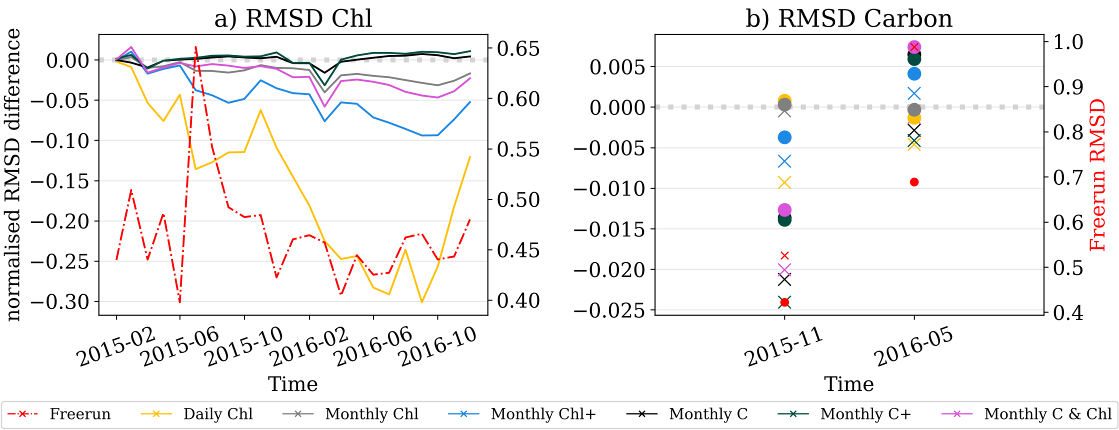

Figure 2: Differences of the RMSD between DA experiments and Freerun normalised by the RMSD of Freerun between the logarithm of monthly model forecast of and composite of the chlorophyll from Argo and field campaign data from NAAMES in left axis. The RMSD of Freerun is shown in dashed red line by right axis. The cross points represent comparisons with modelled total carbon while the dot points represent comparisons with modelled non-diatom carbon.

discussed in the conclusions, alongside identification of independent datasets that could be used in future work: BGC-Argo profiles for subsurface chlorophyll and carbon, and SOCAT for surface $p\mathrm{CO}_2$ are examples of candidates. The observations section (Section 2.2) should also be restructured to clearly separate datasets used for assimilation and experiment setup sections from datasets used for validation.

Answer: We agree that this is a limitation of this work. In the revised manuscript, we will add an assessment of the state from data assimilation using BGC-Argo data for phytoplankton chlorophyll and in situ measurements from the North Atlantic Aerosols and Marine Ecosystems Study (NAAMES) for phytoplankton carbon assessment.

In Sect. 2.2, we will add a new subsection describing the independent observations.

Independent observations

To assess the DA system, we use in situ observations of chlorophyll and carbon data.

The in situ chlorophyll data are obtained from the BGC-Argo program (Group, 2016; Roemmich et al., 2019). The BGC-Argo program extends the Argo program, which is an international program that deploys automatic instruments floating with ocean currents. The BGC-Argo floats provide vertical profiles for both physical variables such as temperature, salinity, and pressure, and biogeochemical variables. In this study, only quality-controlled measurements with delayed-mode corrections from the upper 50 m were used, focusing the evaluation on the near-surface productive layer and the depth range most directly influenced by the DA. During the experiment period, the dataset has, on average, 448 observations per month.

The in situ observations of carbon is less common compared to chlorophyll. Here, we use the analytical measurements of phytoplankton carbon from North Atlantic Aerosols and Marine Ecosystems Study (NAAMES). The projects provide ship-based measurements for plankton stocks, rate processes, and community compositions. These in situ carbon observations are analytically determined for cells less than 64 μm using a BD Influx Flow Cytometer using methods detailed by Graff et al. (2015). Due to the limitation of the field campaign, the in situ datasets is limited to Northwest Atlantic and weeks of data in November 2015 and May 2016.

In Sect. 5.1.2, we will add:

We first compare our results with independent in situ observations in Fig. 2 with the root mean squared distance of the logarithm of chlorophyll and carbon between model forecast and the observations. The evaluation is performed on observation locations by linear interpolation from model grid points. For convenience, the point observations are considered as monthly averaged values, which could cause additional errors in this evaluation. The DA experiments show improvements compared to Freerun experiment as indicated by the negative values. All experiments assimilating chlorophyll show reduced RMSD compared to Freerun for chlorophyll. Experiments assimilating carbon show slightly increased RMSD of the chlorophyll. Because the carbon observations include only cells below 64 μm, we compare the observations with non-diatom phytoplankton in MEDUSA, which is considered small phytoplankton. assimilating chlorophyll has little impact on the RMSD of non-diatom carbon. The RMSD of non-diatom carbon (dot sign in Fig 2b)) is reduced by assimilating carbon in November 2015 but the RMSD is increased in May 2016. The increased RMSD by assimilating carbon in the RMSD of non-diatom carbon could be a result of post-processing and mismatch between the phytoplankton functional types, especially considering the seasonal dependence of the results. When comparing with total modelled carbon (cross sign in Fig 2b)), assimilating carbon better constrains the carbon than assimilating chlorophyll. For both carbon and chlorophyll, Daily Chl shows better performance than monthly assimilations. For monthly assimilation, simultaneous assimilation of chlorophyll and carbon provides a balanced RMSD compared to assimilating a single type of phytoplankton composition.

The assessment against in situ data is limited by the spatial and temporal coverage. Hence, we further evaluate the results by assimilated satellite data.

We will add potential validation datasets for other variables at the end of the conclusion.

Moreover, independent observations will be needed for quantitatively validating the DA system in addition to phytoplankton chlorophyll and carbon. For example, using BGC-Argo floats, Surface Ocean CO_2 Atlas (SOCAT) or World Ocean Atlas datasets for nutrients, oxygen, and carbonate variables.

8. Seasonality deterioration from chlorophyll-only assimilation

   The finding that assimilating chlorophyll alone with post-processing can deteriorate the seasonality of modelled global phytoplankton is currently buried in the seasonality section. Given that most operational biogeochemical DA systems rely exclusively on chlorophyll assimilation, this negative result has direct practical implications for the community and should be elevated, or at least acknowledged, in both results and conclusions.

   Answer: We will add a discussion on this in the conclusions, but we will limit this to the specific modelling system because our results may not be generalised to other models. For example, we see a seasonal discrepancy with Pradhan et al. (2019).

   In L491 of the conclusion, we will add

   However, these adjustments to variables other than the assimilated product could also lead to increased differences with observations and can deteriorate the seasonal variation in some regions. For example, in Monthly Chl+, the seasonality of phytoplankton carbon could deteriorate. This finding could suggest deteriorated seasonality for other DA systems assimilating chlorophyll alone, but, considering the seasonal discrepancy of RMSD with Pradhan et al. (2019), this could also be model and period dependent, which requires further investigation.

9. pCO_2 degradation in the Daily Chl experiment

   The Daily Chl experiment, which most closely resembles current operational biogeochemical DA systems, produces increased pCO_2 growth relative to Freerun, attributed to strong phytoplankton reduction in the eastern equatorial Pacific and increased zooplankton respiration. This is a potentially important and worrying result suggesting that high-frequency chlorophyll-only assimilation, as currently practised operationally, may degrade the carbon cycle representation. This finding deserves deeper discussion, including a mechanistic explanation and an assessment of its implications for operational systems. Similarly, the mechanism behind the highest oxygen increase in the Monthly Chl & C experiment is not explained and should be addressed.

   Answer: The Daily Chl experiment does not closely resemble the operational systems because it does not update phytoplankton carbon and other variables as done in operational systems. Figures 6 and 10 show similar spatial patterns between Daily Chl and Monthly Chl experiments, while Monthly Chl+ shows different patterns. This suggests that the degradation is mostly because phytoplankton carbon is not updated, which is performed by balancing schemes in operational systems. The degradation is less pronounced for Monthly Chl because of the infrequent update of chlorophyll. As Monthly Chl+ is more consistent with operational systems, our results do not imply operational system may degrade carbon cycle representation.

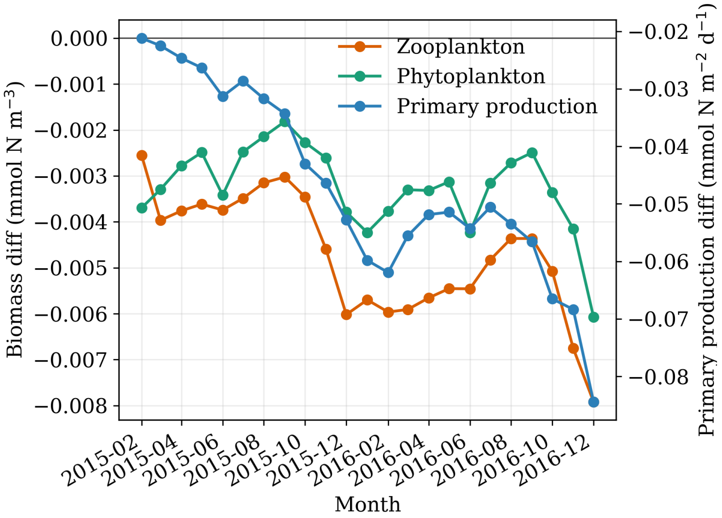

Figure 3: The time series of the differences of the ensemble mean of the spatially averaged phytoplankton carbon, zooplankton and primary production between the Monthly Chl & C and Freerun.

In these experiments, phytoplankton, primary production and zooplankton all decreased compared to the Freerun. Hence, the increased oxygen is likely a result of the reduction in oxygen consumption by zooplankton exceeds the reduction in oxygen production by photosynthesis. This is consistent with Fig. 10 where the zooplankton is reduced.

We will add this explanation in L471:

The increased oxygen in these experiments is a result of less consumption of oxygen from the reduced zooplankton. This reduction of oxygen consumption exceeds the reduction of oxygen generation from photosynthesis due to reduced phytoplankton.

10. One-way coupling, reduced state vector, and post-processing redistribution

    These are reasonable first-implementation choices but they limit how far improved skill can be attributed specifically to the information content of the carbon product rather than to the mapping and redistribution scheme. The manuscript does not discuss these structural limitations critically enough. In particular, the proportionality assumption underlying Equation 1 where DA increments are distributed to diatom and non-diatom PFTs in proportion to their forecast fractions assumes that all PFTs respond proportionally to a perturbation, which is not generally true. Pradhan et al. (2020), already cited in this manuscript, demonstrated that PFTs do not always change proportionally, and this limitation should be acknowledged clearly in the methods.

    Answer: We agree that marine biogeochemical processes can have significant impact on physics (Manizza et al., 2005; Skákala et al., 2022). However, for operational systems that perform data assimilation, it is still uncommon to use two-way coupled models (Gehlen et al., 2015; Yumruktepe et al., 2022; Polton et al., 2023). We adjust the introduction and acknowledge that this could limit the accuracy of the model forecast.

    The coupling to NEMO is one-way; that is, the marine ecosystem model is forced by the physical ocean, but has no feedback from the ecosystem to the physics. Although marine biogeochemistry processes can have significant impact on physics (Manizza et al., 2005; Skákala et al., 2022), the models in current data assimilation systems are still primarily one-way coupled (Gehlen et al., 2015; Yumruktepe et al., 2022; Polton et al., 2023). Hence, we acknowledge that the one-way coupling can limit the accuracy of the model forecast. But following the common practice of using one-way coupled model avoids the need to investigate the impact of two-way coupling is outside the scope of this work.

    We will add the acknowledgement of the limitation of the post-processing scheme and the benefits of using the ensemble covariance in L145 of Sect.3.1.

    Although, in this system, we can technically update non-assimilated variables by ensemble covariance as demonstrated by Pradhan et al. (2020), updating by ensemble covariance requires tuning of ensemble perturbations for ensemble cross-covariance, and were not investigated for MEDUSA. Investigating the update by ensemble covariance deviates from our goal of investigating the impact of the carbon product on the DA system. Hence, even though heuristic and potentially less optimal than relying on ensemble covariance, this work follows the practice in variational operational DA systems where increments in total phytoplankton are distributed into diatom and non-diatom PFTs by post-processing following Ford (2021). Consequently, the impact of DA on PFT is more than a result of the observation product. They also reflect the imposed post-processing assumptions.

11. Cross-product inconsistency

    The combined assimilation does not bring the averaged misfits closer to zero, and the chlorophyll:carbon ratio in observations differs from the model. The joint-assimilation result should be framed explicitly as a compromise under imperfect product consistency, not as a solution.

    Answer: The non-zero misfits could be a result of either model errors or inconsistencies between the observational products. The carbon data are derived based on chlorophyll data and their ratio is validated by Sathyendranath et al. (2020). Hence, it is likely that the inconsistency arises from model errors.

    To further illustrate this, we will add the following in L241 in Sect. 5.1.1.

    ... may struggle to correctly handle corrections for both chlorophyll and carbon. Considering the carbon data are derived based on chlorophyll data and their ratio is validated by Sathyendranath et al. (2020), the inconsistency of different signs may primarily arise from model errors. The assimilation of both products find a compromise between biased model and observation products.

12. Speculative claims about complex models

    Lines 503–506 assert that more complex models such as ERSEM or REcoM2 would likely produce more robust DA responses to perturbations in carbon or chlorophyll. This claim is unsupported by evidence or references and should be reframed as a hypothesis for future work. If the authors wish to maintain it, they should cite studies comparing DA performance across models of different complexity.

    Answer: There are comparisons of different models by assimilating chlorophyll for regional modelling but they do not investigate a carbon product (Ciavatta et al., 2025). We will rephrase the sentence to show it is a hypothesis for future work.

    In MEDUSA, carbon and nitrogen are assumed to have a fixed stoichiometric relationship. By contrast, more complex models such as ERSEM (Butenschön et al., 2016) or quota-based models like REcoM2 (Schourup-Kristensen et al., 2014) may exhibit distinct responses to perturbations in carbon or chlorophyll. For example, Ciavatta et al. (2025) compared responses of assimilating chlorophyll in models of different complexities. It is of interest to make such comparisons for carbon products. We hypothesise that in more complex models, where the representation of carbon and chlorophyll dynamics is more sophisticated, the DA is likely to produce more robust ecosystem responses.

13. Model spinup

    The coupled NEMO-MEDUSA model is spun up for only 15 years prior to the assimilation experiments. While this may be sufficient to equilibrate surface biogeochemical variables, it is generally considered inadequate for full equilibration of deeper ocean carbon cycle variables such as dissolved inorganic carbon and alkalinity, which can require centuries to millennia to reach steady state. The authors should acknowledge this as a limitation and discuss whether residual drift in the deeper carbon cycle variables could influence the surface pCO_2 and oxygen results presented in Section 5.4.

    Answer: We acknowledge that the deep ocean carbon cycle could affect the surface pCO_2 and oxygen by large-scale circulation. However, these processes typically operate on timescales longer than the two-year experimental period. Furthermore, since our DA experiments are compared with FreeRun, they are subject to the same deep ocean conditions. It is unlikely that this could affect our comparisons between different experiments.

    We will add a discussion in L183 of Sect. 4.

    The coupled NEMO-MEDUSA model is spun up without DA for a period of 15 years to create the initial state before applying the ensemble perturbation. The 15-year spinup is sufficient to equilibrate surface biogeochemical variables, but the deeper ocean variables may not be able to reach a steady state within the spinup period. However, all of our experiments are performed under the same deep ocean conditions which should therefore have limited impact on our comparisons between experiments.

Minor Concerns

1. The abstract does not mention the two key negative results identified in this review: the deterioration of phytoplankton seasonality from chlorophyll-only assimilation, and the increased pCO_2 growth in the Daily Chl experiment. Given their practical implications for operational systems, at least a brief acknowledgement of these findings should be included in the abstract.

   Answer: As discussed in point 9 of the major concerns, we are convinced that the pCO_2 degradation is not an important result. But we will report the deterioration in seasonality in Monthly Chl+ in the abstract.

   In L10, we will add:

   We demonstrate that, compared with assimilating only a chlorophyll product, which may degrade the seasonality of phytoplankton carbon, with the new carbon product the DA can...

2. In line 90: the sentence “In MEDUSA, the primary currency of nitrogen drives the model evolution” is unclear. The authors should clarify what is meant and explain more transparently how chlorophyll is diagnosed from nitrogen via a space- and time-dependent scaling factor. As written, this passage is likely to confuse readers unfamiliar with the specific MEDUSA formulation.

   Answer: We will rephrase the sentence:

   In MEDUSA, the phytoplankton biomass is represented by nitrogen. The phytoplankton chlorophyll is also explicitly modelled, influenced by the primary production of the phytoplankton biomass represented by nitrogen.

3. The relationship between PDAF and LESTKF should be stated more explicitly. The authors should clarify that PDAF is the computational framework within which LESTKF operates as the specific ensemble DA algorithm.

   Answer: We will adjust the sentence in L106 to:

   This study adopts the local error subspace transform Kalman filter (LESTKF, Nerger et al., 2012), which is a specific ensemble DA algorithm provided by the computational framework of PDAF.

4. The authors are encouraged to summarise the experimental design in a table, listing for each experiment: the assimilated variable(s), the model variables directly updated, whether post-processing is applied, and the temporal frequency of assimilation.

   Answer:

Table 1: Summary of the data assimilation experiments. “Directly updated” refers to variables modified by the data assimilation analysis, whereas post-processing refers to the additional increment-based adjustment defined in Eq. (1).

Experiment	Observation(s)	Model variable(s) directly updated	Post-processing	Frequency
Freerun	None	None	No	None
Daily Chl	Chlorophyll	Chlorophyll	No	Daily
Monthly Chl	Chlorophyll	Chlorophyll	No	Monthly
Monthly Chl+	Chlorophyll	Chlorophyll	Yes	Monthly
Monthly C	Carbon	Carbon	No	Monthly
Monthly C+	Carbon	Carbon	Yes	Monthly
Monthly Chl & C	Chlorophyll & carbon	Chlorophyll & carbon	No	Monthly

5. Section 5 states that pCO_2 and oxygen are evaluated "without available observations." Observational products for both variables exist for example SOCAT for surface pCO_2, and World Ocean Atlas or BGC-Argo for oxygen. The correct statement is that the authors chose not to validate against these datasets in the present study.

   Answer: We rephrase L210 to:

   We further assess the effects of each experiment on marine ecosystem variables such as PFTs, zooplankton and gases such as pCO_2 and oxygen without validating these variables against in situ datasets in the present study.

6. The citation of Falkowski, even if broadly still used should be updated with more recent ones. In this sense, the statement that phytoplankton 'account for 40% of global carbon uptake' is broadly correct but ambiguous. The authors should rephrase to refer explicitly to global primary production or carbon fixation rather than carbon uptake and update the refrences.

   Answer: We will rephrase the sentence with more recent references:

   In particular, as the base of the food web, phytoplankton sequestrate carbon and emit oxygen, accounting for 50% of total global net primary production estimated on the order of 50 Gt C yr^{-1} (Falkowski, 1994; Buitenhuis et al., 2013; Johnson and Bif, 2021; Ryan-Keogh et al., 2023).

7. Consider adding Skákala et al. 2024, (Progress in Oceanography: `https://www.sciencedirect.com/science/article/pii/S0079661124000557`) to the motivation, as they demonstrated that phytoplankton community composition is the most uncertain and least observable marine ecosystem indicator when inferred from surface chlorophyll alone, which supports the argument for assimilating additional ocean-colour products such as phytoplankton carbon. Add the suggested papers, if necesarry to the reference list

   Answer: We will add this in L46:

   Moreover, Skákala et al. (2024) shows that phytoplankton community compositions is the most uncertain and least observable variables in the marine ecosystem.

Other minor suggestions

1. The authors are encouraged to clarify the distinction between "marine ecosystem model" and "ocean biogeochemical model", which are used somewhat interchangeably throughout. While both refer to MEDUSA, they carry different conceptual emphases: the former typically refers to biological components such as phytoplankton, zooplankton, nutrients and detritus, whereas the latter emphasises chemical cycles including carbon, oxygen and alkalinity. Since MEDUSA encompasses both, the authors should acknowledge this explicitly and be consistent in their use of terminology thereafter.

   Answer: We will change "marine biogeochemical model" to "marine ecosystem model" in the manuscript.

2. The term "primary currency" (line 90) should be replaced with a clearer expression such as "master prognostic variable", or simply clarified that the model is nitrogen-based with all other biogeochemical variables derived from or scaled to nitrogen.

   Answer: We rephrase the sentence:

   The model includes nutrients, phytoplankton, zooplankton and detritus to simulate marine nitrogen, silicon, iron, alkalinity and oxygen cycles with a benthic ecosystem for seafloor organic pools. In the model, the evolution of phytoplankton concentration is primarily represented in terms of nitrogen biomass.

3. The authors should be more precise in their use of "chlorophyll" versus "chlorophyll-a". The OC-CCI product assimilated is specifically chlorophyll-a, and in the text the generic term "chlorophyll" is systematically used.

   Answer: The term "chlorophyll" is used interchangeably with "chlorophyll-a" in ocean biogeochemical modelling.

   We will add in L25:

   The popular use of chlorophyll-*a* is because it is widely used as a proxy for phytoplankton biomass in ocean ecosystem modelling. For brevity, chlorophyll-*a* and chlorophyll are used interchangeably in this study.

4. The word "analysis" is used in two distinct senses throughout: its DA meaning as the corrected model state after assimilation and its general scientific meaning, as in "data analysis". This dual use could confuse readers unfamiliar with DA terminology. The authors should disambiguate where necessary, adopting "DA analysis" or "analysed state"

   Answer: Among all occurrences of the word "analysis", "data analysis" is only used in the code and data availability section. Therefore, we believe the term "analysis" is used exclusively for "DA analysis" in the manuscript.

5. The term "balancing scheme" is used repeatedly but never explicitly defined. The authors should provide a concise self-contained definition upon first use, clarifying that it refers to a method for redistributing DA increments from an observed variable into other unobserved model variables based on assumed stoichiometric relationships.

   Answer: We will rephrase the following sentence in L37:

Hence, operational marine biogeochemical DA requires balancing schemes, which distribute DA increments of chlorophyll to other model variables for a consistent model state across variables, e.g., different phytoplankton compositions and nutrients, after the DA update (Hemmings et al., 2008), or rely on estimates of error covariances from an ensemble of model forecasts (Pradhan et al., 2020).

6. The term “benchmark” (line 120) is slightly misleading. The chlorophyll product is not merely a passive reference, it is itself an active assimilation product being tested and compared throughout the study. The authors should replace it with “reference experiment” or similar.

   Answer: Changed.

Figures

The authors are encouraged to revise all figures to improve readability and quality. Specific comments follow:

1. Figure 1: The boxes and circles should be made larger, with larger text inside them. Several acronyms used in the figure : PDAF and LESTKF are not defined in the caption, forcing the reader to consult the main text. These should be spelled out in the caption so the figure is self-contained. Additionally, the phrase “resulting analysis” in the caption is ambiguous and should be clarified explicitly as referring to the reconstructed state used to reinitialise the following ensemble forecast.

   Answer: Changed

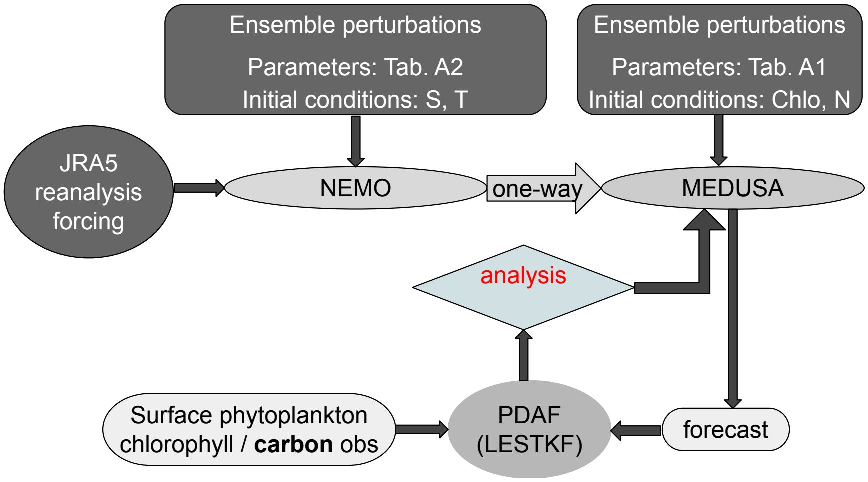

Figure 4: An illustration of the data assimilation system where an ensemble of forecasts is obtained from the one-way coupled NEMO-MEDUSA forced by JRA-55 reanalysis whereas the satellite surface chlorophyll and carbon observations are assimilated. The analysis (reconstructed state) is used to initialise the following ensemble forecast. Perturbed parameters are listed in Tab. A2 for MEDUSA and Tab. A1 for NEMO. Here, Parallel Data Assimilation Framework (PDAF) is the computational framework and local error subspace transform Kalman filter (LESTKF) is the DA algorithm provided by PDAF.

2. Figure 2: The line width of all curves should be increased to improve visibility.

   Answer: Changed.

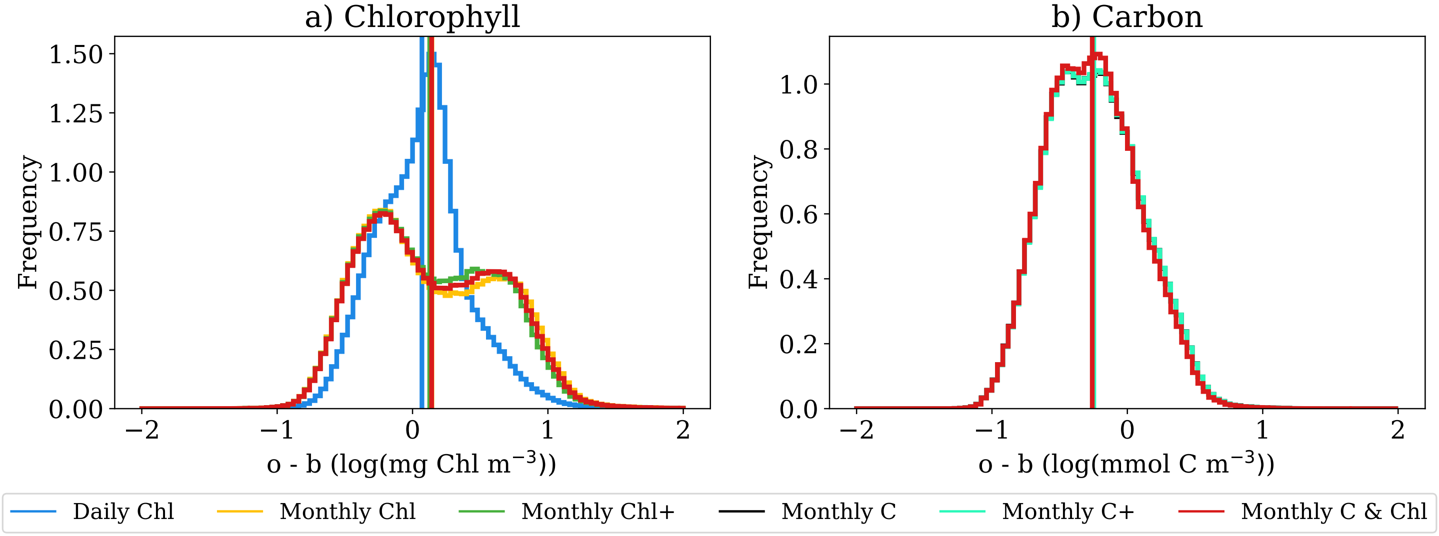

Figure 5: The histogram of the misfit between observations and forecast for log-transformed variables at analysis step. High observation and forecast misfits with low probability in a) are neglected for display purposes. The mean of the misfits for each experiment is presented as vertical lines. This histogram is constructed from all instantaneous global observations assimilated over the entire experimental period.

3. Figure 3: The x-axis date labels overlap and should be adjusted, for example by rotating them or reducing their frequency. The legend is too small and should be enlarged. The use of red for the right-hand axis is discouraged as it is not accessible to colour-blind readers and should be replaced with a colour-blind friendly alternative throughout.

   Answer: Changed.

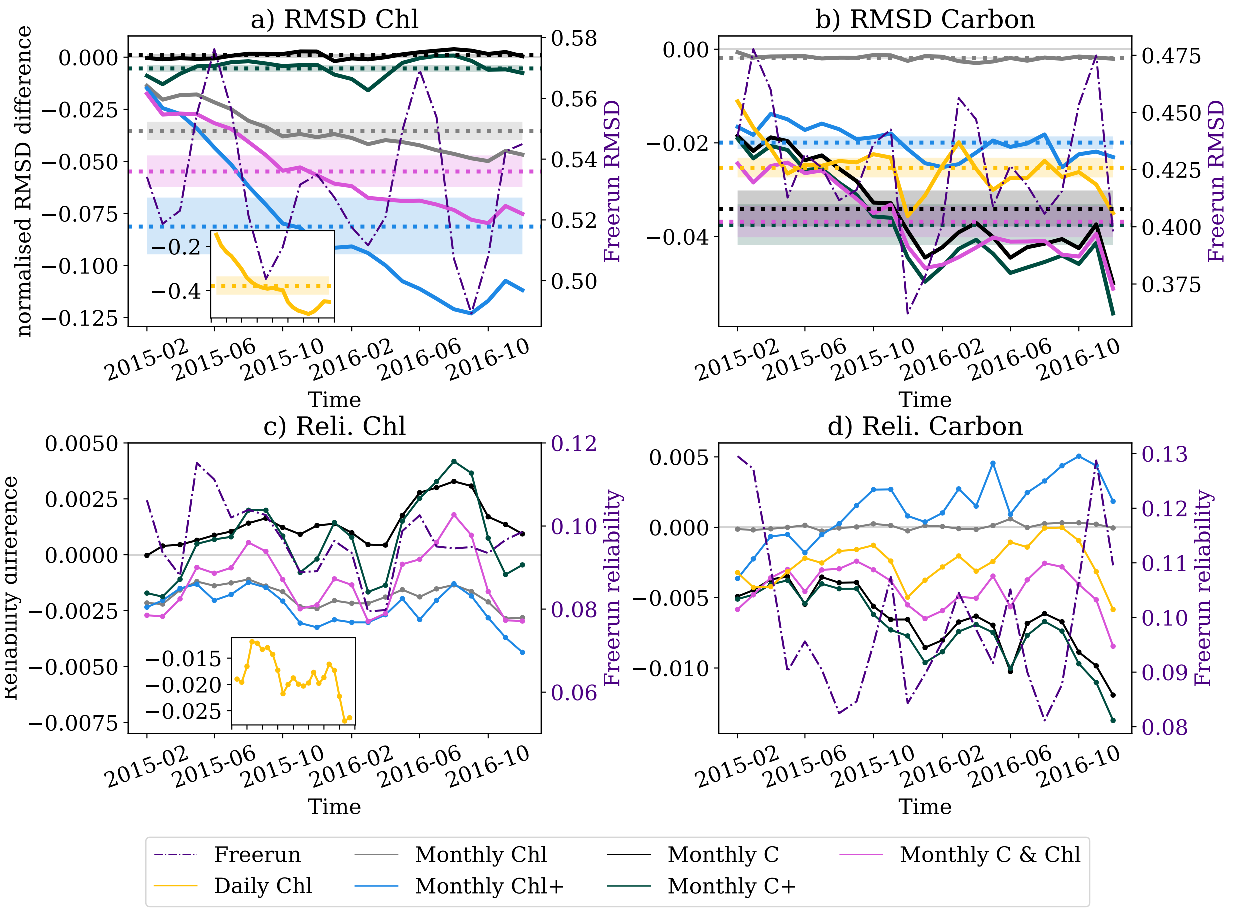

Figure 6: Upper row: time series of the RMSD of the ensemble mean of the logarithmic surface phytoplankton a) chlorophyll and b) carbon. The left axis shows the differences in RMSD between DA experiments and Freerun normalised by the RMSD of Freerun, and the right axis shows the RMSD of Freerun (red line). The light gray horizontal line is the zero line representing no global adjustments compared to Freerun. The inset in a) is the RMSD difference of chlorophyll in the “Daily Chl” experiment. The dotted horizontal lines are time-averaged normalised differences, and the shaded areas are the 95% confidence interval over the experiment period. Lower row: reliability score of the ensemble of surface phytoplankton a) chlorophyll and b) nitrogen compared with the assimilated observations derived from CRPS. The left axis is for the reliability differences between each DA experiment and Freerun, and the right axis is the reliability score of Freerun (red line). The inset in c) is the differences in the reliability of chlorophyll in the “Daily Chl” experiment.

4. Figure 4: The use of red and green is not accessible to colour-blind readers and should be replaced with a more inclusive colour palette. Font sizes for all labels should be increased. The caption should be simplified by structuring it as ”First row… second row…” rather than the current format.

   Answer: Changed.

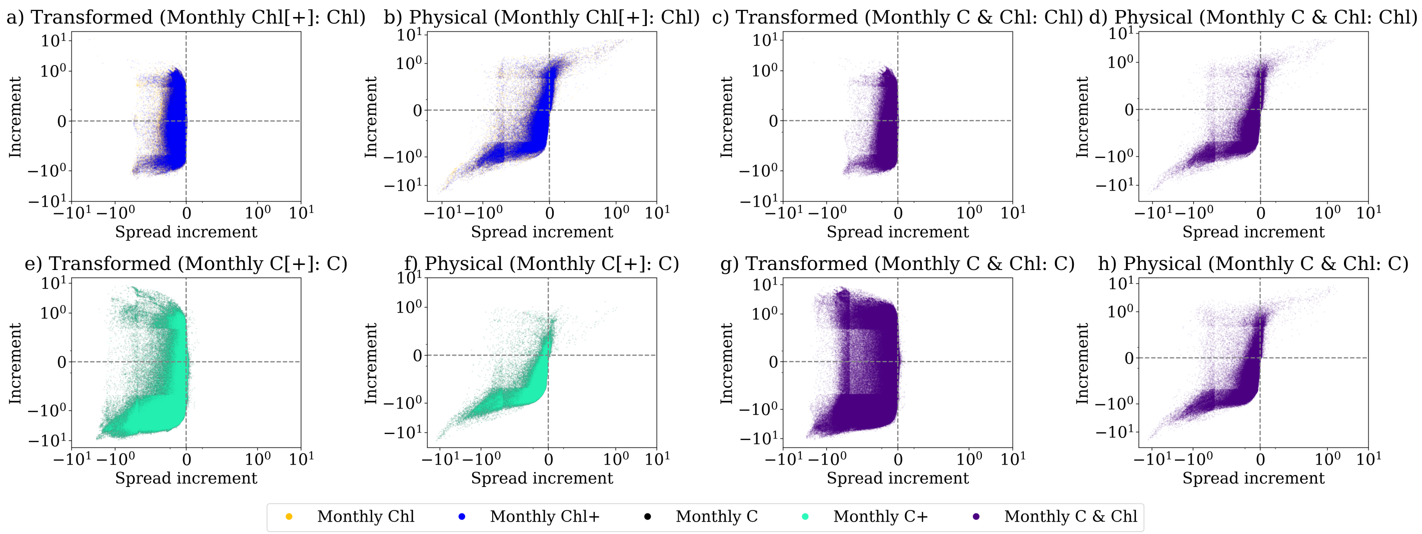

Figure 7: Scatter plot of the ensemble mean increment against the ensemble spread increment in log-scale. Here, the LESTKF is performed on transformed variables following the Gaussian distribution, and physical values are the ensemble mean and standard deviation of the phytoplankton concentrations. First row: Chlorophyll increment and spread increment; second row: carbon increment and spread increment. The “Monthly Chl[+]” represents both “Monthly Chl” and “Monthly Chl+” experiments, and “Monthly C[+]” represents both “Monthly C” and “Monthly C+” experiments; the variable is given after the colon.

5. Figures 5, 6 and 7: The captions are unnecessarily long and repetitive. In particular, the sentence describing how differences are computed appears verbatim across all three figures.

   Answer: We will remove repetitive description in Figs. 6 and 7 by saying that they are similar to Fig 5, but for different variables.

References

Anugerahanti, P., Roy, S., and Haines, K.: A perturbed biogeochemistry model ensemble evaluated against in situ and satellite observations, Biogeosciences, 15, 6685–6711, https://doi.org/10.5194/bg-15-6685-2018, 2018.

Arteaga, L. A., Behrenfeld, M. J., Boss, E., and Westberry, T. K.: Vertical Structure in Phytoplankton Growth and Productivity Inferred From Biogeochemical-Argo Floats and the Carbon-Based Productivity Model, Global Biogeochemical Cycles, 36, e2022GB007 389, https://doi.org/https://doi.org/10.1029/2022GB007389, e2022GB007389 2022GB007389, 2022.

Buitenhuis, E. T., Hashioka, T., and Quéré, C. L.: Combined constraints on global ocean primary production using observations and models, Global Biogeochemical Cycles, 27, 847–858, https://doi.org/https://doi.org/10.1002/gbc.20074, 2013.

Butenschön, M., Clark, J., Aldridge, J. N., Allen, J. I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., Lessin, G., van Leeuwen, S., van der Molen, J., de Mora, L., Polimene, L., Sailley, S., Stephens, N., and Torres, R.: ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels, Geoscientific Model Development, 9, 1293–1339, https://doi.org/10.5194/gmd-9-1293-2016, 2016.

Ciavatta, S., Lazzari, P., Álvarez, E., Bertino, L., Bolding, K., Bruggeman, J., Capet, A., Cossarini, G., Daryabor, F., Nerger, L., Popov, M., Skákala, J., Spada, S., Teruzzi, A., Wakamatsu, T., Yumruktepe, V., and Brasseur, P.: Control of simulated ocean ecosystem indicators by biogeochemical observations, Progress in Oceanography, 231, 103 384, https://doi.org/https://doi.org/10.1016/j.pocean.2024.103384, 2025.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/https://doi.org/10.1256/qj.05.108, 2005.

Falkowski, P. G.: The role of phytoplankton photosynthesis in global biogeochemical cycles, Photosynthesis Research, 39, 235–258, https://doi.org/10.1007/BF00014586, 1994.

Ford, D.: Assimilating synthetic Biogeochemical-Argo and ocean colour observations into a global ocean model to inform observing system design, Biogeosciences, 18, 509–534, https://doi.org/10.5194/bg-18-509-2021, 2021.

Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., Lehodey, P., Perruche, C., Samuelsen, A., and Simon, E.: Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments, Journal of Operational Oceanography, 8, s168–s187, https://doi.org/10.1080/1755876X.2015.1022350, 2015.

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall'Olmo, G., van Dongen-Vogels, V., Reifel, K. M., and Behrenfeld, M. J.: Analytical phytoplankton carbon measurements spanning diverse ecosystems, Deep Sea Research Part I: Oceanographic Research Papers, 102, 16–25, https://doi.org/https://doi.org/10.1016/j.dsr.2015.04.006, 2015.

Group, B.-A. P.: The scientific rationale, design and implementation plan for a Biogeochemical-Argo float array, Report, https://doi.org/10.13155/46601, 2016.

Hemmings, J. C., Barciela, R. M., and Bell, M. J.: Ocean color data assimilation with material conservation for improving model estimates of air-sea CO2 flux, Journal of Marine Research, 2008.

Johnson, K. S. and Bif, M. B.: Constraint on net primary productivity of the global ocean by Argo oxygen measurements, Nature Geoscience, 14, 769–774, https://doi.org/10.1038/s41561-021-00807-z, 2021.

Mamnun, N., Völker, C., Vrekoussis, M., and Nerger, L.: Uncertainties in ocean biogeochemical simulations: Application of ensemble data assimilation to a one-dimensional model, Frontiers in Marine Science, Volume 9 - 2022, https://doi.org/10.3389/fmars.2022.984236, 2022.

Mamnun, N., Völker, C., Vrekoussis, M., and Nerger, L.: Spatially Varying Biogeochemical Parameter Estimation in a Global Ocean Model, Journal of Geophysical Research: Oceans, 130, e2025JC022 752, https://doi.org/https://doi.org/10.1029/2025JC022752, e2025JC022752 2025JC022752, 2025.

Manizza, M., Le Quéré, C., Watson, A. J., and Buitenhuis, E. T.: Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model, Geophysical Research Letters, 32, https://doi.org/https://doi.org/10.1029/2004GL020778, 2005.

Nerger, L., Janjić, T., Schröter, J., and Hiller, W.: A Unification of Ensemble Square Root Kalman Filters, Monthly Weather Review, 140, 2335 – 2345, https://doi.org/https://doi.org/10.1175/MWR-D-11-00102.1, 2012.

Polton, J., Harle, J., Holt, J., Katavouta, A., Partridge, D., Jardine, J., Wakelin, S., Rulent, J., Wise, A., Hutchinson, K., Byrne, D., Bruciaferri, D., O'Dea, E., De Dominicis, M., Mathiot, P., Coward, A., Yool, A., Palmiéri, J., Lessin, G., Mayorga-Adame, C. G., Le Guennec, V., Arnold, A., and Rousset, C.: Reproducible and relocatable regional ocean modelling: fundamentals and practices, Geoscientific Model Development, 16, 1481–1510, https://doi.org/10.5194/gmd-16-1481-2023, 2023.

Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A., and Nerger, L.: Assimilation of Global Total Chlorophyll OC-CCI Data and Its Impact on Individual Phytoplankton Fields, Journal of Geophysical Research: Oceans, 124, 470–490, https://doi.org/https://doi.org/10.1029/2018JC014329, 2019.

Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A., and Nerger, L.: Global Assimilation of Ocean-Color Data of Phytoplankton Functional Types: Impact of Different Data Sets, Journal of Geophysical Research: Oceans, 125, e2019JC015 586, https://doi.org/https://doi.org/10.1029/2019JC015586, e2019JC015586 2019JC015586, 2020.

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W. B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S., Zilberman, N., Bakker, D., Baringer, M., Belbeoch, M., Bittig, H. C., Boss, E., Calil, P., Carse, F., Carval, T., Chai, F., Conchubhair, D. O., d'Ortenzio, F., Dall'Olmo, G., Desbruyeres, D., Fennel, K., Fer, I., Ferrari, R., Forget, G., Freeland, H., Fujiki, T., Gehlen, M., Greenan, B., Hallberg, R., Hibiya, T., Hosoda, S., Jayne, S., Jochum, M., Johnson, G. C., Kang, K., Kolodziejczyk, N., Körtzinger, A., Traon, P.-Y. L., Lenn, Y.-D., Maze, G., Mork, K. A., Morris, T., Nagai, T., Nash, J., Garabato, A. N., Olsen, A., Pattabhi, R. R., Prakash, S., Riser, S., Schmechtig, C., Schmid, C., Shroyer, E., Sterl, A., Sutton, P., Talley, L., Tanhua, T., Thierry, V., Thomalla, S., Toole, J., Troisi, A., Trull, T. W., Turton, J., Velez-Belchi, P. J., Walczowski, W., Wang, H., Wanninkhof, R., Waterhouse, A. F., Waterman, S., Watson, A., Wilson, C., Wong, A. P. S., Xu, J., and Yasuda, I.: On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array, Frontiers in Marine Science, Volume 6 - 2019, https://doi.org/10.3389/fmars.2019.00439, 2019.

Ryan-Keogh, T. J., Thomalla, S. J., Chang, N., and Moalusi, T.: A new global oceanic multi-model net primary productivity data product, Earth System Science Data, 15, 4829–4848, https://doi.org/10.5194/essd-15-4829-2023, 2023.

Sathyendranath, S., Platt, T., Žarko Kovač, Dingle, J., Jackson, T., Brewin, R. J. W., Franks, P., nón, E. M., Kulk, G., and Bouman, H. A.: Reconciling models of primary production and photoacclimation, Appl. Opt., 59, C100–C114, https://doi.org/10.1364/AO.386252, 2020.

Schourup-Kristensen, V., Sidorenko, D., Wolf-Gladrow, D. A., and Völker, C.: A skill assessment of the biogeochemical model REcoM2 coupled to the Finite Element Sea Ice–Ocean Model (FESOM 1.3), Geoscientific Model Development, 7, 2769–2802, https://doi.org/10.5194/gmd-7-2769-2014, 2014.

Skákala, J., Bruggeman, J., Ford, D., Wakelin, S., Akpınar, A., Hull, T., Kaiser, J., Loveday, B. R., O'Dea, E., Williams, C. A., and Ciavatta, S.: The impact of ocean biogeochemistry on physics and its consequences for modelling shelf seas, Ocean Modelling, 172, 101 976, https://doi.org/https://doi.org/10.1016/j.ocemod.2022.101976, 2022.

Skákala, J., Ford, D., Fowler, A., Lea, D., Martin, M. J., and Ciavatta, S.: How uncertain and observable are marine ecosystem indicators in shelf seas?, Progress in Oceanography, 224, 103 249, https://doi.org/https://doi.org/10.1016/j.pocean.2024.103249, 2024.

Skákala, J., Ford, D., Haines, K., Lawless, A., Martin, M. J., Browne, P., Chrust, M., Ciavatta, S., Fowler, A., Lea, D., Palmer, M., Rochner, A., Waters, J., Zuo, H., Banerjee, D. S., Bell, M., Carneiro, D. M., Chen, Y., Kay, S., Partridge, D., Price, M., Renshaw, R., Shapiro, G., and While, J.: Marine data assimilation in the UK: the past, the present, and the vision for the future, Ocean Science, 21, 1709–1734, https://doi.org/10.5194/os-21-1709-2025, 2025.

Tanioka, T., Garcia, C. A., Larkin, A. A., Garcia, N. S., Fagan, A. J., and Martiny, A. C.: Global patterns and predictors of C:N:P in marine ecosystems, Communications Earth & Environment, 3, 271, https://doi.org/10.1038/s43247-022-00603-6, 2022.

Yumruktepe, V. c., Samuelsen, A., and Daewel, U.: ECOSMO II(CHL): a marine biogeochemical model for the North Atlantic and the Arctic, Geoscientific Model Development, 15, 3901–3921, https://doi.org/10.5194/gmd-15-3901-2022, 2022.