Response to reviewers

Yumeng Chen

June 2026

We thank both reviewers 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. Line numbers referred in this response letter are based on the submitted manuscript instead of the revised manuscript.

Reviewer 1

Answer: We thank the reviewer for the useful comments on the manuscript. Before going to the point-by-point comments, we believe it would be beneficial to highlight the scope and aim of this study.

This study uses the widely used biogeochemistry model MEDUSA, which has been used for a variety of studies, including within the UKESM. The performance of MEDUSA is evaluated in the original model description paper and in various studies (e.g., Tagliabue et al., 2016; Sellar et al., 2019; Planchat et al., 2023). These provide a good quantification of the limitations and uncertainties of the model and as such further evaluation is not an objective for this manuscript.

We agree that the quality of different phytoplankton carbon products can differ, however, the purpose of this study is not to seek the best phytoplankton carbon product. Instead, this study investigates the impact of assimilating a phytoplankton carbon product, i.e. asking the question what we can expect from assimilating these data in contrast to assimilating chlorophyll data, which is commonly assimilated. To our knowledge, whilst there are studies assimilating particulate organic carbon (POC), phytoplankton carbon assimilation is not well-studied in the literature.

Data assimilation can indeed assimilate various observations simultaneously. The relationship between different observations is specified by the observation error covariance matrix in DA. For most observations the observation errors are deemed uncorrelated. In our study, we do not assume correlated observation errors between phytoplankton carbon and chlorophyll, but we inflate the observation errors, which is a common approach. Assimilating multiple phytoplankton carbon products from ocean colour could complicate the study even more because we expect stronger observation error correlations between these products.

Major comments

1. Carbon products and phytoplankton product derived from remote sensing have been derived and used for years. There is no novelty in that and I am surprised you chose to only look at one such product. For example, Behrenfeld et al. (2005) showed how additional information can be gleaned from using a backscattering based C_phyto. In particular, through many manuscript, we have been able to show how phytoplankton photo-acclimation is the major forcing on the chl/C_phyto ratio and how it can inform us, for example, on nutrient limitation (`https://egusphere.copernicus.org/preprints/2025/egusphere-2025-4261/` ← it has been accepted). The point here is not to make you cite papers I contributed to but make you aware that the utility of estimate of C_phyto from space has been shown in many works and, in particular, in providing information content additional to Chl.

   Answer: The goal of this study is to investigate the effects of assimilating a phytoplankton carbon product on the ecosystem model. To our knowledge, the phytoplankton carbon derived from ocean colour has not been used in data assimilation for ecosystem modelling and forecasting but there are studies assimilating particulate organic carbon. We included the utility of other carbon products in our revised manuscript.

   However, it is also worth noting that carbon products derived from the backscattering of particles such as Behrenfeld et al. (2005) are based on empirical algorithms. The carbon product used in this study is underpinned by a different algorithm using a photoacclimation model (Sathyendranath et al., 2020). The chlorophyll-to-carbon ratio in the model varies as a function of the average, daily, photosynthetically active radiation available in the mixed layer. The carbon product is based on the derived chlorophyll-to-carbon ratio. Moreover, Sathyendranath et al. (2020) validated the chlorophyll-to-carbon ratio in their study whereas some carbon products, e.g., Aqua MODIS Level-4 Global Mapped Phytoplankton Carbon Data, has limited validations.

   We have rephrased the introduction to include relevant studies from L47:

   Besides chlorophyll data, other phytoplankton products, such as particulate inorganic ocean carbon and particulate organic ocean carbon, can be derived from satellite ocean colour (Siegel et al., 2005; Yang et al., 2024) independently of chlorophyll in their derivation. In particular, satellite ocean colour can be used to derive phytoplankton carbon (Behrenfeld et al., 2005; Graff et al., 2015; Bellacicco et al., 2020) using an empirical relationship between the backscattering of particles and phytoplankton carbon. These products are used to understand the ocean's biological pump and carbon cycles in addition to the chlorophyll products (Siegel et al., 2023; Bourdin et al., 2025). Recently, in the ESA Biological Pump and Carbon Export Processes (BICEP) project, a new phytoplankton carbon (referred to as carbon hereafter) product has been derived from the ocean colour observations (Sathyendranath et al., 2021). The product is derived using an approach different from the backscattering of particles used in other carbon products. The new product is based on the chlorophyll-to-carbon ratio from a photoacclimation model where the ratio varies with the photosynthetically active radiation available in the ocean mixed layer. Compared to other carbon products with limited quality assessments (e.g., NASA Ocean Biology Processing Group, 2025), the product's chlorophyll-to-carbon ratio is validated with field data (Sathyendranath et al., 2020).

In this study, we explore the effects of assimilating the newly derived carbon product so that we can understand whether assimilating phytoplankton carbon can provide additional information compared with assimilating the chlorophyll product alone in the modelling system. Note that, here, as this study focuses on assessing the modelling response of assimilating carbon products, we do not seek the best carbon product for DA, and opt to use the validated off-the-shelf observational dataset.

2. The first test of a good C_phyto product is whether its ratio to Chl is consistent with lab studies of photoacclimation, e.g. is $30 < \text{C\_phyto}/\text{Chl} < 300$ (unless you are dealing with domination by mixotrophy in which case it could go lower, but you don't resolve them in your model.

   Answer: We agree that such an initial test is relevant when assessing the quality of a c_phyto product. However, this work focuses on investigating the effect of the assimilation of observations on the modelling system instead of the validation of the observation data itself, which is carried out by the data producers. Thus, we consider the observational data and the model as given and focus on the effects of the DA. The data are based on Sathyendranath et al. (2020), where the carbon product is derived based on the carbon-to-chlorophyll ratio. In Sathyendranath et al. (2020), the carbon to chlorophyll ratio is also broadly consistent with other studies as shown in their Figs. 7 and 8. We included such information in the revision of the manuscript.

3. Field measurements of Fchl, as done with Argo floats, can have many biases (see Roesler et al., 2017). One has to be careful on how to use them.

   Answer: We have added this information in L30 in the introduction. We have also pointed out the need for bias correction and better uncertainty quantification for data assimilation:

   However, besides its limited spatial and temporal coverage, operational assimilation of BGC-Argo floats can still face other technical challenges. For example, the chlorophyll estimates from BGC-Argo floats can be biased because they are estimated from measurements of fluorescence (Roesler et al., 2017). These biases are not addressed in studies assimilating BGC-Argo floats, e.g., in synthetic experiments performed by Ford (2021). This suggests the need for further research on bias correction for BGC-Argo floats in DA systems.

4. Phytoplankton products under clouds are typically wrong as they interpolate Chl rather than carbon and phytoplankton photo-adapt under cloud (`https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024GL112274`). How they do it for the product you use could bias your model (70% of the ocean is covered by clouds at any given time).

   Answer: Based on the chlorophyll data product user guide of ESA OC-CCI (Jackson, 2020), the product does not contain data under clouds. This leads to spatial gaps in the product. Daily products only contain pixels with available observations. The composite monthly product is computed by averaging the available daily data over each month excluding missing data. The biases and uncertainties of the chlorophyll data product are validated by in situ observations. In data assimilation, we use bias-corrected chlorophyll data as observations and the uncertainties as the observation error. The carbon product is derived based on chlorophyll data and therefore inherits the treatment for data under clouds.

   In data assimilation, these sparse and irregularly spaced observations without interpolation can be assimilated. To obtain an analysis at a grid point, the data assimilation algorithm used in this paper assimilates observations adjacent to it within a radius. The update of the model grid point depends on the spatial correlation between the grid point and the observation locations derived from the ensemble forecast. This approach is well established and widely used in numerical weather prediction.

   We added the information after L127:

   Both observation products do not contain information under clouds. Daily products only contain pixels with available observations. The composite monthly product is computed as the monthly average of available daily data excluding missing data.

   In L108, we added:

   The LESTKF can assimilate sparse and irregularly spaced observations without spatial interpolation of the observation product. In our configuration of the LESTKF, model variables at each grid point assimilate observations adjacent to that point within a radius of 200 km. The DA increment at grid points without observations is computed based on the spatial correlation between the grid point and the observation locations.

5. How do you define the 'effectiveness' of DA is key and need to be provided. Obviously DA will force the model to the data.

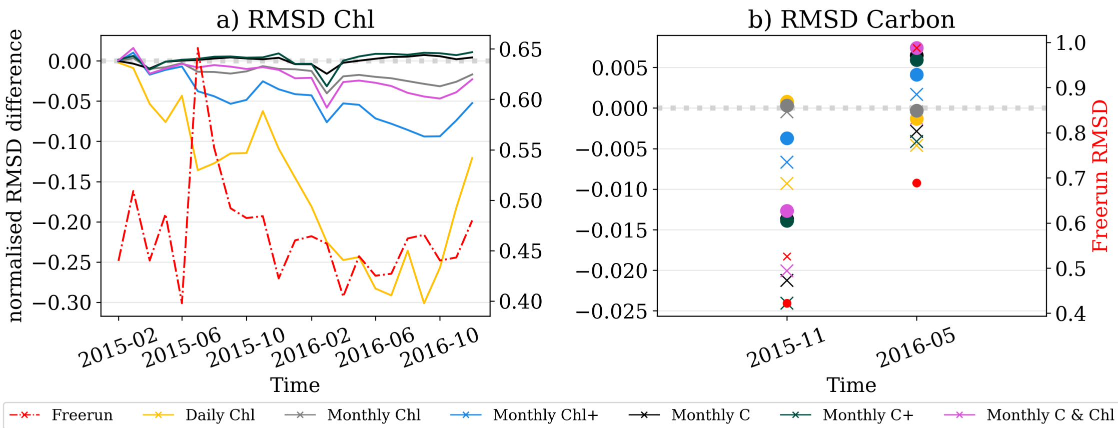

Figure 1: Differences in RMSD between DA experiments and Freerun normalised by the RMSD of Freerun between the logarithm of monthly model forecasts of and composites of chlorophyll from Argo and field campaign data from NAAMES on the left axis. The RMSD of Freerun is shown as a dashed red line on the right axis. The crosses represent comparisons with modelled total carbon while the dots represent comparisons with modelled non-diatom carbon.

Answer: Indeed, by its very definition, DA will force the model to the assimilated data. The initially submitted manuscript focuses on the state adjustments induced by assimilating observations. We removed the word “effectiveness” and instead focusing on the effect of DA on the assimilated variables.

We agree with the necessity of assessing the modelling system with independent observations. In the revised manuscript, we added 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 added a new subsection describing the independent observations.

Independent assessment 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 of chlorophyll 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 are not routinely monitored, and hence are less common compared to chlorophyll. Here, we use the analytical measurements of phytoplankton carbon from the field campaign in the North Atlantic Aerosols and Marine Ecosystems Study (NAAMES). The project provides 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 limitations of the field campaign, the in situ dataset is limited to Northwest Atlantic and weeks of data in November 2015 and May 2016.

In Sect. 5.1.2, we added:

The effect of the DA is here demonstrated by the root mean square difference (RMSD) between the ensemble mean of monthly averaged model forecasts and monthly composite observations. We first compare our model forecasts with independent in situ observations in Fig. 4 using the root mean squared difference of the logarithm of chlorophyll and carbon. The assessment is performed at observation locations using 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. In Fig. 4a), the RMSD of the chlorophyll is calculated for the upper 50 m (around 18 model levels) of the ocean for model forecasts and observations. All experiments assimilating chlorophyll show reduced RMSD of chlorophyll compared to Freerun. Experiments assimilating carbon without chlorophyll show slightly increased RMSD of the chlorophyll. This suggests that assimilating carbon product does not necessarily improve chlorophyll in MEDUSA.

The comparison with in situ carbon observations is more complicated because the carbon observations include only cells smaller than 64 μm. MEDUSA has no explicit definition of the size of diatom and non-diatom phytoplankton. Hence, in Fig. 4b), we compare the observations with model variables of both total phytoplankton (cross signs) and non-diatom phytoplankton (dot signs), which is considered to represent “small” phytoplankton. Assimilating chlorophyll has little impact on the RMSD of non-diatom carbon in November 2015. The RMSD of non-diatom carbon is reduced by assimilating carbon in November 2015 but the RMSD is increased in May 2016. The increased RMSD by assimilating carbon could be a result of post-processing and mismatch between the observations and PFTs, especially considering the seasonal dependence of the results. When comparing with total modelled carbon, assimilating carbon better constrains the carbon than assimilating chlorophyll regardless of season. Nevertheless, the impact of DA on the RMSD of carbon is smaller in November 2015 than in May 2016. For both carbon and chlorophyll, Daily Chl shows better performance than monthly assimilations. Overall, simultaneous assimilation of chlorophyll and carbon provides a balanced RMSD reduction compared to assimilating a single type of phytoplankton composition.

6. How are you converting C to N? The product is carbon and your model currency is N.

   Answer: We use a fixed Redfield ratio of C:N of 6.625:1. This follows the assumption used in the model formulation (Yool et al., 2013). We mentioned it in L96-L98:

   For consistency with the observations the nitrogen fields in this study are represented as carbon, with the conversion between carbon and nitrogen following the model assumption of a fixed ratio between C:N of 6.625:1. For clarity, we describe nitrogen in terms of its carbon equivalent (hereafter referred to simply as "carbon").

7. To evaluate the distribution of parameter (e.g. histogram of distributions, whether of [Chl] or [Cphyto]), you could compare your model distribution to those of estimate from satellite. The near log-normal distribution should arise in both.

   Answer: The log-normal distributions of model and observation products are shown below. This study follows a common practice in biogeochemistry data assimilation, as shown in L140, where the state vector is transformed under a log-normal assumption, e.g., Ford and Barciela (2017); Pradhan et al. (2019). Although this is a basis for our choice of state vector and observations in the data assimilation system, we will not explore this in detail since using the log-transformation is common practice in the assimilation of chlorophyll data.

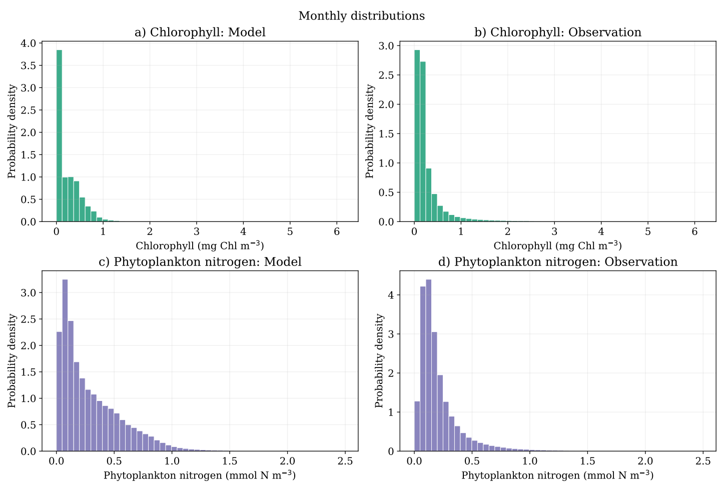

Figure 2: Distribution of phytoplankton in model and observation.

attached comments:

1. L30: Note that they measure fluorescence and not chlorophyll directly. There are many issues with these estimates (e.g., Roesler et al., 2017).

   Answer: See point 3 in major comments.

2. L34: Phytoplankton adjust their intercellular chlorophyll based on light and nutrients as well as species composition.

   Answer: We rephrased the corresponding sentence to:

   Assimilating satellite phytoplankton chlorophyll products (referred to as chlorophyll hereafter) can effectively control errors in modelled chlorophyll. However, even though the chlorophyll provides a proxy for biomass, assimilating chlorophyll does not necessarily correct other compositions of phytoplankton such as silicate, nitrogen and carbon, or factors that constrain chlorophyll such as the availability of light, nutrients, and species composition.

3. L46: nor have they been validated.

   Answer: The data producers may disagree with this statement. For example, the PFT data from Losa et al. (2017) is validated against in situ data.

   We added:

   Recent efforts were made to derive PFTs from satellite data with uncertainty validations (see e.g., Xi et al., 2021).

4. L47: What other relevant products? Are they independent from chlorophyll?

   Answer: We added that the particulate organic and inorganic carbon can be obtained independent of chlorophyll in their derivation.

   We added:

   Besides chlorophyll data, other phytoplankton products, such as particulate inorganic ocean carbon and particulate organic ocean carbon, can be derived from satellite ocean colour (Siegel et al., 2005; Yang et al., 2024) independently of chlorophyll in their derivation.

5. L50: There have been estimates of C_phyto from space produced at least from 2005 (see Behrenfeld et al., 2005). Again, the question is whether it is validated (the one of Behrenfeld was in Grag et al., 2015) and whether it provide additional information to [chl].

   Answer: We described the relevant observation products in the introduction. As mentioned, a major goal of this study is to investigate the potential benefits of assimilating phytoplankton carbon for global ecosystem modelling. Thus, we consider the question of whether there is potential additional information through the use of data assimilation which utilises multivariate relations from the model dynamics. The data is validated by its ratio between chlorophyll and carbon by Sathyendranath et al. (2020).

   See also point 1 and 2 in Major comments.

   In L52, we modify the text into

   In this study, we explore the effects of assimilating the newly derived carbon product so that we can understand whether assimilating phytoplankton carbon can provide additional information compared with assimilating the chlorophyll product alone in the modelling system. Note that, here, as this study focuses on assessing the modelling response of assimilating carbon products, we do not seek the best carbon product for DA, and opt to use the validated off-the-shelf observational dataset.

6. L54: The first thing to do is to see whether you get realistic C_phyto/Chl ratios (should vary between 30 to 300 based on lab studies)

   Answer: See point 2 in Major comments.

7. L85: This is unfortunate as it has been shown that you do not get the correct heating rate in the spring w/o phytoplankton.

   Answer: We agree that marine biogeochemistry 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). Investigating the effects of two-way coupling is out of the scope of the paper.

   We adjusted the text in L83:

   The marine ecosystem model coupled to NEMO is the Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification (MEDUSA), an intermediate complexity model (Yool et al., 2013). Although marine biogeochemical processes can have a strong impact on physics (Manizza et al., 2005; Skákala et al., 2022), the models in current DA systems are still primarily one-way coupled (Gehlen et al., 2015; Yumruktepe et al., 2022; Polton et al., 2023). Hence, even though two-way coupling can provide a better feedback between the ocean and biogeochemistry, following the common practice of using one-way coupled model avoids the need to investigate the impact of two-way coupling, which is outside the scope of this work. Hence, 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.

8. L89: No it is not. Currently phytoplankton role in CO_2 sequestration is similar to before the industrial revolution. It is the physical-chemical pump that is taking the excess anthropogenic CO2.

   Answer: Regardless of the impact of anthropogenic CO_2, a model incapable of simulating CO_2 uptake in marine ecosystems misses important climate processes. The missing processes are expected to impact climate models, which, depending on their complexities, attempt to capture the most prominent processes in the earth system.

   We changed the text to remove mentioning climate prediction here.

   Beyond these components, MEDUSA is capable of simulating the CO_2 uptake of marine ecosystems. The carbon cycle simulation is enabled by the $p\mathrm{CO}_2$, pH, alkalinity, and by carbonate species like H_2CO_3, HCO_3^- and CO_3^{2-}.

9. L94: I don't know if you are aware, but there are is significant portion of diatoms that are $< 20\mu m$.

   Answer: We agree that many diatoms are less than 20 μm so they are not strictly large phytoplankton by definition. In MEDUSA, a silicon biomass variable exists for diatoms but not for non-diatom phytoplankton. The diatom phytoplankton variable is responsible for silicon uptake and biogenic silica production and the non-diatom phytoplankton variable represents the non-silicifying phytoplankton community. However, the model also makes a simplified and idealised assumption that approximates the diatom phytoplankton functional type, which forms a key component of large phytoplankton, as "large phytoplankton". This is partly motivated by the interaction between phytoplankton and zooplankton, where small phytoplankton are strongly controlled by fast-growing microzooplankton whereas large phytoplankton are weakly controlled by slower-growing mesozooplankton.

The manuscript was modified:

Phytoplankton biomass in MEDUSA is divided into diatom and non-diatom PFTs. The diatom phytoplankton variable is responsible for silicon uptake and biogenic silica production and the non-diatom phytoplankton variable represents the non-silicifying phytoplankton community. Hence, diatom phytoplankton silicate is also explicitly modelled besides nitrogen and chlorophyll. MEDUSA models diatom phytoplankton as "large phytoplankton" based on the fact that diatoms are a key component of large phytoplankton even though diatom phytoplankton span a large range of sizes (Yool et al., 2011). This is motivated by the fact that zooplankton graze phytoplankton according to their sizes as zooplankton are are divided into microscopic and mesoscopic size classes.

10. L114: You should check for that and for a reasonable carbon/chl ratio.

    Answer: See point 2 in Major comments.

11. L127: How it interpolates under clouds is important as clouds affect chl/c of phytoplanton (Begouen Demaux et al., 2025) and at any given time cover 70% of the ocean on average. Most product simply average chlorophyll over a month to fill the blanks due to clouds with clouds counting as NaN.

    Answer: See point 4 in major comments.

12. L139: See an earlier paper by Campbell

    Answer: Campbell (1995) is added as reference.

13. L155: Is this an absolute value, a relative value or a mix of both?

    Answer: The errors of chlorophyll concentrations, ε, in the ESA OC-CCI data products are given for the logarithm of phytoplankton chlorophyll concentration. As such, it is an absolute additive error for the logarithmic concentration, which relates to a relative multiplicative error for the actual concentration.

    We rephrased the text:

    For chlorophyll observations, σ is the same as the error provided by the observation product, which can be expressed as:

    $$\mathbf{y}_o^\mu = \mathbf{y}_t^\mu + \sigma, \tag{1}$$

    where $\mathbf{y}_t^\mu$ is the unknown true logarithm of chlorophyll concentration and $\mathbf{y}_o^\mu$ is the observed logarithm of chlorophyll concentration. However, the carbon product does not provide an error estimate. In this study, the carbon observation error is estimated as the chlorophyll observation error inflated by 10%. This means that we assume the carbon error, expressed for logarithmic carbon concentrations, to be 1.1σ. As a result, both observations produce a similar spatial pattern of observation error. This treatment is justified by several considerations. First, both products are derived from the same ocean colour measurements. Hence, we can assume that both products share similar sources of measurement error. Second, this choice of observation error avoids overconfidence in the carbon product that could lead to overly strong analysis increments. Third, phytoplankton biomass is assumed to follow a log-normal distribution. Due to this assumption, observation errors expressed as the corresponding standard deviation of a Gaussian distribution are multiplicative instead of additive to the unknown true observations:

    $$\mathbf{y}_o = 10^{\mathbf{y}_o^\mu} = 10^{\mathbf{y}_t^\mu} 10^\sigma = \mathbf{y}_t 10^\sigma. \tag{2}$$

    Equation (2) means that the errors are scaling factors applied to the observation, regardless of the magnitude and unit of the observation itself.

14. L156: At low values the uncertainties can be O(100%) or more for both chlorophyll and C_phyto.

    Answer: Here, the phytoplankton chlorophyll error is inflated by 10% instead of giving an error of 10% of phytoplankton carbon value. Thus, the error of carbon is $10^{1.1\sigma}$.

15. L171: what to you mean by this? The two are correlated in nature too, though one is not derived from the other. The fact that chlorophyll changes from 0.01 to $50\mu g/l$ in the surface ocean and that the ratio of $30 <$ carbon/chl < 300 alone will cause them to be correlated.

    Answer: Data assimilation uses error covariance matrices of the forecast and the observations. What is relevant here is whether the errors are correlated, not whether the variables are correlated. For two independent measurements the errors of the measurements are not expected to be correlated even for the same physical quantity, which has a physical correlation of one. Here, because the carbon product is derived from chlorophyll, the observation error of chlorophyll is expected to propagate to the carbon product leading to error correlations.

    We rephrased the text:

Lastly, the error correlation between the phytoplankton products that arises because they are derived from the same satellite ocean colour measurements is neglected because the carbon and chlorophyll products are not estimated independently. Otherwise, there are no error correlations between them. The neglected error correlations are assumed to be accounted for by the 10% inflation in the carbon observation error. Despite these potential issues, these results still provide an assessment of the value of the carbon product.

16. L197: There are other product (e.g. backscattering) from which C_phyto can be derived at the same resolution as [chl]. While limit yourself to the databases you did?

    Answer: The limitation results directly from the motivation of our study: To investigate the effects of assimilating a carbon product on the modelling instead of finding the best carbon products for the DA system. Since the carbon product is recently derived via a different approach from the backscattering, and is easily accessible with validated carbon and chlorophyll ratio, we chose the current carbon product.

    See also point 1 in Major comments.

    We adjusted L52 to reflect this aspect.

    Note that, here, as this study focuses on assessing the modelling response of assimilating carbon products, we do not seek the best carbon product for DA, and opt to use the validated off-the-shelf observational dataset.

17. L213: How do you define 'effectiveness'?

    Answer: Here, the word 'effectiveness' is removed and we highlight that we focus on the effect of DA on assimilated model variables against the assimilated observations using biases between model and observations, the root mean squared differences (RMSDs), continuous rank probability score (CRPS), and ensemble spread. These scores are commonly used to assess DA. We also added an assessment with independent in situ observations as discussed in point 5 in Major comment.

    We adjusted the text:

    Here, the effects of the DA on assimilated variables are evaluated using the bias between the model and observations, the root mean squared differences (RMSDs), continuous rank probability score (CRPS), and ensemble spread. These scores can expose potential disagreements between different observations, the uncertainty, and reliability of the ensemble after DA.

18. L219: What do you mean by this. Traceable? Validated?

    Answer: Here, it actually means un-biased observations or anchor observations. This concept comes from data assimilation for numerical weather prediction where some observations can have negligible biases so that one can perform bias correction in DA (Eyre, 2016).

    We rephrased the sentence to:

    Quantifying biases in models and observations is non-trivial considering the limited availability of so-called anchor observations with negligible biases (Eyre, 2016; Fowler et al., 2023).

19. L220: Aren't model parameters (e.g. grazing rates etc) tuned based on data, hence are not independent?

    Answer: We agree that the default model parameters are determined based on observations. However, ensemble Kalman filters in general assume no error correlations in time between observations, and uncorrelated forecast and observation errors. In the sequential DA the forecast is initialised by analyses subjected to observations at the time. If observation errors have time correlations, this could indeed lead to error correlations in observations and model forecast. Even though the assumption that observations have no time correlations is made partly based on computational considerations, it is a justifiable approximation.

    Considering that the model and its parameters were proposed earlier than the experiment period, its parameters have been tuned based on observations prior to the observations assimilated in this study. Hence, this allows to assume no correlations between model forecast and observation errors. Moreover, the misfit between observations and model forecast is widely used in the evaluation of operational DA systems, e.g., Saha et al. (2014); Zuo et al. (2019); Barton et al. (2021); Mignac et al. (2025)

    The manuscript was adjusted as follows:

    Because DA methods assume uncorrelated forecast and observation errors, as a proxy, the bias can be revealed by the misfit between observations and the model forecast over multiple time steps used in the DA (Saha et al., 2014; Zuo et al., 2019; Barton et al., 2021; Mignac et al., 2025). The expectation of the misfit is zero in the absence of biases. Because the model parameters are not determined by the assimilated observations, this diagnostic can expose biases in the DA systems.

20. L230: Isn't this built in? Since your carbon is based on [chl] and since [chl] is biased low in the model, updating the [c] is akin of updating [chl] in the 'right' direction.

    Answer: Yes. This is a result of the model formulation where the nitrogen is the primary currency. This potential limitation was mentioned in the conclusion in L501: "Nevertheless, these adjustments may vary depending on the formulations of individual marine ecosystem models. In MEDUSA, carbon and nitrogen are assumed to have a fixed stoichiometric relationship. "

    We adjusted the manuscript with:

Notably, using the balancing scheme to update the phytoplankton carbon in the "Monthly Chl+" experiment slightly improves the expectation of the misfits in chlorophyll. This suggests that making direct changes in the primary composition of phytoplankton biomass in MEDUSA leads to more sustained changes in the model. This is consistent with the study conducted by (Ford and Barciela, 2017).

21. L241: This may be driven by the fact that loss processes are just as important as processes contributing to growth (in nature).

    Answer: We added the following sentence in the manuscript:

    Considering the carbon data are derived based on chlorophyll data and their ratio is validated by Sathyendranath et al. (2020), in addition to the balancing scheme, the inconsistency of different signs may also arise from model errors. This could be driven by inaccurate modelling of the primary production, the loss of phytoplankton, and/or the coupling of chlorophyll and carbon, where the tendencies of chlorophyll and carbon are erroneous given modelled nutrient conditions. The assimilation of both products finds a compromise between the biased model and observation products.

22. page 9: Here and elsewhere, your model currency is N. Are you assuming a Redfield ratio to convert C:N?

    Answer: Yes. See point 6 in major comments as well.

23. L268: It is hard to evaluate what you mean by 'best'. Both model and data have biases, as you said, and once you assimilate the data the model output will look more like the data. How can you dependently assess what the best strategy is, e.g. by comparing to another INDEPENDENT data source.

    Answer: We agree that it is difficult to evaluate the best assimilation strategy, and we do not recommend a best strategy. The intention of the comparison of RMSD is to show that the DA methods account for uncertainties in both observations when combining them with the model forecast, which has a global effect compared to assimilating a single type of observation. In the case of simultaneous assimilations, we concluded that they provide a more balanced update between carbon and chlorophyll than assimilating one type of observations.

    For further assessment, we added an assessment with Argo data for chlorophyll and with in situ data for carbon. See also point 5 in Major comments.

24. L295: isn't this by construction?; L298: Define 'reliable' ensemble

    Answer: A perfectly reliable ensemble means that the probability distribution function (p.d.f) of the ensemble is identical to the p.d.f of the truth (Leutbecher and Palmer, 2008; Rodwell et al., 2016).

    Data assimilation is constructed to improve the accuracy of state estimation using observations. In an ideal setup, the DA could improve reliability specified with a perfect forecast, observation, and model error covariance matrices without biases. However, it does not necessarily improve the reliability of the ensemble. For example, DA systems routinely have to deal with under-dispersive ensembles due to limited ensemble sizes (Carrassi et al., 2018). Therefore, it is not clear if assimilating multiple observations can improve the overall reliability of the ensemble.

    We changed the text in L282:

    In addition to the RMSD of the ensemble mean, a reliable ensemble DA system is expected to represent the probability of the occurrence of a given state, i.e. the uncertainty of the system. In a perfectly reliable ensemble, by definition, such an ensemble will have an probability distribution function identical to the true probability distribution (Leutbecher and Palmer, 2008; Rodwell et al., 2016). Here, the true probability distribution is represented by the probability distribution of the assimilated observations as calculated by the reliability score of the continuous ranked probability score (CRPS). The CRPS is widely used to diagnose ensemble forecasts (Hersbach, 2000), representing the differences between the cumulative probability of the distribution given by the model ensemble and observations. The CRPS can be decomposed into reliability and resolution scores. In a perfectly reliable ensemble, the reliability score is zero, and the ensemble represents the true uncertainty of the model simulations. The reliability score increases with less reliable ensemble. The reliability score of the CRPS is evaluated on a grid point by grid point basis. As the DA does not guarantee improved reliability by construction, this metric helps us evaluate the accuracy of quantified uncertainty from each experiment.

25. L469: My limited understanding is that the equilibration time of CO_2 and O_2 are very different in the upper. Can this affect what you see?

    Answer: The explanation is added:

    Unlike pCO_2, the ocean surface oxygen concentration shows a strong seasonal variation without an obvious trend, which could be a result of the different equilibration timescales of O_2 from pCO_2.

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 added 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 future studies.

   We added the following sentences in the manuscript:

   Moreover, this choice of the observation error is supported by the statistical 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 spatial and temporal averaged carbon observation error of 0.264 used in this study is similar to the observation error of 0.266 given by the diagnostic. This consistent global error statistic does not address changes in error in space and time, which will require further investigation in future 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 can only occur when the distribution of the spatial and temporal varying C:N ratio is not symmetric (Tanioka et al., 2022) that leads to a ratio that is systematically either too large or too small. Nonetheless, a fixed ratio in general cannot be expected to result in a bias of fixed sign.

   We added this discussion in Sec. 5.1.1:

   The DA algorithm assumes that the forecasts and observations have no biases. However, this assumption is rarely met in real applications. Biases can exist both in observations and model forecasts for both carbon and chlorophyll. In model forecasts, the bias could be a result of inaccurate model parameters, simplified treatment of model processes and biased initial conditions (Mamnun et al., 2025). In addition to model forecast, the application of fixed C:N ratio could affect the biases. This fixed ratio can be viewed

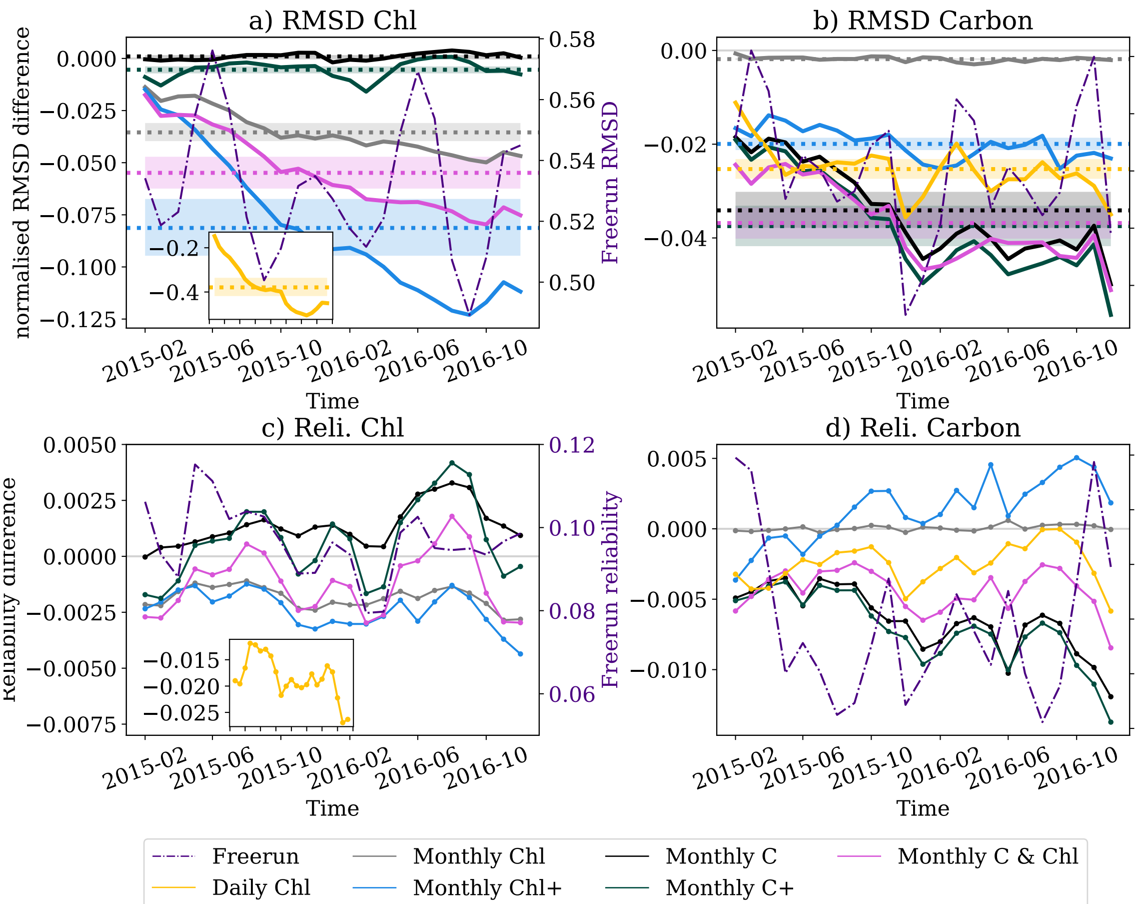

Figure 3: Upper row: time series of the RMSD of the ensemble mean of the logarithmic surface phytoplankton a) chlorophyll and b) carbon. The left axis shows the differences of RMSD between DA experiments and Freerun normalised by the RMSD of Freerun, and the right axis shows the RMSD of Freerun (red 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 horizontal dashed lines with shaded area are the temporal average of the RMSD, and the shaded area represents the 95% confidence interval of the temporal average. Lower row: reliability score of the ensemble of surface phytoplankton a) chlorophyll and b) nitrogen compared against 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 reliability differences of chlorophyll in the "Daily Chl" experiment.

as a global average in both time and space; biases could only occur when the distribution of the spatially and temporally varying C:N ratio is not symmetric that leads to a ratio that is systematically either too large or too small (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 added 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+.

   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 added 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 added in Sect. 5.2.1:

It is also worth noting that a strong El Niño occurred during our experiment period which had a strong impact on the eastern Equatorial Pacific. The differences between observations and Freerun may reflect the modelling capability of the El Niño event as well. This could imply that results from Freerun and other experiments in the eastern Pacific region may not be generalised to years without El Niño 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 added 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) used 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 used 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, our experiments coincide with a strong El Niño event. These could all lead to different responses to the seasonality of the results.

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 added the following in Sec. 5.1.3.

Fig. 4c)-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 benefit from 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 added 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 rephrased:

As our goal is to investigate the effects of assimilating the carbon product instead of constructing the best performing DA system, we follow common practices in operational systems (Skákala et al., 2025). In this study, two datasets are assimilated into NEMO-MEDUSA: a new monthly phytoplankton carbon product and a chlorophyll-*a* dataset for reference. This means that the DA system inherits some challenges in operational DA as we do not assimilate additional observations to improve vertical profiles.

In conclusion, we added:

A reliable ensemble could further lead to multivariate DA increments that could perform better than the post-processing scheme used here as demonstrated by Pradhan et al. (2019). ... Furthermore, this study follows the common practice of operational DA systems by only assimilating satellite ocean colour (Skákala et al., 2025). This means that updates to the subsurface structure are not as accurate as assimilating profiles of marine ecosystem variables such as BGC-Argo data. Hence, it is also of interest to investigate a DA system that assimilates both vertical profiles along with 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 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 added 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 added a new subsection describing the independent observations.

Independent assessment 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 of chlorophyll 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 are not routinely monitored, and hence are less common compared to chlorophyll. Here, we use the analytical measurements of phytoplankton carbon from the field campaign in the North Atlantic Aerosols and Marine Ecosystems Study (NAAMES). The project provides 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 limitations of the field campaign, the in situ dataset is limited to Northwest Atlantic and weeks of data in November 2015 and May 2016.

In Sect. 5.1.2, we added:

The effect of the DA is here demonstrated by the root mean square difference (RMSD) between the ensemble mean of monthly averaged model forecasts and monthly composite observations. We first compare our model forecasts with independent in situ observations in Fig. 4 using the root mean squared difference of the logarithm of chlorophyll and carbon. The assessment is performed at observation locations using 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. In Fig. 4a), the RMSD of the chlorophyll is calculated for the upper 50 m (around 18 model levels) of the ocean for model forecasts and observations. All experiments assimilating chlorophyll show reduced RMSD of chlorophyll compared to Freerun. Experiments assimilating carbon without chlorophyll show slightly increased RMSD of the chlorophyll. This suggests that assimilating carbon product does not necessarily improve chlorophyll in MEDUSA.

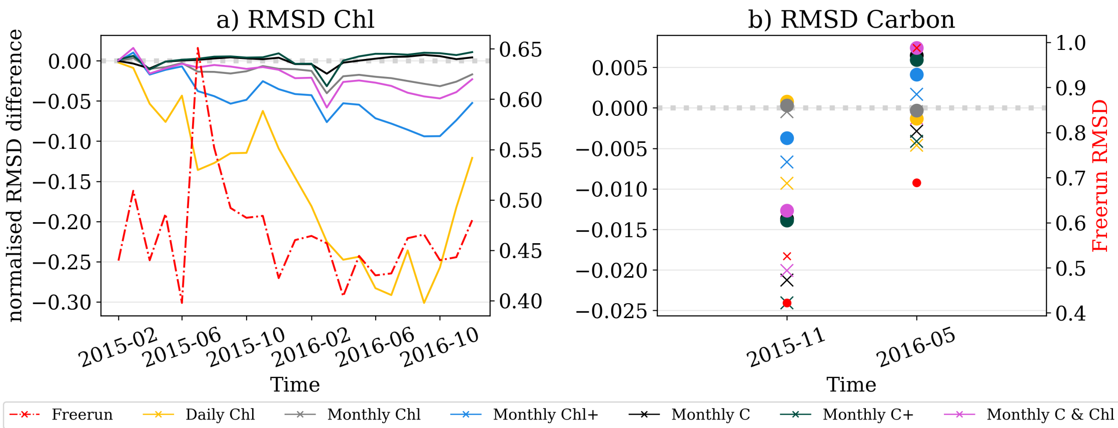

Figure 4: Differences in RMSD between DA experiments and Freerun normalised by the RMSD of Freerun between the logarithm of monthly model forecasts of and composites of chlorophyll from Argo and field campaign data from NAAMES on the left axis. The RMSD of Freerun is shown as a dashed red line on the right axis. The crosses represent comparisons with modelled total carbon while the dots represent comparisons with modelled non-diatom carbon.

The comparison with in situ carbon observations is more complicated because the carbon observations include only cells smaller than 64 μm. MEDUSA has no explicit definition of the size of diatom and non-diatom phytoplankton. Hence, in Fig. 4b), we compare the observations with model variables of both total phytoplankton (cross signs) and non-diatom phytoplankton (dot signs), which is considered to represent "small" phytoplankton. Assimilating chlorophyll has little impact on the RMSD of non-diatom carbon in November 2015. The RMSD of non-diatom carbon is reduced by assimilating carbon in November 2015 but the RMSD is increased in May 2016. The increased RMSD by assimilating carbon could be a result of post-processing and mismatch between the observations and PFTs, especially considering the seasonal dependence of the results. When comparing with total modelled carbon, assimilating carbon better constrains the carbon than assimilating chlorophyll regardless of season. Nevertheless, the impact of DA on the RMSD of carbon is smaller in November 2015 than in May 2016. For both carbon and chlorophyll, Daily Chl shows better performance than monthly assimilations. Overall, simultaneous assimilation of chlorophyll and carbon provides a balanced RMSD reduction compared to assimilating a single type of phytoplankton composition.

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

Moreover, independent observations will be needed to quantitatively validate the DA system in addition to phytoplankton chlorophyll and carbon. For example, BGC-Argo floats, the Surface Ocean CO_2 Atlas (SOCAT) or World Ocean Atlas datasets could be used 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 added a discussion on this in the conclusions, but we have limited 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 added

   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 investigations.

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

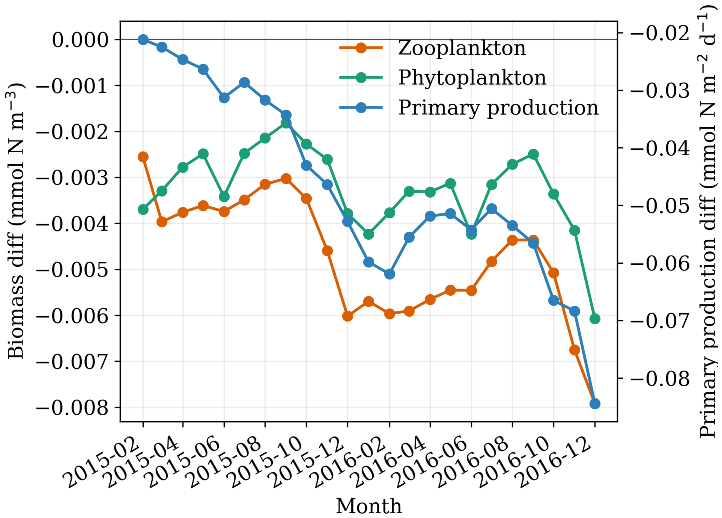

Figure 5: The time series of the differences of the ensemble mean of the spatially averaged phytoplankton carbon, zooplankton and primary production between the Monthly Chl & C and Freerun.

carbon and other variables as done in operational systems. Figures 6 and 10 (in the submitted manuscript) 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.

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 added this explanation:

The increased oxygen in these experiments is a result of less consumption of oxygen from the reduced zooplankton. This reduction in oxygen consumption exceeds the reduction in 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 adjusted the introduction and acknowledge that this could limit the accuracy of the model forecast.

The marine ecosystem model coupled to NEMO is the Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification (MEDUSA), an intermediate complexity model (Yool et al., 2013). Although marine biogeochemical processes can have a strong impact on physics (Manizza et al., 2005; Skákala et al., 2022), the models in current DA systems are still primarily one-way coupled (Gehlen et al., 2015; Yumruktepe et al., 2022; Polton et al., 2023). Hence, even though two-way coupling can provide a better feedback between the ocean and biogeochemistry, following the common practice of using one-way coupled model avoids the need to investigate the impact of two-way coupling, which is outside the scope of this work. Hence, 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.

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

Although, technically feasible in this system, this approach was not investigated in the context of MEDUSA, which requires further tuning of ensemble perturbations for ensemble cross error covariance. 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 to other model variables, including diatom and non-diatom PFT, by post-processing following Ford (2021).

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 added the following in Sect. 5.1.1.

    As suggested by the histogram, the errors of the modelled carbon and chlorophyll have different signs, and the use of a balancing scheme 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), in addition to the balancing scheme, the inconsistency of different signs may also arise from model errors.

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 rephrased 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 to 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 added a discussion in 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 as they may require centuries or millennia to reach steady state. However, all of our experiments are performed under the same deep ocean conditions which should therefore have a 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 reported the deterioration in seasonality in Monthly Chl+ in the abstract.

   In the abstract, we adjusted the text:

We demonstrate that, compared with assimilating only the chlorophyll product, which may degrade the seasonality of phytoplankton carbon, assimilating the new carbon product can provide different patterns of adjustment and seasonal anomalies in phytoplankton concentrations, surface pCO_2, and oxygen.

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 rephrased the sentence:

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

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 adjusted the sentence 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 rephrased L210 to:

   We further assess the influences 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 rephrased 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 added this in the introduction:

   Yet, as discussed by Skákala et al. (2024), phytoplankton community compositions are the most uncertain and least observable variables in the marine ecosystem even when a balancing scheme is applied.

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 changed "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 rephrased 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 added in the introduction:

   Note that, 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 rephrased the following sentence in introduction:

   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

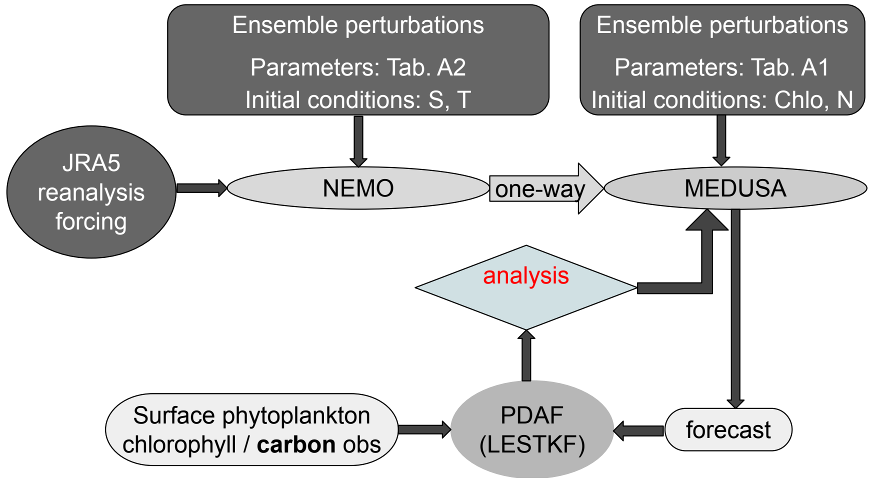

Figure 6: An illustration of the data assimilation system where an ensemble of forecasts is obtained from the one-way coupled NEMO-MEDUSA forced by the JRA-55 reanalysis whereas the satellite surface chlorophyll and carbon observations are assimilated. The analysis (reconstructed state) is used to initialise 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.

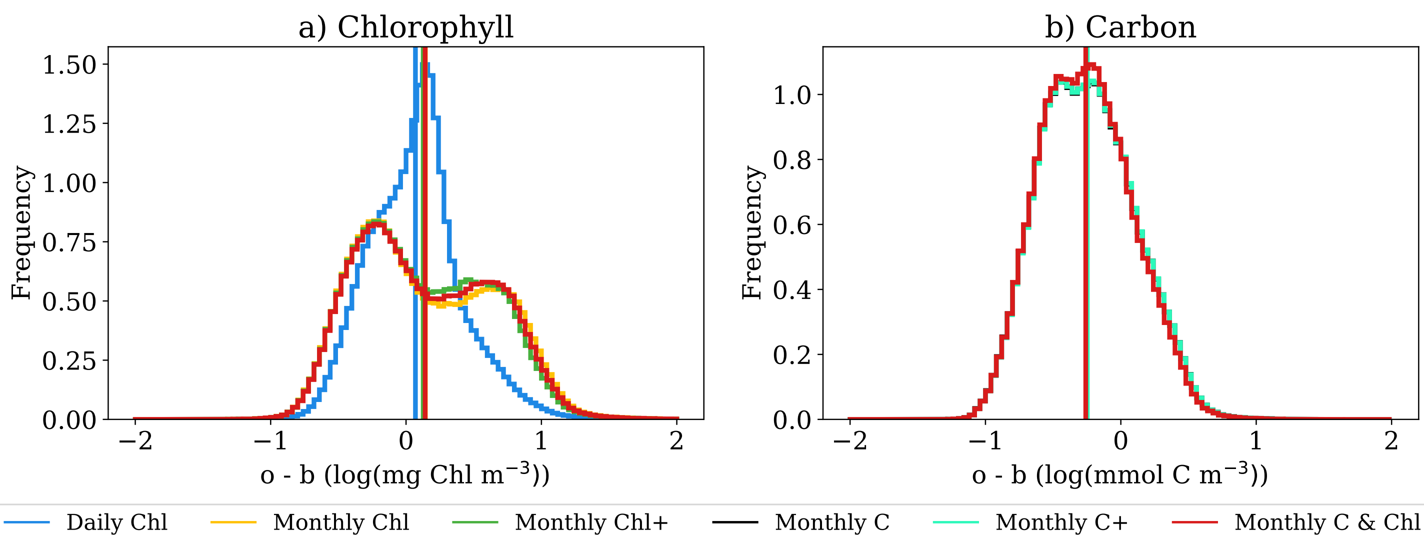

Figure 7: The histogram of the misfit between observations and forecast for log-transformed variables at the analysis step. Large observation and forecast misfits with low probability in a) are neglected for display purposes. The mean of the misfits for each experiment is presented as vertical lines. This histogram was constructed by global instantaneous data used by DA over the entire experiment 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.

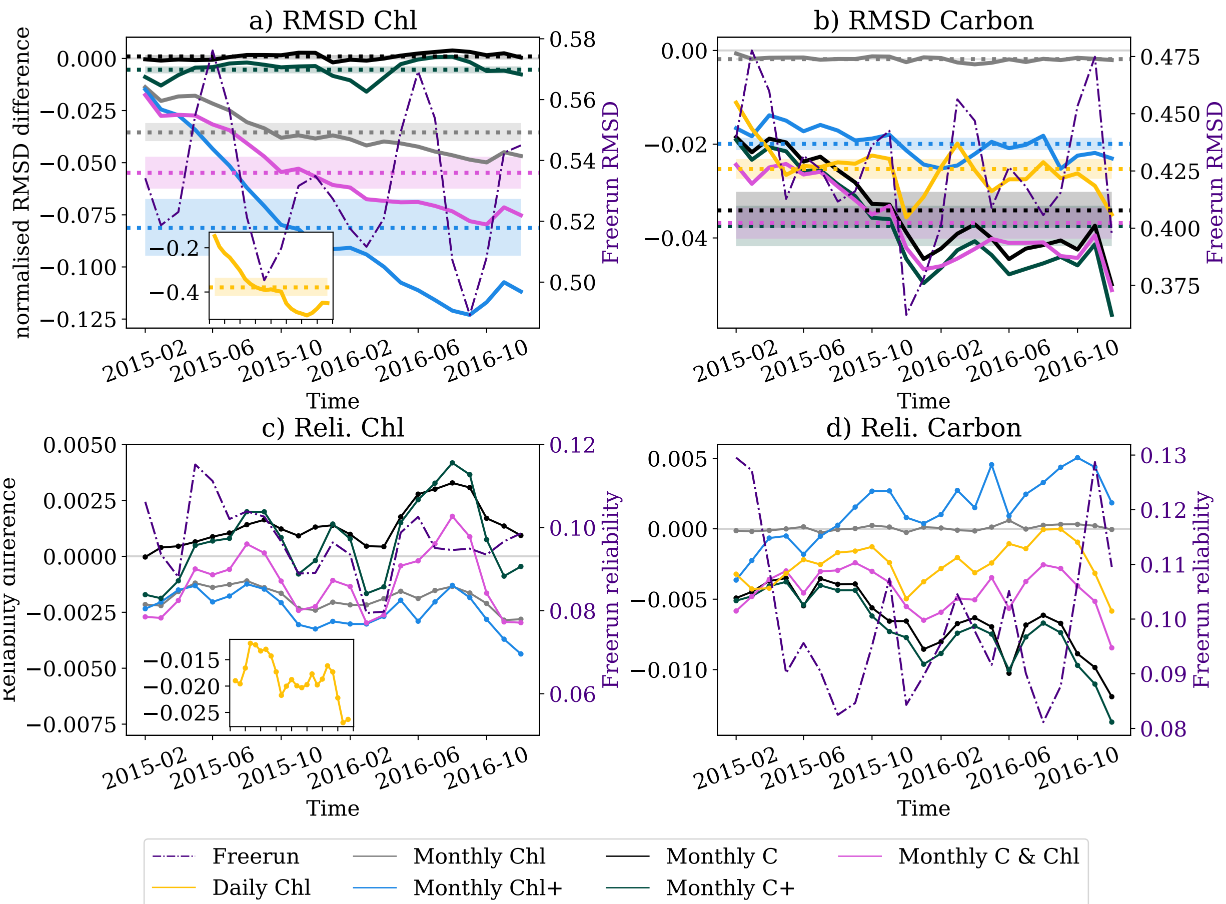

Figure 8: Upper row: time series of the RMSD of the ensemble mean of the logarithmic surface phytoplankton a) chlorophyll and b) carbon. The left axis shows the differences of RMSD between DA experiments and Freerun normalised by the RMSD of Freerun, and the right axis shows the RMSD of Freerun (red 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 horizontal dashed lines with shaded area are the temporal average of the RMSD, and the shaded area represents the 95% confidence interval of the temporal average. Lower row: reliability score of the ensemble of surface phytoplankton a) chlorophyll and b) nitrogen compared against 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 reliability differences 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.

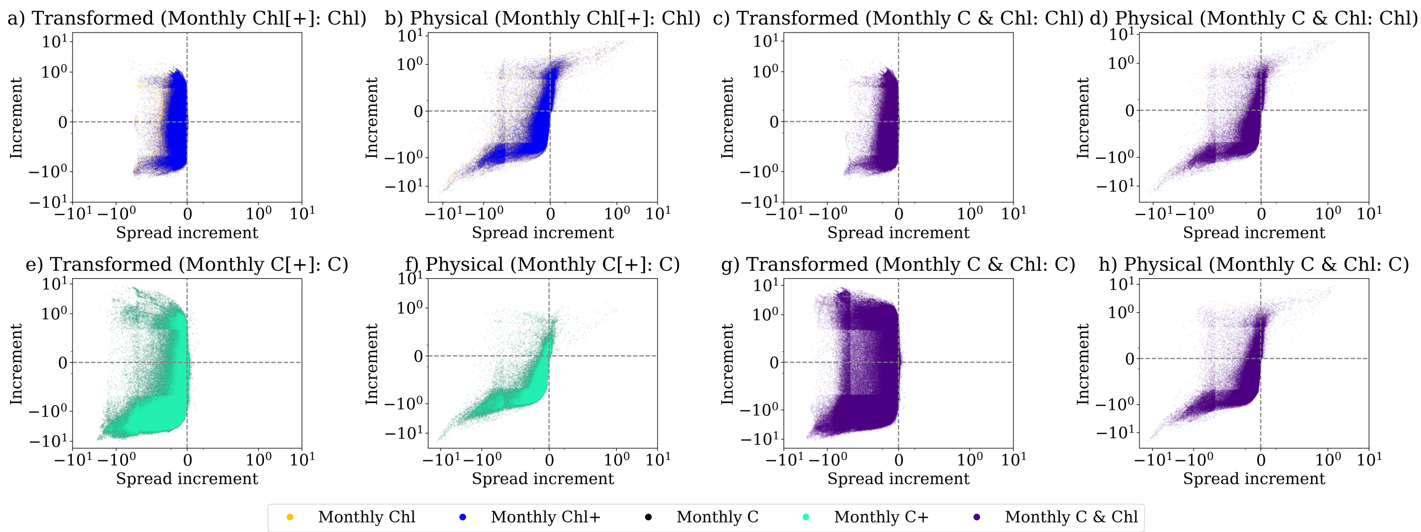

Figure 9: Scatter plot of the ensemble mean increment against the ensemble spread increment on a log scale. Here, the LESTKF is performed on transformed variables following the Gaussian distribution, and the physical values are the ensemble mean and standard deviation of the phytoplankton concentrations. 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 removed 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.

Barton, N., Metzger, E. J., Reynolds, C. A., Ruston, B., Rowley, C., Smedstad, O. M., Ridout, J. A., Wallcraft, A., Frolov, S., Hogan, P., Janiga, M. A., Shriver, J. F., McLay, J., Thoppil, P., Huang, A., Crawford, W., Whitcomb, T., Bishop, C. H., Zamudio, L., and Phelps, M.: The Navy's Earth System Prediction Capability: A New Global Coupled Atmosphere-Ocean-Sea Ice Prediction System Designed for Daily to Subseasonal Forecasting, Earth and Space Science, 8, e2020EA001 199, https://doi.org/https://doi.org/10.1029/2020EA001199, 2021.

Behrenfeld, M. J., Boss, E., Siegel, D. A., and Shea, D. M.: Carbon-based ocean productivity and phytoplankton physiology from space, Global Biogeochemical Cycles, 19, https://doi.org/https://doi.org/10.1029/2004GB002299, 2005.

Bellacicco, M., Pitarch, J., Organelli, E., Martinez-Vicente, V., Volpe, G., and Marullo, S.: Improving the Retrieval of Carbon-Based Phytoplankton Biomass from Satellite Ocean Colour Observations, Remote Sensing, 12, https://doi.org/10.3390/rs12213640, 2020.

Bourdin, G., Karp-Boss, L., Lombard, F., Gorsky, G., and Boss, E.: Dynamics of island mass effect – Part II: Phytoplankton physiological responses, EGUsphere, 2025, 1–56, https://doi.org/10.5194/egusphere-2025-4261, 2025.

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.

Campbell, J. W.: The lognormal distribution as a model for bio-optical variability in the sea, Journal of Geophysical Research: Oceans, 100, 13 237–13 254, https://doi.org/https://doi.org/10.1029/95JC00458, 1995.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, WIREs Climate Change, 9, e535, https://doi.org/https://doi.org/10.1002/wcc.535, 2018.

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.

Eyre, J. R.: Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties, Quarterly Journal of the Royal Meteorological Society, 142, 2284–2291, https://doi.org/https://doi.org/10.1002/qj.2819, 2016.

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.

Ford, D. and Barciela, R.: Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products, Remote Sensing of Environment, 203, 40–54, https://doi.org/https://doi.org/10.1016/j.rse.2017.03.040, earth Observation of Essential Climate Variables, 2017.

Fowler, A. M., Skákala, J., and Ford, D.: Validating and improving the uncertainty assumptions for the assimilation of ocean-colour-derived chlorophyll a into a marine biogeochemistry model of the Northwest European Shelf Seas, Quarterly Journal of the Royal Meteorological Society, 149, 300–324, https://doi.org/https://doi.org/10.1002/qj.4408, 2023.

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.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather and Forecasting, 15, 559 – 570, https://doi.org/10.1175/1520-0434(2000)015⟨0559:DOTCRP⟩2.0.CO;2, 2000.

Jackson, T.: ESA Ocean Colour Climate Change Initiative – Phase 3: Product User Guide for v5.0 Dataset, Product User Guide D4.2, Plymouth Marine Laboratory, `http://www.esa-oceancolour-cci.org/`, issue 1.0, dated 12 October 2020. Prepared for the European Space Agency Ocean Colour Climate Change Initiative, 2020.

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.

Leutbecher, M. and Palmer, T.: Ensemble forecasting, Journal of Computational Physics, 227, 3515–3539, https://doi.org/https://doi.org/10.1016/j.jcp.2007.02.014, predicting weather, climate and extreme events, 2008.

Losa, S. N., Soppa, M. A., Dinter, T., Wolanin, A., Brewin, R. J. W., Bricaud, A., Oelker, J., Peeken, I., Gentili, B., Rozanov, V., and Bracher, A.: Synergistic Exploitation of Hyper- and Multi-Spectral Precursor Sentinel Measurements to Determine Phytoplankton Functional Types (SynSenPFT), Frontiers in Marine Science, Volume 4 - 2017, https://doi.org/10.3389/fmars.2017.00203, 2017.

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.

Mignac, D., Waters, J., Lea, D. J., Martin, M. J., While, J., Weaver, A. T., Vidard, A., Guiavarc'h, C., Storkey, D., Ford, D., Blockley, E. W., Baker, J., Haines, K., Price, M. R., Bell, M. J., and Renshaw, R.: Improvements to the Met Office's global ocean–sea ice forecasting system including model and data assimilation changes, Geoscientific Model Development, 18, 3405–3425, https://doi.org/10.5194/gmd-18-3405-2025, 2025.

NASA Ocean Biology Processing Group: Aqua MODIS Level-4 Global Mapped Carbon Data, Version 2022.0, https://doi.org/10.5067/AQUA/MODIS/L4M/CARBON/2022.0, dataset; accessed 2026-06-20, 2025.

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.

Planchat, A., Kwiatkowski, L., Bopp, L., Torres, O., Christian, J. R., Butenschön, M., Lovato, T., Séférian, R., Chamberlain, M. A., Aumont, O., Watanabe, M., Yamamoto, A., Yool, A., Ilyina, T., Tsujino, H., Krumhardt, K. M., Schwinger, J., Tjiputra, J., Dunne, J. P., and Stock, C.: The representation of alkalinity and the carbonate pump from CMIP5 to CMIP6 Earth system models and implications for the carbon cycle, Biogeosciences, 20, 1195–1257, https://doi.org/10.5194/bg-20-1195-2023, 2023.

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.

Rodwell, M. J., Lang, S. T. K., Ingleby, N. B., Bormann, N., Hólm, E., Rabier, F., Richardson, D. S., and Yamaguchi, M.: Reliability in ensemble data assimilation, Quarterly Journal of the Royal Meteorological Society, 142, 443–454, https://doi.org/https://doi.org/10.1002/qj.2663, 2016.

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.

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., Briggs, N., Bricaud, A., Schmechtig, C., Poteau, A., D'Ortenzio, F., Ras, J., Drapeau, S., Haëntjens, N., and Barbieux, M.: Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors, Limnology and Oceanography: Methods, 15, 572–585, https://doi.org/https://doi.org/10.1002/lom3.10185, 2017.

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.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., ya Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E.: The NCEP Climate Forecast System Version 2, Journal of Climate, 27, 2185 – 2208, https://doi.org/10.1175/JCLI-D-12-00823.1, 2014.

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.

Sathyendranath, S., Platt, T., Kovač, v., Dingle, J., Jackson, T., Brewin, R., Franks, P., Kulk, G., and Bouman, H.: BICEP / NCEO: Monthly global Phytoplankton Carbon, between 1998-2020 at 9 km resolution (derived from the Ocean Colour Climate Change Initiative v5.0 dataset), NERC EDS Centre for Environmental Data Analysis, https://doi.org/https://dx.doi.org/10.5285/6a6ccbb8ef2645308a60dc47e9b8b5fb, 2021.

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.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, https://doi.org/https://doi.org/10.1029/2019MS001739, 2019.

Siegel, D. A., Maritorena, S., Nelson, N. B., and Behrenfeld, M. J.: Independence and interdependencies among global ocean color properties: Reassessing the bio-optical assumption, Journal of Geophysical Research: Oceans, 110, https://doi.org/https://doi.org/10.1029/2004JC002527, 2005.

Siegel, D. A., DeVries, T., Cetinić, I., and Bisson, K. M.: Quantifying the Ocean's Biological Pump and Its Carbon Cycle Impacts on Global Scales, Annual Review of Marine Science, 15, 329–356, https://doi.org/https://doi.org/10.1146/annurev-marine-040722-115226, 2023.

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.

Tagliabue, A., Aumont, O., DeAth, R., Dunne, J. P., Dutkiewicz, S., Galbraith, E., Misumi, K., Moore, J. K., Ridgwell, A., Sherman, E., Stock, C., Vichi, M., Völker, C., and Yool, A.: How well do global ocean biogeochemistry models simulate dissolved iron distributions?, Global Biogeochemical Cycles, 30, 149–174, https://doi.org/https://doi.org/10.1002/2015GB005289, 2016.

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.

Xi, H., Losa, S. N., Mangin, A., Garnesson, P., Bretagnon, M., Demaria, J., Soppa, M. A., Hembise Fanton d'Andon, O., and Bracher, A.: Global Chlorophyll a Concentrations of Phytoplankton Functional Types With Detailed Uncertainty Assessment Using Multisensor Ocean Color and Sea Surface Temperature Satellite Products, Journal of Geophysical Research: Oceans, 126, e2020JC017 127, https://doi.org/https://doi.org/10.1029/2020JC017127, e2020JC017127 2020JC017127, 2021.

Yang, G., Bellacicco, M., Organelli, E., and Xing, X.: Global Variability of Phytoplankton Carbon and Non-Algal Particles From Ocean Color Data Based on a Photoacclimation Model, Journal of Geophysical Research: Oceans, 129, e2023JC019 922, https://doi.org/https://doi.org/10.1029/2023JC019922, e2023JC019922 2023JC019922, 2024.

Yool, A., Popova, E. E., and Anderson, T. R.: Medusa-1.0: a new intermediate complexity plankton ecosystem model for the global domain, Geoscientific Model Development, 4, 381–417, https://doi.org/10.5194/gmd-4-381-2011, 2011.

Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies, Geoscientific Model Development, 6, 1767–1811, https://doi.org/10.5194/gmd-6-1767-2013, 2013.

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.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, Ocean Science, 15, 779–808, https://doi.org/10.5194/os-15-779-2019, 2019.