Response to Referee #1

General comments:
This study tackles an important question regarding the potential benefit of assimilating BGC-Argo profiles in improving the global ocean biogeochemical reanalyses. The manuscript consists of 3 sections: 1) establish an OSSEs framework; 2) assess different strategies of updating BGC model state variables when assimilating ocean color data; 3) evaluate the benefit of assimilating different numbers of BGC-Argo profiles in addition to ocean color data over the assimilation of profiles or ocean color data alone. Overall, I think this work is well-conceived and will make an important contribution to the state-of-art ocean biogeochemical data assimilation combining the routinely available ocean color data and the emerging BGC-Argo observations.

Thank you for your positive assessment and constructive comments. I will address each of these in turn below.

I have two major comments. First, I feel that the second section of the manuscript on different DA strategies, at its present form, does not add much value to the story and the selection of best DA strategy involving a nitrogen balancing scheme before thorough tuning doesn't seem fair. Second, the first section on OSSEs requires a bit more analysis to prove its credibility. I'll provide more detailed explanations below. Aside from that, I have some minor comments, mostly technical, for the author to consider.

Thank you for these comments. On reflection, I agree that the section on ocean colour assimilation strategies does not add much value to the core aims of the manuscript, and is not fully fledged. I have therefore removed this section, just presenting the results of the OC_3D_PHY run alongside the BGC-Argo assimilation. I will potentially develop the ocean colour assimilation work further in a future publication.

I also agree that presenting further analysis of the OSSE framework would help demonstrate its credibility. I have included the types of assessment requested, as detailed in turn below.

Upon appropriately addressing these comments, I'll recommend publication of the manuscript in Biogeosciences.

1. I question on the value of including section 2 on comparing different update strategies when assimilating surface chl data for following reasons:

1) While I acknowledge the efforts and time needed for comparing 6 different DA strategies, I feel that the present comparison is not sufficient for fairly selecting the best DA strategy. I would argue that the more sophisticated nitrogen balancing scheme failing to outperform other strategies is largely because the parameter values used in the scheme are directly adopted from Hemmings et al. (2008) without proper tuning. These parameters reflect the BGC model's inherent relationships between chl and other model state variables. Since the model used here (MEDUSA) has quite different structure from that of Hemmings (HadoCC), a careful calibration of the parameters in the N balancing scheme is needed before its usage. That maybe contribute to a separate manuscript focusing on the benefit of multivariate BGC update over single-variable update.

I agree with this assessment. Recalibrating the parameterisations of the Hemmings et al. (2008) nitrogen balancing scheme for specific use with MEDUSA would be a considerable amount of work, but necessary to achieve the best results. As suggested, I have therefore

removed this section from the current manuscript, and will potentially develop the work further as part of a separate paper. I now restrict discussion of different multivariate balancing options to the following in Section 5:

"An alternative approach could be to use the nitrogen balancing scheme of Hemmings et al. (2008), which explicitly updates several model state variables to try and account for differing errors in phytoplankton growth and loss processes. This has been successfully used in previous ocean colour assimilation studies (Ford et al., 2012; Ford and Barciela, 2017; Ford, 2020) with the HadOCC model (Palmer and Totterdell, 2001). It was originally designed and tuned for use with HadOCC, so requires further development and tuning for use with the more complex MEDUSA, but an initial implementation for MEDUSA has been developed. Such a scheme offers more potential for controlling the wider biogeochemical state, especially if it could be expanded to alter parameter values as well as state variables."

2) At its present form, I didn't see strong connection between section two and three in the manuscript. To me, the most significant findings are from section three and this section stands out even if section two is completely removed. This is because the comparison in section two didn't suggest a clear winner and the ultimate decision of using the DA strategy of an intermediate complexity rather than the most sophisticated one (the N balancing scheme) for section three further reduce the value of including the entire comparison in section two.

I agree that the most significant findings, in relation to the core aims of the study, are in the final section, and that this section deserves the most attention.

3) If section two was removed, the author can have more space to elaborate and focus on the impacts of assimilating BGC-Argo profiles on different variables and suggesting directions for future work to improve. Currently, I feel the discussion on this part is relatively short compared to the emphasis it receives in the title, abstract and Introduction.

I agree. I have therefore removed the ocean colour section, and used the extra space to further develop the assessment and discussion surrounding BGC-Argo assimilation, as detailed below and in response to Referee #2.

2. I think section one on establishing the OSSEs framework is key to the credibility of assessment on assimilation impact. Presently the only analysis provided to show the credibility of OSSEs is a comparison of the errors between FREE and OBS and between FREE and NATURE for surface chl, NO3 and pCO2 in Figure 2. According to the criteria of designing rigorous ocean OSSE system detailed in Halliwell et al. (2014), I would request the author to comment and/or provide some information on following aspects:

1) Can the NATURE run reasonably capture the key features measured by the observing systems (in this case the surface chla, and the BGC profiles)? The author refers the performance of NATURE run to references given in Section 2 which is not very clear to me which one exactly has the same configuration and time period as the one here. A brief summary and/or some figures on the performance of the NATURE run will help.

The references in Section 2 detail assessment of the model components used, but not of the specific NATURE run which is newly presented in this manuscript. I have clarified this in the text:

"Validation of the general performance of the different system components can be found in the references given in Section 2, and validation of the nature run is presented in Section 4.1."

I have also presented some figures and assessment in Section 4.1 of the revised manuscript demonstrating the performance of the NATURE run as requested. The new Fig. 2 and Fig. 4, and surrounding text, compare surface fields and zonal mean sections between observation-based products, NATURE, and FREE.

2) Figure 2 only presents the surface comparisons. Since the assessment involves the vertical profiles, can the author also comment on whether the errors between FREE and NATURE are comparable to those between FREE and OBS in terms of the vertical distribution pattern of observable BGC variables.

I have added a new Fig. 4, with surrounding assessment in Section 4.1, which shows zonal mean sections down to 2000 m for BGC variables for which there is an observation-based product to compare against. The errors between FREE and NATURE and FREE and OBS are broadly comparable, but often slightly smaller between FREE and NATURE. This is noted and discussed in the revised manuscript:

"For regions and variables where the errors between FREE and NATURE were too low, the potential result may be to underestimate the impact of assimilating dense data, in this case ocean colour, and overestimate the impact of assimilating sparse data, in this case BGC-Argo (Halliwell et al., 2014). This should be borne in mind when drawing conclusions."

3) How about the error growth rate? One important criterion of credible OSSE evaluation is that the differences between the FREE and the NATURE ("truth") grow at the same rate as errors that develop between the state-of-the-art ocean models and the true ocean (Halliwell et al. 2014).

I agree this is important to demonstrate, and I have added a new Fig. 5, with surrounding assessment in Section 4.1, demonstrating that this is the case.

Halliwell, G. R., Srinivasan, A., Kourafalou, V., Yang, H., Willey, D., Le Hénaff, M., and Atlas, R.: Rigorous evaluation of a fraternal twin ocean OSSE system for the open Gulf of Mexico, J. Atmos. Ocean Tech., 31, 105–130, https://doi.org/10.1175/JTECH-D-13-00011.1, 2014.

Specific comments:

L21: '… half the planet's primary production.' Reference?

I have added a reference to Field et al. (1998, http://doi.org/10.1126/science.281.5374.237).

L75-77: Could you briefly add the outcome of assimilating these BGC-Argo observations in these two studies?

I have added the following:

"For instance, BGC-Argo observations of O_2 have been assimilated by Verdy and Mazloff (2017), who produced a five-year state estimate of the Southern Ocean using an adjoint method, and were able to capture over 60% of the variance in oxygen profiles at 200 m and 1000 m depth. Furthermore, Cossarini et al. (2019) assimilated BGC-Argo profiles of Chl-a into a model of the Mediterranean Sea, and found this was successful in adjusting the shape

of chlorophyll profiles, and that with the present number of BGC-Argo floats they could constrain phytoplankton dynamics in up to 10% of the Mediterranean Sea."

L87-88: 'Two groups perform … and the Met Office … presented here.' this information may be meaningful for the groups involved but doesn't seem informative for general readers. Do the two groups aim at different perspectives of the BGC assessment? What are they then?

I have rephrased this section to be less focussed on the details of the project, and instead give a brief summary of the results of Germineaud et al. (2019):

"Two groups performed OSSEs assessing biogeochemistry, Germineaud et al. (2019) and this study. Germineaud et al. (2019) presented a probabilistic evaluation at a single assimilation time step, finding that Chl-a from BGC-Argo floats added value at surface locations where ocean colour was unavailable, and at depth."

L116-118: Isn't that oxygen and dissolved inorganic N are also simulated? Or are they implicitly included in the 'coupled carbon cycle'? It's not clear to me what 'coupled' means here.

This sentence was poorly phrased and incomplete. Oxygen and DIN are indeed also simulated, and the word "coupled" was unnecessary and confusing in this context. The revised text reads:

"MEDUSA is of intermediate complexity, representing two phytoplankton and two zooplankton types, and the cycles of nitrogen, silicon, iron, carbon, and oxygen."

L154: This is acceptable, but could you comment if the physics of the NATURE run without DA is reliable to conduct the OSSEs?

The NATURE run does capture key features of the physics. To demonstrate this, I have added an assessment of surface temperature to the new Fig. 2 and Fig. 3, with surrounding assessment in Section 4.1.

L156: Just curious if there is any particular reason for using log10 instead of log-normal transformation.

In practice, it should make no difference to the assimilation whether log-normal or $\log_{10}$ is used. The shape of the distribution is the same (the ratio of log(x) to $\log_{10}$(x) is identical for all values of x), except that $\log_{10}$ gives a smaller variance. It is the shape of the distribution that matters for the assimilation, so as long as the same transformation is applied consistently to both model and observations, it should not matter whether log-normal or $\log_{10}$ is used. In the literature, some studies use $\log_{10}$ (e.g. Gregg, 2008, https://doi.org/10.1016/j.jmarsys.2006.02.015), while others use log-normal (e.g. Ciavatta et al., 2011, https://doi.org/10.1029/2011JC007219). The decision to use $\log_{10}$ here is a historical one, following Ford et al. (2012, https://doi.org/10.5194/os-8-751-2012), but as stated the choice should make no difference.

L160: How large is the correlation length-scale? Water et al. (2015) is on physical DA. Same length scale used for the BGC assimilation here? I'm thinking that BGC fields are more dynamic and thus have a smaller correlation length-scale.

The correlation length-scale is the same as in Waters et al. (2015, https://doi.org/10.1002/qj.2388), and varies with the Rossby radius, from a value of 25 km at low latitudes to 150 km at the Equator (see Fig. 3 of Waters et al., 2015,

https://doi.org/10.1002/qj.2388). For a 1/4° resolution ocean model, which is limited in its resolution of mesoscale features, using the same correlation length-scale for BGC as for physics is probably appropriate for an initial implementation. It is true though that the appropriate correlation length-scale(s) to use for assimilating BGC-Argo is an open question, and this should be addressed in future development of the assimilation.

I have clarified the length-scale in Section 2.2.2:

"In the horizontal, a correlation length-scale based on the first baroclinic Rossby radius was used, varying from a value of 25 km at low latitudes to 150 km at the Equator, consistent with Waters et al. (2015)."

I have also added the following to Section 5:

"Related to this, the correlation length-scales used by the assimilation should be appropriately tuned for biogeochemical variables. In this study, a single horizontal correlation length-scale based on the first baroclinic Rossby radius was used, varying from a value of 25 km at low latitudes to 150 km at the Equator, following the physics implementation of Waters et al. (2015). This may help explain why the BGC-Argo assimilation demonstrated less widespread impact at high latitudes than near the Equator. A different correlation length-scale may be appropriate for biogeochemical variables. Furthermore, NEMOVAR has recently been developed to allow the use of multiple correlation length-scales (Mirouze et al., 2016), so both small- and large-scale corrections can be considered."

L165-170: Can the surface information help constrain the BGC fields below the mixed layer? '… below the mixed layer the vertical length-scale increases with the model's vertical grid resolution.' this is confusing to me.

The length-scales are designed to limit the spreading of information across the base of the mixed layer. For an example see Fig. 4 of Waters et al. (2015, https://doi.org/10.1002/qj.2388) and the surrounding description in their Section 3.6. For surface observations the length-scale is set equal to the mixed layer depth, meaning that information from the surface observations is spread to the base of the mixed layer, but has limited impact on BGC fields below it. This is a deliberate decision based on the lack of correlation between water mass properties in and below the mixed layer. This is likely to be the case as much for BGC as for physics (see e.g. Fontana et al., 2013, https://doi.org/10.5194/os-9-37-2013). The vertical correlation length-scale is set to a minimum value at the mixed layer depth, and then increases with depth (see Fig. 4 of Waters et al., 2015, https://doi.org/10.1002/qj.2388). This increase is proportional to the increase in vertical model grid spacing that occurs with depth.

I have rephrased the description in the text to make this clearer. In Section 2.2.2:

"The vertical correlation length-scale is dependent on the model's mixed layer depth, as determined from a one-day model forecast. At the surface, the vertical correlation length-scale is set to the depth of the mixed layer, so that information from surface observations is spread to the base of the mixed layer but not below it."

In Section 2.2.3:

"The vertical correlation length-scale was flow-dependent and varies with depth, as detailed by Waters et al. (2015). At the surface the vertical correlation length-scale was set to the depth of the mixed layer, decreasing to a minimum value at the base of the mixed layer. This minimised the spread of information across the pycnocline, due to the lack of correlation of water mass properties in and below the mixed layer (Waters et al., 2015; Fontana et al.,

2013). Below the mixed layer, the vertical correlation length-scale increased with depth, proportional to the increase in vertical model grid spacing that occurs with depth."

L172: 'The increments … from the two methods should be similar, though not identical.' Why? Isn't that the two methods have different treatments below the mixed layer?

As I have removed the section comparing ocean colour assimilation strategies, I have also removed this sentence.

L191: are these ratios fixed or time-dependent?

Again, as I have removed the section comparing ocean colour assimilation strategies, I have also removed this section. But to answer the question, the ratios are time-dependent, based on the background ratios in each assimilation cycle.

L216-217: this sentence should be reworded, something like: 'The approach taken to the assimilation of partial pressure of CO2 (pCO2) into HadOCC (While et al., 2012) is therefore adopted here with pH. In HadOCC, pCO2 is a function of temperature, …'

I have reworded the sentence as suggested:

"The approach taken to the assimilation of partial pressure of CO_2 (pCO_2) into HadOCC (While et al., 2012) was therefore adopted here with pH. In HadOCC, pCO_2 is a function of temperature, salinity, DIC, and alkalinity, and at constant temperature and salinity constant lines of pCO_2 are found in DIC/alkalinity space (see Fig. 1 of While et al. (2012))."

L261: Fujii et al. 2019 suggested the assimilative model to be configured either in reduced resolution or sufficiently different physical parameterizations.

I have clarified this in the text:

"The nature run is often run either at higher resolution than the assimilative model, or with significantly different parameterisations (Fujii et al., 2019)."

L272: 'year 5000', is it true or typo?

This is true. Spinning up UKESM1 for CMIP6 was a massive endeavour, as documented by Yool et al. (2020, https://doi.org/10.1029/2019MS001933).

L311: 30% is fine for estuarine and coastal waters, but would it be too large for chl-a profiles in open ocean?

30% is a commonly used value in open ocean chlorophyll assimilation studies (e.g. Pradhan et al., 2020, https://doi.org/10.1029/2019JC015586), and especially for daily products I would expect this to be appropriate. For instance, Krasemann et al. (2017, https://esa-oceancolour-cci.org/sites/esa-oceancolour-cci.org/alfresco.php?file=6d534e45-fbfd-4cc5-8125-d84f0b3abea6&name=OC-CCI-PVIR-v3.20170303.pdf) found an RMSD of 0.31 for ocean colour matchups of $\log_{10}$(chl-a) against in situ observations. Maritorena et al. (2010, https://doi.org/10.1016/j.rse.2010.04.002, Fig. 10) estimated errors to be in excess of 30% across much of the ocean for daily products, though lower for monthly composites.

I have modified the text as follows:

"A standard deviation of 30 % was agreed on within AtlantOS for Chl-a from ocean colour, a value commonly used in assimilation studies (Pradhan et al., 2020)."

L327: for these variables, are the error standard deviations fixed or monthly varying as well?

They are fixed, I have made this clearer in the text:

"Observation error standard deviations were set to a climatological constant equal to the average global observation error specified. These were fixed in time, and specified as …"

L357 & Table 1: would it be clearer to reserve the term 'control run' for the definition in Eq 2 only and call the 'non-assimilative run' the 'free run' throughout the text?

It would, I agree. I have changed that throughout the text.

L391: What's the DA impact for depths below 250 m?

As suggested above, I have removed this section. But the impact reduces quickly below 250 m, as could be seen in the original Fig. 4.

Figure 7 does not include O2 or pH while Figure 8 does. What's the rationale of presenting different set of variables here?

Fig. 5-7 were chosen to match the variables shown in Fig. 2, with extra variables just presented in Fig. 8 to limit the number of figures. I agree that it would be better to expand the range of information presented in this section, as suggested in detail by Referee #2, and so I have presented extra assessment accordingly. The original Fig. 5-7 have now been amalgamated into a new Fig. 7, with O_2 and pH also added. Surrounding assessment is presented in Section 4.2.

L522: Any comment on why O2 is not improved by BGC-Argo data? And why "in situ technologies such as gliders" can play a role?

I have added the following to Section 5:

"All assimilated variables were greatly improved below the mixed layer, but at the surface more limited improvements were seen for Chl-a and O_2, than for NO_3 and pH. This is likely due to the relative importance of top-down versus bottom-up control for these variables, and the density of data required for the assimilation to have a major impact. In the case of NO_3, and DIC which helps control pH, concentrations typically increase with depth, and the supply of NO_3 and DIC from below the mixed layer is a major contribution to surface values. Therefore, changes at depth due to the assimilation will alter surface values through indirect processes. O_2 and Chl-a typically decrease in concentration with depth, and dynamics within the mixed layer are much more important in setting surface values. For O_2, major drivers are temperature and ocean–atmosphere exchange. For Chl-a a major driver is light availability. It seems that the BGC-Argo data was too sparse, even in ARGO_FULL, to have a widespread impact in these circumstances. More dedicated observations within the mixed layer may be likely to have more of an impact on surface values. For Chl-a, this can be successfully provided by ocean colour, as the results of this study show. For O_2 and other variables, alternative in situ observing technologies such as gliders may be able to play a role (Telszewski et al., 2018)."

Response to Referee #2

The manuscript by Ford presents an OSSE experiment to investigate a number of assimilation strategies for ocean colour (OC) and biogeochemical Argo (BGC-Argo) observations using an already published DA method. The simulations are performed using a global model and sets of synthetic observations that resemble the current L3 chlorophyll OC and two potential arrays of BGC-Argo based on the current Argo network.

Thank you for your review and constructive comments. I will answer these in turn below.

The manuscript is well written and the performed modelling experiments allow novel and useful insights on the integration of BGC-Argo and OC data into global model assimilation. However, as presented, results seem rather superficial. The work would be the basis for a very valuable paper, but the results need to be explored further before I can recommend publication. My main concern is that results present a single month of simulation (i.e., the last month of a 2-year simulation: 1-year spinup and 1-year assimilation) and a single global statistics (e.g., Figs. 3, 4 and 10). Conclusion/discussion on assimilation strategies and impact of the observing systems are possibly misled by the limited results. Results of the whole year of the assimilation runs should be presented and the MEAred maps and profiles enriched with spatial and temporal statistics to provide quantitative insights on the impact of BGC-Argo in different seasons and regions of the global ocean. In particular, which are the areas and seasons that could benefit most from BGC-Argo assimilation and the integration of the two observing systems? I feel that the manuscript misses the objective to provide useful indications to design future observing system strategy, as it is proposed in the title.

Thank you for your comments. I agree that more detailed assessment would be useful, and have greatly expanded the assessment presented in Section 4.2, as suggested.

A second issue concerns the comparison between the PHY (phytoplankton update) and NIT (nitrogen balancing scheme) assimilation schemes. While the novelty of the OSSE experiment is related to the integration of OC and BGC Argo, the lack of the NIT implementation for BGC-Argo assimilation is a significant limitation of the manuscript that should be discussed. L425-427 are misleading. In fact, the choice of the PHY update scheme is explained at L229-230 (i.e., apparently, the NIT method has not been implemented for the joint BGC-Argo and OC assimilation). The joint OC and BGC-Argo assimilation strategy should be clearly described at L347-348. Then, it is not clear the objective of the first set of experiments (OC assimilation, which conclusions are mainly already published) if its results are not used for the second set of experiments. Even if the NIT method has not been implemented, the manuscript can be completed by a more unbiased discussion on pro and cons of the two methods and by presenting the work to be done and the the benefits to have the NIT method working with the BGC-Argo assimilation. In fact, some conclusions seem misleading. While the nitrogen balancing scheme is reported as the method with more potentiality (L406-407 in results and L510-L515 in discussion), results on BGC-Argo assimilation runs do not support this conclusion. I suggest to clarify better the proposed assimilation strategy and to detail better what would be needed to have the NIT method working with BGC-Argo assimilation. The conclusion that "only minimally altered for use with MEDUSA, so more specific tuning may help (L516-517)" seems misleading.

Referee #1 also questioned the value of including the comparison of different ocean colour assimilation strategies in this manuscript. Based on the feedback of both referees, I have

removed this section of the paper. This should result in a more focussed manuscript and avoid the confusion my description of the different strategies seems to have caused. It has also allowed more space to further explore the results of the BGC-Argo experiments. I now restrict discussion of different multivariate balancing options to the following in Section 5:

"An alternative approach could be to use the nitrogen balancing scheme of Hemmings et al. (2008), which explicitly updates several model state variables to try and account for differing errors in phytoplankton growth and loss processes. This has been successfully used in previous ocean colour assimilation studies (Ford et al., 2012; Ford and Barciela, 2017; Ford, 2020) with the HadOCC model (Palmer and Totterdell, 2001). It was originally designed and tuned for use with HadOCC, so requires further development and tuning for use with the more complex MEDUSA, but an initial implementation for MEDUSA has been developed. Such a scheme offers more potential for controlling the wider biogeochemical state, especially if it could be expanded to alter parameter values as well as state variables."

Minor points:
Line 173: why should the two methods provide similar increments? One is uniform with depth from surface to MLD depth, while the second method is not limited to the MLD and vertical increments are mediated by covariance.

I have removed this sentence, as it was a source of confusion. It is true that in the 3D method vertical increments are mediated by covariance, which will be a source of differences. To clarify though, in the 3D method the vertical correlation length-scales are defined to allow surface information to be spread to the base of the mixed layer but not below it. I have rephrased the description of the assimilation method to make this clearer. In Section 2.2.2:

"The vertical correlation length-scale is dependent on the model's mixed layer depth, as determined from a one-day model forecast. At the surface, the vertical correlation length-scale is set to the depth of the mixed layer, so that information from surface observations is spread to the base of the mixed layer but not below it."

In Section 2.2.3:

"The vertical correlation length-scale was flow-dependent and varies with depth, as detailed by Waters et al. (2015). At the surface the vertical correlation length-scale was set to the depth of the mixed layer, decreasing to a minimum value at the base of the mixed layer. This minimised the spread of information across the pycnocline, due to the lack of correlation of water mass properties in and below the mixed layer (Waters et al., 2015; Fontana et al., 2013). Below the mixed layer, the vertical correlation length-scale increased with depth, proportional to the increase in vertical model grid spacing that occurs with depth."

Line 335: can the author provide some more details on how the observation and background errors have been matched? and which is the value of inflation?

The background error standard deviations estimated using the Canadian Quick method were output on the model grid, and the global mean value calculated. For each variable, the estimated standard deviations were multiplied by a constant so that the global mean value now matched the constant observation error standard deviations of 0.638 mmol m^{-3} for NO_3, 2.767 mmol m^{-3} for O_2, and 0.006 for pH. This meant that the global mean of each field matched the observation error standard deviations, while maintaining the spatial variation of the original estimates. I have added the following:

"In order to give sufficient weight to the observations for the assimilation to be effective, the background error standard deviations were inflated. This was achieved by multiplying the

gridded field of background error standard deviations for each variable by a constant, so that the global mean background error standard deviation matched the observation error standard deviation used for that variable. This meant that on average, equal weight was given to the background and to the observations, but the ratio of background to observation error varied spatially based on the estimates from the CQ method."

Line 336: this sentence is not clear: the observation error is set from the real global BGC-float array, so what will change when the system is functioning with real BGC-floats? More generally, the discussion missed to tackle this topic: how much are the results affected by the selected observation and background errors? would it be a different impact of the two observing systems using different observation errors?

The observation error will remain the same, but the background error will change. In part, this will be due to the different model parameterisations used in the OSSE framework compared with the standard model setup. Furthermore, the background error should reflect the error in the assimilative system, which will be dependent on the number and locations of BGC-Argo floats in the real ocean. It is likely that a different specification of the observation and background errors would give different quantitative results, but show a similar qualitative impact.

I have modified Section 3.4:

"Once the system is fully functioning with real BGC-Argo data available, the background error estimates can be appropriately refined, based on the errors in the real-world assimilative model, and the actual distribution of BGC-Argo observations."

I have added the following to Section 5:

"The background error standard deviation estimates also need to be refined once real BGC-Argo observations are being assimilated, to reflect the background error in the assimilative system, which will depend on the actual distribution of BGC-Argo floats. The average ratio of background to observation error may also differ from that assumed in this study. It is likely that this would give different quantitative results, but the qualitative impact of the assimilation would remain similar."

Table 3 can be enriched to improve the identification of the runs at a glance. I would suggest to add 2 new columns for the assimilated and updated variables, and to split the note column in two new ones: background error (i.e., vertical propagations: 2D and 3D) and type of increment.

Removing the comparison between ocean colour assimilation strategies means much of this information is now redundant, and I have simplified the table accordingly to aid identification of the runs.

L354-355: explain how MAEosse and MAEcontrol are computed for pixels in the maps of Figures 5-9 and for points of the profiles in Figures 3, 4 and 10. Which are the distribution compared?

For the maps, the MAE is calculated independently for each model grid cell by calculating the absolute difference between the model run and the nature run on each day of the given time period (the 31 days of December in this case), and then calculating the median of those 31 values. For the profiles, at each model depth level the absolute difference between the model run and the nature run on each day of the given time period is calculated, giving a set of values comprising of 31 days x 1442 longitudes x 1207 latitudes (with land points then excluded). The median of this set of values, weighted by the area of each grid cell, is then

calculated to give the global MAE value for that depth level. I have moved detail of metrics to be a new Section 3.6, and added the following:

"Where MAE_{red_abs} or $MAE_{red\_\%}$ is presented as a spatial map, the MAE was calculated independently for each model grid cell. This was done by calculating the absolute difference between the model run and the nature run in that grid cell on each day of the given time period, and calculating the median of those values. Where MAE_{red_abs} or $MAE_{red\_\%}$ is presented as a profile, the MAE was calculated independently for each model depth level. At each depth, the absolute difference between the model run and the nature run on each day of the given time period was calculated for each grid cell in the region of interest. The median of this set of values was calculated, weighted by the area of each grid cell, to give the MAE value for that depth level."

L377: the absolute differences in Atlantic and Indian Oceans appear significant (i.e., 1 order magnitude). Can the author provide more details about which modifications of the FREE run (w.r.t. NATURE run) should have served to increase nutrient concentration? And which modifications compensate it (L382)? I agree that achieving a global appropriate level of error with a complex model (with uniform parameterization) can not be managed, however the author should provide some details on which modifications didn't work as expected. This can be helpful in understanding the effectiveness of the data assimilation in those areas.

The nutrient differences in the Atlantic and Indian Oceans are of an order of magnitude, but with low absolute values, typically increasing from O(0.01) to O(0.1) mmol N m^{-3}. It is difficult to pinpoint the exact cause of any given change, as several parameters have been altered, and these will have complex interactions depending on the underlying concentrations of different variables. One potentially significant change though is that the nutrient uptake half-saturation concentration for phytoplankton was greatly increased for nitrogen, and decreased for iron. In areas which are nitrogen-limited, phytoplankton will therefore be less efficient at taking up nutrients. Furthermore, a decrease in zooplankton grazing half-saturation concentration means zooplankton become more efficient at grazing low phytoplankton populations. In NATURE, the Atlantic and Indian Oceans are the areas with the lowest DIN and phytoplankton concentrations. A first-order explanation may therefore be that the increased nitrogen uptake half-saturation concentration means phytoplankton take up less DIN, resulting in higher DIN concentrations. This then allows greater phytoplankton growth, as more DIN is available, though it is used less efficiently. This is then balanced by an increase in grazing, resulting in slightly elevated DIN and zooplankton concentrations, but largely unchanged phytoplankton concentrations. In other areas, which aren't so nitrogen-limited, the balance of processes is different, leading to different changes.

I have added the following to Section 4.1:

"The modifications introduced in FREE served to increase nutrient concentrations in these regions, but also to suppress phytoplankton growth, resulting in little overall change in Chl-a. This is largely the result of increasing the nitrogen nutrient uptake half-saturation concentration for phytoplankton, and decreasing the zooplankton grazing half-saturation concentration."

L385: the absolute difference of NO3 and pCO2 between FREE and NATURE seems much lower than that between FREE and real observation. Can the author discuss which are the implications of the low difference for the OSSE assimilation results? I would argue that the effectiveness of the assimilation might be limited in some areas by the low differences between the synthetic observations generated from NATURE and the FREE run. Further, I would argue that MAEred might not be a good metric

because of the MAEcontrol at the denominator in the areas where NATURE and FREE are so close.

From previous studies, the conclusion has been that an insufficient level of error would lead to "an overestimation of impact when sparse data are assimilated and an underestimation when dense (e.g., satellite) data are assimilated" (Halliwell et al., 2014, https://doi.org/10.1175/JTECH-D-13-00011.1). This suggests that this study may overestimate the impact of BGC-Argo observations in some regions. I mentioned this in the original manuscript, but agree it's a point which deserves further discussion. I have added the following to Section 4.1:

"For regions and variables where the errors between FREE and NATURE were too low, the potential result may be to underestimate the impact of assimilating dense data, in this case ocean colour, and overestimate the impact of assimilating sparse data, in this case BGC-Argo (Halliwell et al., 2014). This should be borne in mind when drawing conclusions."

I have added the following to Section 5:

"Furthermore, for some variables and regions the error between the free-running model and the nature run was smaller than might be expected in real-world systems, potentially leading to an overestimate of the quantitative impact of BGC-Argo data (Halliwell et al., 2014)."

I have also added some assessment of the absolute as well as percentage reduction in MAE, to avoid the issue of having $MAE_{control}$ in the denominator. This gives generally similar conclusions, but I agree is a useful additional way of looking at the results. I have described the metric in the new Section 3.6, added a new Fig. 8, and mentioned this metric throughout the assessment in Section 4.2.

L395: Since large areas of the global ocean are characterized by a DCM below 60m depth, the degradation of MAE below 60m would deserve a more detailed comment.

Given that phytoplankton biomass is not degraded, I would speculate that depth variations in carbon-to-chlorophyll ratios are not being correctly characterised in these runs. However, as I have removed the comparison of ocean colour assimilation strategies, and the remaining OC_3D_PHY run does not show this degradation, I have also removed this discussion from the revised manuscript.

L405: why should the similar behaviours of OC_2D and OC_3D demonstrate that the use of NEMOVAR to create 3D increments for the combined OC and BGC-Argo float assimilation is fit-for-purpose?

This comment was simply intended to imply that because OC_3D gives similar behaviour to the proven strategy of OC_2D, assimilating ocean colour in this manner (which is a prerequisite for combining OC and BGC-Argo chl-a) should also give acceptable results. As I have removed the comparison between ocean colour assimilation strategies, I have also removed this comment from the revised manuscript.

Figure 4: I would suggest to reduce y-axis to 0-250m depth for the chlorophyll, phytoplankton and zooplankton plots to increase their readability. Are the high positive values below 250 in chla and phytoplankton plots due to the very low values of those variables below the euphotic zone?

I have altered the figure as suggested. The high percentage values are indeed due to the low absolute values, and I have added comment on this in the text:

"Beneath the euphotic zone, where Chl-a was near-zero, the assimilation had little impact. Positive values of $MAE_{red\_\%}$ were seen below 250 m, but values of MAE_{red_abs} (not shown) tended to zero below about 220 m."

L450-451 and L477-478: Can the degradation of Alkalinity be due to an improper working of the smallest combined change of DIC and Alk with pH assimilation? A more detailed analysis is expected to show the pro and cons of the method when pH is used instead of pCO2.

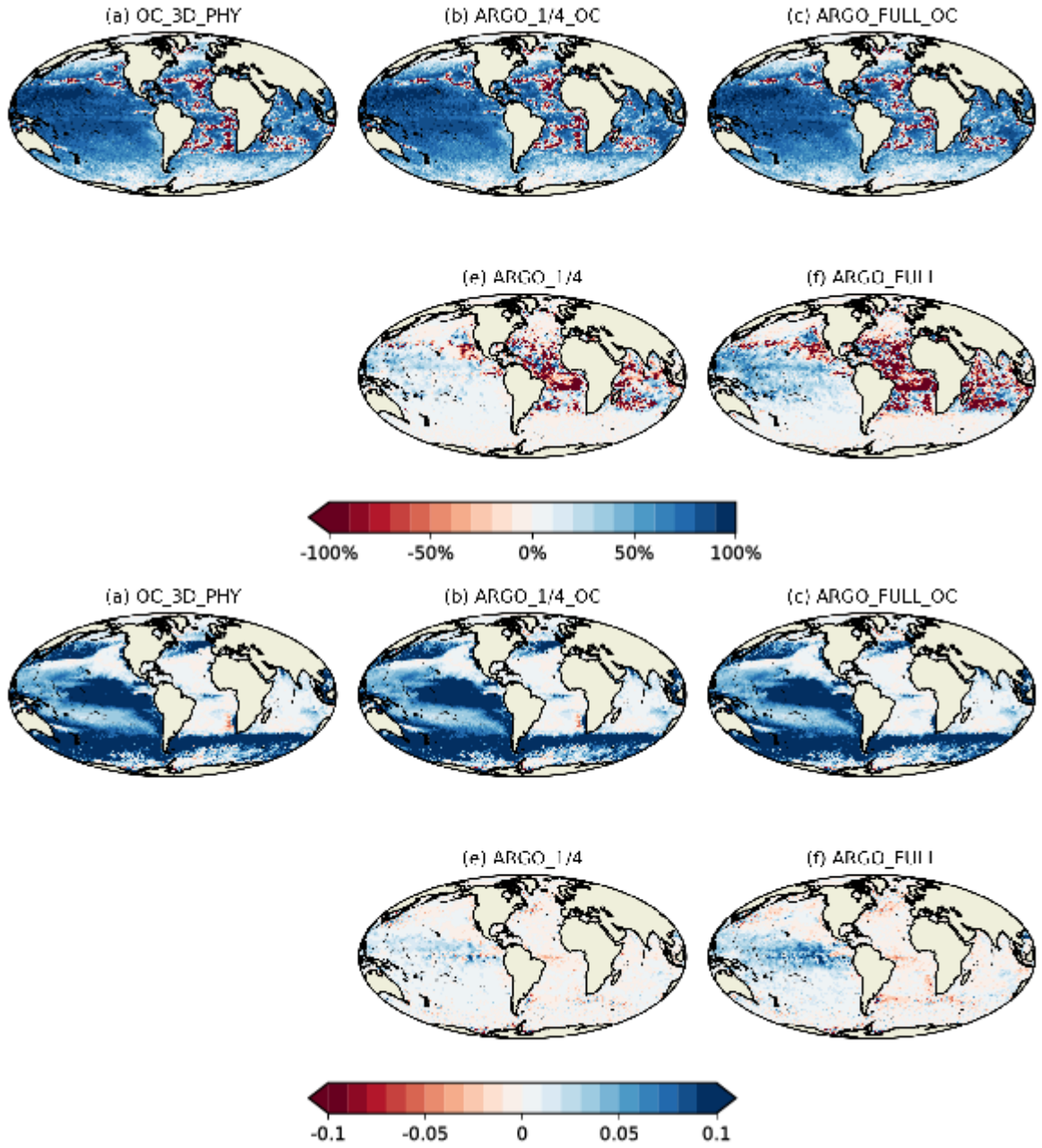
Testing of the scheme shows the calculation is being performed correctly, as also indicated by the overall improvement in pH. The cause is likely to be that making the smallest combined change to DIC and alkalinity is not necessarily the approach that minimises errors in both DIC and alkalinity. In some circumstances it might be more appropriate to e.g. make a smaller or no change to alkalinity, and a larger change to DIC. Or even to make a change of the opposite sign to alkalinity, and an even larger change to DIC to compensate. Unfortunately, without concurrent observations of DIC or alkalinity, this information is not known at the time of assimilation. This is equally the case whether pCO_2 or pH is being assimilated. An assumption therefore needs to be made, and during the development of the original pCO_2 assimilation scheme it was decided that the safest assumption would be to make the smallest combined change in DIC and alkalinity – an assumption adopted for pH in this study. In light of these results it may be worth revisiting this assumption, but to do so effectively would involve a great deal of experimentation which is best left for a future study.

I have added the following to Section 5:

"A novel method for assimilating pH was introduced in this study, following the method for assimilating pCO_2 developed by While et al. (2012). The method corrects pH, a diagnostic variable in the model, by making the smallest combined change to DIC and alkalinity required to reach the target pH value. This was successful in improving both pH and DIC, but alkalinity was often degraded. This highlights that making the smallest combined change to DIC and alkalinity does not guarantee that errors in both DIC and alkalinity are minimised. In some circumstances it might be more appropriate to make a smaller or no change to alkalinity, and a larger change to DIC. Or even to make a change of the opposite sign to alkalinity, and an even larger change to DIC to compensate. Unfortunately, without concurrent observations of DIC or alkalinity, this information is not known at the time of assimilation. This is equally the case whether pCO_2 or pH is being assimilated, and so it was decided that the safest assumption would be to make the smallest combined change to DIC and alkalinity. In light of these results it may be worth revisiting this assumption."

L454-L455: why are the negative values of MAEred in the Atlantic and Indian Oceans (Fig. 5) related to the compensating errors introduced in FREE? Since the FREE and NATURE differences are very low in those areas (Fig. 2), the impact of the assimilating synthetic observations (generated from NATURE) should be negligible. I wonder whether the MAEred is not a good metric because of the MAEcontrol at the denominator for those areas.

Below is the original version of Fig. 5, showing the percentage reduction in MAE, and an alternative version showing the absolute reduction in MAE. It is true that the response in the Atlantic and Indian Oceans is minimal in absolute terms. I have added a new Fig. 8 showing this, and modified the surrounding discussion in Section 4.2 accordingly. I have also used both percentage and absolute reduction in MAE throughout this section.



Figures 8 and 10 seem redundant and not necessary. For example, the MAEred over OC_3D_PHY of ARGO_FULL_OC (Fig. 8a and Fig 10a) provides the same information (except for the normalization of denominator) of the difference between MAEred over FREE of ARGO_FULL_OC and MAEred over FREE of OC_3D_PHY (Fig. 4a and Fig 5a and c). I suggest that the relative impact of adding BGC-Argo can be shown by a new table of the MAEred over FREE numeric values. The table can report values for selected regions and different seasons/months providing indications of which areas of the global ocean and periods of the year can benefit most by the BGC-Argo assimilation.

As suggested, I have removed these figures, and added new Fig. 10-12 showing the impact of the BGC-Argo assimilation throughout the year in different regions. Please see the revised Section 4.2 for the detailed assessment.

L505-508: which parameter settings between NATURE and FREE caused the degradation of the other variables? Can additional details be added?

The interaction between different parameter changes is complex, and varies depending on the underlying concentrations of each of the variables. The biggest impact on zooplankton though is likely to have come from alterations to the grazing half-saturation concentration, which was changed from 0.8 mmol N m^{-3} in NATURE to 0.36 mmol N m^{-3} in FREE for microzooplankton, and from 0.3 mmol N m^{-3} in NATURE to 0.135 mmol N m^{-3} in FREE for mesozooplankton. Other significant changes to the ecosystem dynamics are likely to have come from changing the nutrient uptake half-saturation concentrations for phytoplankton. For nitrogen, this was changed from 0.5 mmol N m^{-3} in NATURE to 2.13 mmol N m^{-3} in FREE for non-diatoms, and from 0.75 mmol N m^{-3} in NATURE to 3.195 mmol N m^{-3} in FREE for diatoms. For iron, this was changed from 0.00033 mmol Fe m^{-3} in NATURE to 0.00011 mmol Fe m^{-3} in FREE for non-diatoms, and from 0.00067 mmol Fe m^{-3} in NATURE to 0.00022 mmol Fe m^{-3} in FREE for diatoms.

I have added the following to Section 4.1:

"The modifications introduced in FREE served to increase nutrient concentrations in these regions, but also to suppress phytoplankton growth, resulting in little overall change in Chl-a. This is largely the result of increasing the nitrogen nutrient uptake half-saturation concentration for phytoplankton, and decreasing the zooplankton grazing half-saturation concentration."

I have added the following to Section 5:

"It seems likely that this degradation occurred due to the changed MEDUSA parameter settings between NATURE and FREE, meaning that the underlying processes were altered such that identical concentrations of phytoplankton now led to different concentrations of other variables. Relevant perturbations included changes to the grazing half-saturation concentration for zooplankton, and nutrient uptake half-saturation concentrations for phytoplankton."

L510-511 please explain: this seems not supported by results or references.

I have added the following:

"This will also be the case for assimilation schemes which use ensembles to generate cross-correlations between Chl-a and other model variables. These schemes are reliant on the model relationships between variables being correct, as it is these model relationships which the cross-correlations are based on. If the response of zooplankton to an increase in phytoplankton in the model ensemble differs from that in the real ocean, then the cross-correlations used in the assimilation will lead to a zooplankton response which follows the (incorrect) model rather than the real ocean, in exactly the same way as seen in this study."

L522: BGC-Argo assimilation has a small and positive impact on O2 as shown in Figure 10f (BGC-Argo assimilation). The degradation of O2 at surface seems due to OC assimilation (see figure 4f).

Agreed. This was clumsy wording on my part, and I have rephrased this in the revised manuscript:

"All assimilated variables were greatly improved below the mixed layer, but at the surface more limited improvements were seen for Chl-a and O_2, than for NO_3 and pH."

L531-L535: DA method improvements are important, but the paper does not really tackle this aspect; thus, those lines may fit better in the introduction and not as the last conclusion.

I believe that the Discussion section is the best place for this discussion, as it details future work and recommendations arising from the results presented. I acknowledge that the original manuscript could have done this more effectively, and have expanded the discussion around assimilation improvements, relating it better to the results. I also take the point that something else, such as recommendations on observing system strategies, would fit better as the last conclusion, and have changed this around accordingly.

Assimilating synthetic Biogeochemical-Argo and ocean colour observations into a global ocean model to inform observing system design

David Ford[1]

[1]Met Office, FitzRoy Road, Exeter, EX1 3PB, UK

Correspondence: David Ford (david.ford@metoffice.gov.uk)

Abstract. A set of observing system simulation experiments ~~has been performedto explore~~ was performed. These assessed the impact on global ocean biogeochemical reanalyses of assimilating chlorophyll from remotely sensed ocean colour, and ~~assess the potential impact of assimilating~~ in situ observations of chlorophyll, nitrate, oxygen, and pH from a proposed array of Biogeochemical-Argo (BGC-Argo) floats. Two ~~different~~ potential BGC-Argo array distributions were tested: one where biogeochemical sensors are placed on all current Argo floats, and one where biogeochemical sensors are placed on a quarter of current Argo floats. ~~This latter approximately corresponds to the proposed BGC-Argo array of 1000 floats. Different strategies for updating model variables when assimilating ocean colour were assessed. All similarly improved the assimilated variable surface chlorophyll, but had a mixed impact on the wider ecosystem and carbon cycle, degrading some key variables of interest.~~ Assimilating BGC-Argo data ~~gave no added benefit over ocean colour in terms of simulating surface chlorophyll, but for most other variables, including sub-surface chlorophyll, adding BGC-Argo greatly improved~~ greatly improved model results throughout the water column. This included surface partial pressure of carbon dioxide (pCO_2), which ~~was not assimilated but~~ is an important output of reanalyses. In terms of surface chlorophyll, assimilating ocean colour effectively constrained the model, with BGC-Argo providing no added benefit at the global scale. The vertical distribution of chlorophyll was improved by assimilating BGC-Argo data. Both BGC-Argo array distributions gave benefits, with greater improvements seen with more observations. From the point of view of ocean reanalysis, it is recommended to proceed with development of BGC-Argo as a priority. The proposed array of 1000 floats will lead to clear improvements in reanalyses, with a larger array likely to bring further benefits. The ocean colour satellite observing system should also be maintained, as ocean colour and BGC-Argo will provide complementary benefits. ~~There is also much potential to improve the use of existing observations, particularly in terms of multivariate balancing, through improving assimilation methodologies.~~

1 Introduction

Throughout the ocean, physical and chemical processes interact with a teeming ecosystem to affect all life on Earth. The upwelling of nutrient-rich waters fuels the growth of phytoplankton, which form the base of the marine food web and contribute half the planet's primary production (Field et al., 1998). Oxygen is required for marine and terrestrial life, and its availability depends on ocean circulation, solubility, and biological activity. Carbon is taken up at the sea surface, at a rate contingent on physics and biology, and transported throughout the ocean. Some is stored for centuries at vast depths, mitigating climate change. Some is quickly released back to the atmosphere. All these phenomena are regulated by an array of processes which display variability on a range of scales from milliseconds to millennia, and from nanometres to ocean basins.

Understanding, monitoring, and predicting these processes is key to addressing some of the biggest challenges facing humanity. Rising carbon dioxide (CO_2) emissions are leading to climatic changes which threaten severe impacts on people and ecosystems (IPCC, 2014). Uptake of carbon by the global ocean acts to mitigate these impacts, but the ocean carbon sink is highly variable and its future magnitude uncertain (McKinley et al., 2017). At the same time, when CO_2 dissolves in seawater it reacts with it, leading to a decrease in pH referred to as ocean acidification (Doney et al., 2009). This could have major consequences for marine life, particularly organisms which form calcium carbonate shells, which become at risk of dissolution if the seawater pH is too low. Changes in climate and eutrophication also appear to be leading to expanding "dead zones" in the ocean (Diaz and Rosenberg, 2008; Altieri and Gedan, 2015), where oxygen concentrations are too low for most organisms to survive. On shorter time-scales, primary production varies considerably due to natural climate variability such as the El Niño Southern Oscillation (ENSO), and changes can have profound impacts on higher trophic levels, and hence the fisheries and aquaculture on which an estimated 12 % of the global population rely for their livelihoods (FAO, 2016). All these factors and more are captured in a drive towards "Good Environmental Status" of national waters, as regulated by policies such as the Marine Strategy Framework Directive (MSFD) of the European Union (EU).

Comprehensively monitoring all relevant processes in the global ocean, and their variability and trends, is not a trivial task. For ocean biogeochemistry, the global observing system consists of various components which, while often sparse and disparate, have allowed fundamental insights. Two decades of routine satellite ocean colour data (Groom et al., 2019) have yielded unprecedented knowledge about phytoplankton variability (Racault et al., 2017), and even helped overturn decades of scientific consensus on bloom formation (Behrenfeld and Boss, 2014). In situ stations such as the Bermuda Atlantic Time Series (BATS) have allowed long-term monitoring of multiple variables at fixed locations (Bates et al., 2014), and various networks of ships, gliders, and moorings give ongoing views of different aspects of the global ocean (Telszewski et al., 2018). These observation networks are vital, and have transformed our understanding of ocean biogeochemistry. But they remain sparse, and coverage is insufficient to address all outstanding scientific questions, or provide comprehensive monitoring on a global scale.

Observation of ocean physics has been revolutionised by the advent of Argo (Roemmich et al., 2019). A global array of around 4000 autonomous floats drift at a typical parking depth of 1000 m, and every ten days descend to 2000 m before rising to the surface, profiling temperature and salinity as they do so. The data are then transmitted to satellites in near-real-time,

before the float returns to its parking depth. Argo has facilitated breakthroughs in climate science (Wijffels et al., 2016), and improvements in physical ocean reanalyses and forecasts (Davidson et al., 2019).

The Argo initiative is now being extended to biogeochemistry through the Biogeochemical-Argo (hereafter BGC-Argo) programme (Biogeochemical-Argo Planning Group, 2016; Roemmich et al., 2019). In the next decade, it is planned to establish a global array of 1000 BGC-Argo floats, which are Argo floats equipped with biogeochemical sensors. The aim is for all these floats to measure six core variables: oxygen concentration (O_2), nitrate concentration (NO_3), pH, chlorophyll-a concentration (~~chl-a~~Chl-*a*), suspended particles, and downwelling irradiance. This promises to transform scientific understanding of ocean biogeochemistry. Thanks to a series of regional programmes, there are already over 300 operational floats measuring one or more biogeochemical variables. Few of these floats yet measure all the core BGC-Argo variables, and spatial coverage is highly uneven, but important scientific discoveries regarding phytoplankton, carbon, and nutrient dynamics have been made (Roemmich et al., 2019).

The value of observations can be further enhanced by combining them with numerical models using data assimilation (Kalnay, 2003). Ocean colour data are increasingly assimilated in state-of-the-art reanalysis (Rousseaux and Gregg, 2015; Ciavatta et al., 2016; Ford and Barciela, 2017) and forecasting (Teruzzi et al., 2014; Skákala et al., 2018) systems. This has consistently been shown to improve simulations of phytoplankton, but the impact on other model variables, especially sub-surface, is limited (Gehlen et al., 2015). Physical data assimilation has the potential to improve biogeochemistry, but has often been found to have the opposite effect, due to spurious impacts on vertical mixing to which biogeochemical variables are particularly sensitive (Park et al., 2018; Raghukumar et al., 2015). Assimilating multivariate in situ biogeochemical data should help address these issues and greatly improve reanalyses and forecasts (Yu et al., 2018), but due to the sparsity of observational coverage, efforts have largely been limited to parameter estimation (Schartau et al., 2017), 1D models (Torres et al., 2006), individual research cruises (Anderson et al., 2000), or surface-only carbon data (Valsala and Maksyutov, 2010; While et al., 2012). Furthermore, in situ biogeochemical observations are rarely available in near-real-time, limiting their suitability for operational applications.

The increasing availability of BGC-Argo data promises to change this, with great potential for improving reanalyses and forecasts (Fennel et al., 2019). For instance, BGC-Argo observations of O_2 ~~in the Southern Ocean~~ have been assimilated by Verdy and Mazloff (2017), ~~and Cossarini et al. (2019) have assimilated profiles of chl-a in the~~ who produced a five-year state estimate of the Southern Ocean using an adjoint method, and were able to capture over 60% of the variance in oxygen profiles at 200 m and 1000 m depth. Furthermore, Cossarini et al. (2019) assimilated BGC-Argo profiles of Chl-*a* into a model of the Mediterranean Sea, and found this was successful in adjusting the shape of chlorophyll profiles, and that with the present number of BGC-Argo floats they could constrain phytoplankton dynamics in up to 10% of the Mediterranean Sea.

This paper describes the development of a scheme to assimilate profiles of ~~chl-a~~Chl-*a*, NO_3, O_2, and pH into an updated version of the Met Office's global ~~physical-biogeochemical~~ physical–biogeochemical ocean reanalysis system. A set of observing system simulation experiments (OSSEs) (Masutani et al., 2010) is presented to assess the potential value of different numbers of BGC-Argo floats.

~~This~~ The work forms part of a coordinated effort within the EU Horizon 2020 research project AtlantOS (https://www.atlantos-h2020.eu). ~~AtlantOS had the aim of transforming various loosely-coordinated components into a "sustainable, efficient, and fit-for-purpose" Integrated Atlantic Ocean Observing System (IAOOS), consistent with the Framework for Ocean Observing (Lindstrom et al., 2012). One work package focussed on observing system design studies, using OSSEs to assess potential future improvements to existing and forthcoming components of the IAOOS.~~ Four groups performed OSSEs assessing physics observations, the results of which have been synthesised by Gasparin et al. (2019). Two groups performed OSSEs assessing biogeochemistry, ~~the Institute of Environmental Geosciences (IGE), and the Met Office. The IGE experiments have been published by Germineaud et al. (2019) , and the Met Office experiments are presented here~~Germineaud et al. (2019) and this study. Germineaud et al. (2019) presented a probabilistic evaluation at a single assimilation time step, finding that Chl-*a* from BGC-Argo floats added value at surface locations where ocean colour was unavailable, and at depth.

The biogeochemistry OSSEs consider two potential scenarios: 1) a global BGC-Argo array equivalent to having biogeochemical sensors on one in four existing Argo floats, which is comparable to the planned target of 1000 floats, and 2) a global BGC-Argo array equivalent to having biogeochemical sensors on all existing Argo floats. The aims were to assess the impact on reanalysis and forecasting systems that might be seen by assimilating multivariate BGC-Argo data, the influence of array size, and the value BGC-Argo would add to the existing ocean colour satellite constellation. Assimilation of physics variables was not included, due to the issues mentioned above, and reflecting the way state-of-the-art biogeochemical reanalyses are run (Fennel et al., 2019).

This paper describes the updated model and newly-developed assimilation scheme, and setup of the OSSEs. Results are ~~presented exploring ways to use ocean colour data assimilation to make multivariate updates, and combine it with in situ chl-a profiles. Results are~~ then presented showing the impact of assimilating the two potential BGC-Argo arrays, with and without ocean colour data. Finally, recommendations are made for the future development of observing and assimilation systems.

2 Model and assimilation

The reanalysis system is an upgraded version of that used in previous biogeochemical data assimilation studies at the Met Office ~~(Ford et al., 2012; While et al., 2012; Ford and Barciela, 2017; ?)~~(Ford et al., 2012; While et al., 2012; Ford and Barciela, 2017; Ford, 202
.

2.1 Model

The physical ocean model used is the GO6 configuration (Storkey et al., 2018) of the Nucleus for European Modelling of the Ocean (NEMO) hydrodynamic model (Madec, 2008), using the extended ORCA025 tripolar grid, which has a horizontal resolution of 1/4° and 75 vertical levels. This is coupled online to the GSI8.1 configuration (Ridley et al., 2018) of the Los Alamos Sea Ice Model (CICE) (Hunke et al., 2015). Together, these form the ocean and sea ice components of the GC3.1 configuration (Williams et al., 2017) of the Hadley Centre Global Environment Model version 3 (HadGEM3), which is used for physical climate simulations submitted to the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al.,

2016). When combined with the physics version of the data assimilation scheme described below, the ocean and sea ice models are also used in version 14 of the Forecasting Ocean Assimilation Model (FOAM), earlier versions of which are described by Blockley et al. (2014) and Storkey et al. (2010). FOAM is run operationally at the Met Office to produce short-range forecasts of the physical ocean and sea ice state. It is also used to initialise the ocean and sea ice components of the Met Office Global Seasonal forecasting system version 5 (GloSea5) (MacLachlan et al., 2015; Scaife et al., 2014), and short-range coupled ocean–atmosphere forecasting system (Guiavarc'h et al., 2019).

The biogeochemical ocean model used in this study is version 2 of the Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification (MEDUSA) (Yool et al., 2013). MEDUSA is of intermediate complexity, representing two phytoplankton and two zooplankton types, ~~with a variable carbon to chlorophyll ratio and a coupled carboncycle~~and the cycles of nitrogen, silicon, iron, carbon, and oxygen. This differs from previous versions of the Met Office ~~physical-biogeochemical~~ physical–biogeochemical ocean reanalysis system (Ford and Barciela, 2017), which used the Hadley Centre Ocean Carbon Cycle Model (HadOCC) (Palmer and Totterdell, 2001). This is because, following an intercomparison (Kwiatkowski et al., 2014) of biogeochemical models developed in the UK, MEDUSA was chosen to be the ocean biogeochemical component of version 1 of the UK Earth System Model (UKESM1) (Sellar et al., 2019). UKESM1 consists of a lower-resolution version of GC3.1, coupled with models of ocean biogeochemistry, atmospheric chemistry and aerosols, and ice-sheets, and is used for Earth system climate simulations submitted to CMIP6. Using the same model versions for forecasting, reanalysis, and climate simulations provides a seamless framework for simulating the Earth system (Martin et al., 2010).

2.2 Assimilation

2.2.1 Overview

The data assimilation scheme used here is version 5 of NEMOVAR (Weaver et al., 2003, 2005; Mogensen et al., 2009, 2012), following the implementation for assimilating physical variables into the global FOAM system (Waters et al., 2015), and for assimilating ocean colour data into HadOCC ~~(?)~~ (Ford, 2020) and the European Regional Seas Ecosystem Model (ERSEM) ~~(Skákala et al., 2018)~~(Skákala et al., 2018, 2020). As detailed in Waters et al. (2015), this implementation of NEMOVAR uses a first guess at appropriate time (FGAT) 3D-Var methodology. A conjugate gradient algorithm is used to iteratively minimise the cost function

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x}) \quad (1)$$

where the increment $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_{\mathrm{b}}$ is the difference between the state vector $\mathbf{x}$ and its background estimate $\mathbf{x}_{\mathrm{b}}$, the innovation vector $\mathbf{d} = \mathbf{y} - H(\mathbf{x}_i)$ is the difference between the observation vector $\mathbf{y}$ and $\mathbf{x}_i = \mathbf{M}_{\mathrm{t_0}\rightarrow\mathrm{t}_i}(\mathbf{x}_{\mathrm{b}})$, where $\mathbf{M}_{\mathrm{t_0}\rightarrow\mathrm{t}_i}$ is the nonlinear propagation model that propagates the background state to the state at time i, operated on by the observation operator H, $\mathbf{H}$ is the linearised observation operator, $\mathbf{B}$ is the background error covariance matrix, and $\mathbf{R}$ is the observation error covariance matrix. A diffusion operator is used to efficiently model spatial correlations (Mirouze et al., 2016). The observation operator

forms part of the NEMO code, and computes model values in observation space by interpolating model fields to observation locations at the closest model time step to the time each observation was made. The observation operator was extended in this study to work for 3D biogeochemical variables as well as physical variables.

When applied to physics data, NEMOVAR decomposes the full multivariate background error covariance matrix into an unbalanced and a balanced component for each variable. The balanced component is derived using a set of linearised balance operators, based on physical relationships (Weaver et al., 2005; Mogensen et al., 2012). In this study, NEMOVAR has been applied to biogeochemical variables with no multivariate relationships applied, and the cost function is minimised separately for each assimilated variable. Development of biogeochemical balance relationships within NEMOVAR could be expected to bring improved results, but this would be a major development to NEMOVAR. The aim of this study was to develop an initial implementation that could be used with BGC-Argo data, and that can be further developed as systems mature.

All increments are applied to the model over one day using incremental analysis updates (IAU) (Bloom et al., 1996), which applies an equal proportion of the increments at each model time step, in order to reduce initialisation shocks.

NEMOVAR is used in this study to assimilate simulated ocean colour and BGC-Argo data, as described in the following sections. NEMOVAR can be used for combined physical–biogeochemical assimilation ~~(?)~~(Ford, 2020), but physics data ~~is~~ was not assimilated in this study.

2.2.2 Ocean colour

NEMOVAR ~~is~~ was used here to assimilate total surface $\log_{10}$(~~chl-a~~Chl-*a*) from ocean colour. Since MEDUSA simulates ~~chl-a~~ Chl-*a* for two phytoplankton types, diatoms and non-diatoms, these ~~are~~ were summed by the observation operator to give total ~~chl-a~~Chl-*a*, to match the input observations. Log-transformation ~~is~~ was performed in order to give a more Gaussian error distribution (Campbell, 1995). The background and observation error covariances used ~~are~~ were the same as in ~~?~~Ford (2020). In the horizontal, a correlation length-scale based on the first baroclinic Rossby radius ~~is used,~~ was used, varying from a value of 25 km at low latitudes to 150 km at the Equator, consistent with Waters et al. (2015).

For surface data, such as ocean colour, NEMOVAR can be applied in one of two ways~~:~~

. The first, which is computationally most efficient and has been used in previous ocean colour assimilation studies (Ford, 2020; Skákala e , is to calculate a set of 2D surface increments ~~. These can then be~~ which are applied equally through the mixed layer~~, as in ? and Skákala et al. (2018).~~

. The second is to calculate a set of 3D increments, with information from the surface observations propagated downwards using vertical correlation length-scales, as described by Waters et al. (2015) for physical variables. The sub-surface background error standard deviations are parameterised based on the vertical gradient of density with depth to allow a flow-dependent vertical structure to the errors. The vertical correlation length-scale is dependent on the model's mixed layer depth~~: at the surfacethe vertical~~ , as determined from a one-day model forecast. At the surface, the vertical correlation length-scale is set to the depth of the mixed layer, ~~at the base of the mixed layer the vertical length-scale is set to a minimum value, and below the mixed layer the vertical length-scale increases with the model's vertical grid resolution. This means~~ so that information from surface observations is spread to the base of the mixed layer ~~, as determined from a one-day model forecast.~~

~~The increments applied to the model from the two methods should be similar, though not identical. The main advantage of the latter method is that it~~ but not below it. The latter method allows satellite and in situ profile observations of a given variable to be consistently combined by NEMOVAR, to produce a single set of 3D increments for that variable, and was therefore the method employed in this study.

~~In each case, NEMOVAR produces~~ This gives a set of 3D $\log_{10}$(~~chl-a~~Chl-*a*) increments on the model grid~~(either 2D or 3D)~~, which must be applied to the model. ~~Various methods can be used to do this, three of which are tested here:~~

~~The simplest method is to convert~~ The $\log_{10}$(~~chl-a) increments to chl-a~~ Chl-*a*) increments were converted to Chl-*a* increments using the background total ~~chl-a~~Chl-*a*, and split ~~this~~ between diatoms and non-diatoms so as to conserve the ratio between the two in the background model field. ~~This updates the model chl-a, but does not directly alter the phytoplankton biomass , effectively just changing the phytoplankton carbon-to-chlorophyll ratio.~~

~~A more common approach (Teruzzi et al., 2014; Skákala et al., 2018) is to use method 1 but also update the phytoplankton biomass, by conserving the stoichiometric ratios in the background field.~~

~~A third method is to use the nitrogen balancing scheme of Hemmings et al. (2008), as has been routinely used with HadOCC (Ford et al., 2012). This uses a principle of conservation of mass to calculate increments to the six HadOCC state variables (phytoplankton, zooplankton, dissolved inorganic nitrogen (DIN), detritus, dissolved inorganic carbon (DIC), alkalinity) at all depths. The scheme was designed and parameterised for use with HadOCC, so is not immediately compatible with the more complex and differently parameterised MEDUSA. In this study it has been extended for use with MEDUSA by summing each of the phytoplankton and zooplankton functional types to obtain total phytoplankton and zooplankton, and using these as inputs to the nitrogen balancing scheme, while maintaining the parameter values of Hemmings et al. (2008). The scheme then calculates 3D increments to phytoplankton, zooplankton, DIN, detritus, DIC, and alkalinity, which are applied to the model, with phytoplankton split into diatoms and non-diatoms, and zooplankton into meso-zooplankton and micro-zooplankton, so that the background ratios are conserved. An increment is applied to silicate that is equal and opposite to the increment applied to diatom silicon biomass, to conserve silicon. Detrital carbon is updated to preserve the ratio between detrital nitrogen and carbon~~ Phytoplankton biomass was then similarly updated, to conserve the stoichiometric ratios in the background field~~. In its original form, the scheme accepts 2D surface chl-a increments, and calculates one set of increments within the mixed layer using mixed layer-averaged values of background phytoplankton biomass, growth and loss rates, then further increments beneath the mixed layer by scaling the mixed layer increments to the local background phytoplankton biomass. In this study the scheme has been further extended to accept 3D chl-a increments, and calculate a corresponding set of multivariate increments for every depth level using the background values at that depth. This allows the scheme to be used with either 2D or 3D chl-a increments from NEMOVAR.~~, following the approach of Teruzzi et al. (2014) and Skákala et al. (2018, 2020).

~~It is not clear which of the above approaches would yield the best results with MEDUSA, so all three multivariate balancing methods have been tested, each with either 2D or 3D chl-a increments from NEMOVAR, giving six combinations as described in Section 3.~~

2.2.3 BGC-Argo

For in situ profiles of biogeochemistry, as might be obtained from BGC-Argo, sets of 3D increments ~~are~~ were calculated for each assimilated variable, following the physics implementation of Waters et al. (2015). The method ~~is~~ was the same as for calculating 3D ocean colour increments, as described above. The vertical correlation length-scale was flow-dependent and varies with depth, as detailed by Waters et al. (2015). At the surface the vertical correlation length-scale was set to the depth of the mixed layer, decreasing to a minimum value at the base of the mixed layer. This minimised the spread of information across the pycnocline, due to the lack of correlation of water mass properties in and below the mixed layer (Waters et al., 2015; Fontana et al., 2013). Below the mixed layer, the vertical correlation length-scale increased with depth, proportional to the increase in vertical model grid spacing that occurs with depth.

In this study ~~chl-a~~Chl-*a*, NO_3, O_2, and pH ~~have been~~ were assimilated, but the methodology is simple to extend to other model variables. As for ocean colour assimilation, ~~chl-a~~ Chl-*a* is the sum of diatom and non-diatom ~~chl-a~~Chl-*a*, and a log-transformation ~~is~~ was performed prior to assimilation. As described above, assimilation of ~~chl-a~~ Chl-*a* from ocean colour and in situ profiles can be combined. NO_3 and O_2 are state variables in MEDUSA, taking NO_3 to be equivalent to the MEDUSA DIN variable, while pH is a diagnostic variable calculated using version 2.0 of the mocsy carbonate package (Orr and Epitalon, 2015), which implements the SolveSAPHE algorithm (Munhoven, 2013) for solving the alkalinity-pH equation.

The ~~chl-a increments can be applied using different multivariate balancing methods, as~~ Chl-*a* increments were applied using the stoichiometric balancing method described for ocean colour above. The NO_3 increments ~~are~~ were directly applied to the MEDUSA DIN variable, and the O_2 increments to the O_2 variable. As pH is a diagnostic variable, the pH increments cannot be applied directly. ~~A similar approach is therefore~~ The approach taken to the assimilation of partial pressure of CO_2 (pCO_2) into HadOCC (While et al., 2012) ~~, which has also been performed with MEDUSA.~~ was therefore adopted here with pH. In HadOCC, pCO_2 is a function of temperature, salinity, DIC, and alkalinity, and at constant temperature and salinity constant lines of pCO_2 are found in DIC/alkalinity space (see Fig. 1 of While et al. (2012)). The scheme of While et al. (2012) assumes that temperature and salinity are error-free (and can be directly updated by physical data assimilation if not), and therefore updates DIC and alkalinity. As there is no unique combination of DIC and alkalinity that gives a specific pCO_2 value, the smallest combined change to DIC and alkalinity is made in order to reach the target ~~pCO2~~ pCO_2 value. The same approach ~~is~~ was taken here with pH, which in MEDUSA is a function of temperature, salinity, nutrients, latitude, depth, DIC, and alkalinity. In DIC/alkalinity space, locally constant lines of pH are found when considering the range of present oceanic conditions (see Fig. 1a of Munhoven (2013)). The scheme developed here therefore updates DIC and alkalinity, assuming the other contributors to pH to be error-free, by making the smallest combined change which would give the target pH.

~~In the case where chl-a is assimilated using the nitrogen balancing scheme of Hemmings et al. (2008), and profiles of NO_3 and pH are also assimilated, this gives two different sets of increments to NO_3, DIC, and alkalinity. This combination is not used in this study, but in the current implementation precedence would be given to the increments derived from profiles of NO_3 and pH, as these are more directly related to the available observations, and just these increments applied in this situation. Ideally though, further balancing between the different variables would be applied, which can be considered as a future development.~~

3 Observing system simulation experiments (OSSEs)

3.1 Overview

As detailed by Masutani et al. (2010), OSSEs provide a way to test the impact on forecasts and reanalyses of assimilating observations which do not yet exist, by using synthetic observations. An OSSE typically comprises the following elements:

- A "nature run", which is a realistic non-assimilative model simulation of the real world, which provides a "truth" against which to validate the assimilative model.
- Synthetic observations representing both current and future observing networks, which are sampled from the nature run with appropriate errors prescribed.
- Optionally, a ~~non-assimilative control~~ free run, which provides an alternative model simulation of the nature run period.
- An assimilative ~~control~~ run, which assimilates synthetic observations representing current observing networks into the alternative model simulation.
- One or more additional versions of the assimilative run which also assimilate synthetic observations representing the future observing networks under consideration.
- Assessment of the impact on reanalysis or forecast skill of assimilating these observations, by validating against the nature run.

One of the keys to obtaining informative results from an OSSE is to ensure that all sources of error are appropriately accounted for (Halliwell et al., 2014, 2017; Hoffman and Atlas, 2016). If the ~~control~~ free run is more similar to the nature run than the real forecasting system is to the real world, then it can become easier for the assimilative system to recover the truth, and the impact of the observing networks may be incorrectly estimated. As such, three general OSSE approaches have been developed, which differ in how the ~~control~~ free run varies from the nature run.

- In "identical twin experiments", the nature and ~~control~~ free runs differ only in their initial conditions. This set-up was frequently used in early OSSEs, but as most sources of model error are neglected, the results were found to be overly optimistic, and the approach is no longer widely recommended (Arnold and Dey, 1986).
- In "fraternal twin experiments", the same model is still used for both the nature and ~~control~~ free run, but more aspects are modified. These could include the initial conditions, lateral and surface boundary conditions, parameterisations, and resolution. This takes much better account of model errors, and the approach is recommended over identical twin experiments (Arnold and Dey, 1986; Masutani et al., 2010; Yu et al., 2019).
- In "full OSSEs", significantly different models are used for the nature and ~~control~~ free runs, in order to make the two more independent. The nature run is often ~~of much~~ run either at higher resolution than the assimilative model, or with

significantly different parameterisations (Fujii et al., 2019). It is recommended to use this approach if possible (Masutani et al., 2010), but it relies on having two appropriately different models available, which is not always the case.

Due to the lack of availability of an appropriate alternative model for the nature run, it was decided within AtlantOS to take a fraternal twin approach for the biogeochemical OSSEs. This is sufficient to account for most sources of error, as long as any limitations of the approach are considered when drawing and acting upon conclusions.

3.2 Model formulation

The nature run in this study was run from 1 January 2008 to 31 December 2009, using the default parameterisations for the model versions used. This is intended to be the best available non-assimilative model representation of the real world~~, and validation~~. Validation of the general performance of ~~different aspects of the system~~ the different system components can be found in the references given in Section ~~2.~~ 2, and validation of the nature run is presented in Section 4.1. Atmospheric boundary conditions were taken from the ERA-Interim reanalysis (Dee et al., 2011). Initial conditions for NEMO were taken from the end of a 30-year hindcast of GO6 (Storkey et al., 2018). Initial conditions for CICE were taken from a pre-operational trial of the FOAM v14 system. Initial conditions for MEDUSA were based on year 5000 of the initial ocean-only phase of the spin-up of UKESM1 for use in CMIP6 projections (Yool et al., 2020). As the UKESM1 spin-up was run at 1° resolution with pre-industrial atmospheric CO_2 concentrations, the UKESM1 fields were interpolated to 1/4° resolution, and the DIC and alkalinity fields replaced by the contemporary model estimates used to initialise the 1/4° resolution experiments in ~~?~~Ford (2020). To allow the model to settle, the first year of the nature run ~~is~~ was treated as spin-up. The period was chosen to match OSSEs of the in situ physics observing system performed at the Met Office (Mao et al., in revision), and more widely as part of AtlantOS (Gasparin et al., 2019).

The ~~non-assimilative control~~ free run was performed for the same period, including spin-up, but differed from the nature run in the following ways:

- Atmospheric boundary conditions were taken from the JRA-55 reanalysis (Kobayashi et al., 2015).
- NEMO initial conditions were taken from an earlier date (1 January 1999) of the hindcast of Storkey et al. (2018).
- MEDUSA initial conditions were taken from an earlier year (1218) of the UKESM1 ocean-only spin-up, with DIC and alkalinity taken from the end of the non-assimilative 1/4° resolution experiment of ~~?~~Ford (2020).
- The NEMO parameter rn_efr, which affects near-inertial wave breaking and therefore vertical mixing (Calvert and Siddorn, 2013), was increased from 0.05 to 0.1.
- The scheme used for advection of biogeochemical variables was changed from Total Variance Dissipation (TVD) (Zalesak, 1979) to the Monotone Upstream Scheme for Conservative Laws (MUSCL) (Van Leer, 1977; Lévy et al., 2001).
- An alternative set of MEDUSA parameters was used, specifically Parameter Set 3 from Table 2 of Hemmings et al. (2015), which was found to give differences of an appropriate magnitude.

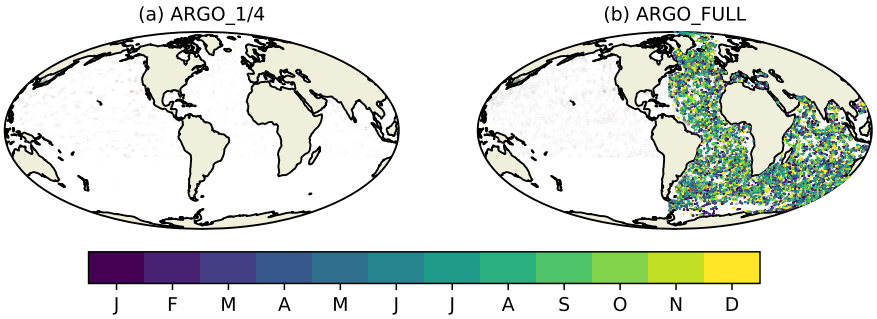


Figure 1. Simulated BGC-Argo float trajectories for 2009 equivalent to having biogeochemical sensors on (a) one in four Argo floats and (b) all Argo floats. Colours represent the month in which each observation is valid.

Together, these modifications generate approximations to the errors that exist in atmospheric fluxes and simulations of ocean physics and biogeochemistry. It is important to modify all of these, as errors in atmosphere and ocean physics have significant impacts on biogeochemical reanalyses and forecasts, and these errors must be accounted for if realistic conclusions are to be drawn from the OSSEs.

3.3 Synthetic observations

Synthetic ocean colour and BGC-Argo observations were generated from the nature run for each day of 2009. Total ~~chl-a~~ Chl-*a* from ocean colour represents the current observing network typically assimilated in biogeochemical reanalyses (Fennel et al., 2019). Observations were simulated at the same locations as were actually observed in 2009 in version 3.1 of the ESA CCI level three daily merged sinusoidal grid product (Sathyendranath et al., 2019), as used in recent Met Office reanalyses (Ford and Barciela, 2017). Whilst the products date from 2009 rather than present day, the observational coverage is similar, with three sensors contributing in 2009 (MERIS, MODIS, and SeaWiFS), and three contributing to recent ocean colour products (MODIS, OLCI, and VIIRS). BGC-Argo float trajectories were based on Argo float trajectories (Argo, 2000) used for physics OSSEs within AtlantOS (Gasparin et al., 2019). These were generated using recent real Argo float trajectories, with modifications to ensure more even geographic coverage - for details see Gasparin et al. (2019). In this study, for testing the scenario equivalent to having biogeochemical sensors on all current standard Argo floats, the same "backbone" float trajectories were used as in the studies synthesised by Gasparin et al. (2019). For the scenario equivalent to having biogeochemical sensors on one in four Argo floats, these were subsampled using the last two digits of the float IDs. The geographic coverage in each case is shown in Fig. 1.

In data assimilation, two components of observation error are typically considered: measurement error and representation error (Janjić et al., 2018). Measurement error has been accounted for in this study by adding unbiased Gaussian noise to the nature run values at observation locations, using standard deviations from the literature. A standard deviation of 30 % was agreed on within AtlantOS for ~~chl-a~~ Chl-*a* from ocean colour, ~~and the~~ a value commonly used in assimilation studies

(Pradhan et al., 2020). The same value was ~~also~~ used for BGC-Argo ~~chl-a~~ Chl-*a* profiles, consistent with Boss et al. (2008). For the remaining variables the values from Johnson et al. (2017) were used: 1 % for O_2, 0.005 for pH, and 0.5 mmol m^{-3} for NO_3. To avoid generating spuriously noisy profiles, a single value of Gaussian noise was calculated per profile, rather than at every depth level. Where the standard deviations used were a percentage, this was calculated using the mean of the profile.

Representation error arises from observations and models representing differing spatial and temporal scales and processes. Since the nature and ~~control~~ free runs were at the same resolution, this was accounted for in the same way as for the physics OSSEs in AtlantOS (Gasparin et al., 2019). For each profile, the equivalent nature run values were calculated either three days before or three days after, chosen at random. The difference between these and the "truth" value were taken to be the representation error, and added on to the observation values. The advantage of this approach is that representation error is higher in more variable regions, as would be expected in real-world data assimilation applications.

3.4 Error covariances

For assimilating ocean colour data, the monthly-varying background and observation error standard deviations from ~~?~~ Ford (2020) were used. To provide consistency between assimilating ~~2D and 3D~~ surface and in situ $\log_{10}$(~~chl-a~~Chl-*a*), these were also used for assimilating $\log_{10}$(~~chl-a~~Chl-*a*) from BGC-Argo.

For other variables, pre-existing error standard deviations were not available, so were calculated for this study. Observation error standard deviations were set to a climatological constant equal to the average global observation error specified. ~~This was~~ These were fixed in time, and specified as 0.638 mmol m^{-3} for NO_3, 2.767 mmol m^{-3} for O_2, and 0.006 for pH. Background error standard deviations were calculated using the Canadian Quick (CQ) method (Polavarapu et al., 2005; Jackson et al., 2008), which uses the variance of the difference between successive days of a free-running model simulation as a proxy for background error variance. Annual background error standard deviations were calculated from the output of the ~~non-assimilative control~~ free run. The CQ method is known to ~~under-estimate~~ underestimate the magnitude of the error standard deviations (Bannister, 2008), and the results in this study were considerably lower than the observation error standard deviations used. In order to give sufficient weight to the observations for the assimilation to be effective, the background error standard deviations were inflated~~so that their average value~~. This was achieved by multiplying the gridded field of background error standard deviations for each variable by a constant, so that the global mean background error standard deviation matched the observation error standard deviation used ~~.~~ for that variable. This meant that on average, equal weight was given to the background and to the observations, but the ratio of background to observation error varied spatially based on the estimates from the CQ method. Once the system is fully functioning with real BGC-Argo data available, ~~these~~ the background error estimates can be appropriately refined, based on the errors in the real-world assimilative model, and the actual distribution of BGC-Argo observations.

3.5 Experiments

Using these inputs, ~~two sets~~ a set of assimilation experiments were performed, in addition to the nature and ~~non-assimilative control~~ free runs, as detailed in Table 1. The nature and ~~non-assimilative control~~ free runs were run from 1 January 2008 to 31 December 2009, with the first year treated as spin-up. Each assimilation experiment was run from 1 January 2009 to 31

Table 1. Experiments performed.

Identifier
NATURE
FREE
OC ~~_2D_CHL~~
~~OC_3D_CHL Ocean colour 3D chl-a increments, only chl-a updated OC_2D_PHY Ocean colour 2D chl-a increments, biomass updated OC_3D_PHY~~
ARGO_FULL_OC
ARGO_1/4
ARGO_FULL

December 2009, using initial conditions from the end of the ~~non-assimilative control~~ free run spin-up, and assimilating the synthetic observations into the version of the model used for the ~~control~~ free run.

~~The first set of experiments~~ Five assimilation experiments were run. One just assimilated ocean colour~~, to find the most appropriate combination of vertical propagation (2D increments applied through the mixed layer or 3D increments) and multivariate balancing (chl-a only, updates to phytoplankton biomass, Hemmings et al. (2008) nitrogen balancing) for use with MEDUSA. This gave six combinations.~~

~~The second set of experiments introduced BGC-Argo, with two runs assimilating~~ . Two assimilated ocean colour in combination with the 1/4 subsampled BGC-Argo array and the full BGC-Argo array~~, in combination with the chosen ocean colour scheme from the first set of experiments~~. A final two runs assimilated the 1/4 subsampled and full BGC-Argo arrays without ocean colour.

All the experiments, with unique identifiers for each, are detailed in Table 1.

4 ~~Results~~

~~The results are presented in three sub-sections below. The first assesses how the differences between NATURE and FREE compare to errors in real-world reanalyses. The second assesses the runs just assimilating ocean colour, while the third assesses the runs assimilating BGC-Argo. The main metric~~

3.1 Metrics

The main metrics used for assessment ~~is the~~ are the absolute and percentage reduction in median absolute error (MAE), defined respectively as:

$$\mathrm{MAE}_{\mathrm{red}\,\mathrm{red\_abs}} = \frac{\mathrm{MAE}_{\mathrm{control}} - \mathrm{MAE}_{\mathrm{OSSE}}}{\mathrm{MAE}_{\mathrm{control}}} \times 100 \mathrm{MAE}_{\mathrm{control}} - \mathrm{MAE}_{\mathrm{OSSE}} \quad (2)$$

$$\mathrm{MAE}_{\mathrm{red\_\%}} = \frac{\mathrm{MAE}_{\mathrm{control}} - \mathrm{MAE}_{\mathrm{OSSE}}}{\mathrm{MAE}_{\mathrm{control}}} \times 100 \quad (3)$$

where $\mathrm{MAE}_{\mathrm{OSSE}}$ is the MAE of each OSSE compared with NATURE, and $\mathrm{MAE}_{\mathrm{control}}$ is the MAE of a control run compared with NATURE. When considering the impact of data assimilation versus a free run, FREE is used as the control run, and when assessing the added value of BGC-Argo ~~, OC_3D_PHY~~ over ocean colour, OC is used as the control run. A positive value of $\mathrm{MAE}_{\mathrm{red\_abs}}$ or $\mathrm{MAE}_{\mathrm{red\_\%}}$ represents a reduction in error in the OSSE compared to the control, and a negative value represents an increase in error. This is a modification of the approach taken by Gasparin et al. (2019), who used the percentage reduction in mean square error. MAE is used instead because the biogeochemical variables being considered are highly non-Gaussian, so it is more appropriate to use a metric such as MAE which is based on robust statistics. $\mathrm{MAE}_{\mathrm{red\_abs}}$ is used in addition to $\mathrm{MAE}_{\mathrm{red\_\%}}$, as this can be more informative in regions where $\mathrm{MAE}_{\mathrm{control}}$ is small.

Where $\mathrm{MAE}_{\mathrm{red\_abs}}$ or $\mathrm{MAE}_{\mathrm{red\_\%}}$ is presented as a spatial map, the MAE was calculated independently for each model grid cell. This was done by calculating the absolute difference between the model run and the nature run in that grid cell on each day of the given time period, and calculating the median of those values. Where $\mathrm{MAE}_{\mathrm{red\_abs}}$ or $\mathrm{MAE}_{\mathrm{red\_\%}}$ is presented as a profile, the MAE was calculated independently for each model depth level. At each depth, the absolute difference between the model run and the nature run on each day of the given time period was calculated for each grid cell in the region of interest. The median of this set of values was calculated, weighted by the area of each grid cell, to give the MAE value for that depth level.

4 Results

The results are presented in two sub-sections below. The first assesses the ability of NATURE to capture key ocean features, and how differences between NATURE and FREE compare to errors in real-world reanalyses. The second assesses the assimilation runs, and the potential impact of assimilating BGC-Argo and ocean colour data.

4.1 Errors in free-running model

As stated in Section 3, OSSEs yield the most reliable conclusions when all sources of real-world error have been appropriately accounted for (Halliwell et al., 2014). This means that the errors between FREE and NATURE should, ideally, be broadly similar to the errors between FREE and the real world. Furthermore, NATURE should be able to represent key features observed in the real ocean. As the real-world ocean is not known exactly, observation-based products must be used to perform this assessment, even though these can have large uncertainties and do not exactly represent the real world. For many biogeochemical variables, coarse climatologies are the only suitable products for a global assessment.

Figure 2 shows ~~the absolute difference between FREE and an observation-based product, and between FREE and NATURE, for surface chl-a,~~ surface fields of temperature, Chl-*a*, NO_3, O_2, pH, and pCO_2, from observation-based products, NATURE, and FREE. These are plotted for the final month of the simulations, December 2009. The observation-based products used are

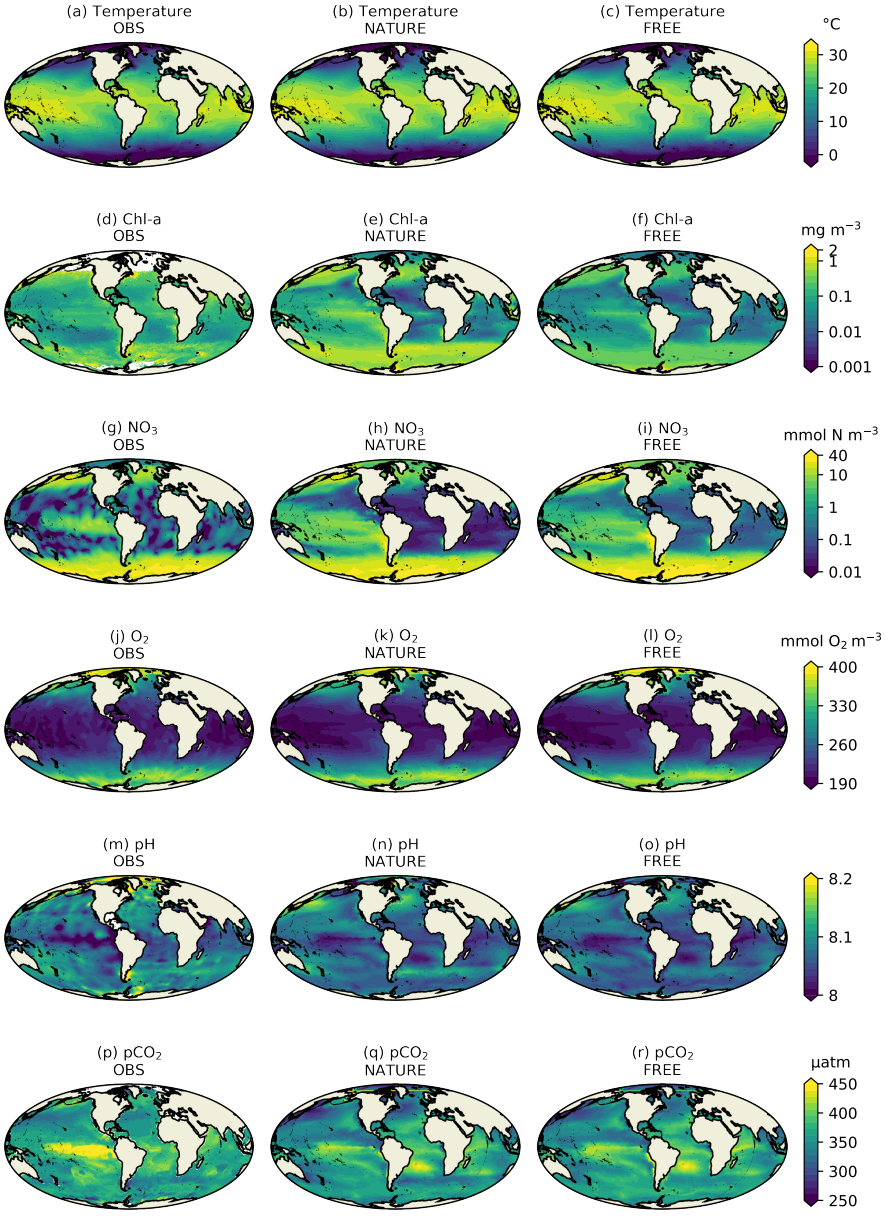


Figure 2. Monthly mean surface (a-c) temperature, (d-f) Chl-*a*, (g-i) NO_3, (j-l) O_2, (m-o) pH, and (p-r) pCO_2, for December 2009, from real-world observation-based products (left column), NATURE (middle column), and FREE (right column).

the ~~monthly~~ EN4.2.1 monthly analysis product for temperature (Good et al., 2013; Gouretski and Reseghetti, 2010), the Ocean Colour CCI v3.1 ~~ocean colour CCI product for chl-a~~ monthly product for Chl-*a* (Sathyendranath et al., 2019), the 2018 World

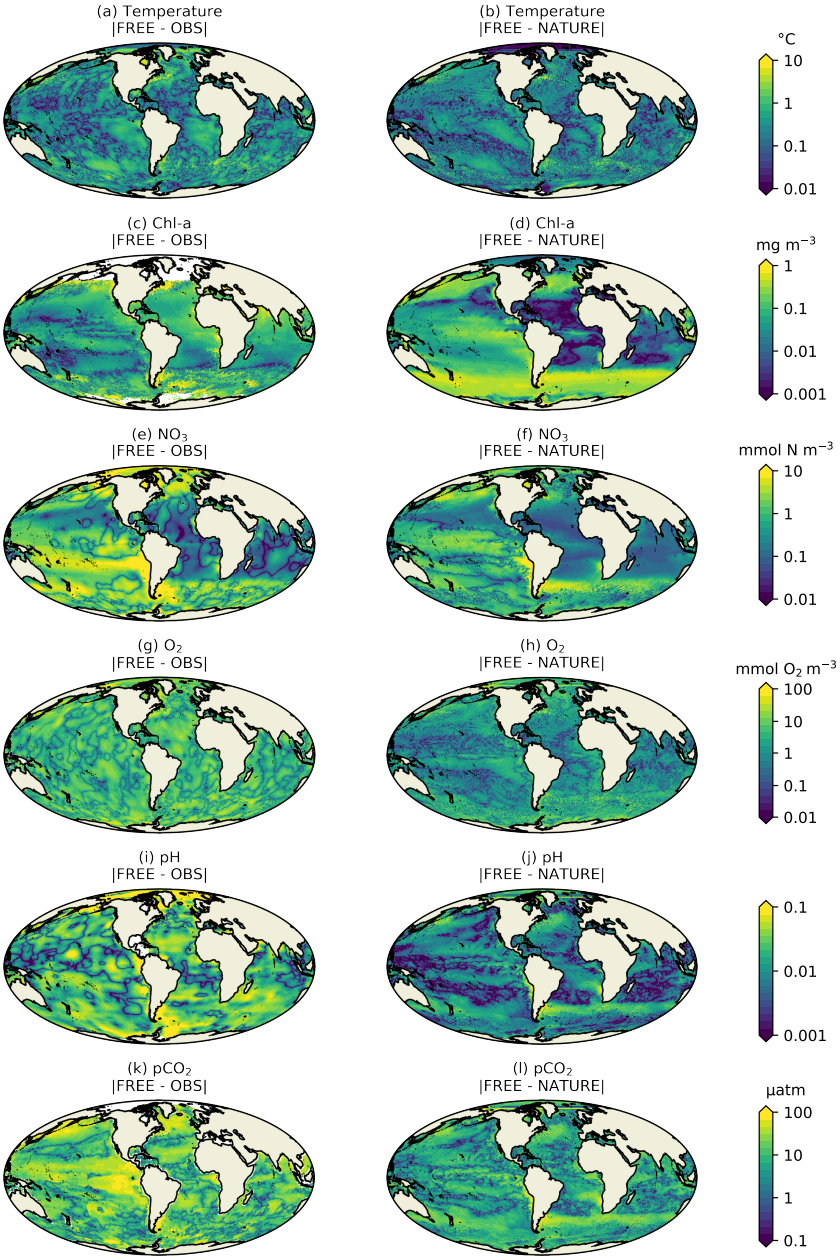


Figure 3. Absolute difference for December 2009 for surface (a-b) temperature, (c-d) Chl-*a*, (e-f) NO_3, (g-h) O_2, (i-j) pH, and (k-l) pCO_2, between FREE and real-world observation-based products (left column), and between FREE and NATURE (right column).

Ocean Atlas (WOA18) monthly climatology for NO_3 ~~(?),~~ and O_2 (Garcia et al., 2018a, b), the GLODAP v2 annual climatology

for pH (Lauvset et al., 2016), and the monthly ~~pCO_2~~ statistical analysis product of Landschützer et al. (2015a, b) for pCO_2. The observation-based products were bilinearly interpolated to the model grid.

For both physical and biogeochemical variables, NATURE captured the broad global distribution, with generally appropriate magnitudes. There were some discrepancies, such as an overestimation of Chl-*a* in parts of the Pacific and Southern Oceans, and an underestimation in the Atlantic and Indian Oceans. NO_3 was too low in the Atlantic and Indian Oceans, and pCO_2 was too low in the Tropical Pacific. Overall though, NATURE matched the observation-based products sufficiently well for it to be used as the "truth" in these experiments.

FREE also broadly captured these features, as expected from state-of-the-art models (Fennel et al., 2019). Differences between FREE and NATURE were generally similar to the differences between FREE and the observation-based products, but with some exceptions. This can be seen more clearly in Fig. 3, which shows the absolute difference between FREE and the observation-based products, and between FREE and NATURE, for the fields plotted in Fig. 2.

~~For chl-a~~

For temperature (Fig. ~~2a-b~~3a-b), the absolute difference between FREE and NATURE was very similar in pattern to that between FREE and the EN4 analysis, but slightly lower in magnitude in some regions. This suggests that the perturbations applied to the physics (different atmospheric fluxes, initial conditions, and vertical mixing) resulted in an error contribution to the biogeochemical model similar to, but slightly smaller than, that seen in state-of-the-art modelling systems. For Chl-*a* (Fig. 3c-d), the two sets of absolute difference ~~are~~ were broadly similar in the Pacific and Southern Oceans, but in the Tropical Atlantic and Indian Oceans the absolute difference between FREE and NATURE ~~is~~ was smaller than between FREE and the CCI data. In NATURE the ~~chl-a~~ Chl-*a* in these regions was ~~spuriously~~ too low compared with observations, linked to low nutrient concentrations ~~.~~ (Fig. 2). The modifications introduced in FREE served to increase nutrient concentrations in these regions, but also to suppress phytoplankton growth, resulting in little overall change in ~~chl-a.~~ Chl-*a*. This is largely the result of increasing the nitrogen nutrient uptake half-saturation concentration for phytoplankton, and decreasing the zooplankton grazing half-saturation concentration. Elsewhere though, the levels of absolute error ~~are~~ were broadly similar, meaning the OSSEs ~~have~~ had realistic levels of model error. ~~While it is not ideal that the errors differ in the Tropical Atlantic and Indian Oceans, achieving globally appropriate levels of error with a complex biogeochemical model with globally uniform parameterisations could not be managed within the resources of the project. Furthermore, the similarity of NATURE and FREE in these regions is due to the introduction of compensating errors in FREE, rather than a lack of model error. This itself is a common feature of reanalyses, which can result in data assimilation increasing overall error by correctly reducing one of a set of compensating errors, as demonstrated by Ford and Barciela (2017).~~

For NO_3, O_2, pH, and pCO_2 (Fig. 3e-l), the global distributions ~~are more similar, although the absolute difference~~ of absolute error were generally similar between FREE and NATURE ~~is slightly smaller than~~ and between FREE and the ~~observation products, especially for pCO_2, which should be borne in mind when drawing conclusions~~observation-based products, although the absolute difference between FREE and NATURE was often smaller. It should be remembered though that the observation-based products used here have large uncertainties themselves, while the comparison between FREE and NATURE does not

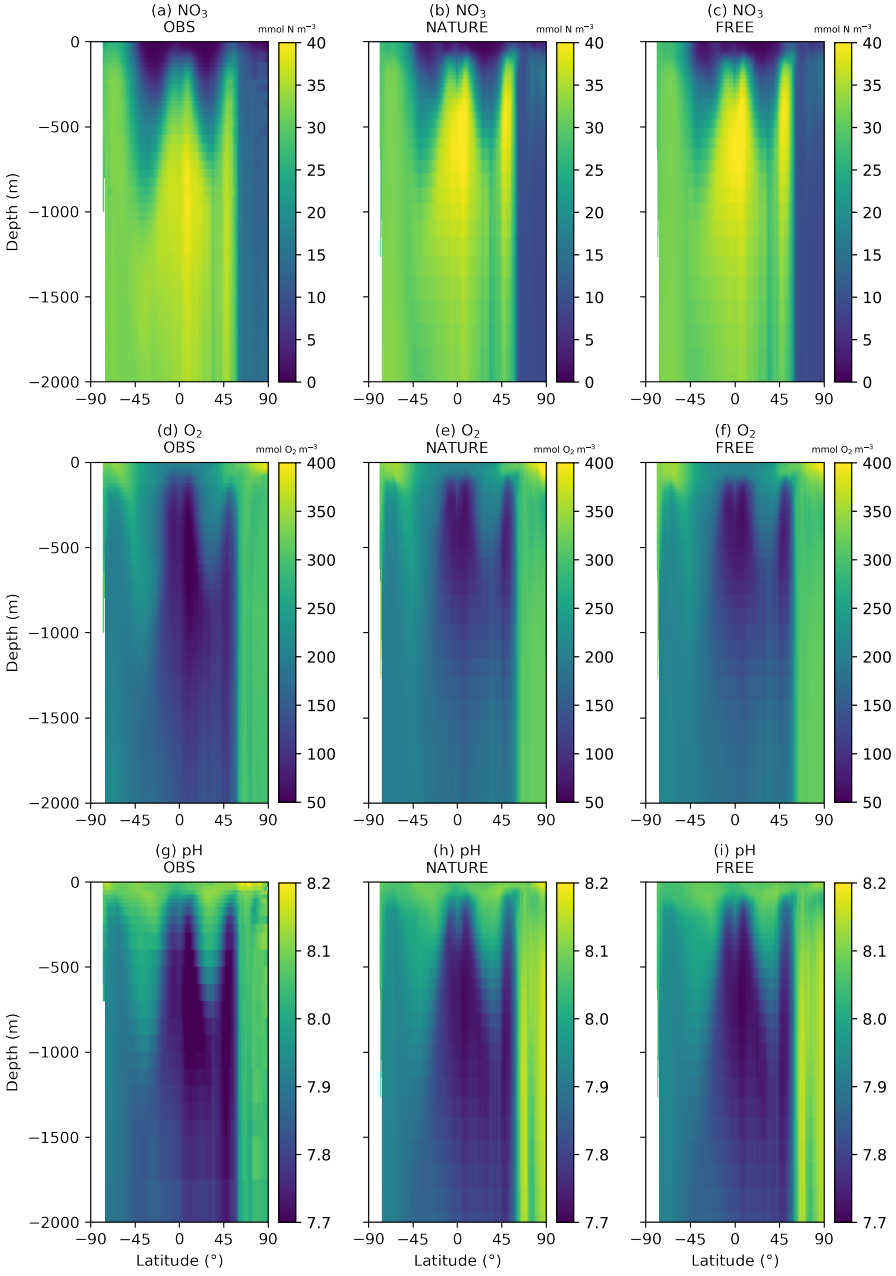


Figure 4. ~~Absolute difference for December 2009 for surface~~ Annual zonal mean sections of (~~a-b~~a-c) ~~chl-a, (c-d)~~ NO_3, ~~and~~ (~~e-f~~d-f) ~~pCO~~O_2 ~~between FREE~~ , and (g-i) pH, from real-world observation-based products (left column), NATURE (middle column), and ~~between~~ FREE ~~and NATURE~~ (right column).

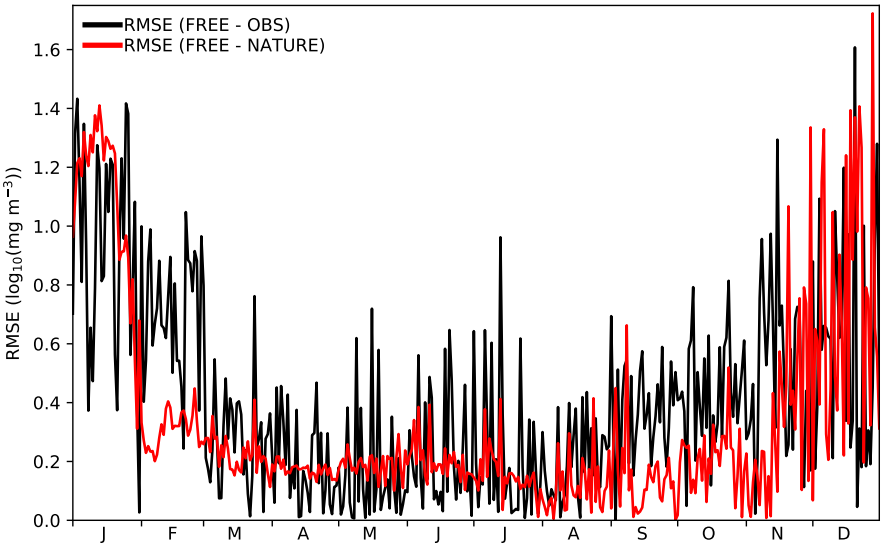


Figure 5. Time series of daily global RMSE for surface $\log_{10}$(Chl-*a*) between FREE and CCI ocean colour observations (black line) and between FREE and NATURE (red line) at CCI ocean colour observation locations.

include observation error. Furthermore, the GLODAP v2 pH climatology is normalised to the year 2005 (Lauvset et al., 2016) , and so due to ocean acidification may slightly overestimate pH in 2009.

4.2 ~~Ocean colour data assimilation~~

~~For each of the six runs just assimilating oceancolour using different techniques, profiles of global MAE_{red} over FREE for the last month of simulation, December 2009, are plotted in Fig. 3 for nine model variables. Profiles are~~ A similar comparison is required for the subsurface ocean, as BGC-Argo profiles were simulated for the upper 2000 m. Fewer observation products are available for such an assessment. No subsurface climatology is available for Chl-*a*, and only annual climatologies covering the full water column are available for NO_3 and O_2 from WOA18, and for pH from GLODAP v2. Annual zonal mean sections of NO_3, O_2, and pH are therefore plotted for the upper ~~250 m , approximately corresponding to the maximum depth the euphotic zone is likely to take.~~

~~Global MAE_{red} over FREE for ocean colour experiments.~~

~~For chl-a, all six runs perform very similarly at the surface, with MAE_{red} values between 67–76 %. Beneath the surface OC_2D_PHY and OC_3D_PHY perform best, reducing the MAE at all depths. OC_2D_CHL and OC_3D_CHL perform least well with some degradations to MAE below about 60 m, though it should be noted that chl-a values are extremely low at depths with insufficient light for growth, so the percentage differences between runs become less meaningful at these depths~~2000 m in Fig. 4, from the observation-based products, NATURE, and FREE. For all three variables, NATURE captured well the broad zonal and depth variations of the climatologies, with appropriate magnitudes. NO_3 concentrations in the upper 1000 m were

slightly too high in the low- and mid-latitudes, as was O_2 and pH below 500 m, but there were no major discrepancies. FREE also captured the observed distibutions of all three variables, though differences between FREE and NATURE were often slighly smaller than between FREE and the observation-based products.

~~For phytoplankton biomass, least impact is seen in OC_2D_CHL and OC_3D_CHL, due to the biomass not being directly updated by the assimilation, with MAE_{red} values of around 7 % at the surface. The other four runs behave similarly as for chl-a, though with a smaller impact at the surface (MAE_{red} of 32–44 %).~~ A further consideration is that the growth of errors with time between FREE and NATURE should be comparable to that between FREE and real-world observations (Halliwell et al., 2014). The only assimilated variable for which there are sufficient daily observations to assess this is Chl-*a* from ocean colour. In Fig. 4 time series of root mean square error (RMSE) for surface $\log_{10}$(Chl-*a*) are plotted between FREE and NATURE and between FREE and observations. The observations are the daily CCI data used to define the locations of the synthetic ocean colour observations. For the comparison, model values were bilinearly interpolated to the observation locations, and a daily global RMSE value calculated. The magnitude of the RMSE between FREE and NATURE and between FREE and the observations was very similar, and ~~less variation in MAE_{red} with depth. OC_2D_PHY and OC_3D_PHY give slightly better results than OC_2D_NIT and OC_3D_NIT.~~

~~For the other seven variables plotted, which are either not directly updated by the assimilation or are only updated by the Hemmings et al. (2008) balancing scheme, results are mixed. OC_2D_CHL and OC_3D_CHL have near-zero MAE_{red} values for all variables, demonstrating that just updating chl-a has very little impact on the wider model state. There are only small differences between OC_2D_PHY and OC_3D_PHY, and between OC_2D_NIT and OC_3D_NIT, suggesting that the use of NEMOVAR to create 3D increments, as required for combining assimilation of chl-a from ocean colour and BGC-Argo, is fit for purpose. Similarly, the extension of the Hemmings et al. (2008) balancing scheme to accept 3D chl-a increments as an input appears successful. The remainder of this sub-section will therefore focus on comparing OC_3D_PHY and OC_3D_NIT~~remained comparable throughout the year, with higher RMSE in austral summer in both cases. The RMSE variability was typically smaller between FREE and NATURE though, suggesting this to be a source of error not fully captured in FREE.

~~OC_3D_PHY results in a large degradation of surface zooplankton biomass and NO_3, with MAE_{red} values of -358 % and -92 % respectively. This is reduced in OC_3D_NIT to -106 % and -15 % respectively. In both runs, for zooplankton biomass MAE_{red} increases towards zero with depth. For NO_3, more complex variation in MAE_{red} is seen with depth, likely reflecting regional variations in the impact of the two assimilation strategies around the nutricline in particular. Both runs improve detrital nitrogen, with MAE_{red} values at the surface of 35 % for OC_3D_PHY and 23 % for OC_3D_NIT. MAE_{red} decreases more quickly with depth in OC_3D_PHY, but remains positive in both cases. O_2 is degraded in both runs, with a larger surface degradation in OC_3D_NIT. With the carbon cycle, DIC, alkalinity, and pH are all degraded in OC_3D_PHY. In OC_3D_NIT, DIC is further degraded near the surface, but alkalinity is now improved, with positive MAE_{red} through the water column. The result on pH, a diagnostic which is a function of DIC and alkalinity, is that OC_3D_PHY and OC_3D_NIT have a near-identical degradation in MAE_{red} of around -157 % at the surface, but OC_3D_NIT gives better results with depth.~~

From these results, there is not a clear indication of which multivariate balancing method gives the best overall results. It could be argued that just updating chl-a is the safest strategy, as this improves the assimilated variable (chl-a), slightly improves phytoplankton biomass, and does not degrade any other variable. It is commonly agreed though that it is highly desirable to try and use ocean colour data to improve the wider model state (Gehlen et al., 2015; Fennel et al., 2019), and this clearly cannot be achieved by solely updating chl-a. Present and future reanalyses do and will make multivariate updates, which are difficult to validate due to the sparsity of in situ observations, and this should be accounted for when considering the potential impact of While it is not ideal that the Chl-*a* errors differ in the Tropical Atlantic and Indian Oceans, and errors between FREE and NATURE were too low for some variables, achieving globally appropriate levels of error with a complex biogeochemical model with globally uniform parameterisations could not be managed within the resources of the project. Furthermore, the similarity of Chl-*a* between NATURE and FREE in some regions was due to the introduction of compensating errors in FREE, rather than a lack of model error. This itself is a common feature of reanalyses, which can result in data assimilation increasing overall error by correctly reducing one of a set of compensating errors, as demonstrated by Ford and Barciela (2017). For regions and variables where the errors between FREE and NATURE were too low, the potential result may be to underestimate the impact of assimilating dense data, in this case ocean colour, and overestimate the impact of assimilating sparse data, in this case BGC-Argo on such reanalyses. The most commonly used method is to update the phytoplankton biomass to preserve stoichiometry (Teruzzi et al., 2014; Skákala et al., 2018). Since using the Hemmings et al. (2008) balancing scheme did not show a clear overall improvement in these tests, it was therefore decided to use OC_3D_PHY as the basis for OSSEs introducing the assimilation of BGC-Argo data, and to use this method when assimilating profiles of chl-a(Halliwell et al., 2014). This should be borne in mind when drawing conclusions.

Global MAE_{red} over FREE for BGC-Argo experiments.

4.2 BGC-Argo assimilationAssimilation runs

For each of the four runsassimilating BGC-Argo data, plus OC_3D_PHY, five assimilation runs, profiles of global $MAE_{red\_\%}$ over FREE for December 2009 are plotted in Fig. 4. Profiles are 6, for nine model variables. The results for MAE_{red_abs} are very similar as for $MAE_{red\_\%}$ (not shown). For Chl-*a* (Fig. 6a), phytoplankton biomass (Fig. 6b), zooplankton biomass (Fig. 6c), and detrital nitrogen (Fig. 6d), which only have significant concentrations in the euphotic zone, profiles are plotted for the upper 2500 m250 m. For NO_3 (Fig. 6e), O_2 (Fig. 6f), DIC (Fig. 6g), alkalinity (Fig. 6h), and pH (Fig. 6i), profiles are plotted for the upper 2500 m. Recall that Chl-*a*, with chl-a, NO_3, O_2, and pH observationshaving been , assimilated in the BGC-Argo experiments, were produced for the upper 2000 m.

Surface MAE_{red} over FREE for chl-a for December 2009.

Surface MAE_{red} over FREE for NO_3 for December 2009.

Surface MAE_{red} over FREE for pCO_2 for December 2009.

Surface MAE_{red} over OC_3D_PHY for chl-a, NO_3, O_2, pH, pCO_2 for December 2009.

MAE_{red} over OC_3D_PHY for chl-a, NO_3, O_2, pH for December 2009 at 100 m depth.

Global MAE_{red} over OC_3D_PHY for BGC-Argo experiments.

~~For chl-a, OC_3D_PHY~~For Chl-*a*, OC, ARGO_1/4_OC, and ARGO_FULL_OC all ~~have an MAE~~had an $MAE_{red\_\%}$ value of 72 % at the surface~~, suggesting that~~. This suggests that ocean colour was very successful at improving surface Chl-*a*, while BGC-Argo ~~is~~ was unable to add much information to that gained from the much denser ocean colour ~~chl-a~~ observations, at least at the global scale. When only BGC-Argo ~~is assimilated, ARGO_1/4 and ARGO_FULL result in~~ was assimilated, a small improvement at the surface of 7 % and 15 % was seen in ARGO_1/4 and ARGO_FULL respectively. Beneath the surface, at depths likely to see a deep chlorophyll maximum, BGC-Argo ~~has~~ had much greater impact, with all four runs outperforming OC~~_3D_PHY~~. ARGO_FULL ~~outperforms~~ outperformed ARGO_1/4, demonstrating benefit from extra in situ observations. Combining BGC-Argo and ocean colour ~~gives~~ gave better results at this depth in ARGO_1/4_OC (which is the proposed observing system), but ARGO_FULL and ARGO_FULL_OC ~~perform~~ performed similarly. Beneath the euphotic zone, where ~~chl-a is~~ Chl-*a* was near-zero, the assimilation ~~has~~ had little impact. Positive values of $MAE_{red\_\%}$ were seen below 250 m, but values of MAE_{red_abs} (not shown) tended to zero below about 220 m.

The results for phytoplankton biomass ~~are~~ were very similar as for ~~chl-a~~Chl-*a*. For zooplankton biomass, ~~the large surface degradation in OC_3D_PHY is still present in ARGO_1/4_OC and ARGO_FULL_OC, and much reduced~~ which was not directly updated by the assimilation, a large degradation in surface values was seen in all three runs assimilating ocean colour data, the impact reducing to near-zero at around 100 m. A smaller degradation was seen in ARGO_FULL, reduced further in ARGO_1/~~4 and ARGO_FULL. Detrital nitrogen is~~ 4. Detrital nitrogen was improved in the upper ~~260–280~~ 100 m in all runs, ~~and degraded beneath that depth. Including ocean colourassimilation increases the magnitude of both the improvement and degradation, as does increasing the number of BGC-Argo floats~~especially those assimilating ocean colour, with little absolute impact of the assimilation beneath that depth.

For NO_3 and O_2, which ~~are~~ were assimilated from the BGC-Argo floats, there ~~is~~ was a clear improvement throughout the water column to 2500 m in ARGO_1/4 and ARGO_FULL, with greater improvement when more floats ~~are~~ were assimilated. Maximum $MAE_{red\_\%}$ ~~is~~ was seen at 100–120 m depth, with less impact at the surface, particularly for O_2. In OC~~_3D_PHY~~, NO_3 and O_2 ~~are degraded throughout the water column~~were degraded in the upper 1000 m. Adding BGC-Argo to ocean colour assimilation partially ~~mitigates~~ mitigated this, with positive $MAE_{red\_\%}$ at some depths and negative $MAE_{red\_\%}$ at others.

With the carbon cycle, ~~throughout most of the water column~~ DIC, alkalinity, and pH were all degraded in OC. In ARGO_1/4 and ARGO_FULL~~improve DIC and degrade alkalinity~~, throughout most of the water column DIC was improved and alkalinity degraded, with an overall improvement in the assimilated variable pH. ~~Including ocean colour assimilation reduces the impact of~~ Combining ocean colour and BGC-Argo ~~, and results in~~ assimilation gave mixed results, with a degradation in pH in the surface layers, but an improvement at depth.

Spatial maps of surface $MAE_{red\_\%}$ over FREE for December 2009 ~~for the same five runs are plotted for chl-a~~ are plotted in Fig. ~~5~~7 for the five assimilation runs for Chl-*a*, NO_3~~in Fig. 6,~~, O_2, pH, and pCO_2~~in Fig. 7. In OC_3D_PHY surface chl-a is~~. In OC surface Chl-*a* was almost universally improved (Fig. 7a), apart from a few small areas in the Atlantic, North Pacific, and Indian Ocean. These correspond to areas where NATURE and FREE ~~are~~ were almost identical, as seen in Fig. ~~2b. As suggested in Section 4.1, it is likely that compensating errors have been introduced in FREE, which the assimilation is not fully able to address.~~ 3d. Very little difference ~~is~~ was made by adding BGC-Argo to ocean colour assimilation, while assimilating

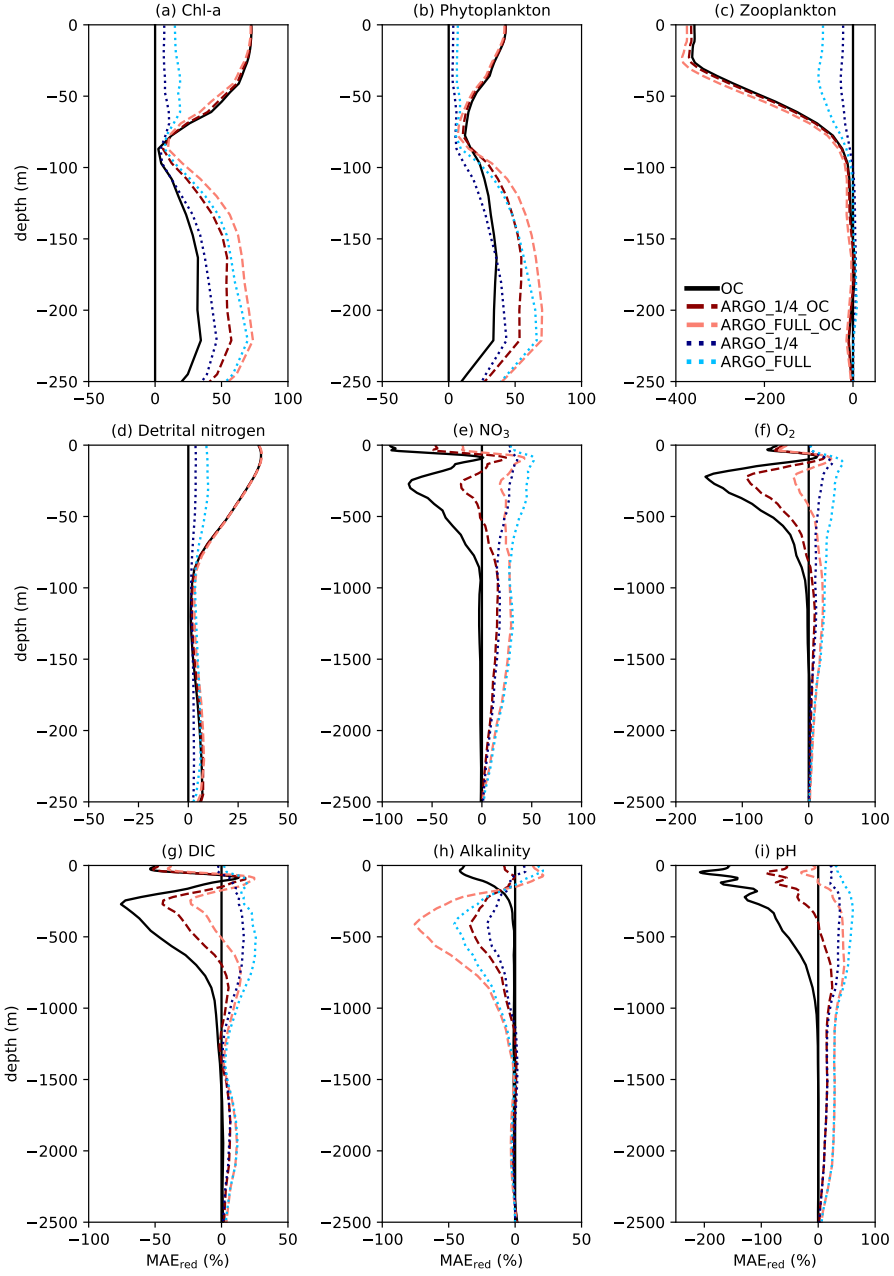


Figure 6. Profiles of global MAE_{red} % over FREE for December 2009. Note that axis scales differ between subplots.

just BGC-Argo data ~~gives~~ gave mixed results for ~~chl-a~~ Chl-*a* at the surface. In the Pacific Ocean, ~~chl-a is~~ Chl-*a* was slightly improved in ARGO_1/4 and further improved in ARGO_FULL, while in the Atlantic and Indian Oceans a degradation ~~is~~ was seen. This again corresponds to regions where there ~~is~~ was little absolute difference between NATURE and FREE (Fig. ~~2b).~~

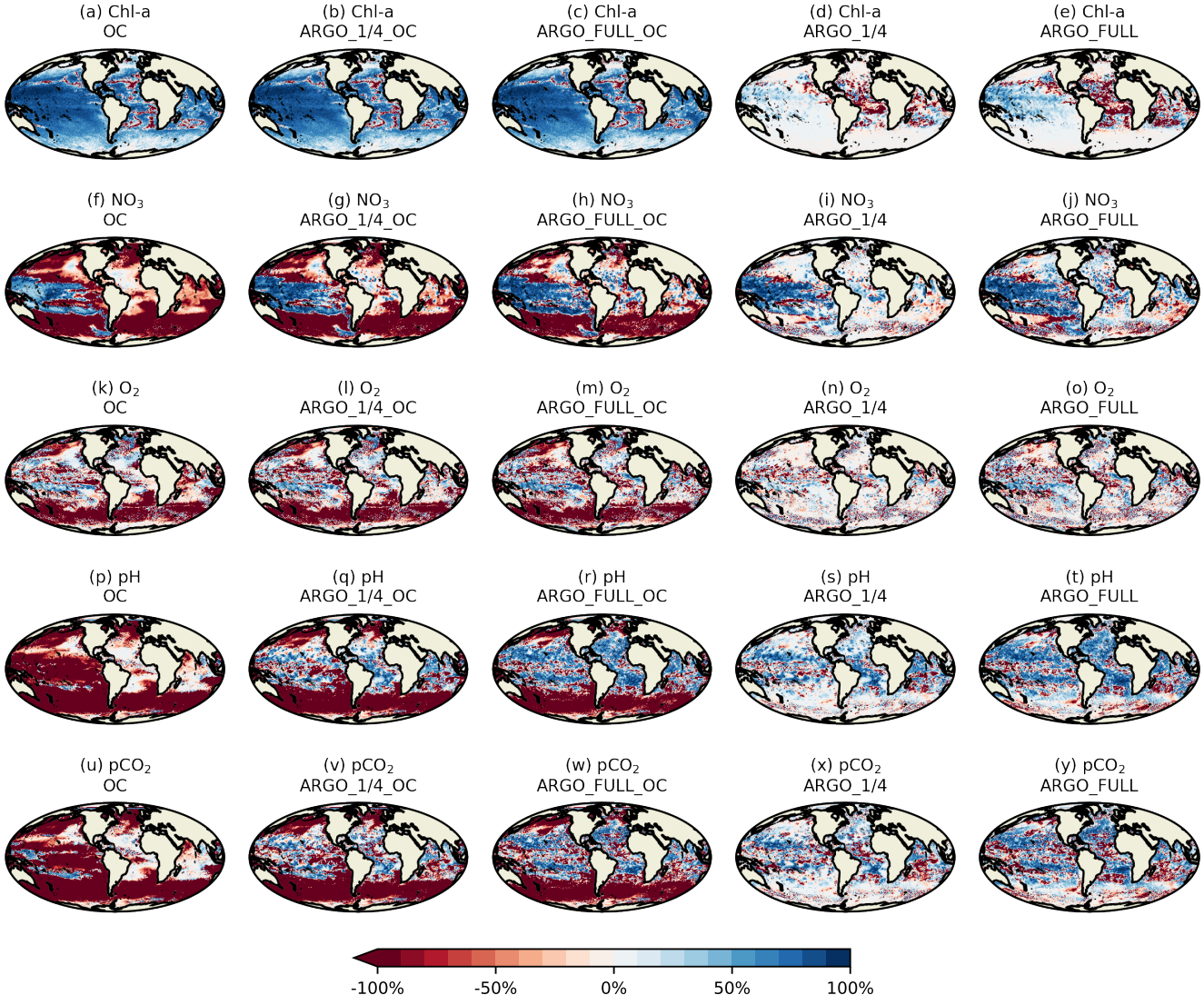


Figure 7. Surface $MAE_{red\_\%}$ over FREE for Chl-*a*, NO_3, O_2, pH, pCO_2 for December 2009.

3d). As such, $MAE_{red\_\%}$ is not the most informative metric in these regions, and it is more appropriate to examine MAE_{red_abs}. This is plotted for Chl-*a* in Fig. 8, which shows the absolute value of the degradation to be very small.

~~At the surface, NO_3 is~~ Surface NO_3 was degraded almost everywhere in OC ~~_3D_PHY~~ (Fig. ~~6a~~7f), apart from the Tropical Pacific, which is where some of the largest differences ~~are~~ were seen between NATURE and FREE (Fig. ~~2d~~3f). Adding BGC-Argo assimilation ~~increases~~ increased this improvement and ~~starts~~ started to reverse the degradation in other regions, particularly in ARGO_FULL_OC. Assimilating just BGC-Argo ~~improves~~ improved NO_3 in most areas, with more impact seen with more floats, but the results ~~are very patchy~~ were patchy in places.

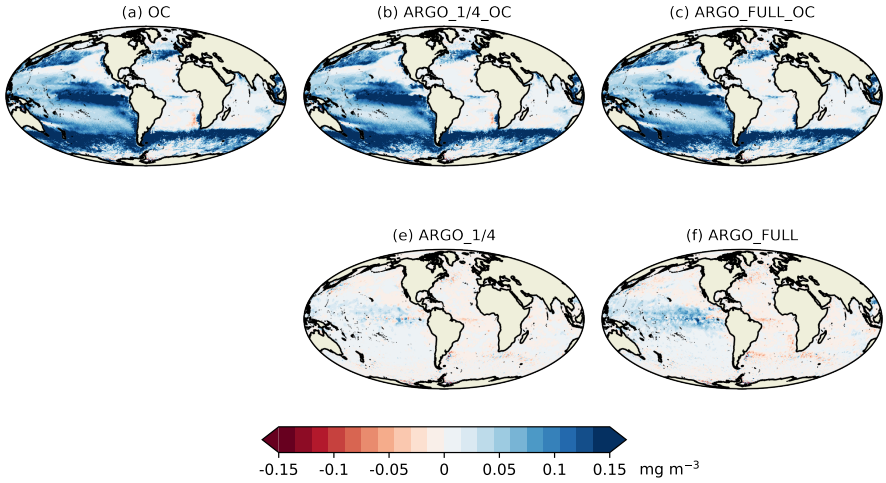


Figure 8. Surface MAE_{red_abs} over FREE for Chl-*a* for December 2009.

The story for ~~pCO~~O_2 (Fig. ~~7) is~~ 7k-o) was very similar as for NO_3, but with a smaller degradation introduced by ocean colour assimilation, and a smaller improvement brought about by BGC-Argo assimilation. For pH (Fig. 7p-t) and pCO_2 (Fig. 7u-y), the patterns were also similar, but with a greater degradation introduced by ocean colour assimilation, and a greater improvement brought about by BGC-Argo assimilation. While pCO_2 ~~is~~ was not directly assimilated, improvements to DIC and alkalinity when assimilating pH should ~~also improve~~ have a similar impact on pCO_2.

Current ~~start-of-the-art~~ state-of-the-art reanalyses typically assimilate ocean colour data (Fennel et al., 2019), so to demonstrate the additional impact BGC-Argo might have in these systems~~spatial plots of surface MAE~~, the remainder of this section focusses on $MAE_{red\_\%}$ over OC, rather than over FREE. Spatial plots of $MAE_{red\_\%}$ over OC ~~_3D_PHY~~ for December 2009 at 100 m depth are shown in Fig. ~~8~~ 9 for the assimilated variables ~~and pCO$_2$. Clear benefit is seen for surface~~ Chl-*a*, NO_3, ~~pH, and pCO~~O_2~~, with widespread positive MAE_{red} values, especially for ARGO_FULL_OC. For $O_2$the situation is more mixed, while for chl-a a small degradation is seen.~~

~~Beneath the surface there is greater benefit for most of the assimilated variables, as seen in the global maps of MAE_{red} over OC_3D_PHY at 100 m depth in Fig. 9. The improvement is~~ , and pH. Clear benefit was found for all variables, with greater improvements than at the surface. The improvement was largest for pH and smallest for ~~chl-a~~Chl-*a*, and in all cases a greater impact ~~is~~ was seen in ARGO_FULL_OC than ARGO_1/4_OC.

~~This is further demonstrated in the profiles of MAE~~To investigate the impact of the BGC-Argo assimilation over the full year, a Hovmöller diagram (Hovmöller, 1949) of daily global $MAE_{red\_\%}$ over OC ~~_PHY_3D plotted~~ with depth is plotted for each of the assimilated variables, for ARGO_1/4_OC and ARGO_FULL_OC, in Fig. 10. ~~Apart from a small degradation in zooplankton above 250 m , detrital nitrogen below 250–350 m , and alkalinity below 175 m, all variables are improved~~ For Chl-*a* (Fig. 10a-b), ARGO_1/4_OC and ARGO_FULL_OC both displayed a very small degradation compared with OC in the surface layers, throughout the year. This is consistent with the small difference seen between the runs in the profiles in Fig. 6a.

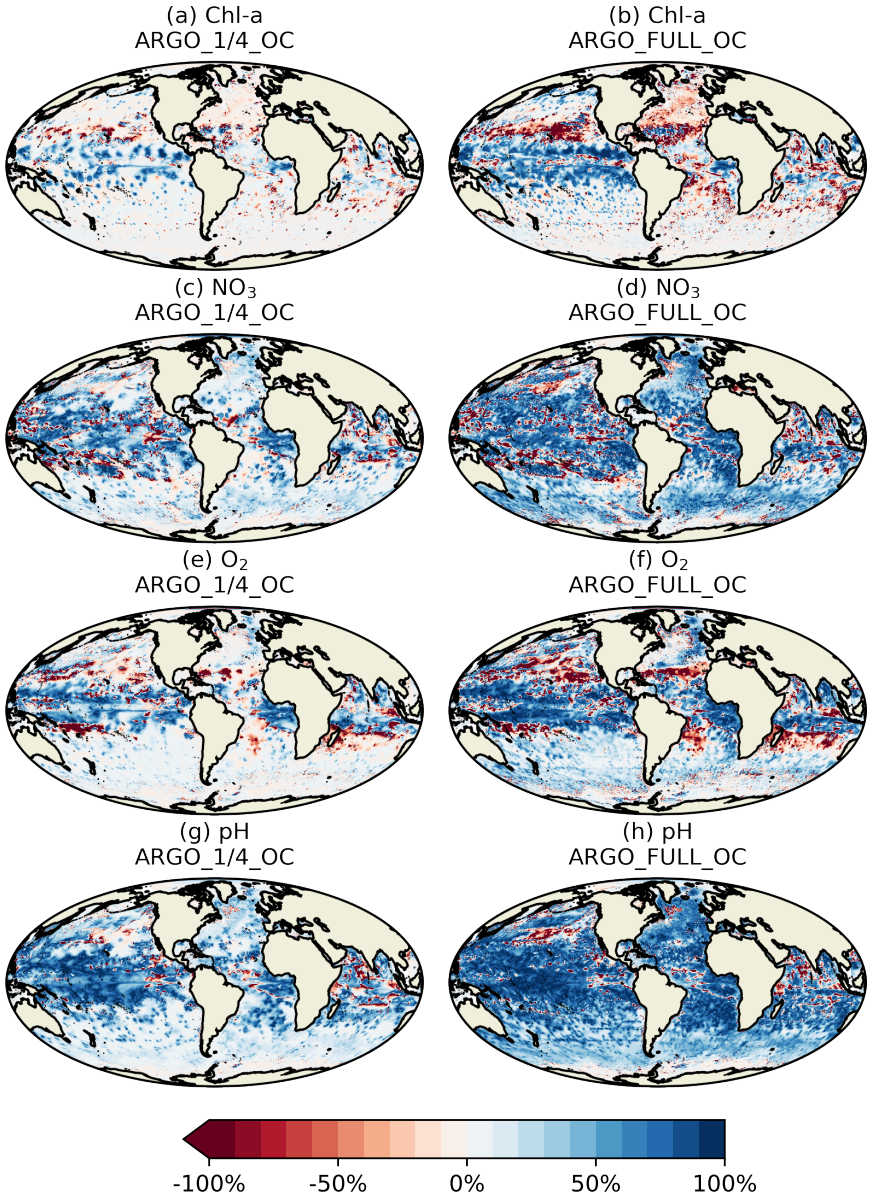


Figure 9. $MAE_{red\_\%}$ over OC for Chl-*a*, NO_3, O_2, pH for December 2009 at 100 m depth.

Between about 80–300 m depth, where deep chlorophyll maxima may be found, a strong positive impact was found on Chl-*a*, strongest in ARGO_FULL_OC. In both runs this took a few weeks to spin up, then remained reasonably consistent throughout the year. Beneath about 300 m depth, values of $MAE_{red\_\%}$ were near-zero, as Chl-*a* was negligible. For NO_3, O_2, and pH, an almost universal improvement was seen throughout the water column. ~~The degradation in alkalinity is compensated for by~~

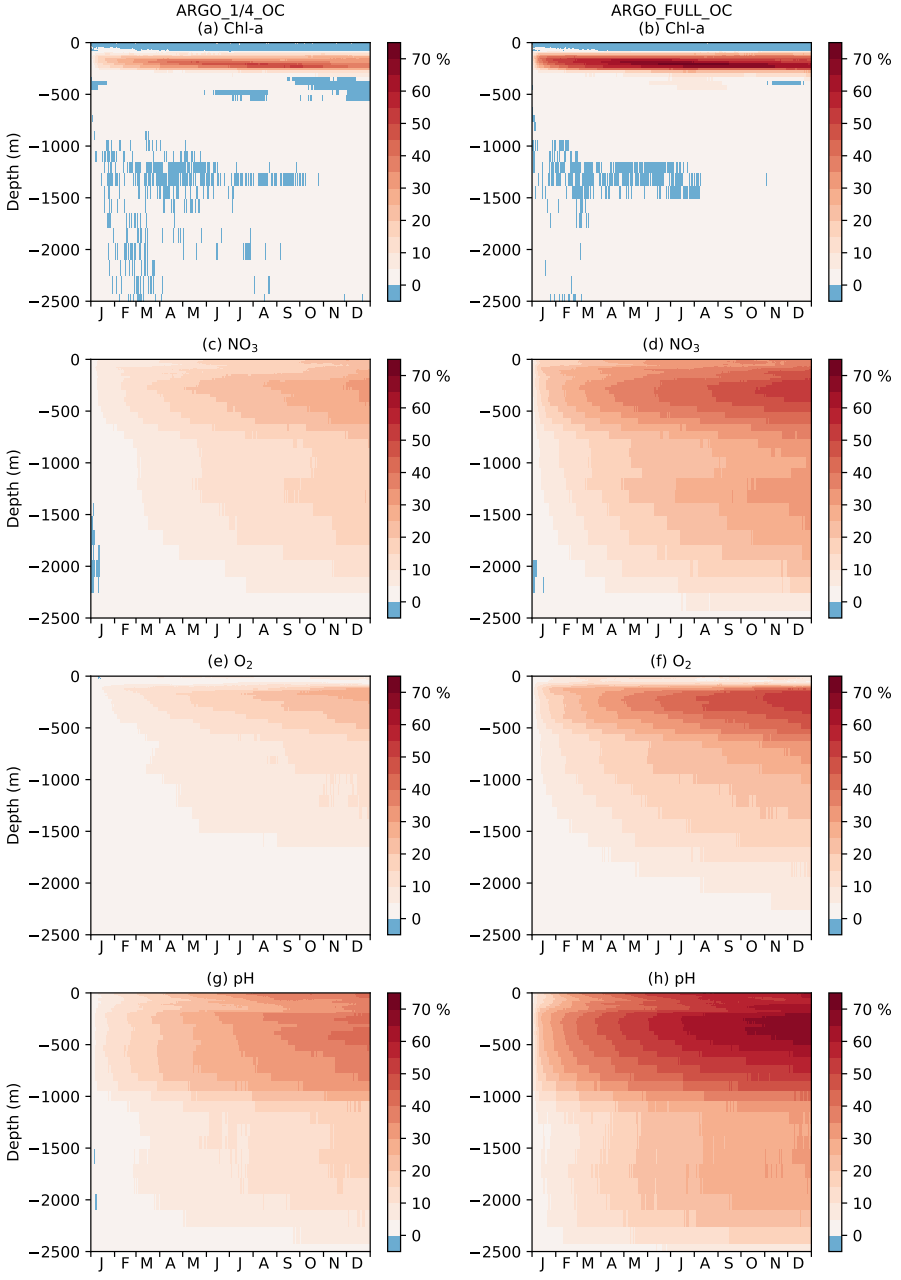


Figure 10. Hovmöller diagram of daily global $MAE_{red\_\%}$ over OC with depth for Chl-*a*, NO_3, O_2, and pH.

~~the improvement in DIC, with an overall improvement in pH. For all variables, a greater impact is seen in ARGO_FULL_OC than ARGO_~~For all three variables, values of $MAE_{red\_\%}$ were highest in the upper 1000 m, with a maximum at a similar depth or slightly deeper than that for Chl-*a*. $MAE_{red\_\%}$ was consistently higher in ARGO_FULL_OC than in ARGO_1/4_OC,

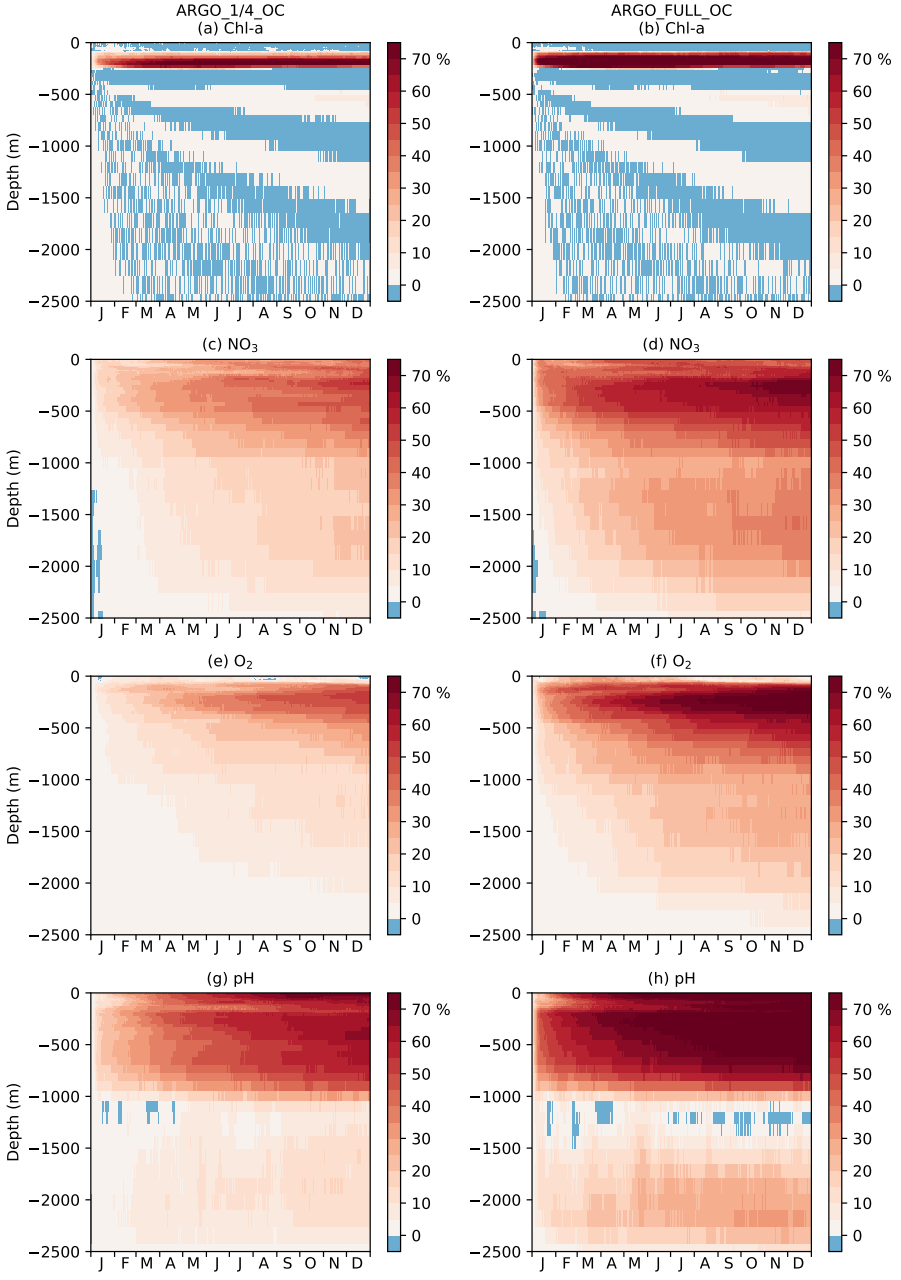


Figure 11. Hovmöller diagram of daily $MAE_{red\ \%}$ over OC with depth in the Tropical Pacific for Chl-*a*, NO_3, O_2, and pH.

615 again demonstrating a positive impact from having a greater number of BGC-Argo floats. In both runs, $MAE_{red\ \%}$ increased throughout the year, suggesting that the impact of the assimilation was still spinning up, and the full potential would not be realised until the system had been run for longer than a year.

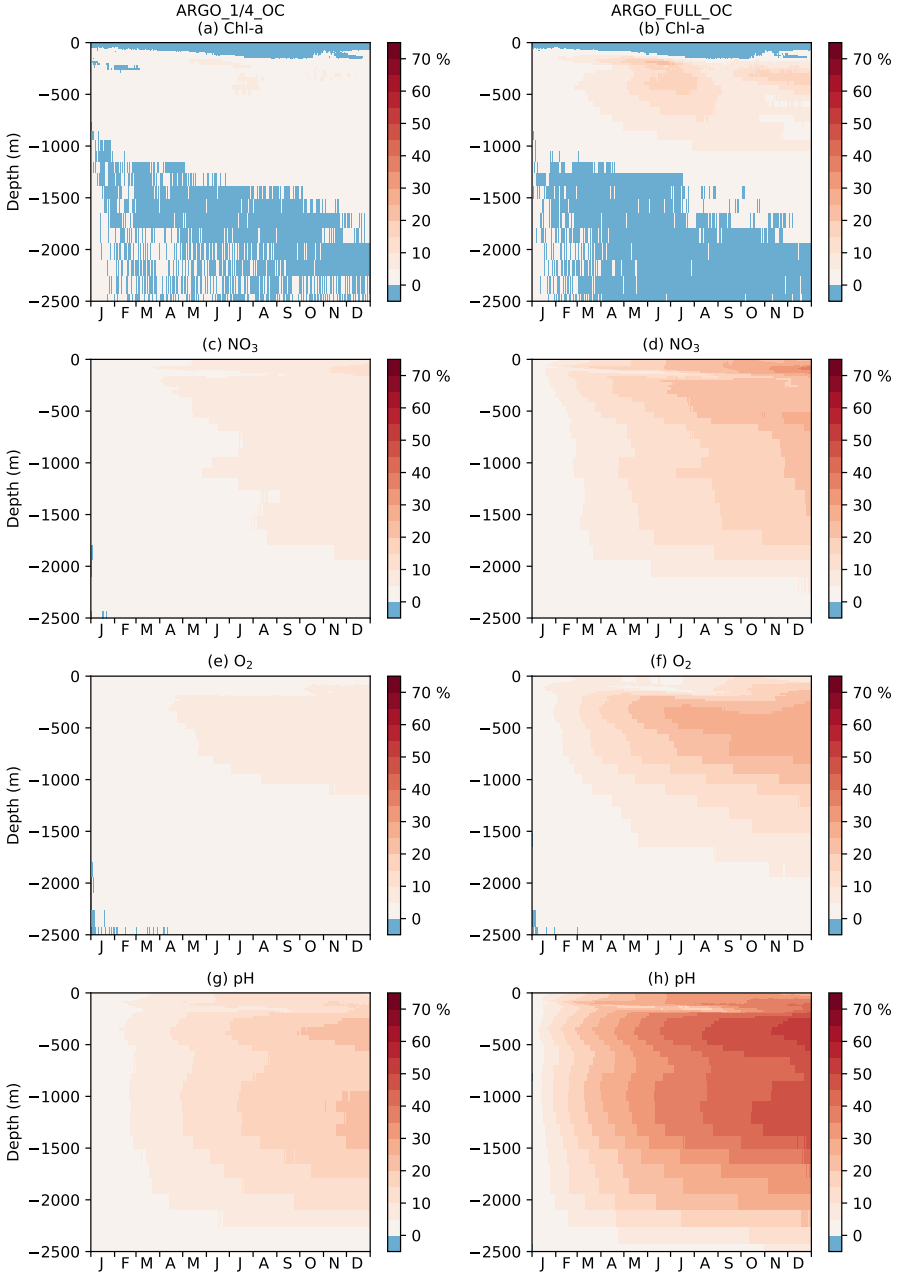


Figure 12. Hovmöller diagram of daily MAE_{red} over OC with depth in the Southern Ocean for Chl-*a*, NO_3, O_2, and pH.

In Fig. 10, global $MAE_{red\_\%}$ is shown. As can be seen in Fig. 9, the sub-surface impact of the BGC-Argo assimilation was not globally uniform, with a stronger impact typically seen at low latitudes than at high latitudes. To demonstrate this further, equivalent versions of Fig. 10 are shown for the Tropical Pacific in Fig. 11, and for the Southern Ocean in Fig. 12. In the

Tropical Pacific, the patterns for all the assimilated variables were very similar to the global average, but with higher $MAE_{red\_\%}$ values, showing a stronger positive impact of the assimilation. This was especially large for Chl-*a* between 80–300 m depth, and pH in the upper 1000 m. $MAE_{red\_\%}$ values for Chl-*a* were largely negative below 300 m, but MAE_{red_abs} values were near-zero (not shown), due to Chl-*a* concentrations being negligible. For NO_3, O_2, and pH, the impact of the assimilation was still spinning up after a year. In the Southern Ocean, $MAE_{red\_\%}$ values were much lower, with especially limited impact from the BGC-Argo assimilation in ARGO_1/4_OC. $MAE_{red\_\%}$ was typically largest for pH and lowest for Chl-*a*. The impact of the assimilation continued to increase with time, so may just have been taking longer to spin up than at lower latitudes. Similar results were seen in the Arctic Ocean (not shown).

5 Summary and discussion

A set of observing system simulation experiments (OSSEs) ~~has been~~ was performed to explore the impact on global ocean biogeochemical reanalyses of assimilating currently-available ocean colour data, and assess the potential impact of assimilating BGC-Argo data. Two different potential BGC-Argo array distributions were tested: one where biogeochemical sensors are placed on all current Argo floats, and one where biogeochemical sensors are placed on a quarter of current Argo floats. This latter approximately corresponds to the proposed BGC-Argo array of 1000 floats (Roemmich et al., 2019). ~~Three different strategies for updating model variables when assimilating ocean colour were assessed: all similarly~~ Assimilating the synthetic ocean colour data greatly improved the assimilated variable surface ~~chl-a~~Chl-*a*, but had a mixed impact on the wider ecosystem and carbon cycle. Assimilating the synthetic BGC-Argo data gave no added benefit over ocean colour in terms of simulating surface ~~chl-a~~Chl-*a*, but for most other variables, including sub-surface ~~chl-a~~Chl-*a*, adding BGC-Argo improved results throughout the water column. This included surface pCO_2, which was not assimilated but is an important output of reanalyses. Both BGC-Argo array distributions gave benefits, with greater improvements seen with increased numbers of observations. The impact of the BGC-Argo assimilation was also increasing with time, and had not yet fully spun up after a year.

Real-world experiments assimilating ~~chl-a~~ Chl-*a* from ocean colour have widely found benefits when validating surface ~~chl-a~~ Chl-*a* against independent observations (Gehlen et al., 2015; Fennel et al., 2019), a conclusion echoed in this study. The impact of ocean colour assimilation on the wider model state has always been more ambiguous, with various studies reporting largely neutral or sometimes negative results, with some evidence of positive impacts highlighted (e.g. Gregg, 2008; Ciavatta et al., 2011; Fontana et al., 2013; Ford and Barciela, 2017). The sparsity of in situ observations, especially for variables such as phytoplankton and zooplankton biomass, has always made it difficult to validate results, or compare conclusions from different studies. Many studies have used inherently multivariate assimilation methods such as the ensemble Kalman filter (Evensen, 2003), while others have employed balance relationships (Ford et al., 2012; Rousseaux and Gregg, 2012; Teruzzi et al., 2014; Skákala et al., 2018). This study ~~tested three variations~~ used a form of the latter~~: just updating chl-a, also updating phytoplankton biomass , and the nitrogen balancing scheme of Hemmings et al. (2008). Solely updating chl-a essentially just acted to change the phytoplankton carbon-to-chlorophyll ratio. This improved chl-a and slightly improved phytoplankton biomass, but had minimal impact on the wider model state. While no damage was done, it was unable to fulfil~~

~~the ambition of extracting maximum information from the assimilated data.~~ , with phytoplankton biomass variables updated to maintain background stoichiometric ratios. Updating phytoplankton biomass, a simple extra step which improved phytoplankton biomass itself and so should be expected to improve other model variables, resulted in a degradation of all other variables examined except for detritus. Zooplankton biomass was especially affected. It seems likely that this degradation occurred due to the changed MEDUSA parameter settings between NATURE and FREE, meaning that the underlying processes were altered such that identical concentrations of phytoplankton now led to different concentrations of other variables. Relevant perturbations included changes to the grazing half-saturation concentration for zooplankton, and nutrient uptake half-saturation concentrations for phytoplankton. This suggests that unless a given biogeochemical model can accurately describe all relevant biogeochemical processes in the ocean, which has not yet been demonstrated, simply improving phytoplankton concentrations might be as likely to degrade as improve other variables. This will also be the case for assimilation schemes which use ensembles to generate cross-correlations between ~~chl-a~~ Chl-*a* and other model variables. ~~Using the~~ These schemes are reliant on the model relationships between variables being correct, as it is these model relationships which the cross-correlations are based on. If the response of zooplankton to an increase in phytoplankton in the model ensemble differs from that in the real ocean, then the cross-correlations used in the assimilation will lead to a zooplankton response which follows the (incorrect) model rather than the real ocean, in exactly the same way as seen in this study. An alternative approach could be to use the nitrogen balancing scheme of Hemmings et al. (2008), which explicitly updates several model state variables to try and account for differing errors in phytoplankton growth and loss processes~~, generally gave an improvement over just updating phytoplankton biomass, but resulted in further degradation of some variables~~. This has been successfully used in previous ocean colour assimilation studies (Ford et al., 2012; Ford and Barciela, 2017; Ford, 2020) with the HadOCC model (Palmer and Totterdell, 2001). It was originally designed and tuned for use with HadOCC, so requires further development and tuning for use with the more complex MEDUSA, but an initial implementation for MEDUSA has been developed. Such a scheme offers more potential for controlling the wider biogeochemical state, especially if it could be expanded to alter parameter values as well as state variables. ~~Furthermore, it was designed for use with the simpler HadOCC model (Palmer and Totterdell, 2001), with which it has been successfully demonstrated (Hemmings et al., 2008; Ford et al., 2012), and only minimally altered for use with MEDUSA, so more specific tuning may help.~~

Adding assimilation of BGC-Argo profiles of ~~chl-a~~Chl-*a*, NO_3, O_2, and pH brought clear improvements to all assimilated variables, and some unassimilated ones. The impact was increased with a larger BGC-Argo array, suggesting that benefits may be seen up to and beyond the target array size of 1000 floats. The observations added important sub-surface information which cannot be obtained from satellite data, but which can yield improvements in simulations of variables such as air-sea CO_2 fluxes. ~~Where~~ All assimilated variables were greatly improved below the mixed layer, but at the surface more limited improvements were seen for Chl-*a* and O_2, than for NO_3 and pH. This is likely due to the relative importance of top-down versus bottom-up control for these variables, and the density of data required for the assimilation to have a major impact. In the case of NO_3, and DIC which helps control pH, concentrations typically increase with depth, and the supply of NO_3 and DIC from below the mixed layer is a major contribution to surface values. Therefore, changes at depth due to the assimilation will alter surface values through indirect processes. O_2 and Chl-*a* typically decrease in concentration with depth, and dynamics within the

mixed layer are much more important in setting surface values. For O_2, major drivers are temperature and ocean–atmosphere exchange. For Chl-*a* a major driver is light availability. It seems that the BGC-Argo ~~data did not result in improvements was for surface chl-a and~~ data was too sparse, even in ARGO_FULL, to have a widespread impact in these circumstances. More dedicated observations within the mixed layer may be likely to have more of an impact on surface values. For Chl-*a*, this can be successfully provided by ocean colour, as the results of this study show. For O_2 ~~. For chl-a, ocean colour observations were able to provide this information, while for O_2~~ and other variables, alternative in situ observing technologies such as gliders may be able to play a role (Telszewski et al., 2018). It should be noted though that the OSSE framework used here did not consider possible real-world issues such as observation biases and inconsistencies between ocean colour and BGC-Argo ~~chl-a observations.~~ Chl-*a* observations. Furthermore, for some variables and regions the error between the free-running model and the nature run was smaller than might be expected in real-world systems, potentially leading to an overestimate of the quantitative impact of BGC-Argo data (Halliwell et al., 2014).

~~From the point of view of ocean data assimilation, the development of~~ For NO_3, O_2, and pH, the overall impact of the assimilation increased with time, and appeared to still be spinning up after a year. The likely explanation is that increments from individual BGC-Argo floats only influenced a relatively small local area, but this influence was long-lasting. Therefore, as further increments in different locations were added with time, the overall error of the system continued to decrease. This suggests that either the BGC-Argo ~~will bring significant advances in reanalysis and forecasting skill. The proposed array of 1000 floats will be enough to deliver clear improvements, but a larger array would be likely to bring further benefits still. Ocean colour and BGC-Argo provide complementary information, so maintaining and developing the existing ocean colour satellite constellation should also be a priority. Technologies such as gliders may also bring additional benefits, especially for nutrients and oxygen in the mixed layer~~observations will take a while to show their full benefit for assimilation, or a greater number of floats will be required, or changes to the assimilation method will be required to make better use of sparse observations. For Chl-*a*, which is typically more dynamic, this appeared to be less of an issue.

There is ~~also~~ much scope for improving data assimilation methodologies to better use existing satellite data, and sparse in situ observations, which could bring at least as much benefit as expanding observing systems. Multivariate balancing, and better integration with physics data assimilation, may help improve unassimilated variables. More effective ways of spreading information from sparse data, such as cross-covariances based on empirical orthogonal functions or derived from an ensemble assimilation scheme, should also be considered. Related to this, the correlation length-scales used by the assimilation should be appropriately tuned for biogeochemical variables. In this study, a single horizontal correlation length-scale based on the first baroclinic Rossby radius was used, varying from a value of 25 km at low latitudes to 150 km at the Equator, following the physics implementation of Waters et al. (2015). This may help explain why the BGC-Argo assimilation demonstrated less widespread impact at high latitudes than near the Equator. A different correlation length-scale may be appropriate for biogeochemical variables. Furthermore, NEMOVAR has recently been developed to allow the use of multiple correlation length-scales (Mirouze et al., 2016), so both small- and large-scale corrections can be considered. The background error standard deviation estimates also need to be refined once real BGC-Argo observations are being assimilated, to reflect the background error in the assimilative system, which will depend on the actual distribution of BGC-Argo floats. The average

ratio of background to observation error may also differ from that assumed in this study. It is likely that this would give different quantitative results, but the qualitative impact of the assimilation would remain similar.

A novel method for assimilating pH was introduced in this study, following the method for assimilating pCO_2 developed by While et al. (2012). The method corrects pH, a diagnostic variable in the model, by making the smallest combined change to DIC and alkalinity required to reach the target pH value. This was successful in improving both pH and DIC, but alkalinity was often degraded. This highlights that making the smallest combined change to DIC and alkalinity does not guarantee that errors in both DIC and alkalinity are minimised. In some circumstances it might be more appropriate to make a smaller or no change to alkalinity, and a larger change to DIC. Or even to make a change of the opposite sign to alkalinity, and an even larger change to DIC to compensate. Unfortunately, without concurrent observations of DIC or alkalinity, this information is not known at the time of assimilation. This is equally the case whether pCO_2 or pH is being assimilated, and so it was decided that the safest assumption would be to make the smallest combined change to DIC and alkalinity. In light of these results it may be worth revisiting this assumption.

From the point of view of ocean data assimilation, BGC-Argo will bring significant advances in reanalysis and forecasting skill, and it is recommended to proceed with its development as a priority. The proposed array of 1000 floats will be enough to deliver clear improvements, and a larger array would be likely to bring further benefits still. Ocean colour and BGC-Argo provide complementary information, so maintaining and developing the existing ocean colour satellite constellation should also be a priority. Technologies such as gliders may also bring additional benefits, for instance for O_2 in the mixed layer.

Data availability. The nature of the 4D data generated in running the model experiments requires a large tape storage facility. These data are in excess of 100 terabytes (TB). However, the data can be made available upon request from the author.

Competing interests. The author declares that they have no conflict of interest.

Acknowledgements. ~~This study received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement 633211 (AtlantOS).~~ The author would like to thank Susan Kay and Matt Martin for useful discussions and comments on the draft manuscript~~.~~, as well as the two anonymous reviewers for their helpful comments in the Biogeosciences discussion forum. This study used synthetic observations based on Argo data. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System.

Financial support. This study received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement 633211 (AtlantOS).

References

Altieri, A. H. and Gedan, K. B.: Climate change and dead zones, Global Change Biology, 21, 1395–1406, https://doi.org/10.1111/gcb.12754, https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12754, 2015.

Anderson, L. A., Robinson, A. R., and Lozano, C. J.: Physical and biological modeling in the Gulf Stream region:: I. Data assimilation methodology, Deep Sea Research Part I: Oceanographic Research Papers, 47, 1787 – 1827, https://doi.org/10.1016/S0967-0637(00)00019-4, http://www.sciencedirect.com/science/article/pii/S0967063700000194, 2000.

Argo: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC), https://doi.org/10.17882/42182, https://www.seanoe.org/data/00311/42182/, 2000.

Arnold, C. P. and Dey, C. H.: Observing-Systems Simulation Experiments: Past, Present, and Future, Bulletin of the American Meteorological Society, 67, 687–695, https://doi.org/10.1175/1520-0477(1986)067<0687:OSSEPP>2.0.CO;2, https://doi.org/10.1175/1520-0477(1986)067<0687:OSSEPP>2.0.CO;2, 1986.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, Quarterly Journal of the Royal Meteorological Society, 134, 1951–1970, https://doi.org/10.1002/qj.339, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.339, 2008.

Bates, N. R., Astor, Y. M., Church, M. J., Currie, K., Dore, J. E., González-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santana-Casiano, J. M.: A Time-Series View of Changing Surface Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO_2 and Ocean Acidification, Oceanography, 27, 126–141, http://www.jstor.org/stable/24862128, 2014.

Behrenfeld, M. J. and Boss, E. S.: Resurrecting the Ecological Underpinnings of Ocean Plankton Blooms, Annual Review of Marine Science, 6, 167–194, https://doi.org/10.1146/annurev-marine-052913-021325, https://doi.org/10.1146/annurev-marine-052913-021325, pMID: 24079309, 2014.

Biogeochemical-Argo Planning Group: The scientific rationale, design and Implementation Plan for a Biogeochemical-Argo float array, Edited by Ken Johnson and Hervé Claustre, https://doi.org/10.13155/46601, 2016.

Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A., and Storkey, D.: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts, Geoscientific Model Development, 7, 2613–2638, https://doi.org/10.5194/gmd-7-2613-2014, https://www.geosci-model-dev.net/7/2613/2014/, 2014.

Bloom, S. C., Takacs, L. L., da Silva, A. M., and Ledvina, D.: Data Assimilation Using Incremental Analysis Updates, Monthly Weather Review, 124, 1256–1271, https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2, https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2, 1996.

Boss, E., Swift, D., Taylor, L., Brickley, P., Zaneveld, R., Riser, S., Perry, M. J., and Strutton, P. G.: Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite, Limnology and Oceanography, 53, 2112–2122, https://doi.org/10.4319/lo.2008.53.5_part_2.2112, https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.2008.53.5_part_2.2112, 2008.

Calvert, D. and Siddorn, J.: Revised vertical mixing parameters for the UK community standard configuration of the global NEMO ocean model, Met Office Hadley Centre Technical Note, 95, https://library.metoffice.gov.uk/Portal/Default/en-GB/RecordView/Index/207517, 2013.

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/10.1029/95JC00458, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95JC00458, 1995.

Ciavatta, S., Torres, R., Saux-Picart, S., and Allen, J. I.: Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?, Journal of Geophysical Research: Oceans, 116, https://doi.org/10.1029/2011JC007219, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JC007219, 2011.

Ciavatta, S., Kay, S., Saux-Picart, S., Butenschön, M., and Allen, J. I.: Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem, Journal of Geophysical Research: Oceans, 121, 1824–1845, https://doi.org/10.1002/2015JC011496, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JC011496, 2016.

Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., Teruzzi, A., and D'Ortenzio, F.: Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea, Ocean Modelling, 133, 112 – 128, https://doi.org/10.1016/j.ocemod.2018.11.005, http://www.sciencedirect.com/science/article/pii/S146350031830060X, 2019.

Davidson, F., Alvera-Azcárate, A., Barth, A., Brassington, G. B., Chassignet, E. P., Clementi, E., De Mey-Frémaux, P., Divakaran, P., Harris, C., Hernandez, F., Hogan, P., Hole, L. R., Holt, J., Liu, G., Lu, Y., Lorente, P., Maksymczuk, J., Martin, M., Mehra, A., Melsom, A., Mo, H., Moore, A., Oddo, P., Pascual, A., Pequignet, A.-C., Kourafalou, V., Ryan, A., Siddorn, J., Smith, G., Spindler, D., Spindler, T., Stanev, E. V., Staneva, J., Storto, A., Tanajura, C., Vinayachandran, P. N., Wan, L., Wang, H., Zhang, Y., Zhu, X., and Zu, Z.: Synergies in Operational Oceanography: The Intrinsic Need for Sustained Ocean Observations, Frontiers in Marine Science, 6, 450, https://doi.org/10.3389/fmars.2019.00450, https://www.frontiersin.org/article/10.3389/fmars.2019.00450, 2019.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.828, 2011.

Diaz, R. J. and Rosenberg, R.: Spreading Dead Zones and Consequences for Marine Ecosystems, Science, 321, 926–929, https://doi.org/10.1126/science.1156401, https://science.sciencemag.org/content/321/5891/926, 2008.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean Acidification: The Other CO_2 Problem, Annual Review of Marine Science, 1, 169–192, https://doi.org/10.1146/annurev.marine.010908.163834, https://doi.org/10.1146/annurev.marine.010908.163834, pMID: 21141034, 2009.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dynamics, 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, https://www.geosci-model-dev.net/9/1937/2016/, 2016.

FAO: The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all, Rome, 2016.

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A. M., Gregoire, M., Jones, E., Kim, H.-C., Lamouroux, J., Murtugudde, R., Perruche, C., and the GODAE OceanView Marine Ecosystem Analysis and Prediction Task Team: Advancing Marine Biogeochemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring

and Managing Ecosystem Health, Frontiers in Marine Science, 6, 89, https://doi.org/10.3389/fmars.2019.00089, https://www.frontiersin.org/article/10.3389/fmars.2019.00089, 2019.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P.: Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components, Science, 281, 237–240, https://doi.org/10.1126/science.281.5374.237, https://science.sciencemag.org/content/281/5374/237, 1998.

Fontana, C., Brasseur, P., and Brankart, J.-M.: Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data, Ocean Science, 9, 37–56, https://doi.org/10.5194/os-9-37-2013, https://www.ocean-sci.net/9/37/2013/, 2013.

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/10.1016/j.rse.2017.03.040, http://www.sciencedirect.com/science/article/pii/S0034425717301438, earth Observation of Essential Climate Variables, 2017.

Ford, D. A.: Assessing the role and consistency of satellite observation products in global physical–biogeochemical ocean reanalysis, Ocean Science, 16, 875–893, https://doi.org/10.5194/os-16-875-2020, https://os.copernicus.org/articles/16/875/2020/, 2020.

Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., and Demaria, J.: Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model, Ocean Science, 8, 751–771, https://doi.org/10.5194/os-8-751-2012, https://www.ocean-sci.net/8/751/2012/, 2012.

Fujii, Y., Rémy, E., Zuo, H., Oke, P., Halliwell, G., Gasparin, F., Benkiran, M., Loose, N., Cummings, J., Xie, J., Xue, Y., Masuda, S., Smith, G. C., Balmaseda, M., Germineaud, C., Lea, D. J., Larnicol, G., Bertino, L., Bonaduce, A., Brasseur, P., Donlon, C., Heimbach, P., Kim, Y., Kourafalou, V., Le Traon, P.-Y., Martin, M., Paturi, S., Tranchant, B., and Usui, N.: Observing System Evaluation Based on Ocean Data Assimilation and Prediction Systems: On-Going Challenges and a Future Vision for Designing and Supporting Ocean Observational Networks, Frontiers in Marine Science, 6, 417, https://doi.org/10.3389/fmars.2019.00417, https://www.frontiersin.org/article/10.3389/fmars.2019.00417, 2019.

Garcia, H. E., Weathers, K., Paver, C. R., Smolyar, I., Boyer, T. P., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Baranova, O. K., Seidov, D., and Reagan, J. R.: World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation, A. Mishonov Technical Ed.; NOAA Atlas NESDIS 83, 2018a.

Garcia, H. E., Weathers, K., Paver, C. R., Smolyar, I., Boyer, T. P., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Baranova, O. K., Seidov, D., and Reagan, J. R.: World Ocean Atlas 2018, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate), A. Mishonov Technical Ed.; NOAA Atlas NESDIS 84, 2018b.

Gasparin, F., Guinehut, S., Mao, C., Mirouze, I., Rémy, E., King, R. R., Hamon, M., Reid, R., Storto, A., Le Traon, P.-Y., Martin, M. J., and Masina, S.: Requirements for an Integrated in situ Atlantic Ocean Observing System From Coordinated Observing System Simulation Experiments, Frontiers in Marine Science, 6, 83, https://doi.org/10.3389/fmars.2019.00083, https://www.frontiersin.org/article/10.3389/fmars.2019.00083, 2019.

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, https://doi.org/10.1080/1755876X.2015.1022350, 2015.

Germineaud, C., Brankart, J.-M., and Brasseur, P.: An Ensemble-Based Probabilistic Score Approach to Compare Observation Scenarios: An Application to Biogeochemical-Argo Deployments, Journal of Atmospheric and Oceanic Technology, 36, 2307–2326, https://doi.org/10.1175/JTECH-D-19-0002.1, https://doi.org/10.1175/JTECH-D-19-0002.1, 2019.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research: Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JC009067, 2013.

Gouretski, V. and Reseghetti, F.: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database, Deep Sea Research Part I: Oceanographic Research Papers, 57, 812 – 833, https://doi.org/10.1016/j.dsr.2010.03.011, http://www.sciencedirect.com/science/article/pii/S0967063710000671, 2010.

Gregg, W. W.: Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, Journal of Marine Systems, 69, 205 – 225, https://doi.org/10.1016/j.jmarsys.2006.02.015, http://www.sciencedirect.com/science/article/pii/S0924796307000607, physical-Biological Interactions in the Upper Ocean, 2008.

Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., Brockmann, C., Chauhan, P., Choi, J.-k., Chuprin, A., Ciavatta, S., Cipollini, P., Donlon, C., Franz, B., He, X., Hirata, T., Jackson, T., Kampel, M., Krasemann, H., Lavender, S., Pardo-Martinez, S., Mélin, F., Platt, T., Santoleri, R., Skákala, J., Schaeffer, B., Smith, M., Steinmetz, F., Valente, A., and Wang, M.: Satellite Ocean Colour: Current Status and Future Perspective, Frontiers in Marine Science, 6, 485, https://doi.org/10.3389/fmars.2019.00485, https://www.frontiersin.org/article/10.3389/fmars.2019.00485, 2019.

Guiavarc'h, C., Roberts-Jones, J., Harris, C., Lea, D. J., Ryan, A., and Ascione, I.: Assessment of ocean analysis and forecast from an atmosphere–ocean coupled data assimilation operational system, Ocean Science, 15, 1307–1326, https://doi.org/10.5194/os-15-1307-2019, https://www.ocean-sci.net/15/1307/2019/, 2019.

Halliwell, G. R., Srinivasan, A., Kourafalou, V., Yang, H., Willey, D., Le Hénaff, M., and Atlas, R.: Rigorous Evaluation of a Fraternal Twin Ocean OSSE System for the Open Gulf of Mexico, Journal of Atmospheric and Oceanic Technology, 31, 105–130, https://doi.org/10.1175/JTECH-D-13-00011.1, https://doi.org/10.1175/JTECH-D-13-00011.1, 2014.

Halliwell, G. R., Mehari, M. F., Hénaff, M. L., Kourafalou, V. H., Androulidakis, I. S., Kang, H. S., and Atlas, R.: North Atlantic Ocean OSSE system: Evaluation of operational ocean observing system components and supplemental seasonal observations for potentially improving tropical cyclone prediction in coupled systems, Journal of Operational Oceanography, 10, 154–175, https://doi.org/10.1080/1755876X.2017.1322770, https://doi.org/10.1080/1755876X.2017.1322770, 2017.

Hemmings, J. C. P., 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, 66, 87–126, https://doi.org/doi:10.1357/002224008784815739, https://www.ingentaconnect.com/content/jmr/jmr/2008/00000066/00000001/art00004, 2008.

Hemmings, J. C. P., Challenor, P. G., and Yool, A.: Mechanistic site-based emulation of a global ocean biogeochemical model (MEDUSA 1.0) for parametric analysis and calibration: an application of the Marine Model Optimization Testbed (MarMOT 1.1), Geoscientific Model Development, 8, 697–731, https://doi.org/10.5194/gmd-8-697-2015, https://www.geosci-model-dev.net/8/697/2015/, 2015.

Hoffman, R. N. and Atlas, R.: Future Observing System Simulation Experiments, Bulletin of the American Meteorological Society, 97, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00200.1, https://doi.org/10.1175/BAMS-D-15-00200.1, 2016.

Hovmöller, E.: The Trough-and-Ridge diagram, Tellus, 1, 62–66, https://doi.org/10.3402/tellusa.v1i2.8498, https://doi.org/10.3402/tellusa.v1i2.8498, 1949.

Hunke, C., E., Lipscomb, H., W., Turner, K., A., Jeffery, N., Elliott, and S.: CICE: the Los Alamos sea ice model documentation and software users' manual, Version 5.1, LA-CC-06-012, Los Alamos National Laboratory, N.M., 2015.

IPCC: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)], IPCC, Geneva, Switzerland, 2014.

Jackson, D. R., Keil, M., and Devenish, B. J.: Use of Canadian Quick covariances in the Met Office data assimilation system, Quarterly Journal of the Royal Meteorological Society, 134, 1567–1582, https://doi.org/10.1002/qj.294, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.294, 2008.

Janjić, T., Bormann, N., Bocquet, M., Carton, J. A., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., and Weston, P.: On the representation error in data assimilation, Quarterly Journal of the Royal Meteorological Society, 144, 1257–1278, https://doi.org/10.1002/qj.3130, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3130, 2018.

Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., Swift, D. D., Williams, N. L., Boss, E., Haëntjens, N., Talley, L. D., and Sarmiento, J. L.: Biogeochemical sensor performance in the SOCCOM profiling float array, Journal of Geophysical Research: Oceans, 122, 6416–6436, https://doi.org/10.1002/2017JC012838, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JC012838, 2017.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 2003.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., Butenschön, M., Enright, C., Halloran, P. R., Le Quéré, C., de Mora, L., Racault, M.-F., Sinha, B., Totterdell, I. J., and Cox, P. M.: iMarNet: an ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework, Biogeosciences, 11, 7291–7304, https://doi.org/10.5194/bg-11-7291-2014, https://www.biogeosciences.net/11/7291/2014/, 2014.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: A 30 years observation-based global monthly gridded sea surface pCO_2 product from 1982 through 2011 (NCEI Accession 0160558), https://doi.org/10.3334/cdiac/otg.spco2_1982_2011_eth_somffn, 2015a.

Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., van Heuven, S., Hoppema, M., Metzl, N., Sweeney, C., Takahashi, T., Tilbrook, B., and Wanninkhof, R.: The reinvigoration of the Southern Ocean carbon sink, Science, 349, 1221–1224, https://doi.org/10.1126/science.aab2620, https://science.sciencemag.org/content/349/6253/1221, 2015b.

Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., Suzuki, T., and Watelet, S.: A new global interior ocean mapped climatology: the $1^\circ \times 1^\circ$ GLODAP version 2, Earth System Science Data, 8, 325–340, https://doi.org/10.5194/essd-8-325-2016, https://essd.copernicus.org/articles/8/325/2016/, 2016.

Lévy, M., Estublier, A., and Madec, G.: Choice of an advection scheme for biogeochemical models, Geophysical Research Letters, 28, 3725–3728, https://doi.org/10.1029/2001GL012947, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001GL012947, 2001.

Lindstrom, E., Gunn, J., Fischer, A., McCurdy, A., and Glover, L.: A Framework for Ocean Observing. By the Task Team for an Integrated Framework for Sustained Ocean Observing, UNESCO 2012, IOC/INF-1284, https://doi.org/10.5270/OceanObs09-FOO, 2012.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system,

Quarterly Journal of the Royal Meteorological Society, 141, 1072–1084, https://doi.org/10.1002/qj.2396, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2396, 2015.

Madec, G.: NEMO ocean engine, Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No, 1288-1619, 2008.

Mao, C., King, R. R., Reid, R., Martin, M., and Good, S.: Impact of in situ Observations in FOAM: An Observing System Simulation Experiments Perspective, in revision.

Martin, G. M., Milton, S. F., Senior, C. A., Brooks, M. E., Ineson, S., Reichler, T., and Kim, J.: Analysis and Reduction of Systematic Errors through a Seamless Approach to Modeling Weather and Climate, Journal of Climate, 23, 5933–5957, https://doi.org/10.1175/2010JCLI3541.1, https://doi.org/10.1175/2010JCLI3541.1, 2010.

Masutani, M., Schlatter, T. W., Errico, R. M., Stoffelen, A., Andersson, E., Lahoz, W., Woollen, J. S., Emmitt, G. D., Riishøjgaard, L.-P., and Lord, S. J.: Observing System Simulation Experiments, in: Data Assimilation: Making Sense of Observations, edited by Lahoz, W., Khattatov, B., and Menard, R., pp. 647–679, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-74703-1_24, https://doi.org/10.1007/978-3-540-74703-1_24, 2010.

McKinley, G. A., Fay, A. R., Lovenduski, N. S., and Pilcher, D. J.: Natural Variability and Anthropogenic Trends in the Ocean Carbon Sink, Annual Review of Marine Science, 9, 125–150, https://doi.org/10.1146/annurev-marine-010816-060529, https://doi.org/10.1146/annurev-marine-010816-060529, pMID: 27620831, 2017.

Mirouze, I., Blockley, E. W., Lea, D. J., Martin, M. J., and Bell, M. J.: A multiple length scale correlation operator for ocean data assimilation, Tellus A: Dynamic Meteorology and Oceanography, 68, 29 744, https://doi.org/10.3402/tellusa.v68.29744, https://doi.org/10.3402/tellusa.v68.29744, 2016.

Mogensen, K. S., Balmaseda, M. A., Weaver, A., Martin, M., and Vidard, A.: NEMOVAR: a variational data assimilation system for the NEMO ocean model, ECMWF Newsletter, 120, 17–21, https://doi.org/10.21957/3yj3mh16iq, https://www.ecmwf.int/node/17483, 2009.

Mogensen, K. S., Balmaseda, M. A., and Weaver, A.: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4, Tech. Rep. 668, ECMWF, https://doi.org/10.21957/x5y9yrtm, https://www.ecmwf.int/node/11174, 2012.

Munhoven, G.: Mathematics of the total alkalinity-pH equation - pathway to robust and universal solution algorithms: the SolveSAPHE package v1.0.1, Geoscientific Model Development, 6, 1367–1388, https://doi.org/10.5194/gmd-6-1367-2013, https://www.geosci-model-dev.net/6/1367/2013/, 2013.

Orr, J. C. and Epitalon, J.-M.: Improved routines to model the ocean carbonate system: mocsy 2.0, Geoscientific Model Development, 8, 485–499, https://doi.org/10.5194/gmd-8-485-2015, https://www.geosci-model-dev.net/8/485/2015/, 2015.

Palmer, J. and Totterdell, I.: Production and export in a global ocean ecosystem model, Deep Sea Research Part I: Oceanographic Research Papers, 48, 1169 – 1198, https://doi.org/10.1016/S0967-0637(00)00080-7, http://www.sciencedirect.com/science/article/pii/S0967063700000807, 2001.

Park, J.-Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J., and Zhang, S.: Modeling Global Ocean Biogeochemistry With Physical Data Assimilation: A Pragmatic Solution to the Equatorial Instability, Journal of Advances in Modeling Earth Systems, 10, 891–906, https://doi.org/10.1002/2017MS001223, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001223, 2018.

Polavarapu, S., Ren, S., Rochon, Y., Sankey, D., Ek, N., Koshyk, J., and Tarasick, D.: Data assimilation with the Canadian middle atmosphere model, Atmosphere-Ocean, 43, 77–100, https://doi.org/10.3137/ao.430105, https://doi.org/10.3137/ao.430105, 2005.

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/10.1029/2019JC015586, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JC015586, e2019JC015586 2019JC015586, 2020.

Racault, M.-F., Sathyendranath, S., Brewin, R. J. W., Raitsos, D. E., Jackson, T., and Platt, T.: Impact of El Niño Variability on Oceanic Phytoplankton, Frontiers in Marine Science, 4, 133, https://doi.org/10.3389/fmars.2017.00133, https://www.frontiersin.org/article/10.3389/fmars.2017.00133, 2017.

Raghukumar, K., Edwards, C. A., Goebel, N. L., Broquet, G., Veneziani, M., Moore, A. M., and Zehr, J. P.: Impact of assimilating physical oceanographic data on modeled ecosystem dynamics in the California Current System, Progress in Oceanography, 138, 546 – 558, https://doi.org/10.1016/j.pocean.2015.01.004, http://www.sciencedirect.com/science/article/pii/S0079661115000063, combining Modeling and Observations to Better Understand Marine Ecosystem Dynamics, 2015.

Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., and Schroeder, D.: The sea ice model component of HadGEM3-GC3.1, Geoscientific Model Development, 11, 713–723, https://doi.org/10.5194/gmd-11-713-2018, https://www.geosci-model-dev.net/11/713/2018/, 2018.

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., Le Traon, P.-Y., 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, 6, 439, https://doi.org/10.3389/fmars.2019.00439, https://www.frontiersin.org/article/10.3389/fmars.2019.00439, 2019.

Rousseaux, C. S. and Gregg, W. W.: Climate variability and phytoplankton composition in the Pacific Ocean, Journal of Geophysical Research: Oceans, 117, https://doi.org/10.1029/2012JC008083, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012JC008083, 2012.

Rousseaux, C. S. and Gregg, W. W.: Recent decadal trends in global phytoplankton composition, Global Biogeochemical Cycles, 29, 1674–1688, https://doi.org/10.1002/2015GB005139, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GB005139, 2015.

Sathyendranath, S., Brewin, R. J., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A. B., Dingle, J., Doerffer, R., Donlon, C., Dowell, M., Farman, A., Grant, M., Groom, S., Horseman, A., Jackson, T., Krasemann, H., Lavender, S., Martinez-Vicente, V., Mazeran, C., Mélin, F., Moore, T. S., Müller, D., Regner, P., Roy, S., Steele, C. J., Steinmetz, F., Swinton, J., Taberner, M., Thompson, A., Valente, A., Zühlke, M., Brando, V. E., Feng, H., Feldman, G., Franz, B. A., Frouin, R., Gould, R. W., Hooker, S. B., Kahru, M., Kratzer, S., Mitchell, B. G., Muller-Karger, F. E., Sosik, H. M., Voss, K. J., Werdell, J., and Platt, T.: An Ocean-Colour Time Series for Use in Climate Studies: The Experience of the Ocean-Colour Climate Change Initiative (OC-CCI), Sensors, 19, https://doi.org/10.3390/s19194285, https://www.mdpi.com/1424-8220/19/19/4285, 2019.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, Geophysical Research Letters, 41, 2514–2519, https://doi.org/10.1002/2014GL059637, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL059637, 2014.

Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A.: Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling, Biogeosciences, 14, 1647–1701, https://doi.org/10.5194/bg-14-1647-2017, https://www.biogeosciences.net/14/1647/2017/, 2017.

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/10.1029/2019MS001739, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001739, 2019.

Skákala, J., Ford, D., Brewin, R. J., McEwan, R., Kay, S., Taylor, B., de Mora, L., and Ciavatta, S.: The Assimilation of Phytoplankton Functional Types for Operational Forecasting in the Northwest European Shelf, Journal of Geophysical Research: Oceans, 123, 5230–5247, https://doi.org/10.1029/2018JC014153, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JC014153, 2018.

Skákala, J., Bruggeman, J., Brewin, R. J. W., Ford, D. A., and Ciavatta, S.: Improved Representation of Underwater Light Field and Its Impact on Ecosystem Dynamics: A Study in the North Sea, Journal of Geophysical Research: Oceans, 125, e2020JC016 122, https://doi.org/10.1029/2020JC016122, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JC016122, e2020JC016122 10.1029/2020JC016122, 2020.

Storkey, D., Blockley, E. W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M. J., Barciela, R. M., Hines, A., Hyder, P., and Siddorn, J. R.: Forecasting the ocean state using NEMO:The new FOAM system, Journal of Operational Oceanography, 3, 3–15, https://doi.org/10.1080/1755876X.2010.11020109, https://doi.org/10.1080/1755876X.2010.11020109, 2010.

Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., Calvert, D., Graham, T., Hewitt, H. T., Hyder, P., Kuhlbrodt, T., Rae, J. G. L., and Sinha, B.: UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions, Geoscientific Model Development, 11, 3187–3213, https://doi.org/10.5194/gmd-11-3187-2018, https://www.geosci-model-dev.net/11/3187/2018/, 2018.

Telszewski, M., Palacz, A., and Fischer, A.: Biogeochemical In-situ Observations - Motivation, Status and New Frontiers, in: New Frontiers in Operational Oceanography, edited by Chassignet, E. P., Pascual, A., Tintoré, J., and Verron, J., chap. 6, pp. 131–160, GODAE OceanView, 2018.

Teruzzi, A., Dobricic, S., Solidoro, C., and Cossarini, G.: A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties, Journal of Geophysical Research: Oceans, 119, 200–217, https://doi.org/10.1002/2013JC009277, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JC009277, 2014.

Torres, R., Allen, J., and Figueiras, F.: Sequential data assimilation in an upwelling influenced estuary, Journal of Marine Systems, 60, 317 – 329, https://doi.org/10.1016/j.jmarsys.2006.02.001, http://www.sciencedirect.com/science/article/pii/S0924796306000315, 2006.

Valsala, V. and Maksyutov, S.: Simulation and assimilation of global ocean pCO2 and air2̌013sea CO2 fluxes using ship observations of surface ocean pCO2 in a simplified biogeochemical offline model, Tellus B: Chemical and Physical Meteorology, 62, 821–840, https://doi.org/10.1111/j.1600-0889.2010.00495.x, https://doi.org/10.1111/j.1600-0889.2010.00495.x, 2010.

Van Leer, B.: Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection, Journal of Computational Physics, 23, 276–299, 1977.

Verdy, A. and Mazloff, M. R.: A data assimilating model for estimating Southern Ocean biogeochemistry, Journal of Geophysical Research: Oceans, 122, 6968–6988, https://doi.org/10.1002/2016JC012650, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JC012650, 2017.

Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A., and While, J.: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model, Quarterly Journal of the Royal Meteorological Society, 141, 333–349, https://doi.org/10.1002/qj.2388, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2388, 2015.

Weaver, A. T., Vialard, J., and Anderson, D. L. T.: Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks, Monthly Weather Review, 131, 1360–1378, https://doi.org/10.1175/1520-0493(2003)131<1360:TAFVAW>2.0.CO;2, https://doi.org/10.1175/1520-0493(2003)131<1360:TAFVAW>2.0.CO;2, 2003.

Weaver, A. T., Deltel, C., Machu, E., Ricci, S., and Daget, N.: A multivariate balance operator for variational ocean data assimilation, Quarterly Journal of the Royal Meteorological Society, 131, 3605–3625, https://doi.org/10.1256/qj.05.119, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.119, 2005.

While, J., Totterdell, I., and Martin, M.: Assimilation of pCO_2 data into a global coupled physical-biogeochemical ocean model, Journal of Geophysical Research: Oceans, 117, https://doi.org/10.1029/2010JC006815, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006815, 2012.

Wijffels, S., Roemmich, D., Monselesan, D., Church, J., and Gilson, J.: Ocean temperatures chronicle the ongoing warming of Earth, Nature Climate Change, 6, 116–118, https://doi.org/10.1038/nclimate2924, 2016.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations, Journal of Advances in Modeling Earth Systems, 10, 357–380, https://doi.org/10.1002/2017MS001115, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001115, 2017.

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, https://www.geosci-model-dev.net/6/1767/2013/, 2013.

Yool, A., Palmiéri, J., Jones, C. G., Sellar, A. A., de Mora, L., Kuhlbrodt, T., Popova, E. E., Mulcahy, J. P., Wiltshire, A., Rumbold, S. T., Stringer, M., Hill, R. S. R., Tang, Y., Walton, J., Blaker, A., Nurser, A. J. G., Coward, A. C., Hirschi, J., Woodward, S., Kelley, D. I., Ellis, R., and Rumbold-Jones, S.: Spin-up of UK Earth System Model 1 (UKESM1) for CMIP6, Journal of Advances in Modeling Earth Systems, n/a, e2019MS001 933, https://doi.org/10.1029/2019MS001933, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001933, e2019MS001933 2019MS001933, 2020.

Yu, L., Fennel, K., Bertino, L., Gharamti, M. E., and Thompson, K. R.: Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling, Ocean Modelling, 126, 13 – 28, https://doi.org/10.1016/j.ocemod.2018.04.005, http://www.sciencedirect.com/science/article/pii/S1463500318301318, 2018.

Yu, L., Fennel, K., Wang, B., Laurent, A., Thompson, K. R., and Shay, L. K.: Evaluation of nonidentical versus identical twin approaches for observation impact assessments: an ensemble-Kalman-filter-based ocean assimilation application for the Gulf of Mexico, Ocean Science, 15, 1801–1814, https://doi.org/10.5194/os-15-1801-2019, https://www.ocean-sci.net/15/1801/2019/, 2019.

Zalesak, S. T.: Fully multidimensional flux-corrected transport algorithms for fluids, Journal of Computational Physics, 31, 335–362, 1979.