https://doi.org/10.5194/os-2021-35
Preprint. Discussion started: 3 May 2021

Ocean Science
Discussions
Open Access
EGU

Can assimilation of satellite observations improve subsurface biological properties in a numerical model? A case study for the Gulf of Mexico

Bin Wang[1], Katja Fennel[1], Liuqian Yu[1,2]
[1]Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada
[2]Department of Ocean Science, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

Correspondence to: Bin Wang (Bin.Wang@dal.ca)

Abstract

Given current threats to ocean ecosystem health, there is a growing demand for accurate biogeochemical hindcasts, nowcasts, and predictions. Provision of such products requires data assimilation, i.e., a comprehensive strategy for incorporating observations into biogeochemical models, but current data streams of biogeochemical observations are generally considered insufficient for the operational provision of such products. This study investigates to what degree the satellite observations in combination with sparse BGC Argo profiles can improve subsurface biogeochemical properties. The multivariate Deterministic Ensemble Kalman Filter (DEnKF) has been implemented to assimilate physical and biological observations into a biogeochemical model of the Gulf of Mexico. First, the biogeochemical model component was tuned using BGC-Argo observations. Then, observations of sea surface height, sea surface temperature, and surface chlorophyll were assimilated, and profiles of both physical and biological variables were updated based on the surface information. We assessed whether this leads to improved subsurface distributions, especially of biological properties, using observations from five BGC-Argo floats that were not assimilated, but used in the *a priori* tuning. Results show that assimilation of the satellite data improves model representation of major circulation features, which translate into improved three-dimensional distributions of temperature and salinity. The multivariate assimilation also improves the agreement of subsurface nitrate through its tight correlation with temperature, but the improvements in subsurface chlorophyll were modest initially due to suboptimal choices of the model's optical module. Repeating the assimilation run after adjusting light attenuation

parameterization through further *a priori* tuning greatly improved the subsurface distribution of chlorophyll. Therefore, even sparse BGC-Argo observations can provide substantial benefits to biogeochemical prediction by enabling *a priori* model tuning. Given that, so far, the abundance of BGC-Argo profiles in the Gulf of Mexico and elsewhere is insufficient for sequential assimilation, updating 3D biological properties in a model that has been well calibrated is an intermediate step toward full assimilation of the new data types.

1. Introduction

Given the multiple and increasing pressures of ocean warming, acidification, deoxygenation, and changes in primary productivity on ocean ecosystem health, accurate model simulations are urgently needed to assess past and current states of marine ecosystems, forecast future trends, and predict the ocean's response to different scenarios of climate change and management policies. In practice, numerical models are imperfect representations of the natural system and their accuracy is limited by many factors including insufficient model resolution, inaccuracies in discretion schemes and model formulations, parameterization of unresolved processes, and uncertainties in model inputs. Data assimilation is a practical approach used to compensate for these model deficiencies. It can be viewed as a statistical method to interpolate and extrapolate the sparse observations into the regular model space in a dynamically consistent way. Its success critically depends on well-resolved observations

Data assimilation is well developed in physical oceanography (Edwards et al. 2015) but less mature in biogeochemical ocean modelling, largely due to insufficient observations (Fennel et al., 2019). Thus far, satellite data of ocean color have been the major source of observations to be assimilated (Gregg, 2008; Hu et al., 2012; Mattern et al., 2013; Teruzzi et al., 2018) because of their relatively high resolution and routine availability. More recent advances have focused on the incorporation of other satellite-derived products including size-fractionated chlorophyll (Ciavatta et al., 2018; Pradhan et al., 2019; Skákala et al., 2018; Xiao and Friedrichs, 2014a, 2014b) and optical properties (Ciavatta et al., 2014; Shulman et al., 2013). However, these measurements are limited to the surface ocean and provide little information about the subsurface and ocean interior. In addition, it has been acknowledged that assimilating satellite data of ocean color often fails to improve and even degrades simulation of unobserved biological variables (Ciavatta et al., 2018; Fontana et al., 2013; Ford and Barciela, 2017; Skákala et al., 2018; Teruzzi et al.,

https://doi.org/10.5194/os-2021-35
Preprint. Discussion started: 3 May 2021

2018). Problems also remain in accounting for the co-dependencies or covariances between biological variables. For instance, Fontana et al. (2013) found subsurface nitrate was barely impacted by assimilating the satellite surface chlorophyll because of its weak correlations with surface chlorophyll. Although BGC-Argo floats may ultimately provide us with abundant subsurface measurements of multiple key biogeochemical properties (Biogeochemical-Argo Planning Group, 2016; Chai et al., 2020; Roemmich et al., 2019), the profiling observations will likely remain insufficient for three-dimensional data assimilation for a number of years, making satellite data the main observation streams for sequential data assimilation in biogeochemical models (Ford, 2021).

The insufficient availability of subsurface and interior ocean biogeochemical observations is not only reflected in the immaturity of biogeochemical data assimilation but also its skill assessment. To date, skill assessments of biogeochemical data assimilation systems have rarely looked below the surface. Although a few studies exist (Ourmières et al., 2009; Teruzzi et al., 2014), the simulated vertical structures are compared with observations on basin- or climatological scales. This makes the below-surface impacts of biogeochemical data assimilation poorly known at present.

Finally, since physical processes affect biological properties through advection and diffusion of biological tracers as well as some temperature-dependent biological activities (e.g. phytoplankton growth), deficiencies in biological models can arise from imperfect simulation of the physics (Doney, 1999; Doney et al., 2004; Oschlies and Garcon, 1999). Although there have been studies demonstrating a positive effect of physical data assimilation on biological properties (Fiechter et al., 2011; Ourmières et al., 2009), often this approach degrades biological distributions because of elevated vertical velocities and violation of consistency between physical and biological properties (Anderson et al., 2000; Raghukumar et al., 2015; Yu et al., 2018). To address these issues, joint assimilation of physical and biological observations (Song et al., 2016b, 2016a) or multivariate updates based on the cross-covariances between physical and biological properties (Goodliff et al., 2019; Yu et al., 2018) have been suggested.

In this study, a multivariate physical-biological data assimilation scheme is applied to a coupled physical-biological model in the Gulf of Mexico. The rationale for choosing the Gulf of Mexico is that the dominant circulation, including the Loop Current and its associated mesoscale eddies, is stochastic and can influence the subsurface biological distributions, e.g. deep chlorophyll maximum (Fommervault et al., 2017). By comparing forecast results from the assimilative model with independent observations from five BGC-Argo floats, we rigorously evaluate whether the main biological observation stream (satellite estimates of surface chlorophyll) in combination with physical observations (satellite estimates of sea surface height and sea surface temperature) can inform the 3D ocean distributions in high spatial and temporal resolution.

2. Tools and methods

2.1 Coupled physical and biological model

The coupled physical and biological model used in this study is based on the Regional Ocean Modeling System (Haidvogel et al., 2008, ROMS, https://www.myroms.org) and configured in the Gulf of Mexico (Figure 1) with a horizontal resolution of ~5 km and 36 vertical sigma levels (Wang et al., 2020; Yu et al., 2019). The model used a Multidimensional Positive Definitive Advection Transport Algorithm (MPDATA, Smolarkiewicz and Margolin 1998) to solve the horizontal and vertical advection of tracers, a Smagorinsky-type formula (Smagorinsky, 1963) to parameterize horizontal viscosity and diffusivity, and the Mellor-Yamada 2.5-level closure scheme (Mellor and Yamada, 1982) to calculate the vertical turbulent mixing. Atmospheric forcing is provided by the European Centre for Medium-Range Weather Forecast ERA-Interim product (ECMWF reanalysis, https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim) with a horizontal resolution of 1/8° to calculate the surface wind stress as well as the net heat fluxes and freshwater fluxes.

The biological model uses a nitrogen-based model (Fennel et al., 2006) to simulate transportation and transformation of seven pelagic variables, i.e. nitrate (NO3), ammonium (NH4), chlorophyll (Chl), phytoplankton (Phy), zooplankton (Zoo), small detritus (SDet), and large detritus (LDet). As a separate state variable, chlorophyll accounts for photoacclimation based on Geider et al. (1997). In our coupled model, the biological tracers are advected and diffused as part of the 3D circulation but provide no feedback to the physical model. Biological parameters are from the parameter optimization study of Wang et al. (2020) except that the half-saturation constant of nitrate was re-tuned from 0.5 mmol N m^{-3} to about 1.4 mmol N m^{-3} because the previous model underestimated the nitrate in the euphotic zone.

https://doi.org/10.5194/os-2021-35
Preprint. Discussion started: 3 May 2021

The coupled model receives freshwater and nutrients inputs from the Mississippi-Atchafalaya river systems which are specified by the daily measurements from the US Geological Survey river gauges and those from other major rivers which utilize the climatological estimates (Xue et al., 2013). To ensure a dynamically consistent biological field, a one-year spin-up is performed in 2014 where the physical model is initialized from the output of the 1/12° data-assimilative global HYCOM/NCODA (https://www.hycom.org) and the biological model starts from a regressed 3D field of nitrate based on its climatological relationship with temperature (see Figure S1). A semi-prognostic method is used during the spin-up period to reduce model drift by replacing model density with a linear combination of model and climatological density fields when calculating the horizontal pressure gradient (Greatbatch et al., 2004; Sheng et al., 2001). After the spin-up, experiments are performed for a year from January 2015 to December 2015.

2.2 Data assimilation technique

In this study, the data assimilation scheme uses the deterministic formulation of the Ensemble Kalman Filter (DEnKF) which was first introduced by Sakov and Oke (2008). The approach consists of two steps: 1) the forecast step in which an ensemble of state variables is integrated forward in time by the model, and 2) the analysis step in which observations are assimilated to update the forecasted ensemble following the Kalman Filter equations

$$x^a = x^f + K(d - Hx^f), \tag{1}$$

$$\mathrm{K} = P^f H^T (HP^f H^T + R)^{-1} \tag{2}$$

where x represents the model state estimate, d represents the available observations, H represents the measurement operator mapping the model state onto observations, and K represents the Kalman gain matrix which is determined by the model error matrix P and observation error matrix R (Equ. 2). Superscripts a and f represent analysis (i.e. updated) and forecast (i.e. prior to the update) estimates, and T represents the matrix transpose. Unlike the original stochastic EnKF which updates each ensemble member with perturbed observations, the DEnKF updates ensemble mean ($\bar{x}$) and anomalies ($A = x - \bar{x}$) separately without perturbating observations, i.e. the former is updated as in equ. 1 while the latter is updated by

$$A^a = A^f - \frac{1}{2} K H A^f \,. \tag{3}$$

More details of the DEnKF can be referred to Sakov and Oke (2008) and Yu et al. (2018).

2.3 Observations

In this study, physical and biological observations are jointly assimilated to constrain the coupled model. The observations assimilated include sea surface height (SSH), sea surface temperature (SST), Argo T-S profiles, and satellite estimates of surface chlorophyll.

The SSH observations for assimilation are obtained by adding the 1/4° mapped Sea level anomaly (SLA) from Archiving Validation and Interpretation of Satellite Oceanographic Data (AVISO, http://marine.copernicus.eu) to a mean dynamic topography (MDT) from Rio et al. (2013), and are adjusted by removing the spatially averaged mismatches between assimilated and forecasted SSH to account for differences in reference time between the SLA data (1993-2012) and our coupled model (2015) (Haines et al., 2011; Song et al., 2016b; Xu et al., 2012). This is equivalent to assimilating the SSH gradient into the model, as it is the only dynamically meaningful quantity for driving the geostrophic component of ocean currents and adjusting subsurface thermohaline structures. The SST observations are Advanced Very High Resolution Radiometer (AVHRR, https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1) product with a horizontal resolution of 0.01°. Observation errors are specified as 0.02 m for SSH and 0.3°C for SST.

The surface chlorophyll is provided by the Ocean-Colour Climate Change Initiative project (OC-CCI, https://www.oceancolour.org) at a daily frequency with a spatial resolution of 1/24°. However, for the daily chlorophyll field, a large portion of data can be missing due to cloud cover and inter-orbit gaps. In 2015 for the Gulf of Mexico, the spatial coverage of surface chlorophyll varies from 0 to 63% with a mean coverage of 9.5±9.0%. Hence, to increase the availability of observations, an asynchronous data assimilation method (Sakov et al., 2010) is applied by assimilating the daily records of surface chlorophyll within a 7-day window before the update. Errors associated with the surface chlorophyll are set to be 35% of the measured concentrations.

Profiling observations are from the International Argo project (hereafter referred to as Argo floats; http://www.usgodae.org/) and five BGC-Argo floats which were funded by the Bureau of Ocean Energy

Management (hereafter referred to as BOEM floats). In 2015, the Argo floats provided nearly 800 T-S profiles extending from the surface to 2,000 m depth in the Gulf of Mexico. These are treated either as independent observations for model skill assessment or, in the DAargo experiment (see Section 2.4), assimilated with uncertainties being 0.3°C for temperature and 0.01 for salinity. The BOEM floats collected more than 500 profiles of temperature, salinity, chlorophyll, and backscatter at a bi-weekly frequency from 2011 to 2015, 114 profiles of which were collected in 2015 (see Figure 1 for their locations) and are used as independent observations. Backscatter is converted into phytoplankton and particulate organic carbon (POC) concentrations following Wang et al. (2020).

2.4 Simulation strategy

For the sake of keeping our data assimilation experiments computationally affordable, we chose an ensemble size of 20 which has been successfully used in previous studies, e.g. an idealized channel (Yu et al., 2018), the Middle Atlantic Bight (Hu et al., 2012; Mattern et al., 2013), and the Gulf of Mexico (Yu et al., 2019). Spurious correlations, which can arise with relatively small ensembles, are avoided here by applying a distance-based localization with a radius of 50 km (Evensen, 2003). In addition, ensemble anomalies are inflated by 1.05 at each update step to account for the possible underestimation of ensemble spread (Anderson and Anderson, 1999).

In order to account for uncertainties in the model's initial, boundary and atmospheric forcing conditions and biological parameters, the ensemble is initialized from 20 different daily outputs, centered on the initial date of 1 January 2015, from a previous deterministic model simulation (as described above in Section 2.1) and is forced by open boundary conditions which are lagged by up to ±10 days for the different ensemble members. Furthermore, each ensemble member is forced by a perturbed version of the wind forcing. Specifically, the wind forcing from the deterministic run is decomposed into empirical orthogonal functions (EOFs) and then the first 4 EOFs are perturbed by multiplying random numbers with zero mean variance of 1 as in Li et al. (2016) and Thacker et al. (2012). In addition, four sensitive biological parameters, namely the mortality rate of phytoplankton, the maximum ratio of chlorophyll to carbon, the grazing rate of zooplankton, and the growth rate of phytoplankton at 0 °C, are subject to a Gaussian perturbation with a relative variance of 75% and each parameter is resampled from the distribution before each forecast step to prevent some extreme parameter values being used throughout the whole data assimilation experiment.

We performed three 1-year simulations in 2015: 1) a deterministic model simulation without data assimilation (henceforth referred to as **Free** simulation), 2) an ensemble run assimilating satellite data (SSH, SST, and satellite surface chlorophyll) only (henceforth **DAsat**), and 3) and an ensemble run assimilating Argo T-S profiles in addition to satellite data (henceforth **DAargo**).

A two-step update is used on a weekly data assimilation cycle in both assimilative experiments, where the physical observations are first assimilated to update both physical and biological state variables through the multivariate covariance, and chlorophyll observations are assimilated next to update only biological state variables. Given the computational expense, we limit updates to two physical variables (temperature and salinity) and four biological variables (nitrate, chlorophyll, phytoplankton, and zooplankton). All these state variables are updated throughout the whole water column.

To evaluate the prediction skill, we calculate the root-mean-square-errors (RMSE) of model forecast (*M*) with respect to assimilated and independent observations (*O*)

$$RMSE = \sqrt{\frac{1}{N}\sum(M - O)^2} \tag{4}$$

where *N* represents the number of model-data pairs available. In the absence of direct measurements for nitrate, we estimate nitrate profiles along the BOEM float trajectories based on their climatological relationship with temperature (Figure S1). To account for the overestimation of nitrate in warm waters which typically occurs in the euphotic zone, an unbiased root-mean-square-error (unbiased RMSE) is used to quantify the model-data misfit of nitrate.

$$unbiased\ RMSE = \sqrt{\frac{1}{N}\sum\left(M - O - \frac{1}{N}\sum(M - O)\right)^2} \tag{5}$$

3. Results

3.1 Assimilation impacts on physical properties

The dominant circulation features in the Gulf of Mexico, the Loop Current and Loop Current Eddies, are assessed by comparing their fronts, defined here as the 10-cm SSH contour, from satellite data, the free run, and the two data-assimilative runs. In the first two months, all model estimates of the Loop Current are different from satellite observations due to the influence of initial conditions (Figure 2). After March, the SSH field shows a similar northward and westward extension of the Loop Current intrusion between two assimilative runs and satellite observations, but large deviations from observations remain in the Free run. In addition, all estimates except for the Free run well reproduce the satellite-observed timing of eddy shedding as well as the size, shape, and position of Loop Current eddies.

For a more quantitative assessment, the daily output of SSH and SST fields from the three runs are compared with the satellite estimates (Figure 3). This shows large reductions of RMSEs for SSH and SST in most of the model domain in two data-assimilative runs. The most significant reductions for SSH are located in the regions influenced by Loop Current and Loop Current eddies. In contrast, the reductions in SST RMSEs are more spatially homogeneous with less dispersion in their probability density function. A summary of the overall RMSEs for physical variables from the free run and data assimilative runs is shown in Table 1. In general, the two data-assimilative runs both significantly improved SSH (by 54%) and SST (by 36%). Differences between the DAsat and DAargo runs are small.

The correction of mesoscale features by data assimilation was not limited to the surface but extend to the subsurface and even deep waters. Specifically, the two assimilative runs corrected the position, the amplitude, and the polarity of mesoscale eddies, and hence better represented the elevated and depressed thermoclines within these eddies (Figure 4). The most noticeable improvement (by 60~61%) was witnessed by the Float 287 which captured a newly detached Loop Current eddy with features of high SSH and depressed thermoclines during July and October. In addition, assimilation of Argo T-S profiles in the DAargo run led to slight further improvements in the subsurface temperature distributions when compared with the DAsat run. For instance, although the DAsat run greatly improved subsurface temperature distributions along the trajectory of Float 285, an underestimation of temperature at about 200 m depth remains within the peak of the anticyclonic eddy. Corrections imposed by assimilating Argo profiles increased temperature here and decreased the bias from observations. These small but localized further improvements can also be observed by other floats, e.g. in July-October for Float 289 and February for Float 290.

In general, assimilating the satellite data in the DAsat run resulted in large reductions in RMSEs of 3D temperature (by 46%~48%; Table 1) and salinity (by 36%~39%; Table 1) with respect to Argo floats and BOEM floats (Figure 5). The reductions extend to over 1,000 m and about 800 m depth for temperature and salinity, respectively. It should be noted again that data from both Argo and BOEM floats are independent in the DAsat run. Additionally assimilating the Argo profiles in DAargo run yields marginal further improvements in RMSEs of about 3% for temperature and 5% for salinity (Figure 5 and Table 1).

3.2 Assimilation impacts on biological properties

Assimilating satellite observations in the DAsat run reduced RMSEs of surface chlorophyll almost everywhere with only 3% of the model domain experiencing degradation (Figure 6a, c). Although large reductions in RMSE were achieved in the coastal regions, e.g., in the northern Gulf of Mexico, on Campeche Bank, and in Campeche Bay, the simulated chlorophyll concentrations remained much lower than the satellite estimates because of high observational uncertainties and a large background misfit in the Free run. This was expected because the biological model was optimized for the open Gulf (Wang et al., 2020). In the open Gulf, encompassed by the 1,000-m isobath, the overall RMSE of surface chlorophyll was reduced by 19% from 0.13 mg m^{-3} in the Free run to 0.11 mg m^{-3} in the DAsat run (Table 2). Assimilating Argo T-S profiles in the DAargo run led to lower reductions in the overall RMSEs of surface chlorophyll (Table 2) and even more deteriorations (Figure 6b-c).

In order to evaluate the impacts of data assimilation on subsurface biological properties, the temporal evolution of nitrate superimposed with the isoline of nitrate being 1 mmol N m^{-3} and SSH is shown in Figure 7 for the BEOM floats. Because of its high correlation with temperature, the nitrate distribution was modulated in the two assimilative runs along with the improvement in density fields. For instance, the two assimilative runs reproduce the newly detached Loop Current eddy observed by Float 287, and hence capture the depressed thermoclines and nitraclines that are not present in the Free run. As a result, the unbiased RMSE of nitrate following this float is reduced by 40% in the DAsat run and 38% in the DAargo run. In general, data assimilation improved the overall agreement of subsurface nitrate by 28% and 30% in the DAsat and DAargo runs relative to the Free run (Table 2).

https://doi.org/10.5194/os-2021-35
Preprint. Discussion started: 3 May 2021

The impacts of assimilation on subsurface chlorophyll are more complicated to assess because of the high nonlinearity of the model with regard to chlorophyll. Although the mean vertical profiles of chlorophyll are well reproduced in all three experiments (Figure S2), all failed to resolve the high spatiotemporal variability in subsurface chlorophyll which is at least partly due to the presence of mesoscale eddies (Figure 8). As a result, assimilation improved subsurface chlorophyll RMSEs marginally even in the Loop Current eddy of Float 287 where the most noticeable improvements of temperature (~60%) and nitrate (~40%) RMSEs were obtained. Results for phytoplankton and POC are similar as for chlorophyll although the reductions in their RMSEs are larger because assimilating the satellite data improved their mean concentrations in the upper layer (Figure S2, Table 2).

The model's inability to reproduce the spatiotemporal variability of subsurface chlorophyll is also reflected by the positions of deep chlorophyll maximum (DCM, denoted by red lines in Figure 8). As a ubiquitous phenomenon in the oligotrophic regions, a distinct DCM is observed throughout the whole year in the open Gulf of Mexico and its depth is inversely correlated with SSH (correlation coefficient = -0.6). Although the mean position and magnitude of the DCM are well reproduced by the model with and without data assimilation (Figure S2), the simulated DCM depth is much more stable and less sensitive to SSH variations. As a result, the reduction in the RMSE of DCM depth is limited to 18% in DAsat run but is significant (Table 2).

3.3 Sensitivity of subsurface chlorophyll to the light attenuation parameterization

We surmised that the model's inability to simulate interactions between mesoscale eddies and subsurface chlorophyll was due to inaccuracies in the light attenuation module where the attenuation coefficient, *Att*, was more strongly determined by the depth and less sensitive to the chlorophyll concentration ($Att = 0.04+0.025\times chl$). Therefore, we performed a sensitivity test with an alternative light attenuation parameterization ($Att = 0.027+0.075\times chl^{1.2}$) that more emphasizes the self-shading effect of chlorophyll on attenuation.

Both with and without data assimilation, the alternative parameterization led to higher correlations between simulated SSH and DCM depth with correlation coefficient of -0.60 in Free-alt run and -0.67 in DAsat-alt run (here -alt is added to indicate the modified light attenuation parameterization). As a result, the alternative parameterization produces slightly lower RMSEs of DCM depth (Table 2) and yields larger improvements in chlorophyll and phytoplankton within the Loop Current eddy of Float 287. As shown in Figure 9c, when using the original parameterization, assimilating the satellite data depresses the DCM depth from 70 m in the Free run to 90 m in the DAsat run, yet with a considerable bias of 20 m when compared to the observations. However, the chlorophyll is underestimated in the DAsat run and as a result its RMSEs are barely improved. In contrast, in the DAsat-alt run the DCM depth is corrected to 120 m, in agreement with the observations, and represents the vertical chlorophyll distribution more accurately although the nitrate profile is almost the same as in DAsat run. This was because the alternative parameterization accounted for the elevated intensity of photosynthetically active radiation (PAR) as a response to reduced chlorophyll concentrations in the upper layer (Figure 9b), which in turn facilitated the synthesis of chlorophyll and hence corrected their concentrations toward the observations.

4. Discussion

We implemented a coupled data assimilation scheme for jointly assimilating physical and biological observations in a biogeochemical model and evaluated to what degree satellite observations can inform subsurface distributions, especially of biological properties. Although biological data assimilation has received much attention in recent years, observations that are assimilated and used in skill assessment are typically limited to the surface ocean. The increasing availability of BGC-Argo data now makes it possible to validate and improve model performance below the surface (Cossarini et al., 2019; Terzić et al., 2019; Wang et al., 2020) but so far they are too sparse for sequential assimilation in three dimensions; hence, relevant applications are limited to idealized twin experiments (Ford, 2021; Yu et al., 2018) and a few specific regions with high float densities, e.g. the Mediterranean Sea (Cossarini et al., 2019). In addition, since a biogeochemical model is coupled to a physical model, assimilating physical observations theoretically should confer improvements in the biological model through correcting the circulation (Fiechter et al., 2011; Raghukumar et al., 2015; Song et al., 2016b, 2016a) and potentially by providing additional constraints via multivariate updates to biological variables (Goodliff et al., 2019; Yu et al., 2018). This is particularly important when the physical model is biased (Yu et al., 2018).

Our study shows that assimilating satellite data (DAsat run) can constrain the main circulation features in the Gulf of Mexico, i.e. the Loop Current and its associated mesoscale eddies. Temperature and salinity are also improved down to ~1,000 m depth because of the correction of mesoscale eddies. When calculating the reductions in RMSE for SSH and each single profile of temperature and salinity, we find that the improvement in SSH is highly correlated with those in temperature (correlation coefficient = 0.96) and salinity (correlation coefficient = 0.92, Figure S4). Assimilating the satellite data also improves subsurface nitrate because it is tightly correlated with the density structure expressed by SSH and temperature profiles. However, improvements in temperature and nitrate do not necessarily yield better simulations of chlorophyll or phytoplankton because they tend to be light-limited below the surface. In our biogeochemical model, the light intensity is attenuated by water and chlorophyll, and is not directly updated by the data assimilation scheme but only adjusted indirectly through changes in chlorophyll during forecast steps. This, in turn, impacts the synthesis of chlorophyll and growth of phytoplankton. However, in the original parameterization, the light attenuation is mainly controlled by the water and much less sensitive to chlorophyll concentrations than it appears to be in reality. A sensitivity test showed that by applying an alternative light parameterization with more pronounced self-shading effects of chlorophyll, the subsurface chlorophyll and phytoplankton distributions are further improved after assimilating the satellite data.

The improvement on biological variables as a result of a model's dynamical response to data assimilation has already been demonstrated in previous studies but they focused on the influence of correcting nutrient fields on surface chlorophyll. For instance, Fiechter et al. (2011) improved simulated iron concentrations inside eddies and therefore surface chlorophyll by correcting the timing and position of mesoscale eddies in the Gulf of Alaska. Ourmières et al. (2009) also reported that when the prior nitrate distributions were correct, assimilating the physical observations can improve surface chlorophyll because of a more accurate simulation of mixed layer depth and nutrient inputs into the euphotic zone. However, the efficiency of this mechanism depends on the accuracy of biological models and can be further improved by addressing the model representations of key biological processes. For example, the usage of suboptimal biological parameters can yield a substantial degradation of data assimilation efficiency, especially with respect to unobserved variables (Song et al., 2016a). Although BGC-Argo profiles so far are insufficient for sequential assimilation, they can provide substantial benefits to the biogeochemical prediction by enabling *a priori* model tuning, e.g. of biological parameter values (Wang et al., 2020) and the key parameterization schemes (Terzić et al., 2019).

Assimilating Argo T-S profiles in the DAargo run yields slightly further improvements with respect to independent profiles of temperature and salinity, similar to the twin experiments in Yu et al. (2019). Such additional benefits in physical properties are also translated into the simulation of subsurface nitrate but not into other biological fields, i.e. chlorophyll, phytoplankton, and POC. Moreover, assimilating the Argo T-S profiles can even degrade surface chlorophyll. This issue has been also reported in a recent study (Goodliff et al., 2019) which assimilated sea surface temperature to update both physical and biological variables and the issue was alleviated by muting the multivariate update between the sea surface temperature and chlorophyll.

In addition to the model dynamical response, the biological fields can be directly updated by physical and biological observations through multivariate covariances. In order to distinguish their influence, we show the increments obtained from assimilating each observation type in the DAsat run (Figure 10). As shown in Figure 10a, b, assimilating physical observations has a much stronger impact than biological observations on nitrate and therefore we conclude that the improvement of nitrate in this study is mainly obtained from assimilating physical observations. This is consistent with previous studies (Ciavatta et al., 2018; Skákala et al., 2018; Teruzzi et al., 2018) where assimilating surface chlorophyll had little impact on nitrate and even degraded it by both variational and sequential data assimilation. In variational data assimilation, it is hard to define the background errors accurately (Mattern et al., 2017; Teruzzi et al., 2018) and the biological model can fit itself to observed chlorophyll by many different pathways, e.g. direct changes of biomass or an indirect way through nitrate. However, observations are often insufficient to provide this information (Mattern et al., 2017). In sequential data assimilation, the multivariate covariance between surface chlorophyll and subsurface nitrate can be considered but typically this covariance is not linear or constant. For instance, Fontana et al. (2013) assimilated satellite surface chlorophyll into a biological model in the North Atlantic and found that subsurface nitrate was barely influenced because it was weakly correlated with surface chlorophyll, leading the authors to suggested that it is impossible to fully constrain a 3D biogeochemical model by only assimilating the surface chlorophyll. This issue remains when assimilating the surface chlorophyll to update other biological variables (Yu et al., 2018), e.g. phytoplankton functional groups (Ciavatta et al., 2018).

In contrast to the nitrate, assimilating satellite data of physical and biological observations have a comparable influence on subsurface chlorophyll (Figure 10c-f). Specifically, they can change subsurface chlorophyll concentrations even below 100 m depth and vertical structures of chlorophyll by adjusting the DCM depth, e.g., there are 10% and 5% of profiles with changes in DCM depth exceeding ±20 m due to the update of physical and biological observations, respectively. Because currently BGC-Argo profiles are sparse, i.e. only 14 profiles are available at all update steps, it is hard to draw definitive conclusions about these impacts on chlorophyll and DCM depth.

In general, coupled data assimilation of both physical and biological satellite observations is able to improve subsurface biological properties because it benefits from the high correlations of some biological distributions, especially nutrients, with the vertical density structure and because of the dynamical responses to improvements in circulation in the forecast step. However, this is preconditioned on the coupled model being well calibrated *a priori*. Therefore, this study provides an intermediate step toward 3D updates of biological properties before the BGC-Argo profiles will ultimately become more abundant.

5. Conclusion

In this study, a coupled data assimilation scheme for both physical and biological satellite observations was implemented to investigate whether these observations can inform subsurface distributions. In addition, Argo T-S profiles were assimilated to assess their impact beyond satellite observations. The multivariate update was applied by using the covariance structure between physical and biological variables. The Gulf of Mexico was selected as the study region because the dominant physical features, the Loop Current and its associated mesoscale eddies, are stochastic and can influence the biological properties in three dimensions substantially. Our results show that assimilating satellite data leads to significant improvements in the simulation of SSH and SST, and also projects these improvements from the surface to about 1,000 m depth for temperature and salinity as shown by an assessment of the independent BGC-Argo profiles. With respect to biological fields, the subsurface nitrate distribution benefits greatly from the tight correlation with density and the improved fidelity of mesoscale features. However, initially there were only slight improvements in other biological variables below the surface, i.e. chlorophyll, phytoplankton, and POC, because a suboptimal light parameterization did not react to the changed chlorophyll concentrations appropriately and failed to provide accurate feedbacks on the synthesis of chlorophyll and growth of phytoplankton. We tested an alternative light parameterization with a larger relative contribution from chlorophyll to light attenuation. As a result, the subsurface chlorophyll and phytoplankton were further improved. This highlights the importance of *a priori* tuning to achieve better assimilation performance. Finally, assimilating the Argo T-S profiles on top of satellite observations yields slight further improvements with respect to independent vertical profiles of temperature and salinity, which also translated into improvements in subsurface nitrate.

Code and data availability: The ROMS model code can be accessed at http://www.myroms.com (last access: 16 June 2016). HYCOM data can be downloaded at http://tds.hycom.org/thredds/dodsC/datasets (last access: 16 August 2018). Profiling data from the BGC-Argo floats are available at the National Oceanographic Data Center (NOAA), https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:159562 (Hamilton and Leidos, 2017)

Author contributions. BW and KF conceived the study. BW carried out data assimilation experiments and analyses. LY provided data assimilation techniques. BW and KF discussed the results and wrote the paper with contributions from all coauthors.

Competing interests. The authors declare that they have no conflict of interest.

Financial support. This research was funded by the Gulf of Mexico Research Initiative (GoMRI-V-487).

Reference

Anderson, J. L. and Anderson, S. L.: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts, Monthly Weather Review, 127(12), 2741–2758, 1999.

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(10), 1787–1827, doi:https://doi.org/10.1016/S0967-0637(00)00019-4, 2000.

Biogeochemical-Argo Planning Group: The scientific rationale, design and implementation plan for a Biogeochemical-Argo float array, edited by Ken Johnson and Hervé Claustre., 2016.

Chai, F., Johnson, K. S., Claustre, H., Xing, X., Wang, Y., Boss, E., Riser, S., Fennel, K., Schofield, O. and Sutton, A.: Monitoring ocean biogeochemistry with autonomous platforms, Nature Reviews Earth & Environment, 1(6), 315–326, doi:10.1038/s43017-020-0053-y, 2020.

Ciavatta, S., Torres, R., Martinez-vicente, V., Smyth, T., Olmo, G. D., Polimene, L. and Allen, J. I.: Progress in Oceanography Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling, Progress in Oceanography, 127, 74–95, doi:10.1016/j.pocean.2014.06.002, 2014.

Ciavatta, S., Brewin, R. J. W., Skakala, J., Polimene, L., de Mora, L., Artioli, Y. and Allen, J. I.: Assimilation of Ocean-Color Plankton Functional Types to Improve Marine Ecosystem Simulations, Journal of Geophysical Research : Oceans, 123, 834–854, doi:10.1002/2017JC013490, 2018.

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, doi:https://doi.org/10.1016/j.ocemod.2018.11.005, 2019.

Doney, S. C.: Major challenges confronting marine biogeochemical modeling, Global Biogeochemical Cycles, 13(3), 705–714, doi:10.1029/1999GB900039, 1999.

Doney, S. C., Lindsay, K., Caldeira, K., Campin, J.-M., Drange, H., Dutay, J.-C., Follows, M., Gao, Y., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Madec, G., Maier-Reimer, E., Marshall, J. C., Matear, R. J., Monfray, P., Mouchet, A., Najjar, R., Orr, J. C., Plattner, G.-K., Sarmiento, J., Schlitzer, R., Slater, R., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y. and Yool, A.: Evaluating global ocean carbon models: The importance of realistic physics, Global Biogeochemical Cycles, 18(3), doi:10.1029/2003GB002150, 2004.

Edwards, C. A., Moore, A. M., Hoteit, I. and Cornuelle, B. D.: Regional Ocean Data Assimilation, Annual Review of Marine Science, 7(1), 21–42, doi:10.1146/annurev-marine-010814-015821, 2015.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynamics, 53(4), 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Fennel, K., Wilkin, J., Levin, J., Moisan, J., Reilly, J. O. and Haidvogel, D.: Nitrogen cycling in the Middle Atlantic Bight : Results from a three-dimensional model and implications for the North Atlantic nitrogen budget, GLOBAL BIOGEOCHEMICAL CYCLES, 20, 1–14, doi:10.1029/2005GB002456, 2006.

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., the GODAE OceanView Marine Ecosystem Analysis, , and Team, P. T.: Advancing Marine Biogeochemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing Ecosystem Health, Frontiers in Marine Science, 6, 89, doi:10.3389/fmars.2019.00089, 2019.

Fiechter, J., Broquet, G., Moore, A. M. and Arango, H. G.: A data assimilative, coupled physical–biological model for the Coastal Gulf of Alaska, Dynamics of Atmospheres and Oceans, 52(1), 95–118, doi:https://doi.org/10.1016/j.dynatmoce.2011.01.002, 2011.

Fommervault, O. P. De, Perez-brunius, P., Damien, P., Camacho-ibar, V. F. and Sheinbaum, J.: Temporal variability of chlorophyll distribution in the Gulf of Mexico: bio-optical data from profiling floats, Biogeosciences, 14, 5647–5662, doi:10.5194/bg-14-5647-2017, 2017.

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 Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998 – 2006 based on, Ocean Science, 9, 37–56, doi:10.5194/os-9-37-2013, 2013.

Ford, D.: Assimilating synthetic Biogeochemical-Argo and ocean colour observations into a global ocean model to inform observing system design, Biogeosciences, 18(2), 509–534, doi:10.5194/bg-18-509-2021, 2021.

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

Geider, R. J., MacIntyre, H. L. and Kana, T. M.: Dynamic model of phytoplankton growth and acclimation : responses of the balanced growth rate and the chlorophyll a : carbon ratio to light , nutrient-limitation and temperature, Marine ecology progress series, 148, 187–200, 1997.

Goodliff, M., Bruening, T., Schwichtenberg, F., Li, X., Lindenthal, A., Lorkowski, I. and Nerger, L.: Temperature assimilation into a coastal ocean-biogeochemical model: assessment of weakly and strongly coupled data assimilation, Ocean Dynamics, 69(10), 1217–1237, doi:10.1007/s10236-019-01299-7, 2019.

Greatbatch, R. J., Sheng, J., Eden, C., Tang, L., Zhai, X. and Zhao, J.: The semi-prognostic method, Continental Shelf Research, 24(18), 2149–2165, doi:10.1016/j.csr.2004.07.009, 2004.

Gregg, W. W.: Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, Journal of Marine Systems, 69, 205–225, doi:10.1016/j.jmarsys.2006.02.015, 2008.

Haidvogel, D. B., Arango, H., Budgell, W. P., Cornuelle, B. D., Curchitser, E., Lorenzo, E. Di, Fennel, K., Geyer, W. R., Hermann, A. J., Lanerolle, L., Levin, J., McWilliams, J. C., Miller, A. J., Moore, A. M., Powell, T. M., Shchepetkin, A. F., Sherwood, C. R., Signell, R. P., Warner, J. C. and Wilkin, J.: Ocean forecasting in terrain-following coordinates : Formulation and skill assessment of the Regional Ocean Modeling System, Journal of computational physics, 227, 3595–3624, doi:10.1016/j.jcp.2007.06.016, 2008.

Haines, K., Johannessen, J., Knudsen, P., Lea, D., Rio, M.-H., Bertino, L., Davidson, F. and Hernandez, F.: An ocean modelling and assimilation guide to using GOCE geoid products, Ocean Science, 7, 151–164, doi:10.5194/os-7-151-2011, 2011.

Hamilton, P. and Leidos: Ocean currents, temperatures, and others measured by drifters and profiling floats for the Lagrangian Approach to Study the Gulf of Mexico Deep Circulation project 2011-07 to 2015-06 (NCEI Accession 0159562), Version 1.1, NOAA National Centers for Environmental

Information. Dataset, available at https://accession.nodc.noaa.gov/0159562, last access: 25 October 2017.

Hu, J., Fennel, K., Mattern, J. P. and Wilkin, J.: Data assimilation with a local Ensemble Kalman Filter applied to a three-dimensional biological model of the Middle Atlantic Bight, Journal of Marine Systems, 94, 145–156, doi:10.1016/j.jmarsys.2011.11.016, 2012.

Li, G., Iskandarani, M., Hénaff, M. Le, Winokur, J., Le Maître, O. P. and Knio, O. M.: Quantifying initial and wind forcing uncertainties in the Gulf of Mexico, Computational Geosciences, 20(5), 1133–1153, doi:10.1007/s10596-016-9581-4, 2016.

Mattern, J. P., Dowd, M. and Fennel, K.: Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations, Journal of Geophysical Research: Oceans, 118(May), 2746–2760, doi:10.1002/jgrc.20213, 2013.

Mattern, J. P., Song, H., Edwards, C. A., Moore, A. M. and Fiechter, J.: Data assimilation of physical and chlorophyll a observations in the California Current System using two biogeochemical models, Ocean Modelling, 109, 55–71, doi:https://doi.org/10.1016/j.ocemod.2016.12.002, 2017.

Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, Reviews of Geophysics and Space Physics, 20(4), 851–875, doi:10.1029/RG020i004p00851, 1982.

Oschlies, A. and Garcon, V.: An eddy-permitting coupled physical-biological model of the North Atlantic: 1. Sensitivity to advection numerics and mixed layer physics, Global Biogeochemical Cycles - GLOBAL BIOGEOCHEM CYCLE, 13, 135–160, doi:10.1029/98GB02811, 1999.

Ourmières, Y., Brasseur, P., Lévy, M., Brankart, J.-M. and Verron, J.: On the key role of nutrient data to constrain a coupled physical–biogeochemical assimilative model of the North Atlantic Ocean, Journal of Marine Systems, 75(1), 100–115, doi:https://doi.org/10.1016/j.jmarsys.2008.08.003, 2009.

Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A. and Nerger, L.: Assimilation of Global Total Chlorophyll OC-CCI Data and Its Impact on Individual Phytoplankton Fields, Journal of Geophysical Research: Oceans, 124(1), 470–490, doi:10.1029/2018JC014329, 2019.

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, doi:https://doi.org/10.1016/j.pocean.2015.01.004, 2015.

Rio, M.-H., Mulet, S. and Picot, N.: New global mean dynamic topography from a goce geoid model, altimeter measurements and oceanographic in-situ data, ESA Living Planet Symposium, Proceedings of the conference held on 9-13 September 2013 at Edinburgh in United Kingdom. ESA SP-722. 2-13, p.27, 2013(September), 2–13, 2013.

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. Ó., D'Ortenzio, F., Dall'Olmo, G., Desbruyeres, D., Fennel, K., Fer, I., Ferrari, R., Forget, G., Freeland, H., Fujiki, T., Gehlen, M., Greenan, B., Hallberg, R., Hibiya, T., Hosoda, S., Jayne, S., Jochum, M., Johnson, G. C., Kang, K., Kolodziejczyk, N., Körtzinger, A., Traon, P.-Y. Le, 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 [online] Available from: https://www.frontiersin.org/article/10.3389/fmars.2019.00439, 2019.

Sakov, P. and Oke, P. R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters, Tellus A: Dynamic Meteorology and Oceanography, 60(2), 361–371, doi:10.1111/j.1600-0870.2007.00299.x, 2008.

Sakov, P., Evensen, G. and Bertino, L.: Asynchronous data assimilation with the EnKF, Tellus A: Dynamic Meteorology and Oceanography, 62(1), 24–29, doi:10.1111/j.1600-0870.2009.00417.x, 2010.

Sheng, J., Greatbatch, R. J. and Wright, D. G.: Improving the utility of ocean circulation models through adjustment of the momentum balance, Journal of Geophysical Research: Oceans, 106(C8), 16711–16728, doi:10.1029/2000JC000680, 2001.

Shulman, I., Frolov, S., Anderson, S., Penta, B., Gould, R., Sakalaukus, P. and Ladner, S.: Impact of bio-optical data assimilation on short-term coupled physical , bio-optical model predictions, Journal of Geophysical Research: Oceans, 118, 2215–2230, doi:10.1002/jgrc.20177, 2013.

Skákala, J., Ford, D., Brewin, R. J. W., 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(8), 5230–5247, doi:10.1029/2018JC014153, 2018.

Smagorinsky, J.: General circulation experiments with the primitive equations: I. the basic experiment, Monthly Weather Review, 91(3), 99–164, 1963.

Smolarkiewicz, P. K. and Margolin, L. G.: MPDATA : A Finite-Difference Solver for Geophysical Flows, Journal of computational physics, 140, 459–480, 1998.

Song, H., Edwards, C. A., Moore, A. M. and Fiechter, J.: Data assimilation in a coupled physical-biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 2—Joint physical and biological data assimilation twin experiments, Ocean Modelling, 106, 146–158, doi:https://doi.org/10.1016/j.ocemod.2016.09.003, 2016a.

Song, H., Edwards, C. A., Moore, A. M. and Fiechter, J.: Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: Part 3—Assimilation in a realistic context using satellite and in situ observations, Ocean Modelling, 106, 159–172, doi:https://doi.org/10.1016/j.ocemod.2016.06.005, 2016b.

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(1), 200–217, doi:10.1002/2013JC009277, 2014.

Teruzzi, A., Bolzon, G., Salon, S., Lazzari, P. and Solidoro, C.: Assimilation of coastal and open sea biogeochemical data to improve phytoplankton simulation in the Mediterranean Sea, Ocean Modelling, 132(September), 46–60, doi:10.1016/j.ocemod.2018.09.007, 2018.

Terzić, E., Lazzari, P., Organelli, E., Solidoro, C., Salon, S., D'Ortenzio, F. and Conan, P.: Merging bio-optical data from Biogeochemical-Argo floats and models in marine biogeochemistry, Biogeosciences, 16(12), 2527–2542, doi:10.5194/bg-16-2527-2019, 2019.

Thacker, W. C., Srinivasan, A., Iskandarani, M., Knio, O. M. and Hénaff, M. Le: Propagating boundary uncertainties using polynomial expansions, Ocean Modelling, 43–44, 52–63, doi:https://doi.org/10.1016/j.ocemod.2011.11.011, 2012.

Wang, B., Fennel, K., Yu, L. and Gordon, C.: Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico, Biogeosciences, 17(15), 4059–4074, doi:10.5194/bg-17-4059-2020, 2020.

Xiao, Y. and Friedrichs, M. A. M.: The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model, Journal of Geophysical Research: Oceans, 119, 2691–2712, doi:10.1002/2013JC009433.Received, 2014a.

Xiao, Y. and Friedrichs, M. A. M.: Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models in the Mid-Atlantic Bight : effects of increasing the complexity of the planktonic food web, Biogeosciences, 11, 3015–3030, doi:10.5194/bg-11-3015-2014, 2014b.

Xu, D., Zhu, J., Qi, Y., Li, X. and Yan, Y.: The impact of mean dynamic topography on a sea-level anomaly assimilation in the South China Sea based on an eddy-resolving model, Acta Oceanologica Sinica, 31(5), 11–25, doi:10.1007/s13131-012-0232-x, 2012.

Xue, Z., He, R., Fennel, K., Cai, W., Lohrenz, S. and Hopkinson, C.: Modeling ocean circulation and biogeochemical variability in the Gulf of Mexico, Biogeosciences, 10, 7219–7234, doi:10.5194/bg-10-7219-2013, 2013.

Yu, L., Fennel, K., Bertino, L., Gharamti, M. El 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, doi:https://doi.org/10.1016/j.ocemod.2018.04.005, 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(6), 1801–1814, doi:10.5194/os-15-1801-2019, 2019.

Table list

Table 1. The root-mean-square-error (RMSE) for SSH, SST, as well as vertical profiles of temperature and salinity from Argo and BOEM floats. Percentages in the parentheses represent the relative reductions of RMSE values.

	SSH	SST	Argo		Boem	
	(m)	(°C)	Temp (°C)	salt	Temp (°C)	salt
Free	0.17	0.88	1.70	0.22	1.55	0.18
DAsat	0.08 (54%)	0.55 (37%)	0.89 (48%)	0.14 (36%)	0.83 (46%)	0.11 (39%)
DAargo	0.08 (54%)	0.56 (36%)	0.86 (49%)	0.13 (41%)	0.79 (49%)	0.10 (44%)

Table 2. The root-mean-square-error (RMSE) for surface chlorophyll in the open gulf, vertical profiles of NO_3 based on its climatological relationship with temperature, chlorophyll, phytoplankton, and POC from BOEM floats, as well as the depth of deep chlorophyll maximum. Percentages in the parentheses represent the relative reductions in RMSE values. Only a reduction in RMSE larger than or equal to 10% is considered as a significant improvement (highlighted in **bold**).

	SChl (mg m^{-3})	NO3 (mmol N m^{-3})	Chlorophyll (mmol N m^{-3})	Phytoplankton (mmol N m^{-3})	POC (mmol N m^{-3})	DCM depth (m)
Free	0.13	3.71	0.18	0.11	18.62	25.48
DAsat	**0.11 (19%)**	**2.66 (28%)**	0.17 (6%)	0.10 (9%)	**16.46 (12%)**	**21.08 (18%)**
DAargo	0.12 (9%)	**2.58 (30%)**	0.17 (6%)	0.10 (9%)	**16.77 (10%)**	**22.39 (12%)**
Free_alt	0.17	3.71	0.18	0.11	17.55	24.35
DAsat_alt	**0.13 (26%)**	**2.63 (29%)**	0.17 (6%)	0.10 (9%)	**15.53 (12%)**	**20.42 (16%)**

Figure captions

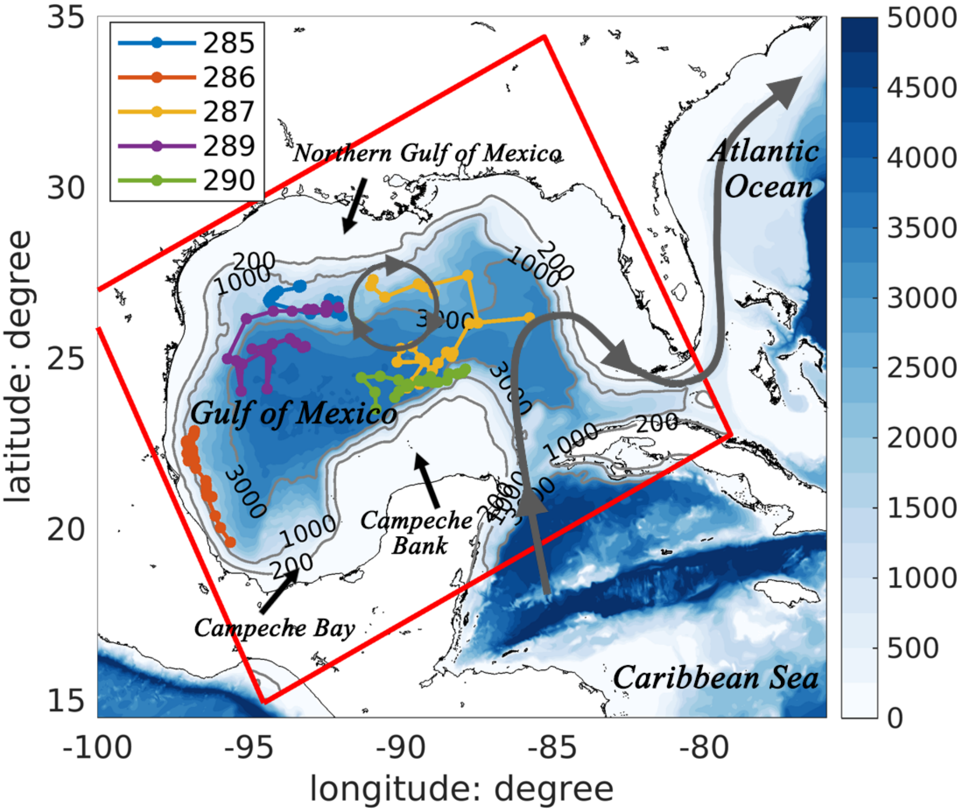


Figure 1. Bathymetric map of the Gulf of Mexico with a schematic pattern of Loop Current and Loop Current Eddies. Trajectories of five BGC-Argo floats (colored lines) in 2015 were also shown in the figure.

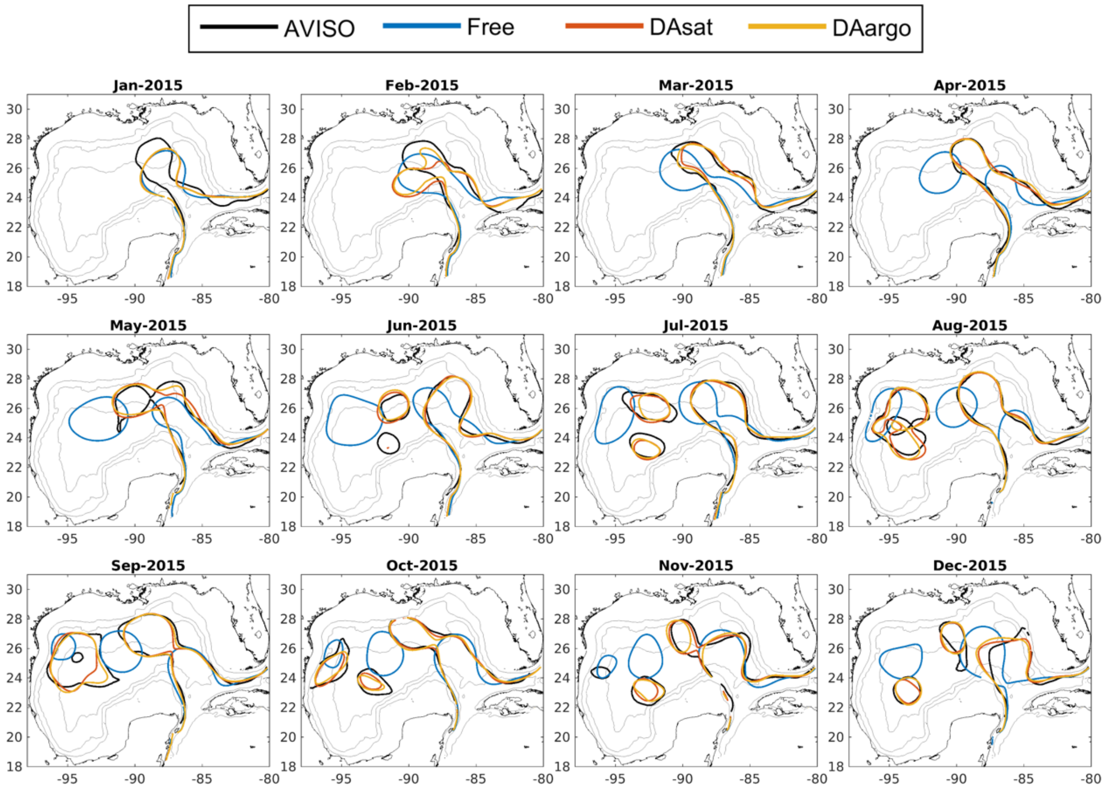


Figure 2. Monthly averaged Loop Current and Loop Current eddies based on 10-cm SSH contour from satellite data (black), free run (blue), DAsat run (orange), and DAargo run (yellow). The gray contours represent the isobath of 200 m, 1000 m, and 3000 m.

https://doi.org/10.5194/os-2021-35
Preprint. Discussion started: 3 May 2021

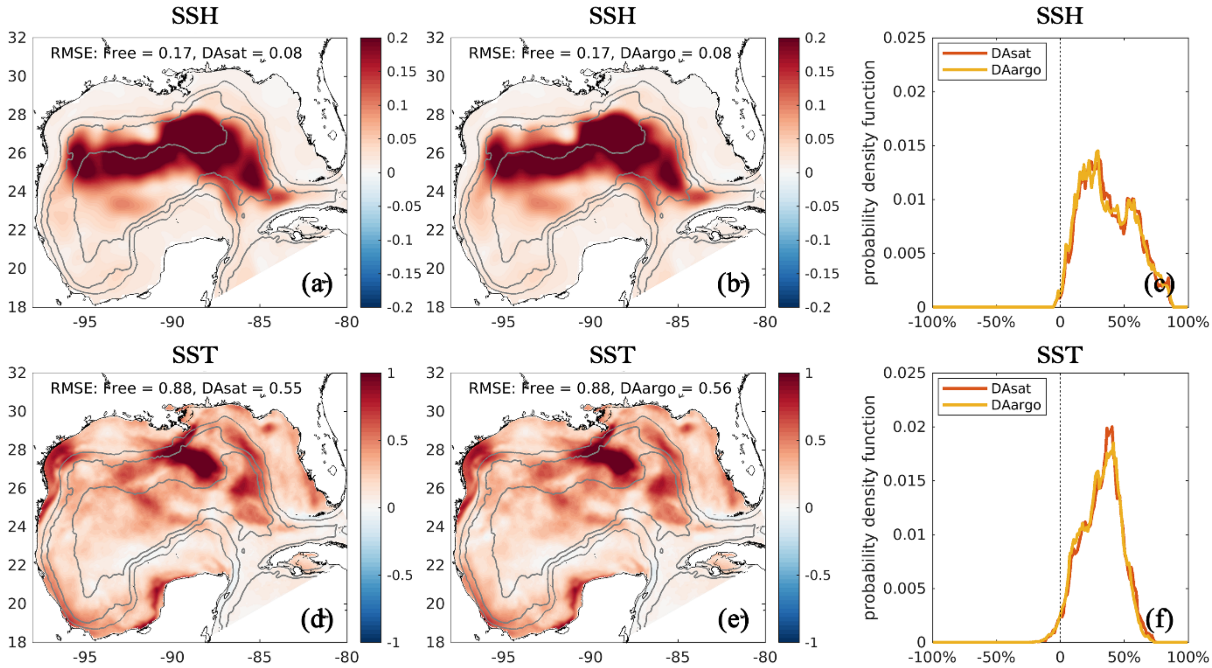


Figure 3. Spatial map of differences (a-e) and histogram of relative differences (c, f) in root-mean-square-error (RMSE) between the free run and the two data-assimilative runs for SSH and SST. Positive values represent improvements while negative values represent deteriorations by data assimilation. Gray contours represent the 300-, 1000-, and 3000-m isobaths.

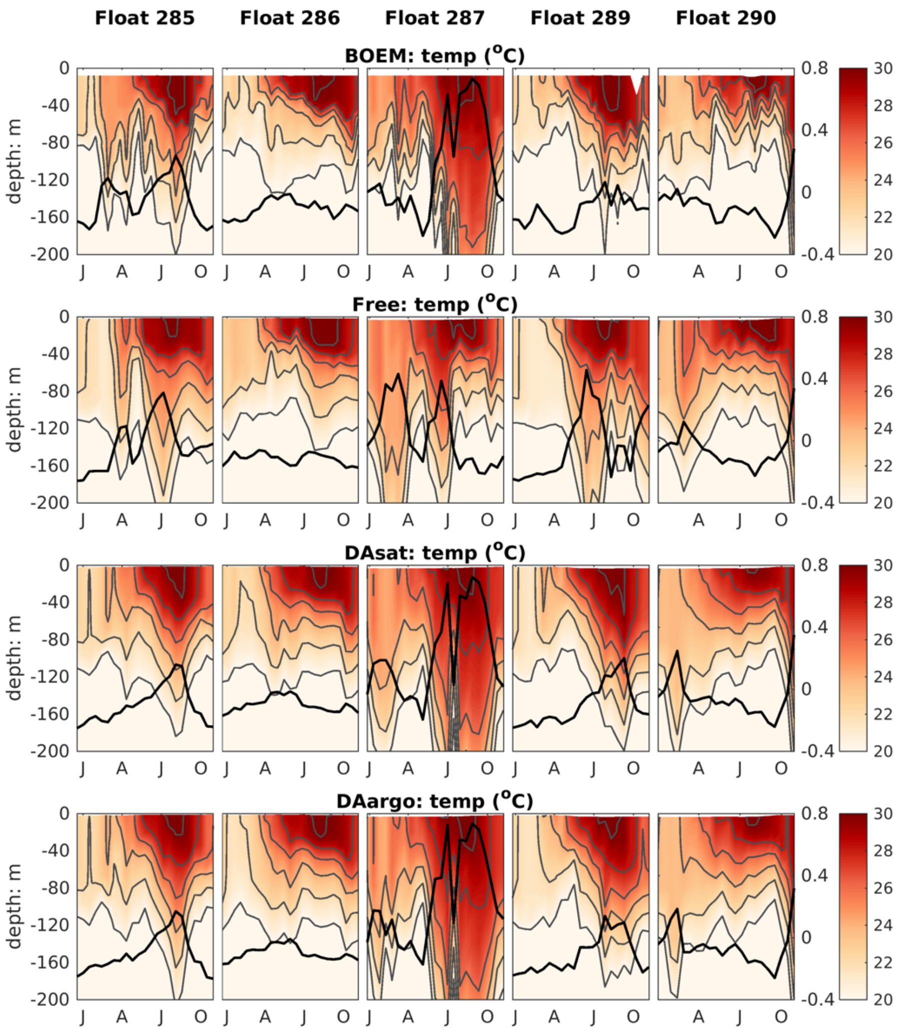


Figure 4. Vertical distributions of temperature from BOEM floats, the Free run, the DAsat run, and the DAargo run. Gray lines represent isothermal lines with an interval of 2°C. Thick black lines represent SSH which is obtained from the matching record of altimeter observations.

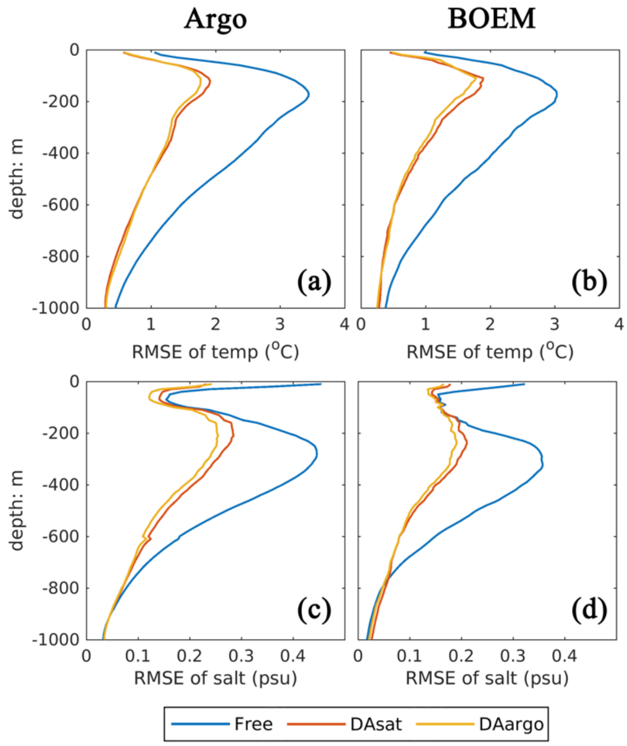


Figure 5. Vertical profiles of root-mean-square-error (RMSE) for temperature and salinity with respect to Argo and BOEM floats.

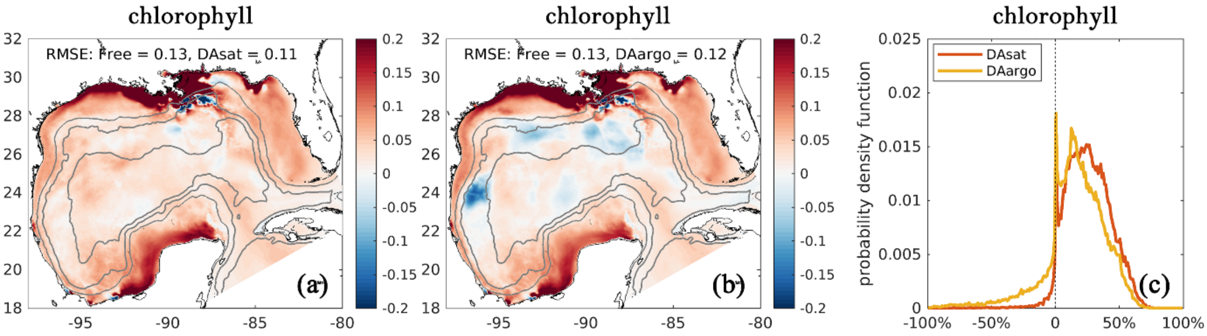
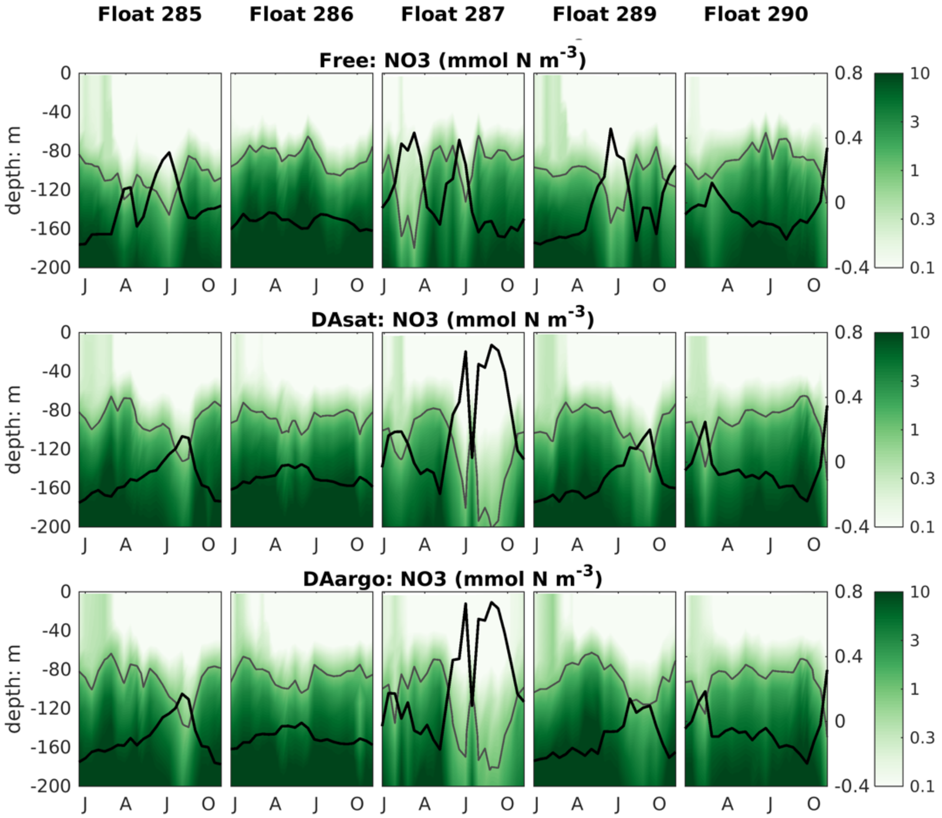


Figure 6. The same as Figure 3 except for surface chlorophyll.



80

81 Figure 7. Same as Figure 4 but for nitrate. Gray contours (thin lines) represent the isoline of nitrate being 82 1 mmol N m^{-3}. Thick black lines represent SSH which is obtained from the matching record of altimeter 83 observations.

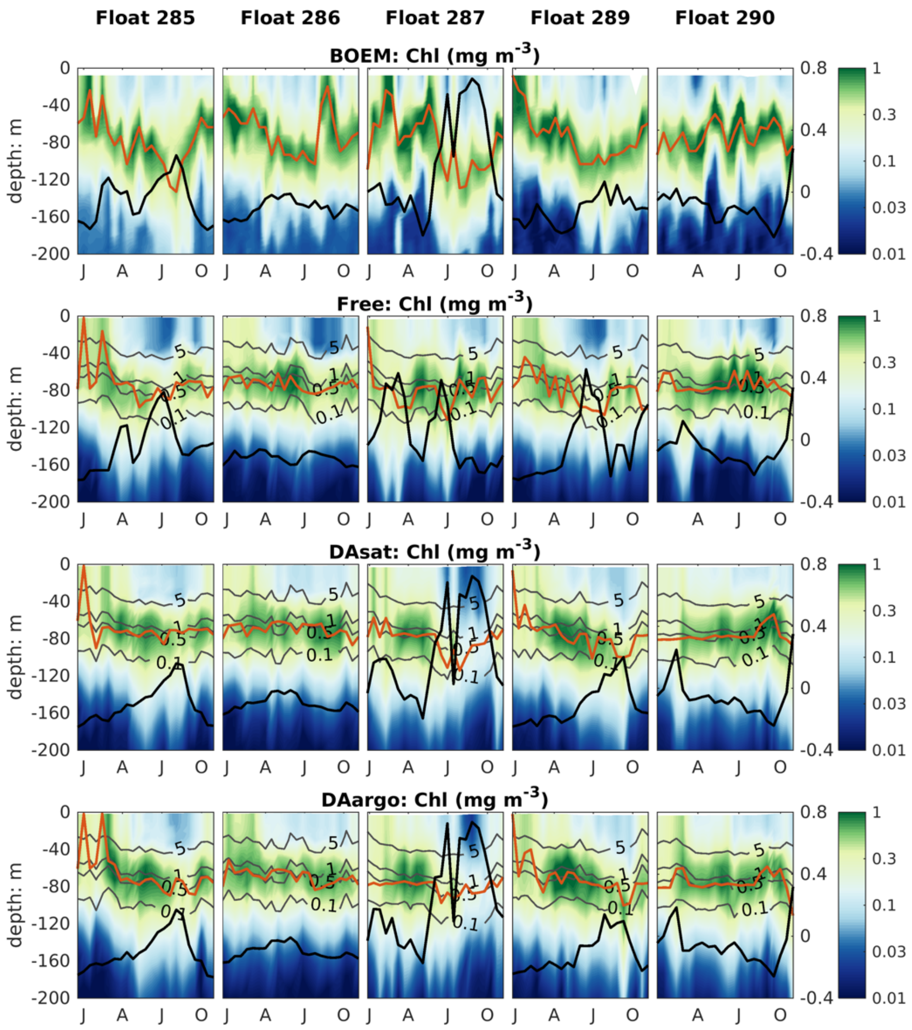


Figure 8. Same as Figure 4 but for chlorophyll. Gray contours represent the simulated isolumes and red lines represent the depth of deep chlorophyll maximum. Thick black lines represent SSH which is obtained from the matching record of altimeter observations.

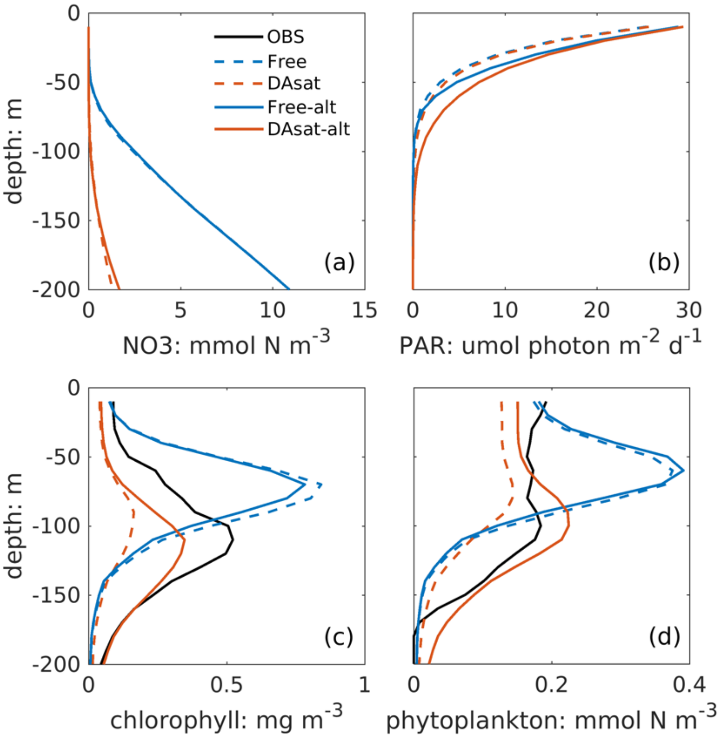


Figure 9. Mean vertical profiles of nitrate, light intensity (photosynthetically active radiation, PAR), chlorophyll, and phytoplankton within the center of the newly detached Loop Current eddy from the Free run, the DAsat run, the Free-alt run, and the DAsat-alt run.

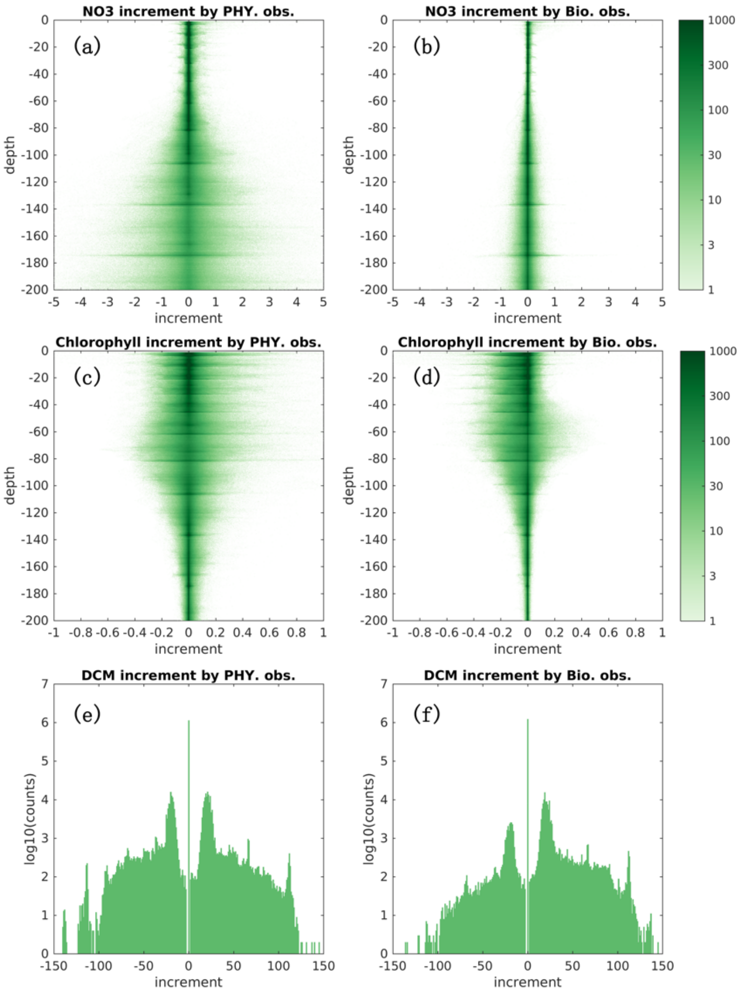


Figure 10 Histogram of increment in nitrate, chlorophyll, and DCM depth obtained by assimilating physical and biological observations.