https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

Earth System Dynamics
Discussions
Open Access

Ocean biogeochemical reconstructions to estimate historical ocean CO_2 uptake

Raffaele Bernardello[1], Valentina Sicardi[1], Vladimir Lapin[1], Pablo Ortega[1], Yohan Ruprich-Robert[1], Etienne Tourigny[1], and Eric Ferrer[1]

[1]Barcelona Supercomputing Center

Correspondence: Raffaele Bernardello (raffaele.bernardello@bsc.es)

Abstract. Given the role of the ocean in mitigating climate change through CO_2 absorption, it is important to improve our ability to quantify the historical ocean CO_2 uptake, including its natural variability, for carbon budgeting purposes. In this study we present an exhaustive intercomparison between two ocean modelling practices that can be used to reconstruct the historical ocean CO_2 uptake. By comparing the simulations to a wide array of ocean physical and biogeochemical observational datasets, we show how constraining the ocean physics towards observed temperature and salinity results in a better representation of global biogeochemistry. We identify the main driver of this improvement to be a more realistic representation of large scale meridional overturning circulation together with improvements in mixed layer depth and sea surface temperature. Nevertheless, surface chlorophyll was rather insensitive to these changes, and, in some regions, its representation worsened. We identified the causes of this response to be a combination of a lack of robust parameter optimization and limited changes in environmental conditions for phytoplankton. We conclude that although the direct validation of CO_2 fluxes is challenging, the pervasive improvement observed in most aspects of biogeochemistry when applying data assimilation of observed temperature and salinity is encouraging; therefore, data assimilation should be included in multi-method international efforts aimed at reconstructing the ocean CO_2 uptake.

1 Introduction

The ocean is responsible for absorbing approximately 25% of CO_2 emissions derived from human activities (Gruber et al., 2023). However, a growing body of evidence highlights the need to understand better the links between climate variability and ocean carbon cycle dynamics, pointing to the ocean carbon sink being more variable than previously assumed (DeVries et al., 2023; Gruber et al., 2019; McKinley et al., 2017). Understanding the mechanisms behind this variability can lead to better estimates of the ocean carbon sink. This becomes particularly important in the context of a future decline of global CO_2 emissions and the UN Framework Convention on Climate Change stocktaking activities. For this reason, the global carbon cycle scientific community has devoted significant efforts over the past few decades to refine our model-based estimates of past ocean carbon uptake. These estimates are hindered by the scarcity of year-round observations in vast global ocean regions and by natural variability in air-sea CO_2 fluxes. The natural variability is superimposed on a long-term trend driven by the increase of atmospheric CO_2 concentration. Moreover, since climate change is affecting the ocean's physical state, it is reasonable to

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

Earth System Dynamics Discussions

expect that this will, in turn, also affect the ocean's ability to absorb carbon. However, since the observational record spans only 3 decades, detecting trends in air-sea CO_2 fluxes that are driven by climate change is challenging. As an example, large variability in the Southern Ocean was in the past interpreted to possibly be an effect of climate change (Le Quéré et al., 2007; Lovenduski et al., 2007), while a decade later, these variations are being explained as a result of natural variability in regional atmospheric circulation (Landschützer et al., 2015; Keppler and Landschützer, 2019).

Coordinated international efforts, like the global carbon project (Friedlingstein et al., 2022) try to quantify and possibly predict the global carbon budget (GCB) by estimating the amount of CO_2 emitted each year and the fractions being absorbed by the ocean and land vegetation. Because of the scarcity of observations, these efforts rely heavily on modelling work. For the oceans, traditionally, the evolution of the air-sea CO_2 flux has been estimated using Ocean Biogeochemical General Circulation Models (OBGCMs). These are forced with atmospheric reanalysis (based on observations of physical atmospheric variables) for a given period, usually spanning around 60 years. In these simulations the ocean physics and biogeochemistry are left free to evolve in response to the atmospheric forcing and the prescribed atmospheric CO_2 concentrations.

In parallel, over the last two decades, climate models have been increasingly used to predict climatic conditions from a few months up to a decade ahead (Merryfield et al., 2020; Bilbao et al., 2021), with experiments commonly referred to as decadal climate predictions. This field of research lies in between weather forecasts and climate projections because it relies on both available observations to initialise the models to leverage the predictability from internal variability sources and future emissions, prescribed as boundary conditions, to faithfully capture the expected human-driven trends. Moreover, even more recently, decadal climate prediction has been extended to global biogeochemical properties, including the ocean carbon cycle (Séférian et al., 2018; Li et al., 2019; Lovenduski et al., 2019). In climate predictions, available observations are assimilated in both the atmosphere and the ocean to drive the model to an initial state consistent with the observed climate. This is done for the historical period up to present, to provide initial conditions also for predictions of the past, known as retrospective predictions, which are needed to verify the skill of the predictions, as well as to diagnose the forecast drift, which is needed to correct the future predictions. These climate simulations of the historical period in which available observations are assimilated are known as reconstructions.

When reconstructions are performed with Earth System Models (ESMs), the ocean biogeochemistry is also expected to evolve according to the observed variability. In this paper, we use the EC-Earth3 Earth System Model (Döscher et al., 2021) to explore whether and how the methodology used to perform reconstructions impacts the simulated representation of ocean biogeochemistry. In particular, we explore the differences between the standard GCB approach, which exclusively relies on prescribing boundary conditions from atmospheric reanalyses, and the additional assimilation of observed ocean physical variables. Recent work has already highlighted the advantage of using climate reconstructions to complement the GCB (Li et al., 2023). Moreover, past work has investigated the impact of assimilating biogeochemical observations to ocean simulations with uncertain results, mostly due to the scarcity of such observations (Valsala and Maksyutov, 2010; While et al., 2012).

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics Discussions

However, few efforts have been focused on investigating the impact on biogeochemistry of assimilating only physical variables, for which we have a far more complete dataset than for biogeochemical variables (Visinelli et al., 2016; Raghukumar et al., 2015). Here, we provide a detailed evaluation of the improvement of the representation of biogeochemical variables when observations of temperature and salinity are assimilated. We make use of several observation-based products that encompass surface pCO_2 air-sea CO_2 fluxes, nutrients, and surface chlorophyll to quantify the improvement of the biogeochemistry simulated by the model when observed physical fields are assimilated.

2 Methodology

We used the ocean component of the Earth System Model EC-Earth3-CC (Döscher et al., 2021). This is composed by the NEMO ocean general circulation model v3.6, (Madec et al., 2017), coupled with the ocean biogeochemical model PISCESv2 (Aumont et al., 2015). We designed two types of simulations in which we apply atmospheric forcing from reanalysis products. In the first type, in line with the usual GCB practice, we apply the omip protocol, (Griffies et al., 2016), where only sea surface salinity (SSS) restoring towards observed climatological values is applied, besides the atmospheric forcing (hereafter, omip). The second type is a reconstruction, where we also apply surface restoring of sea surface temperature (SST) and three-dimensional nudging of temperature and salinity towards time varying observations (hereafter, Data Assimilation or DA). This two-tier approach is then duplicated using two different combinations of atmospheric reanalysis to assess the impact of observational uncertainty. Details of the simulations and references for the data products used are given in Table 1.

All simulations were first equilibrated by repeating 4 times the historical period encompassed by the respective atmospheric forcing. This procedure allows the equilibration of the thermohaline circulation for the two omip simulations (Tsujino et al., 2020). In the case of the data assimilated (DA) reconstructions, a steady-state of the circulation is already achieved at the first cycle due to the 3D nudging of temperature and salinity towards observations. For all simulations, the ocean biogeochemistry is left free to evolve responding to the ocean physics evolution. Ocean physical fields (temperature and salinity) were initialised from EN4.2.2 (Good et al., 2013) in all cases, while dissolved inorganic carbon (DIC) and total alkalinity (TALK) were initialised from GLODAPv2 (Olsen et al., 2016; Lauvset et al., 2016), macronutrients (nitrate, phosphate, silicate) and oxygen were initialised from the World Ocean Atlas 2013 (Garcia et al., 2013b, a). Moreover, dissolved organic carbon (DOC) was initialised from the fields provided by an adjoint model (Hansell et al., 2009) while dissolved iron (Fe) was initialised using the median model results from the Iron Model Intercomparison Project (Tagliabue et al., 2016). The rest of biogeochemical tracers were initialised using low uniform values.

Since this first spinup period was not enough to fully equilibrate the ocean biogeochemical fields, an extension of the spinup was performed by repeating cyclically the physics of the 4^{th} cycle but letting the ocean biogeochemical fields free to evolve. The total spinup time was 525 years for JRA55 simulations and 513 years for the ERA5 simulations. To be consistent with the simulation protocol designed for the Global Carbon Budget 2022 (Friedlingstein et al., 2022), during the spinup phase,

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics
Discussions

atmospheric CO_2 concentration was held constant at 278 ppm, corresponding to the value in the year 1777. The spinup phase was enough to bring the air-sea CO_2 flux drift in all simulations to within 0.1 PgC yr^{-1} on a long-term average (Jones et al., 2016). At the end of the spinup, the historical period (1778-2021) was simulated by repeating the atmospheric forcing and by prescribing the observed atmospheric CO_2 time-series used in the GCB2022.

In DA simulations, the procedure includes restoring of SST and SSS as well as 3D temperature and salinity Newtonian dumping below the mixed layer. We modified the restoring timescale distribution of Sanchez-Gomez et al. (2016) below the mixed layer, to provide a smooth vertical variation between 10 days (above 800m) and 360 days (below 800m). This relaxation is applied everywhere except for the equatorial band between 15^{o}S-15^{o}N (where we leave a 10-times weaker nudging) due to the highly dynamical nature of this region which makes nudging problematic, resulting in spurious vertical velocities that introduce unrealistic injections of nutrients into the surface layers (Sanchez-Gomez et al., 2016; Park et al., 2018). At the surface, SST is restored using a feedback coefficient between flux and temperature of -200 W/m_2/K while the feedback parameter for freshwater fluxes is set at -750 mm/day.

In all model simulations, river nutrient input was prescribed as a climatology based on the GLOBAL-NEWS2 dataset (Mayorga et al., 2010), while DIC and alkalinity river input are based on the output of the Global Erosion Model (Ludwig et al., 1996). We note here that this procedure is in contrast with the GCB protocol, which recommends river fluxes of nutrients and carbon to be switched off. However, in agreement with the GCB procedure (Hauck et al., 2020), for every simulation, we also performed a control simulation, where atmospheric CO_2 concentration was kept constant at the preindustrial value. When calculating global air-sea CO_2 fluxes, we fit a linear trend to the global air-sea CO_2 flux time-series of the control simulation and then subtract this linear trend from the respective historical simulation. With this approach, we do not remove the interannual variability of the historical but we remove the drift (assuming it's the same in control and historical), any long-term trend in the natural carbon flux due to climate variability and change, as well as the outflux caused by the imbalance between river flux of carbon and sediment burial. The latter is slightly higher in the two omip simulations (0.26 – 0.28 Pg C yr^{-1}) than in the two DA simulations (0.21-0.23 Pg C yr^{-1}).

We use several observational datasets to evaluate the performance of our simulations. Details of the datasets used are given in Table 2. For SOCAT and GLODAP variables we used the point-values (i.e. not interpolated) and matched the model's output in space and time to calculate evaluation metrics. GCB2022 provides a central estimate of the global ocean CO_2 flux which is an average of 7 observation-based products and 10 OBGCM's estimates. The latter are produced with a suite of ocean biogeochemical models and using omip-like simulations (i.e. no data assimilation). For NOAA ERSST and SEANOE-MLD climatology, we used the gridded versions to calculate evaluation metrics. For OC-CCIv6.0 we used the level-3 gridded monthly data and subsampled model's output to match only valid points in the satellite images, before calculating differences. Finally, we used the most recent RAPID-MOCHA-WBTS (RAPID – Meridional Overturning Circulation and Heatflux Array- West

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

Boundary time series, hereafter RAPID array) monitoring time series and compared it with our modelled vertically integrated transport at 26^{o}N. References for these datasets are reported in Table 2.

3 Results

We compare the total carbon uptake in our simulations with the estimate from the GCB2022 (Table 3). The average uptake per decade is in general, lower in our simulations than for the GCB2022 estimate, however, when comparing the omip simulations with their respective DA counterparts, the uptake is generally increased in the latter, bringing it closer to the GCB estimate. This is confirmed by the time series of yearly integrated ocean uptake showing that data assimilation moves both the JRA55 and the ERA5 estimates upward, closer to the GCB2022 estimate (Fig. 1). In particular, it is worth noticing how the omip simulations are very close to the multi-model mean of the GCB2022 while the DA simulations separate from this, moving upward and getting closer to the GCB2022 estimate that also includes observation-based products. To further compare our simulations with the GCB2022, we provide a correlation matrix where all our simulations and all the individual GCB2022 models are correlated with the GCB2022 central estimate as well as with the observation-based products that contributed to it. We have ranked the models from high to low depending on their correlation with the central GCB2022 estimate and we notice that our simulations are overall comparable with the rest of models but, more importantly, the DA simulations have a higher correlation with GCB2022 than the omip simulations (Fig. 2). These results indicate that data assimilation is beneficial to improving the trajectory of the yearly globally integrated time-series, when assuming as benchmark the central GCB2022 estimate. However, this estimate is also dependent on models that share similar characteristics to our model and thus, likely, the same biases.

To provide an independent evaluation of the effect data assimilation has on the representation of the ocean carbon cycle we turn to the most comprehensive observational dataset of surface pCO_2. We sample the model's output in time and space to match available observations in SOCAT. These are averaged globally and then in time to give annual averages values (Hauck et al., 2020). From these time series we calculate the root mean square error (RMSE) and correlation coefficients between each simulation and SOCAT (Fig. 3). The differences among the model time series are small and barely discernible. Nevertheless, everywhere except in the tropics we see an increase of the correlation coefficient and a decrease of the RMSE when moving from the omip to the DA simulations, confirming the beneficial effect of data assimilation on the representation of the carbon cycle.

In a similar effort, we compared our simulations to other available observations, besides surface pCO_2. We used the GLODAPv2 database and repeated the same method we used for SOCAT to calculate the RMSE between each simulation and the observations, for six biogeochemical variables. In Fig. 4 we show the relative reduction of RMSE for every variable, when moving from omip to DA. Depending on the variable, the reduction of RMSE ranges from approximately 40% for DIC to close to 10% for nitrate, phosphate and oxygen. Despite this variability, the representation of all variables is systematically improved when using DA with respect to omip. Such a pervasive and consistent improvement is likely related to a better rep-

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

resentation of the three-dimensional large scale circulation. Although the 500 years biogeochemical spinup of our simulations (cf. methodology) may not be enough to equilibrate ocean biogeochemical tracers completely, we consider it sufficient for the ocean dynamic to influence their large-scale distribution. As a further confirmation, we separate the ocean volume in two layers, the upper layer (0-1000m) and the deep layer (1000m-bottom) and repeat the same procedure to calculate again the RMSE between the available observations and the model's output (Table 4). Even when considering the two portions of the ocean's volume separately, the error reduction is generalised to all variables and has similar values to those observed for the global assessment done in Fig. 4.

To verify that indeed large-scale circulation is improved when using data assimilation, we compare the maximum transport at 26 ^{o} N in the Atlantic Ocean with the measurements taken by the RAPID array as a proxy for the strength of the Atlantic Meridional Overturning Circulation (AMOC; Fig.5). Again, we observe how data assimilation is associated with a reduced distance with respect to the observational reference. The omip simulations are characterised by low AMOC values that are strengthened when using data assimilation. Finally, we have evaluated the impact of data assimilation on Sea Surface Temperature (SST), using The Extended Reconstructed Sea Surface Temperature (ERSSTv5) dataset (Huang et al., 2017), a product different from those used for the data assimilation itself (Good et al., 2013). Although we notice that all SST products must share the majority of the observations on which they are based, and therefore cannot be considered independent from each other, we use this exercise to assess once more whether the data assimilation is pushing the model's solution towards the observed state in a consistent way between the two different atmospheric forcings. In Fig. 6 we see how JRA55-omip and ERA5-omip are on opposite sides of the observed values and how data assimilation brings both simulations' estimates closer to it.

We also evaluated the effect data assimilation has on Mixed Layer Depth (MLD), an important metric for mass and energy exchanges between the atmosphere and the ocean. In this case, we use the climatology from de Boyer Montégut et al. (2004) as a reference. In Fig. 7 we can see how the bias is reduced in DA simulations with respect to the omip ones. This is particularly true in regions that are important hot-spots of CO_2 exchange between the ocean and the atmosphere, like the Southern Ocean, the North Atlantic and the North-West Pacific.

To complete our evaluation, we also compared surface chlorophyll produced by our simulations with the OC-CCIv6.0 dataset (Fig. 8). Similarly to what was done for the Mixed Layer Depth, we show the difference between model and observations, using a yearly climatology. In this case, we can observe how the effect of data assimilation is overall negligible except in a narrow band between 30^o N-40^o N in both the North Atlantic and the North Pacific, where the bias is actually slightly increased.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics
Discussions

4 Discussion

We have compared two pairs of simulations performed with different atmospheric forcing reanalysis to evaluate the impact of assimilating observations of temperature and salinity on the overall representation of ocean biogeochemistry. The evaluation of CO_2 fluxes is problematic because there is no accepted observation-based product to be used as a single benchmark. All the existing estimates begin from the same surface pCO_2 dataset and each uses its own method to fill the blanks. We have shown that data assimilation consistently produces estimates of CO_2 fluxes that are better aligned with the central estimate of the Global Carbon Budget 2022 than their omip counterpart. However, the estimate of the GCB2022 is a combination of observation-based products and omip-type simulations performed with a suite of ocean biogeochemical models. That is, only atmospheric forcing is provided as a surface boundary condition to the ocean model and no data assimilation is done. For this reason, showing that our estimates produced with data assimilation correlate better with the GCB22 estimate is informative but not enough to determine with confidence whether one practice (DA) is better than the other (omip).

For the above reasons, we decided to evaluate the performance of our simulations using the most comprehensive observational datasets available for several biogeochemical variables. When evaluating our simulations directly against the in-situ observations of pCO_2 we observed a consistent improvement when applying data assimilation. Similarly, for other biogeochemical variables the evaluation gives consistent results going in the same direction. It is important to remember that no direct data assimilation is provided for these variables and the degree to which they are impacted by the representation of the physical state of the ocean varies depending on the variable. Given this, it is hard to pinpoint a single cause for the improvements we see in biogeochemical variables when we apply data assimilation of temperature and salinity.

The distribution of macronutrients (nitrate, phosphate and silicate) is controlled by large-scale three-dimensional circulation (e.g. AMOC), vertical mixing (e.g. MLD) as well as by primary productivity at the surface (e.g. chlorophyll here is used as a proxy for phytoplankton biomass). The same considerations apply to surface pCO_2 and related CO_2 fluxes because these are impacted by both the large scale distribution of DIC and Alk but also by vertical mixing and, in some regions by primary productivity. The same is true for oxygen as intermediate and deep water ventilation, together with vertical mixing, represent the main input of this gas into the interior of the ocean. However, the solubility of both O_2 and CO_2 is strongly dependent on temperature, and thus, the data assimilation of this variable is likely to have a positive direct impact in their representation.

Based on these considerations, it is reasonable to assume that an improvement in the representation of the large-scale circulation is the main responsible for an improved distribution between the upper and lower layers of all the tracers considered here. We have shown how data assimilation led to AMOC values that are closer to observations with respect to the omip simulations. This result is in line with Karspeck et al. (2017), who also found that subsurface constraining resulted in a greater AMOC mean strength and enhanced variance with respect to reference simulations with no data assimilation. Such improvement is also backed up by the reduced bias in MLD in the subpolar North Atlantic, a key region for the formation of North Atlantic

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

deep water, which, in turn, is directly linked to the strength of the AMOC.

The improvement observed in the representation of the MLD is due to the ameliorated density profile obtained with data assimilation as the wind stress doesn't change between omip and DA simulations. The MLD can have a significant impact on the flux of CO_2, especially in those regions where we observed a bias reduction in the MLD itself. However, with no accepted observation-based product as a reference for CO_2 fluxes, a deeper analysis on the impact of MLD on the CO_2 fluxes in those regions would be somewhat speculative. The same consideration applies to most of the variables considered, as the direct influence of an improved representation of the MLD on any specific variable is difficult to disentangle.

Certainly more difficult to explain is the limited response of surface chlorophyll, despite an overall better representation of nutrients distribution and MLD. In fact, both nutrients availability and MLD have a direct impact on primary production and therefore on surface chlorophyll concentration. It is often the case that the default parameter set in an ocean biogeochemical model is chosen to reasonably reproduce both the large scale distribution of nutrients and that of surface chlorophyll. In this study, the default configuration of PISCESv2 (Aumont et al., 2015) was used without any further adjustment of parameters. Because of the improvement in the large-scale distribution of nutrients and in their input into the productive layer, related to more realistic MLD, the model is presented with a different nutrient availability, when applying data assimilation, with respect to the omip simulations. Similarly, the average light exposure of phytoplankton changes with changes in MLD. In some regions, the bias in the chlorophyll surface fields is actually increased with data assimilation. This is the case for the North Atlantic and North Pacific regions where a shallower MLD seems to coincide with an increase in surface chlorophyll, between 30^o N-40^o N. In these regions, the model responds by increasing the distance with respect to the reference chlorophyll observations because the parameter set used was somehow selected to reproduce the same chlorophyll fields under different nutrient and light availability conditions. In the rest of the ocean, the chlorophyll field seems rather insensitive to the changes brought by data assimilation. For some regions, this is likely due to upper oligotrophic waters experiencing changes in nutrient input that are too little to significantly impact primary production. For regions with higher surface chlorophyll, like the equatorial Pacific and the Southern Ocean, the reason for the weak response probably resides in the availability of iron not changing significantly with the changes in circulation and MLD. In fact, a significant part of iron input in these regions is from atmospheric deposition that is left unchanged in all simulations.

5 Conclusions

We conclude that the assimilation of observations for temperature and salinity has a beneficial effect in the representation of large-scale circulation and mixed layer depth, and this, in turn, translates into an improved representation of most of the ocean biogeochemical variables evaluated. Additionally, in the case of CO_2 and O_2, the improvements are most likely driven also by the direct beneficial effect that an ameliorated temperature field has on the solubility of these gases. Because of this overall beneficial effect on the representation of ocean biogeochemistry, we conclude that CO_2 fluxes are most likely improved

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

as well, although their direct validation is not straightforward. We have shown how not all aspects of biogeochemistry are improved as the surface chlorophyll field's representation is actually rather insensitive or even degraded when using data assimilation. We impute this result to the choice of parameters for the biogeochemical model that was based also on a realistic representation of surface chlorophyll as a reference. Because of this, we suggest that whenever possible, ocean biogeochemical models be fine-tuned using simulations that include some degree of data assimilation of the physical fields. Finally, since simple data assimilation practices, like the one presented here, can be included in simulations at negligible computational cost, we recommend that efforts like the Global Carbon Budget take into account this type of simulations in the future.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics
Discussions
Open Access

Figures

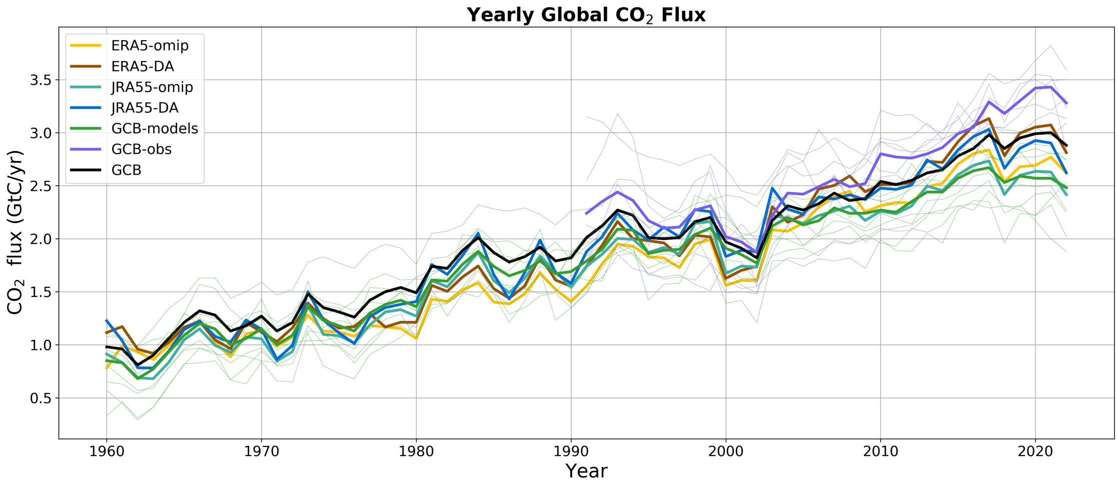

Figure 1. Globally integrated ocean CO_2 flux estimates for omip (orange and light blue) and DA (brown and dark blue) simulations, together with the central estimate of the GCB2022 (black) and the average estimate of both models (green) and observation-based products (purple) from the GCB2022. For the last two, individual estimates are also shown along the average estimates (thin lines of with same color code)

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

Earth System Dynamics
Discussions
Open Access

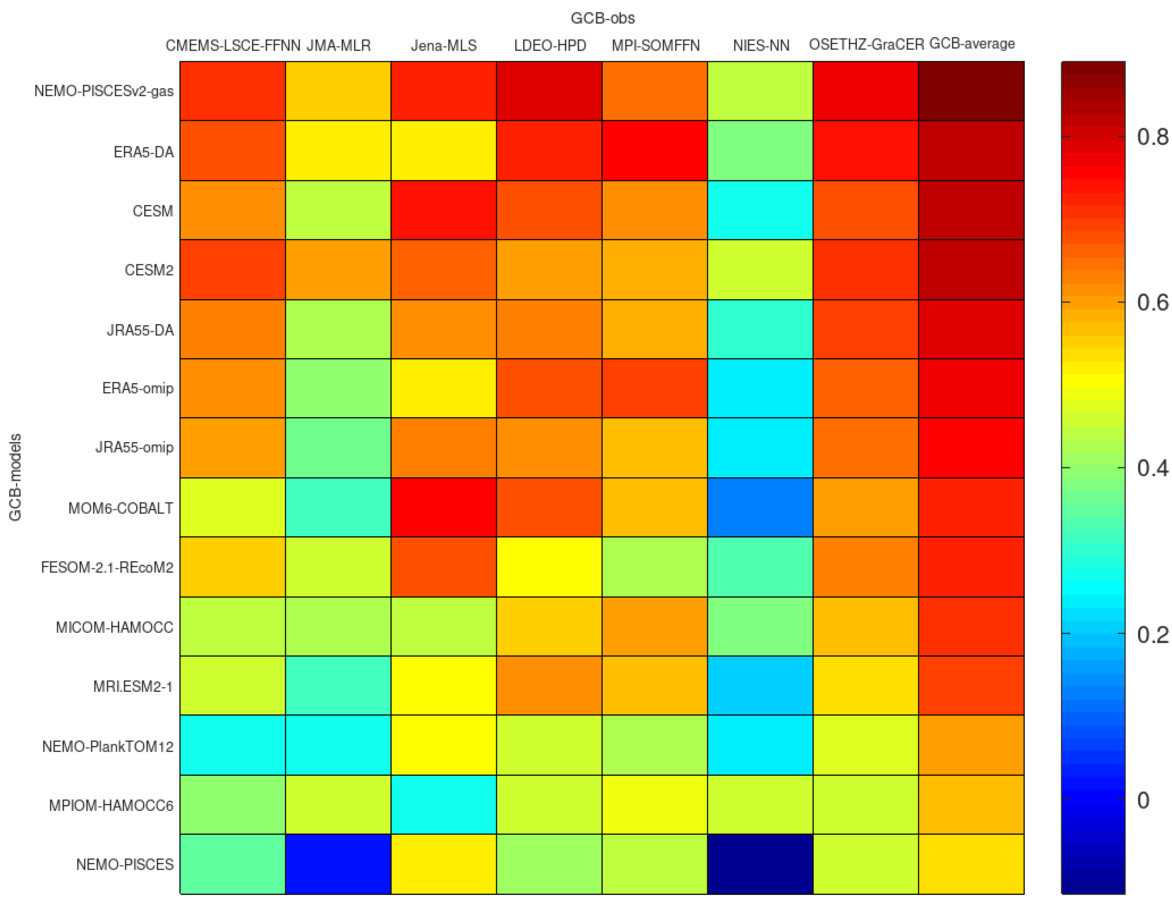

Figure 2. Correlation matrix between GCB2022 model's estimates of global CO_2 flux and observation-based estimates from the same exercise. Models are ranked from high to low based on their correlation with the central GCB2022 estimate (last column). Both DA and omip simulations are also ranked among the GCB models.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics
Discussions

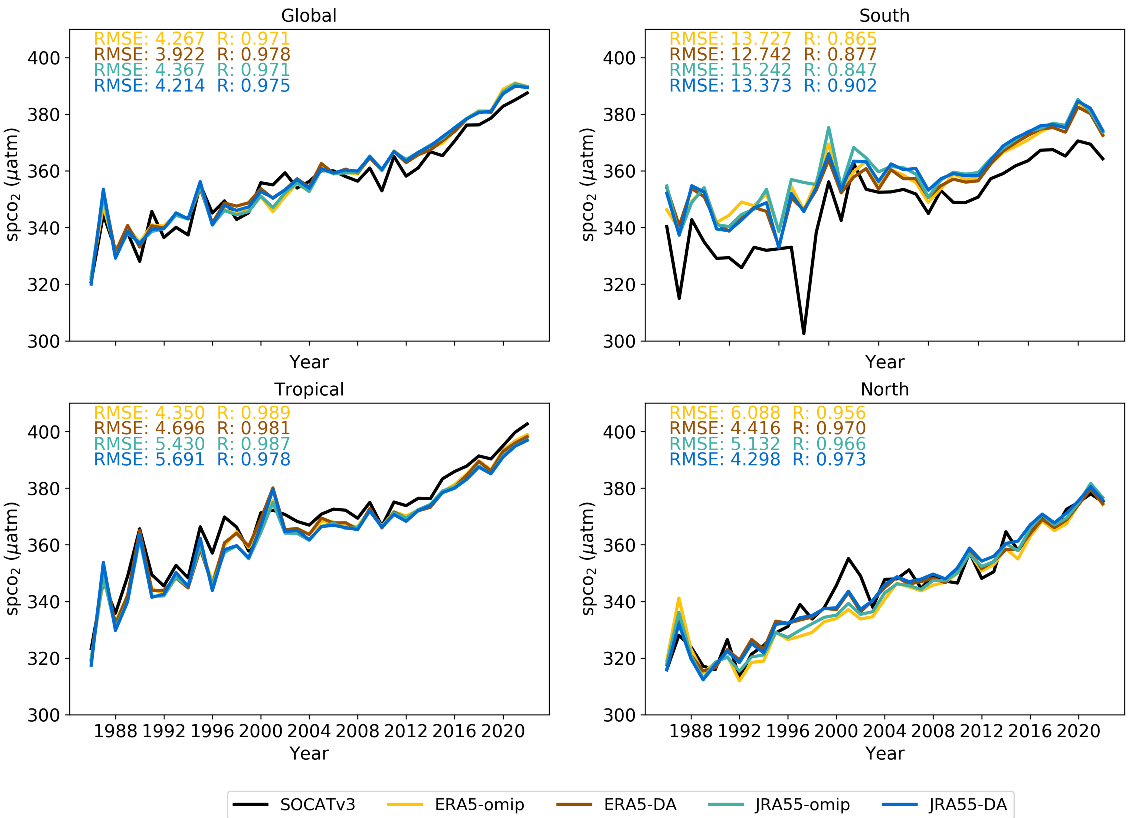

Figure 3. Yearly averages of surface pCO_2 values from the SOCATv3 database, compared to the four model's estimates of $spCO_2$ sampled to match in space and time the SOCATv3 values. The global ocean has been divided in three regions with boundaries at 20º N and 20º S.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics
Discussions

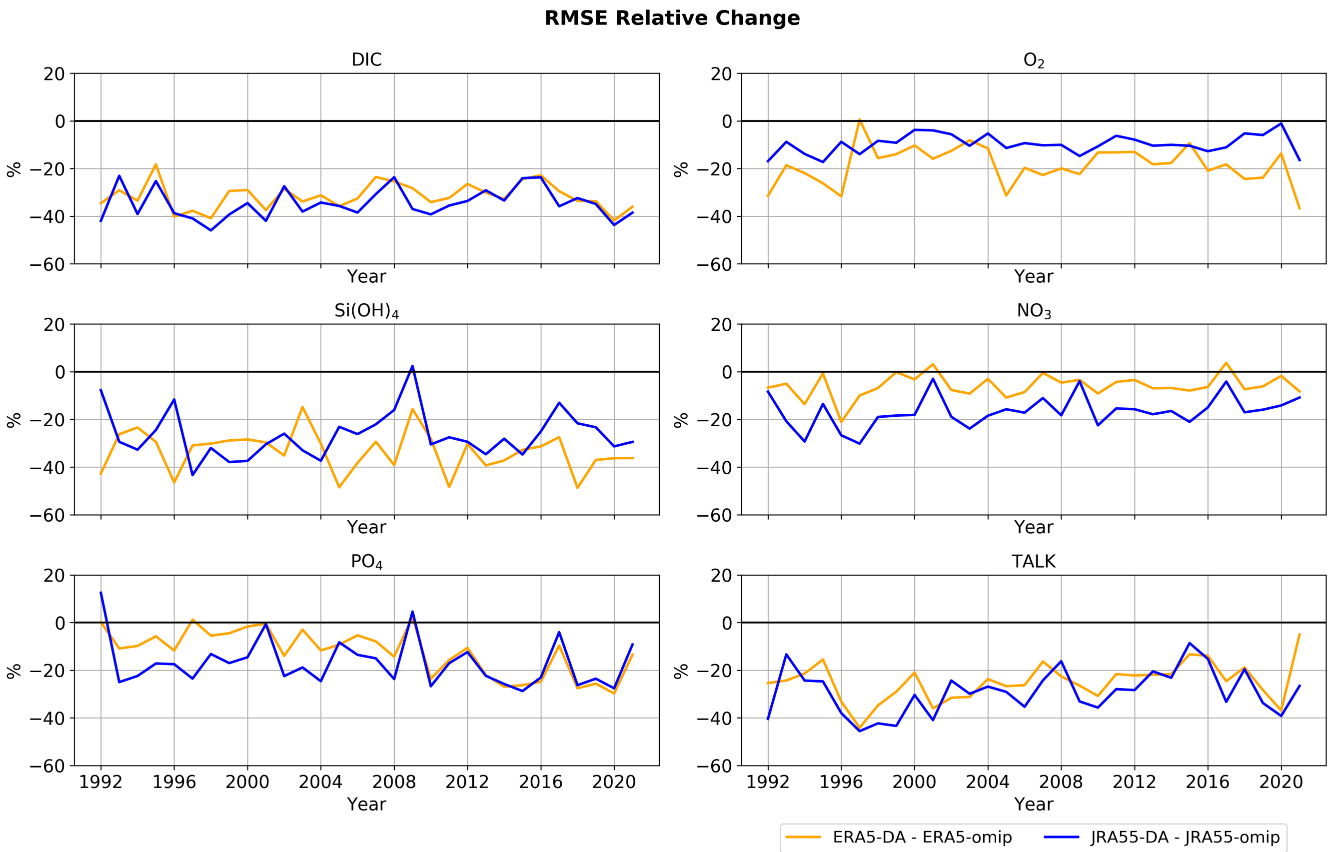

Figure 4. Relative change of RMSE when applying DA with respect to omip. For every variable, the available GLODAP observations were matched in time and space with the corresponding model's estimates to calculate RMSE.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

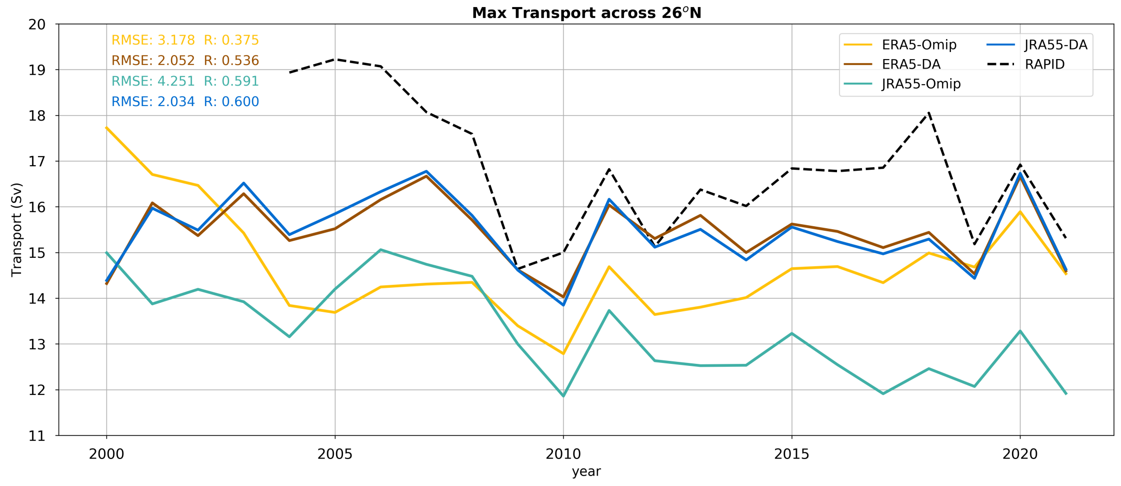

Figure 5. Maximum transport at 26.5 ^{o}N from model's output, compared to the observations from the RAPID array.

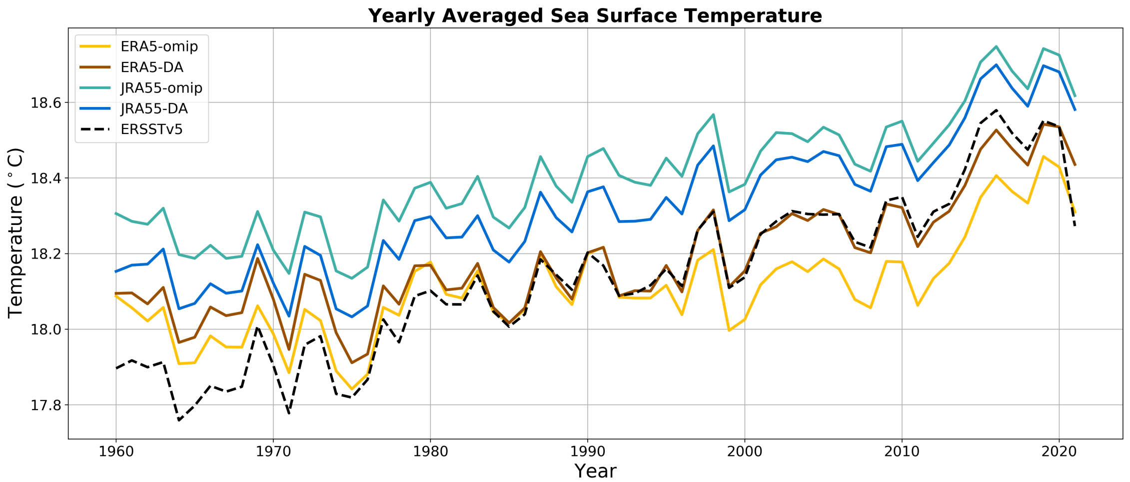

Figure 6. Globally averaged SST from the four simulations, compared with NOAA ERSSTv5.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

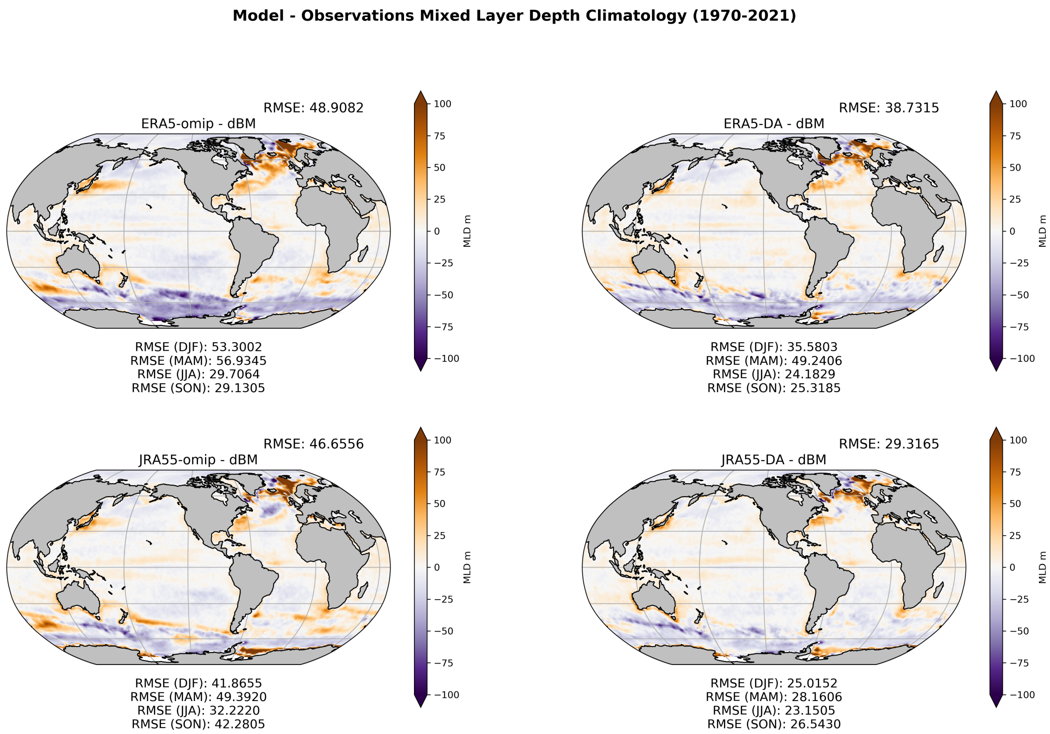

Figure 7. Yearly climatological (1970-2021) mean of mixed layer depth. The four maps show the difference between each simulation and the observation-based gridded product from de Boyer Montégut et al. (2004). RMSE for the year average as well as for each season are reported on top and below of each map, respectively.

© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics Discussions
Open Access

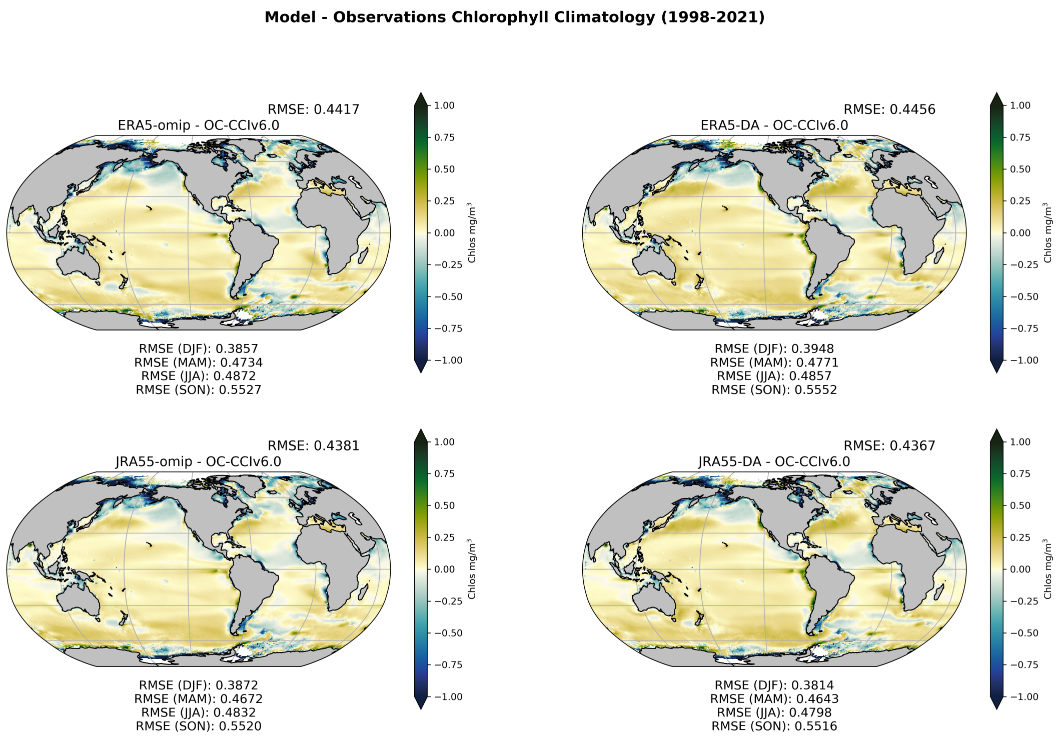

Figure 8. Yearly climatological (1998-2021) mean of surface chlorophyll. The four maps show the difference between each simulation and the OC-CCIv6.0 product. RMSE for the year average as well as for each season are reported on top and below of each map, respectively.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

Earth System Dynamics
Discussions

Tables

Simulation	**SST restoring**	**T&S 3D nudging**	**SSS restoring**	**Atmospheric forcing**	**Period**
JRA55-omip	None	None	JRA55-do-v1.5 (Tsujino et al., 2018)	JRA55-do-v1.5 (Tsujino et al., 2018)	1958-2021
JRA55-DA	COBE-SST (Ishii et al., 2005)	EN4.2.2 (Good et al., 2013)			
ERA5-omip	None	None	ORAS5 (Zuo et al., 2019)	ERA5 (Hersbach et al., 2020)	1959-2021
ERA5-DA	ORAS5 (Zuo et al., 2019)	EN4.2.2 (Good et al., 2013)			

Table 1. Re-analysis and observation-based products used in the two kinds of simulations here performed: omip-like (omip) and Data Assimilation (DA)

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.
Earth System Dynamics
Discussions

Dataset	Version	Variables	Period and frequency	Reference
SOCAT	v3-v2022	surface pCO2	1970-2021 (grouped by month)	Bakker et al. (2016)
GCB2022	v2022	CO2 flux	1959-2021 (models) 1990-2021 (obs.-based) Yearly global integral	Friedlingstein et al. (2022)
GLODAP	v2.2022	NO_3, PO_4, DIC, TAlk, $Si(OH)_4$, O_2	1972-2021 (grouped by month)	Lauvset et al. (2016)
NOAA-ERSST	v5.2023	SST	1960-2021 (monthly)	Huang et al. (2017)
SEANOE-MLD	v2023	MLD	1970-2021 (monthly climatology)	de Boyer Montégut et al. (2004)
RAPID AMOC	v2022.1	Transport at 26.5 ^{o} N	2004-2021 (yearly average)	Moat et al. (2022)
OC-CCI	v6.0	Surface chlorophyll	1997-2021 (monthly)	Sathyendranath et al. (2019)

Table 2. Observation products used for the validation of simulations results

	1960s	1970s	1980s	1990s	2000s	2012-2021	2021
GCB2022	1.1 ± 0.4	1.4 ± 0.4	1.8 ± 0.4	2.1 ± 0.4	2.3 ± 0.4	2.9 ± 0.4	2.9 ± 0.4
ERA5-omip	1.0	1.1	1.5	1.8	2.1	2.7	2.6
ERA5-DA	1.1	1.2	1.6	1.9	2.3	2.9	2.8
JRA55-omip	0.9	1.1	1.7	1.9	2.1	2.6	2.4
JRA55-DA	1.0	1.2	1.7	2.1	2.3	2.8	2.6

Table 3. Global carbon uptake (Pg C yr^{-1}) averaged over each decade from the 1960s to 2021 for the four simulations and the estimate of the GCB2022.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

	ERA5-omip	ERA5-DA	Rel. Change (%)	JRA55-omip	JRA55-DA	Rel. Change (%)
TAlk	50.87	38.10	-25.11%	54.15	39.12	-27.76%
	50.77	38.01	-25.12%	53.81	38.92	-27.68%
DIC	62.93	43.69	-30.58%	66.29	43.71	-34.06%
	62.95	43.45	-30.98%	65.44	43.30	-33.84%
O_2	41.80	33.67	-19.47%	34.55	30.89	-10.58%
	42.27	33.69	-20.31%	33.81	30.51	-9.77%
NO_3	04.08	3.77	-7.01%	4.54	3.68	-19.00%
	04.06	3.79	-6.68%	4.50	3.65	-18.82%
PO_4	0.36	0.31	-14.21%	0.38	0.30	-19.45%
	0.35	0.31	-12.61%	0.37	0.30	-17.78%
$Si(OH)_4$	21.98	14.58	-33.67%	20.10	14.40	-28.34%
	22.67	14.49	-36.05%	19.90	14.27	-28.31%

Table 4. RMSE calculated between each simulation and GLODAP. For each variable the RMSE is calculated for the upper 1000m (upper row of each variable) and below (lower row of each variable). The 4th and 7th columns show the relative change in RMSE between omip and DA, where a negative percentage value means a reduction of the error in DA with respect to omip.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

Code and data availability. The data used in this study have been made publicly available on Zenodo: 10.5281/zenodo.10233501. Further information on the data or extra files will be available upon request. All the simulations have been run with EC-Eeath3-CC: https://ec-earth.org/ (Döscher et al., 2021), using the workflow management Autosubmit (https://autosubmit.readthedocs.io/en/master/introduction/index.html, (Manubens-Gil et al., 2016; Uruchi et al., 2021)). The codes used for the analysis and plots, including jupyter notebooks, will be available upon request to the author. They will be put in the following repository: https://earth.bsc.es/gitlab/es/bsc-ocean-reconstructions. All the analysis and plots have been realized with open source codes: Octave (octave.org/), Python3 (python.org/), Xarray (xarray.dev), CDO (code.mpimet.mpg.de/projects/cdo) and Earthdiagnostics, in house tool for EC-EARTH model postprocessing (https://earthdiagnostics.readthedocs.io/en/latest/). All the observational data are publicly available on their corresponding websites.

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

Author contributions. R.B. and V.S. conceived the study and designed the experiments. R.B., V.S. and V.L. performed the experiments. R.B., V.S., P.O., V.L. and Y.R. performed sensitivity analysis and validation of ocean reconstructions that led to the configuration used in this study. E.T., V.L. and E.F. provided support for all computing aspects of this study. R.B and V.S. performed the analysis. All authors contributed to the writing of the manuscript.

Acknowledgements. The research leading to these results has received funding from the EU H2020 Framework Programme under grant agreement n. GA 821003.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

References

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, Geoscientific Model Development, 8, 2465–2513, https://doi.org/10.5194/gmd-8-2465-2015, 2015.

Bakker, D. C., Pfeil, B., Landa, C. S., Metzl, N., O'brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., et al.: A multi-decade record of high-quality fCO 2 data in version 3 of the Surface Ocean CO 2 Atlas (SOCAT), Earth System Science Data, 8, 383–413, 2016.

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Cruz-García, R., Cvijanovic, I., et al.: Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth, Earth System Dynamics, 12, 173–196, 2021.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, Journal of Geophysical Research: Oceans, 109, 2004.

DeVries, T., Yamamoto, K., Wanninkhof, R., Gruber, N., Hauck, J., Müller, J. D., Bopp, L., Carroll, D., Carter, B., Chau, T.-T.-T., et al.: Magnitude, trends, and variability of the global ocean carbon sink from 1985 to 2018, Global Biogeochemical Cycles, 37, e2023GB007 780, 2023.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., et al.: The EC-earth3 Earth system model for the climate model intercomparison project 6, Geoscientific Model Development Discussions, 2021, 1–90, 2021.

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., et al.: Global carbon budget 2022, Earth System Science Data Discussions, 2022, 1–159, 2022.

Garcia, H., Locarnini, R., Boyer, T., Antonov, J., Baranova, O., Zweng, M., Reagan, J., and Johnson, D.: Dissolved inorganic nutrients (phosphate, nitrate, silicate), World Ocean Atlas, 4, 25, 2013a.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., Reagan, J. R., Johnson, D. R., Mishonov, A. V., and Levitus, S.: World ocean atlas 2013. Volume 4, Dissolved inorganic nutrients (phosphate, nitrate, silicate), https://doi.org/10.7289/V5J67DWD, 2013b.

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, 2013.

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., Chassignet, E. P., Curchitser, E., Deshayes, J., Drange, H., et al.: OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project, Geoscientific Model Development, p. 3231, 2016.

Gruber, N., Clement, D., Carter, B. R., Feely, R. A., Van Heuven, S., Hoppema, M., Ishii, M., Key, R. M., Kozyr, A., Lauvset, S. K., et al.: The oceanic sink for anthropogenic CO2 from 1994 to 2007, Science, 363, 1193–1199, 2019.

Gruber, N., Bakker, D. C., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., McKinley, G. A., and Müller, J. D.: Trends and variability in the ocean carbon sink, Nature Reviews Earth & Environment, 4, 119–134, 2023.

Hansell, D. A., Carlson, C. A., Repeta, D. J., and Schlitzer, R.: Dissolved organic matter in the ocean: A controversy stimulates new insights, Oceanography, 22, 202–211, 2009.

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C., Bopp, L., Chau, T. T. T., Gürses, Ö., Ilyina, T., Landschützer, P., et al.: Consistency and challenges in the ocean carbon sink estimate for the global carbon budget, Frontiers in Marine Science, 7, 571 720, 2020.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., Zhang, H.-M., et al.: NOAA extended reconstructed sea surface temperature (ERSST), version 5, NOAA National Centers for Environmental Information, 30, 25, 2017.

Ishii, M., Shouji, A., Sugimoto, S., and Matsumoto, T.: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection, International Journal of Climatology: A Journal of the Royal Meteorological Society, 25, 865–879, 2005.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., et al.: C4MIP–The coupled climate–carbon cycle model intercomparison project: Experimental protocol for CMIP6, Geoscientific Model Development, 9, 2853–2880, 2016.

Karspeck, A., Stammer, D., Köhl, A., Danabasoglu, G., Balmaseda, M., Smith, D., Fujii, Y., Zhang, S., Giese, B., Tsujino, H., et al.: Comparison of the Atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products, Climate Dynamics, 49, 957–982, 2017.

Keppler, L. and Landschützer, P.: Regional wind variability modulates the Southern Ocean carbon sink, Scientific reports, 9, 7384, 2019.

Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C., Van Heuven, S., Hoppema, M., Metzl, N., Sweeney, C., Takahashi, T., et al.: The reinvigoration of the Southern Ocean carbon sink, Science, 349, 1221–1224, 2015.

Lauvset, S. K., Key, R. M., Olsen, A., Van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., et al.: A new global interior ocean mapped climatology: The 1× 1 GLODAP version 2, Earth System Science Data, 8, 325–340, 2016.

Le Quéré, C., Rodenbeck, C., Buitenhuis, E. T., Conway, T. J., Langenfelds, R., Gomez, A., Labuschagne, C., Ramonet, M., Nakazawa, T., Metzl, N., et al.: Saturation of the Southern Ocean CO2 sink due to recent climate change, science, 316, 1735–1738, 2007.

Li, H., Ilyina, T., Müller, W. A., and Landschützer, P.: Predicting the variable ocean carbon sink, Science advances, 5, 64–71, 2019.

Li, H., Ilyina, T., Loughran, T., Spring, A., and Pongratz, J.: Reconstructions and predictions of the global carbon budget with an emission-driven Earth system model, Earth System Dynamics, 14, 101–119, 2023.

Lovenduski, N. S., Gruber, N., Doney, S. C., and Lima, I. D.: Enhanced CO2 outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, Global Biogeochemical Cycles, 21, 2007.

Lovenduski, N. S., Yeager, S. G., Lindsay, K., and Long, M. C.: Predicting near-term variability in ocean carbon uptake, Earth System Dynamics, 10, 45–57, 2019.

Ludwig, W., Probst, J.-L., and Kempe, S.: Predicting the oceanic input of organic carbon by continental erosion, Global Biogeochemical Cycles, 10, 23–41, 1996.

Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D., et al.: NEMO ocean engine, 2017.

Manubens-Gil, D., Vegas-Regidor, J., Prodhomme, C., Mula-Valls, O., and Doblas-Reyes, F. J.: Seamless management of ensemble climate prediction experiments on HPC platforms, in: 2016 International Conference on High Performance Computing & Simulation (HPCS), pp. 895–900, IEEE, 2016.

Mayorga, E., Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H., Bouwman, A., Fekete, B. M., Kroeze, C., and Van Drecht, G.: Global nutrient export from WaterSheds 2 (NEWS 2): model development and implementation, Environmental Modelling & Software, 25, 837–853, 2010.

© Author(s) 2023. CC BY 4.0 License.

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, 2017.

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A., Danabasoglu, G., Dirmeyer, P. A., Doblas-Reyes, F. J., Domeisen, D. I., et al.: Current and emerging developments in subseasonal to decadal prediction, Bulletin of the American Meteorological Society, 101, E869–E896, 2020.

Moat, B., Frajka-Williams, E., Smeed, D., Rayner, D., Johns, W., Baringer, M., Volkov, D., and Collins, J.: Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2020 (v2020.2), Tech. rep., NERC EDS British Oceanographic Data Centre NOC, doi:10.5285/e91b10af-6f0a-7fa7-e053-6c86abc05a09, 2022.

Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., et al.: The Global Ocean Data Analysis Project version 2 (GLODAPv2)–an internally consistent data product for the world ocean, Earth System Science Data, 8, 297–323, 2016.

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, 2018.

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, 2015.

Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., and Terray, L.: Drift dynamics in a coupled model initialized for decadal forecasts, Climate Dynamics, 46, 1819–1840, 2016.

Sathyendranath, S., Brewin, R. J., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A. B., Dingle, J., Doerffer, R., et al.: An ocean-colour time series for use in climate studies: the experience of the ocean-colour climate change initiative (OC-CCI), Sensors, 19, 4285, 2019.

Séférian, R., Berthet, S., and Chevallier, M.: Assessing the decadal predictability of land and ocean carbon uptake, Geophysical Research Letters, 45, 2455–2466, 2018.

Tagliabue, A., Aumont, O., DeAth, R., Dunne, J. P., Dutkiewicz, S., Galbraith, E., Misumi, K., Moore, J. K., Ridgwell, A., Sherman, E., et al.: How well do global ocean biogeochemistry models simulate dissolved iron distributions?, Global Biogeochemical Cycles, 30, 149–174, 2016.

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., et al.: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do), Ocean Modelling, 130, 79–139, 2018.

Tsujino, H., Urakawa, L. S., Griffies, S. M., Danabasoglu, G., Adcroft, A. J., Amaral, A. E., Arsouze, T., Bentsen, M., Bernardello, R., Böning, C. W., et al.: Evaluation of global ocean–sea-ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2), Geoscientific Model Development, 13, 3643–3708, 2020.

Uruchi, W., Castrillo, M., and Beltrán, D.: Autosubmit GUI: A Javascript-based graphical user interface to monitor experiments workflow execution, Journal of Open Source Software, 6, 2021.

Valsala, V. and Maksyutov, S.: Simulation and assimilation of global ocean pCO2 and air–sea CO2 fluxes using ship observations of surface ocean pCO2 in a simplified biogeochemical offline model, Tellus B: Chemical and Physical Meteorology, 62, 821–840, 2010.

Visinelli, L., Masina, S., Vichi, M., Storto, A., and Lovato, T.: Impacts of data assimilation on the global ocean carbonate system, Journal of Marine Systems, 158, 106–119, 2016.

https://doi.org/10.5194/esd-2023-40
Preprint. Discussion started: 22 December 2023
© Author(s) 2023. CC BY 4.0 License.

While, J., Totterdell, I., and Martin, M.: Assimilation of pCO2 data into a global coupled physical-biogeochemical ocean model, Journal of Geophysical Research: Oceans, 117, 2012.

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