Reconstructing ocean carbon storage with CMIP6 models and synthetic Argo observations

Katherine E. Turner[1], Doug M. Smith[2], Anna Katavouta[1,3], and Richard G. Williams[1]

[1]Department of Earth, Ocean, and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom
[2]UK Met Office, Exeter, United Kingdom
[3]National Oceanography Centre, Liverpool, United Kingdom

Correspondence: Katherine E. Turner (K.E.Turner2@liverpool.ac.uk)

Abstract. The ocean carbon store plays a vital role in setting the carbon response to emissions and variability in the carbon cycle. However, due to the ocean's strong regional and temporal variability, sparse carbon observations limit our understanding of historical carbon changes. We explore how widespread Argo temperature and salinity profiles can provide information to reconstruct ocean carbon inventories with Ensemble Optimal Interpolation. Ensemble Optimal Interpolation draws upon

5   first-order relationships between variables, and high-complexity climate models been used previously to reconstruct interior ocean heat content and salinity changes from Argo profiles. The CMIP6 model ensemble shows coherent relationships between upper-ocean carbon, temperature, salinity, and atmospheric CO_2 that result in solutions that reflect the controls of undersaturation, solubility, and alkalinity. The Ensemble Optimal Interpolation method is tested with synthetic reconstructions of upper-ocean carbon fields using perfect information of temperature and salinity changes, along with atmospheric CO_2 con-

10   centrations. Sensitivity tests of show that, in most regions, the trend in ocean carbon and over 60% of detrended variability can be reconstructed using local temperature and salinity measurements. Expanding the synthetic reconstructions to include irregular sampling consistent with Argo profile locations results in a similar capacity to reconstruct ocean carbon variability, as the increased information provided from multiple sampling locations compensates for the propagation of errors within the CMIP6 covariance fields. Overall, our work indicates that hydrographic measurements can provide valuable information about changes

15   in the ocean carbon inventory. The flexibility of optimal interpolation indicates a strong potential for new, independent, and mechanistic estimates of the historical carbon inventory that can be used to compare with other estimates of the ocean carbon system.

1   Introduction

The global ocean plays a important role in the carbon cycle, being both a major reservoir of carbon and substantial sink of

20   anthropogenic carbon transferred from geological storage to the coupled atmosphere-ocean-terrestrial system. Recent studies estimate that the ocean has taken up around 26% of anthropogenic emissions from fossil fuels and land use changes since 1850 (Friedlingstein et al., 2022). In addition to the long-term uptake of anthropogenic carbon, ocean carbon uptake exhibits interannual and decadal variability on global and regional scales (Landschützer et al., 2016; McKinley et al., 2017; Gruber et al., 2019; McKinley et al., 2020). This variability impacts the ability to detect trends in observations of the partial pressure of

25 CO_2 in seawater (McKinley et al., 2016) and, through oceanic transport, can lead to regionally enhanced acidification (Gruber et al., 2021). The ocean carbon inventory is thus an important aspect of current climate change, and up-to-date estimates on its behaviour are vital for understanding its evolution and impacts on the climate system.

In order to characterise the ocean carbon system with its strong regional and temporal variability, extensive spatial and temporal coverage in observations are required. Whereas there are millions of surface pCO_2 measurements to provide a surface
30 view of the ocean carbon system and air-sea carbon fluxes (Bakker et al., 2016), complementary observations of ocean interior dissolved inorganic carbon (DIC) are limited by the logistics of ship-based bottle measurements. Campaigns for ocean interior DIC observations use repeat transects to provide high-quality observations for many regions (Sloyan et al., 2019). The transects resolve spatial variations in carbon but are limited in their ability to resolve temporal variability as they are repeated on decadal timescales. Alternatively, ocean DIC time series such as those found at the Bermuda Atlantic Time Series and the Hawaii Ocean
35 Timeseries can resolve seasonal, interannual, and decadal variability, but are limited in how well they represent variability on larger spatial scales (Bates et al., 2014). The Bio-Argo programme has allowed for more autonomous sampling of ocean interior carbonate system variables such as pH (Claustre et al., 2020), from which interior DIC can be estimated; currently, however, autonomous measurement methods for interior DIC remain in development.

By itself, the sparsity of interior observations hinders the ability to produce a coherent, global picture of recent ocean carbon
40 changes from a storage perspective. A mapping technique is necessary to expand points of observations into more coherent spatial patterns of change and behaviour. Mapping techniques employ statistics to propagate information from observations to unobserved regions. Time-invariant climatologies of ocean carbon have been created using mapping procedure on data from repeat transects and other ship-based observations (Lauvset et al., 2021). For observations of surface pCO_2, non-linear machine learning procedures have been able to reconstruct spatial patterns in interannual and decadal variability (Landschützer et al.,
45 2016; Landschützer et al., 2019; Gloege et al., 2022), although these algorithms can lead to substantial biases if observations are too sparse for certain regions. Data assimilative methods that use observations to constrain model physics in forward experiments have also been expanded to include biogeochemistry (Verdy and Mazloff, 2017; Carroll et al., 2020).

For any mapping or data assimilative technique, it is imperative to use accurate statistics to avoid the erroneous propagation of information from observations. These statistical relationships can be parameterised or calculated from observations;
50 however, parameterised statistics fail to reflect regional differences in the ocean, and for poorly observed variables, averaging over large length or timescales can lead to overly coarse covariance fields. State-of-the-art climate models provide complete pseudo-data in both space and time and therefore may be used to calculate these statistics, though model biases may lead to errors in the covariances. Nevertheless, fully-coupled climate models have been used to reconstruct ocean heat content trends and variability from observations, with clear improvements in recent years due to the expansion of the Argo programme (Smith
55 and Murphy, 2007; Smith et al., 2015; Cheng and Zhu, 2016).

Ocean carbonate chemistry is controlled by both physical and biogeochemical processes. There are well-understood first-order principles that relate to ocean temperature and salinity to upper ocean carbon; using these relationships, the observations used in previous heat content mappings may be useful for carbon mappings. Increases in temperature reduce the ocean's ability to take up CO_2 through gas solubility laws, while increases in salinity increase the ocean's ability to take up CO_2 by increasing

60   alkalinity (Williams and Follows, 2011). Additionally, temperature and salinity provide constraints on ocean circulation, which alters the vertical gradients of both heat and carbon (Thomas et al., 2018; Williams et al., 2021). The relationships between temperature, salinity, and carbon are regionally dependent as ocean dynamics and biology can set different drivers of CO_2 uptake (Lauderdale et al., 2016). If these relationships can be exploited, the increase in ocean observations from the Argo programme may provide valuable information that can help reconstruct ocean carbon fields alongside temperature and salinity

65   fields. While the ocean carbonate system can also be approximated using other observations such as pH and salinity, for this study we focus on the readily available and widespread observations present within the Argo dataset.

In this study, we apply an Ensemble Optimal Interpolation approach to reconstruct upper-ocean dissolved inorganic carbon from temperature and salinity observations. To calculate the covariance fields, we construct an ensemble of Earth system models from the Climate Model Intercomparison Project Phase 6 (CMIP6) and examine their relationships between upper-

70   ocean carbon and other climate variables. In this "proof of concept" study we aim to show the potential skill available in using model covariance fields and Argo-style measurements. We create synthetic reconstructions of modelled ocean carbon using perfect pseudo-observations of temperature and salinity similar to the ocean heat content synthetic reconstructions in Allison et al. (2019). We then compare the errors within these reconstructions against those from the climatological fields to see where and how the method best works in the model world.

75   We set out the work in this study as follows. In Section 2 we describe the Ensemble Optimal Interpolation scheme for ocean DIC using an ensemble of CMIP6 models, with an overview of the experiments we conduct to test the reconstruction skill. In Section 3 we present the ensemble correlation fields between DIC and temperature and salinity at the same location, with a discussion as to how these correlations relate to physical controls on the ocean carbon response. In Sections 4 and 5 we present sensitivity results to discuss the reconstruction potential of different temperature and salinity sampling distributions, ranging

80   from perfect spatial coverage to coverage more representative of ARGO observations, with global and regional comparisons of the different synthetic Argo reconstructions in Section 6. Lastly, in Section 7 we discuss the potential and shortcomings of this method, both in terms of the setup with temperature and salinity and further expansions with other oceanographic observations.

2   The Ensemble Optimal Interpolation method for ocean carbon

Optimal interpolation is a non-dynamical mapping approach that uses weights to propagate information from observations to

85   regions without observations (Daley, 1991; Smith and Murphy, 2007). The optimal interpolation method involves creating an analysis A at locations i and times t from the sum of a background state B_i and a weighted sum of the difference between the observed and background values, also known as the observation increments. The observations can be the same property as the final analysis, or can be of other properties, i.e. temperature and salinity observations can be used to reconstruct DIC if there is a physical link between them. The inclusion of these sorts of physically relevant data produces a *multivariate analysis*. For

90   observations O at locations k within a sampling region K, the optimal interpolation method can be represented as

$$A_i = B_i + \sum_{k \in K} w_k (O_k - B_k). \tag{1}$$

The crux of the optimal interpolation problem is thus finding a suitable solution for the weights w_k. The optimal weights are those that minimise the expected analysis error at each gridpoint i, calculated as the root mean squared error (RMSE) between the analysis and the truth T_i:

$$RMSE(A_i, T_i) = \sqrt{\sum_t (A_i(t) - T_i(t))^2}. \tag{2}$$

In this work we explore how the relationships between DIC and observed ocean variables such as temperature and salinity can be used to reconstruct upper ocean carbon stocks. The use of optimal interpolation to reconstruct ocean carbon from available carbon measurements is inherently limited by the poor temporal and spatial coverage of existing DIC observations. However, temperature and salinity observations are far more plentiful, particularly in the Argo era, so we explore the extent to which these observations could be used to reconstruct DIC. To avoid the problems associated with sparse input data, we take a multivariate analysis approach to reconstruct carbon from extensive synthetic observations of ocean temperature T and salinity S consistent with observations from the Argo programme (Wong et al., 2020), as well as annual average atmospheric CO_2 concentrations. On annual and longer timescales, atmospheric CO_2 is well-mixed, and so the use of annual atmospheric CO_2 measurements can allow the analysis to capture longer-term DIC changes from changes in the global carbon budget. With background fields for DIC, T, and S taken to be their global climatologies (i.e., $DIC(i,t) = \overline{DIC}(i) + DIC'(i,t)$, where $\overline{DIC}(i)$ is the temporal average DIC concentration at location i over the period 1955-2014), the optimal interpolation scheme is formulated to calculate the residual DIC' from T', S' within a region K:

$$DIC'(i,t) = \sum_{k \in K} (w_{C,k} pCO_2'(t) + w_{T,k} T'(k,t) + w_{S,k} S'(k,t)), \tag{3}$$

where $w_{C,k}$ is the local weighting for the annual global averaged atmospheric pCO_2 and $w_{T,k}$ and $w_{S,k}$ are the weightings for observed ocean temperatures and salinities within the sampling region K.

2.1 Choices for calculating background error covariances

The optimal weights w_k in (1) and (3) are determined by the covariances between the background errors $B_i - T_i$. The covariance fields describe how information should be propagated from areas with observations to those without; the optimal solution will consider how much new information observations provide to the reconstruction and depend on local and larger-scale relationships within the climate system. Any ensemble optimal interpolation method must involve decisions made on how the background error covariances are calculated, and which covariances are included to solve for the weights w_k. In the following we describe how we have made these decisions using a CMIP6 multi-model ensemble and with various assumptions as to which observations are used for the multivariate DIC analyses.

2.1.1 Background covariances from a CMIP6 ensemble

Climate model outputs can be used to provide covariance fields for optimal interpolation solutions. By providing complete pseudo-data in both space and time, climate models avoid some of the errors that arise from the coarseness of observational or

parameterised covariance fields. Background covariance fields from global climate models have been used to reconstruct ocean temperatures and salinities from observations as well as to initialise decadal forecasts (Smith and Murphy, 2007; Smith et al., 2015; Cheng and Zhu, 2016). However, the model ensemble background covariance field will still contain errors that will need

125 to be evaluated using sensitivity testing before the fields can be used with real-world observations.

To construct the background covariance fields, we obtained ocean potential temperature, ocean salinity, and ocean DIC model output from 6 CMIP6 Earth System models with a nominal horizontal resolution around $1°$ (Table 1). The output was taken from the historical experiments and covers the period of year 1955 to year 2014. This period was chosen as it has consistent behaviour in atmospheric CO_2 concentrations and is long enough to allow some multidecadal variability to be captured in the

130 covariance fields. For each model, output from 5 realisations was used so that the models' internal variability was captured by the ensemble.

Table 1. CMIP6 Earth System Models and realisations used for reconstruction

Model (Reference)	Realisations
ACCESS-ESM-1.5 (Ziehn et al., 2020)	r1i1p1f1, r2i1p1f1, r4i1p1f1, r5i1p1f1, r6i1p1f1
CanESM5 (Swart et al., 2019)	r10i1p1f1, r11i1p1f1, r12i1p1f1, r13i1p1f1, r14i1p1f1
CESM2 (Danabasoglu et al., 2020)	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
IPSL-CM6A-LR (Boucher et al., 2020)	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r32i1p1f1
MPI-ESM1.2-LR (Mauritsen et al., 2019)	r1i1p1f1, r2i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
UKESM1 (Sellar et al., 2019)	r1i1p1f2, r2i1p1f2, r3i1p1f2, r4i1p1f2, r8i1p1f2

For all variables, annual averages were calculated from monthly mean outputs, and outputs were bilinearly regridded from their native horizontal grids to a $1° × 1°$ grid using the Python package `xESMF` (https://doi.org/10.5281/zenodo.1134365). Oceanic variables were further integrated vertically, with the results in this study focusing on the results from the integrated

135 top 100m.

2.1.2 Spatial limits on background covariances

The sampling region K for temperature and salinity observations in (3) can vary from co-located observations to global observations; the choice of sampling region requires balancing the extra information provided by additional observations with spurious propagation of information through errors in the covariance fields. In this study we use two idealised sampling meth-

140 ods to explore the limits of this first-order reconstruction of ocean carbon:

1. Observations are taken from at each ocean grid cell at each year, therefore assuming perfect coverage of temperature and salinity profiles. The reconstruction is then set up to use local relationships between ocean carbon, temperature, and salinity, as well as globally averaged atmospheric pCO_2. With this method, we explore how the local relationships between upper-ocean DIC, temperature, and salinity can be used to reconstruct carbon with perfect hydrographic knowl-
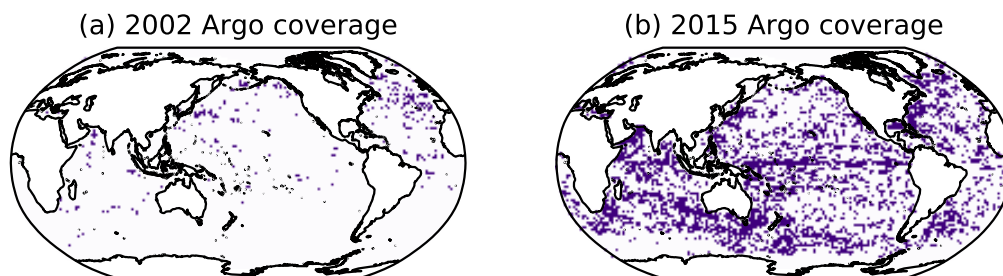
Figure 1. Argo profile distributions consistent with sampling for (a) year 2002 and (b) year 2015. Coloured regions indicate bins that have at least 6 months of observations in the given year.

145   edge. Since observations are complete in this test, the presence of errors can be attributed to poorly known relationships or nonlinear dynamics within the model ensemble.

2. Observations are taken irregularly, consistent with Argo profiles (Wong et al., 2020). While coverage of ocean temperature and salinity observations is significantly higher than those for ocean carbon, sampling remains irregular in space and time, so real-world reconstruction methods must account for this irregularity. We take Argo data from years 2002

150   and 2015 and bin the profiles onto the $1° \times 1°$ horizontal grid. Any grid cell that has at least 6 months of profiles in a given year is taken to be sufficiently observed, and the modelled annual average profile there is used as an observation for the reconstruction. For year 2002, this leads to a distribution of synthetic profiles that is heavily concentrated in the Northern Hemisphere and the North Atlantic in particular (Fig. 1a). By year 2015, the scheme allows for sampling of most of the global ocean outside the Southern Ocean and the Arctic Ocean (Fig. 1b).

155   As the Argo profiles have incomplete global coverage, we conduct additional sensitivity tests to see how nearby Argo observations can be used to construct near-global reconstructions. For these reconstructions, the reconstruction at any given point is made to be a linear combination of observations within a certain radius. We choose radii of 1°, 2°, or 5°: for a radius of 5° and current Argo coverage like that for year 2015, most regions of the ocean have observations within the search radius, allowing for nearly globally complete DIC reconstructions.

160 **2.2   Synthetic reconstructions and tests**

Evaluating how well an analysis reproduces variability requires a comparison with a field truth. As the ocean carbon field is not known to a high accuracy, we conduct synthetic reconstructions in which we reconstruct other models' ocean carbon fields using various distributions of modelled temperature and salinity (Allison et al., 2019). For these tests, the observations come from the models and therefore contain no errors outside of small errors possible from sampling and regridding. As the

165 observations can be assumed to have little error, the weights w_k in (2) can be solved for through a least-squares algorithm, without having to consider observational errors (Smith and Murphy, 2007).

The models used in the synthetic reconstructions are the same models used to calculate the ensemble covariance fields. The inclusion of a model in the ensemble can over-fit the results and produce spuriously accurate reconstructions; therefore, for a synthetic reconstruction of, for instance, UKESM1 DIC inventories, we eliminate all the UKESM1 realisations from

170   the ensemble. We then calculate the covariances and optimal weights calculated from the remaining out-of-sample ensemble members. Then, these optimal weights are used with the UKESM1 temperature and salinity profiles to create a DIC analysis.

To compare the reconstructions, we calculate the improvement in the RMSE, relative to the RMSE of the first-guess background:

$$\varepsilon_i(A) = \frac{RMSE(B_i, T_i) - RMSE(A_i, T_i)}{RMSE(B_i, T_i)}. \tag{4}$$

175   Here the conventions follow those in (1), where $RMSE(B_i, T_i)$ is the RMSE between the modelled truth T and the background climatology field B (equivalent to the variance of T), and $RMSE(A_i, T_i)$ is the RMSE between the modelled truth and the multivariate analysis A. Each RMSE is calculated for the same historical period used in the creation of the model ensemble, i.e. t is taken from output within the modelled years 1955 to 2014. The maximum value of 1 indicates a perfect reconstruction A, whereas values below 0 indicate that the errors are larger for the analysis than they are if the solution were to be taken as

180   the climatological first-guess.

3   Model correlation fields between pCO$_2$, temperature, salinity, and DIC

The least-squares solution for the weights for temperature, salinity, and atmospheric CO$_2$ is a function of the covariances between the input variables and DIC. Thus, the structure of the covariance fields provides insight into how temperature and salinity can be used to reconstruct carbon. For simplicity, we illustrate these relationships through correlation fields taken for

185   the entire model ensemble, which normalise the relationships using the variances of each input variable.

The correlation fields of DIC, temperature, and salinity with atmospheric pCO$_2$ reflect the impact of emissions on the upper ocean. Atmospheric pCO$_2$ is strongly positively correlated with upper-ocean DIC, with strongest correlations (near 1) in the mid-latitude ocean (Fig. 2a). The correlation between atmospheric pCO$_2$ and upper-ocean integrated temperature is weaker than that with DIC but still positive almost everywhere (Fig. 2b). Lastly, the correlation between atomospheric pCO$_2$ and

190   upper-ocean integrated salinity is spatially heterogeneous and near-zero in most regions of the ocean, although smaller regions such as the Gulf Stream show stronger positive correlations.

Correlations between upper-ocean temperature and DIC are positive in most of the ocean, with strong negative correlations in the eastern equatorial Pacific and Indian Oceans (Fig. 2d). A positive correlation appears to contradict the negative correlations that would be expected from the control of temperature on solubility. However, the residuals for temperature and DIC contain

195   forced and climate variability responses. We can approximate the correlation as the sum of pCO$_2$ and non-pCO$_2$ components to separate the various responses:

$$\rho(DIC', T') \approx \underbrace{\alpha\gamma\frac{\mathrm{var}(pCO_2)}{\sigma_{DIC}\sigma_T}}_{\text{pCO}_2\text{ term}} + \underbrace{\frac{\mathrm{cov}(DIC_a, T_a)}{\sigma_{DIC}\sigma_T}}_{\text{non-pCO}_2\text{ term}}, \tag{5}$$
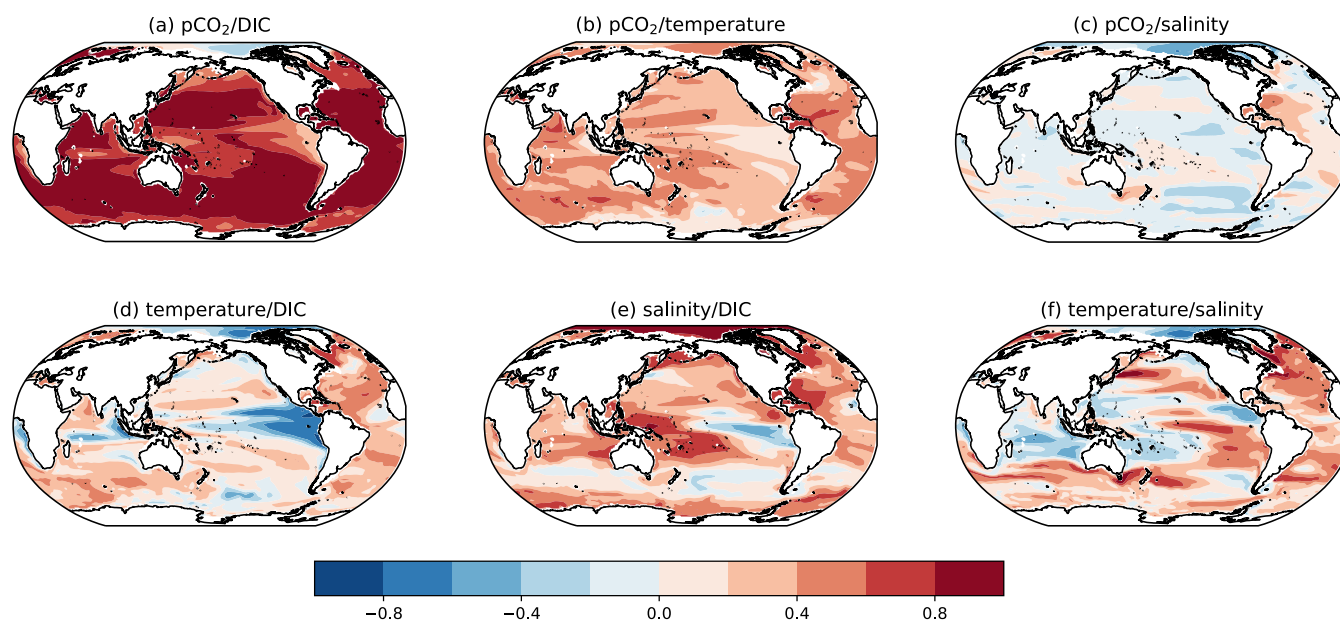
Figure 2. Correlations for the entire CMIP6 historical ensemble between changes in (a) atmospheric pCO_2 and upper-ocean (0-100m integrated) DIC, (b) atmospheric pCO_2 and upper-ocean temperature, (c) atmospheric pCO_2 and upper-ocean salinity, (c) upper-ocean temperature and DIC, (e) upper-ocean salinity and DIC, and (f) upper-ocean temperature and salinity.

where α, γ describe the respective spatial patterns of changes in T', DIC' with changes in pCO_2, T_a, DIC_a are the anomalies of T', DIC' after a regression against pCO_2 is removed, and σ_{DIC}, σ_T are the standard deviations of DIC' and T'. A more detailed derivation of this breakdown and a discussion of the approximation can be found in Appendix A.

The pCO_2 component of the temperature/DIC correlation field is positive almost everywhere, which corresponds with the addition of heat and carbon to the ocean system through CO_2 emissions (Fig. 3a). Conversely, the anomaly term for the temperature/DIC correlation field is negative almost everywhere, which corresponds with the solubility argument for DIC (Fig. 3b). Thus the heterogenity found in the overall correlation between DIC and temperature can be understood as the sum of two separate relationships that arise from the forced and natural variability in temperature and DIC.

The correlations between upper-ocean salinity and upper-ocean DIC are positive in most regions, reflecting the alkalinity control on carbon solubility (Fig. 2e). Correlations are weak and slightly negative in the Southern Ocean and eastern equatorial Pacific, respectively. These correlations may be imprints of dynamical changes; for the Southern Ocean, the salinity in the frontal zone impacts the strength of mode and intermediate water formation (Terhaar et al., 2022), while for the equatorial Pacific enhanced wind-driven upwelling transports fresher, carbon-rich waters to the surface (Williams et al., 2021). Lastly, correlations between upper-ocean temperature and salinity are moderate and show strong regional variability (Fig. 2f). Overall, the correlation fields show first-principle controls relating to the increase of ocean heat and carbon under emissions, as well as solubility and alkalinity controls of temperature and salinity anomalies on ocean DIC.
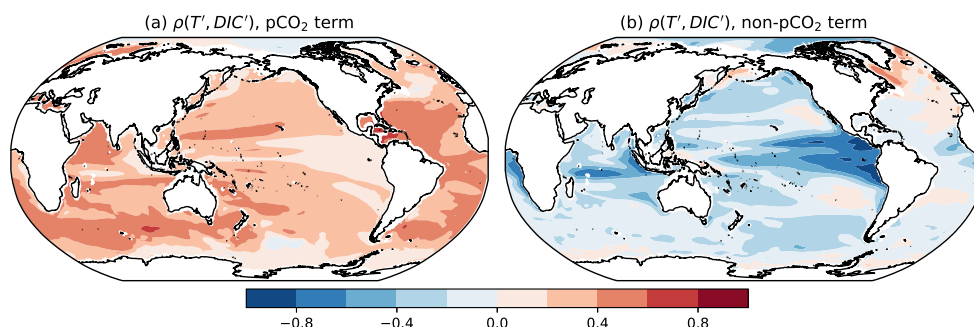
Figure 3. Breakdown of the correlation between upper-ocean DIC and temperature, $\rho(DIC', T')$, from Fig. 2d: (a) a term proportional to the variance of atmospheric pCO_2, and (b) a term consisting of the covariance of the residuals calculated after removing a linear fit against atmospheric pCO_2. For discussion of the breakdown of the correlation into terms and on the approximation, see Appendix A.

3.1 Ensemble optimal weights for pCO_2, temperature, and salinity

215 Translating covariance and correlation fields to optimal mapping parameters involves both the relationships between the input variables and the output variables as well as the relationships between input variables. Therefore, optimal solutions are non-trivial; for instance, regions with similar correlations between DIC and temperature and salinity (such as the equatorial Pacific) may have coefficients with different signs depending on how the system fits the linear model to the data. We thus continue by comparing the least square coefficients fit to atmospheric pCO_2, integrated ocean temperature, and integrated ocean salinity to

220 the correlations in Section 3 and the first-principle controls on carbon solubility.

Increases in atmospheric pCO_2 are translated to increases in upper-ocean carbon (Fig. 4a). The weight magnitudes reach local maxima in the subtropics, in accordance with regional variability in the Revelle buffer factor, which describes the ratio between increases in DIC and increases in atmospheric pCO_2 (Williams and Follows, 2011) and thermocline ventilation. Temperature coefficients are negative almost everywhere (Fig. 4b), while salinity coefficients are positive everywhere (Fig.

225 4c). Thus, while the correlations between upper-ocean DIC, temperature, and salinity show regional variability, the ensemble ultimately reveals a consistent forced control of atmospheric CO_2, solubility control by temperature, and alkalinity control by salinity.

Thus, the least-squares solution for DIC as a combination of atmospheric pCO_2, temperature, and salinity changes is able to capture the controls of the Revelle buffer factor, solubility, and alkalinity on upper-ocean carbon. While the correlations

230 between input variables exhibit strong regional variability and structure in their cross-correlations, ocean temperature and salinity observations end up acting in complementary ways. Thus the optimal weight solutions indicate that the information provided by these observations are consistent with our hypothesised first-order controls.
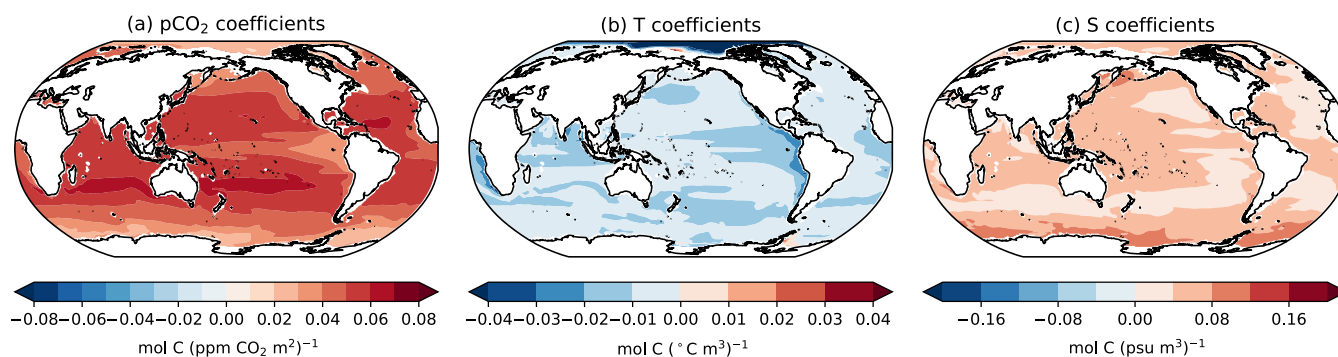
Figure 4. Ensemble coefficients for reconstructing upper-ocean (0-100m integrated) DIC with co-located observations. (a) Coefficients for annual average atmospheric CO_2 concentrations, in units mol C (ppm CO_2 m^2)$^{-1}$, (b) coefficients for upper-ocean integrated temperature, in units mol C ($^{\circ}$C m^3)$^{-1}$, and (c) coefficients for upper-ocean integrated salinity, in units mol C (psu m^3)$^{-1}$.

4 Reconstruction potential using co-located observations

We continue by conducting the sensitivity tests to provide initial bounds for the ability to reconstruct ocean carbon, as well
235 as explore how the makeup of the ensemble impacts the reconstruction. In total, 6 reconstructions were created, in which one model was removed from the ensemble, the covariance fields between DIC, T, S, and atmospheric pCO_2 are recalculated using the remaining models, and then DIC from the excluded model is reconstructed using the new covariance fields and its own T and S observations (Section 2). The coefficients for atmospheric CO_2, temperature, and salinity are similar across the sensitivity setups with each model removed (Appendix B), so we continue by comparing the RMSE improvements ε from (4)
240 for each sensitivity test.

We calculate the ensemble minimum, average, and maximum relative RMSE improvements within the 6 reconstructions created by eliminating specific models from the CMIP6 ensemble. The minimum and maximum error improvements $min(\varepsilon)$ and $max(\varepsilon)$ reflect the makeup of the ensemble. A negative minimum error improvement at any point indicates where one of the models in the ensemble has covariances that are substantially different than the others; therefore, including the model within
245 the ensemble adds uncertainty to the reconstruction and pushes the reconstruction towards the first-guess climatology field. A high maximum error improvement indicates that the solution weights for the sensitivity results are similar. This similarity can arise from strong physical constraints on the upper ocean carbon system but may also arise because the ensemble members are spuriously similar in their architecture or representations of climate processes.

We consider the RMSE improvement of the reconstruction relative to the climatological first-guess as in (4) for both the full
250 upper-ocean carbon response and the detrended carbon response (Fig. 5). Over multiple decades, we expect the full integrated carbon response to be dominated by an upwards trend due to continuing carbon emissions; thus, the improvement in the detrended response provides insight into how interannual and decadal variability in ocean carbon storage is reproduced. Areas where the RMSE increases relative to the climatological first guess are noted in red. For this analysis we focus on the open-

ocean RMSE reductions, as the CMIP6 models have different coastlines after being regridded, leading to small areas of RMSE
255 increases near the land/sea boundary.

Outside of some coastal and Arctic regions, each sensitivity test reduces the upper-ocean RMSE for upper-ocean carbon
(Fig. 5a). This improvement is not due solely to the reproduction of the long-term increase in ocean carbon, as most regions
show improvements in detrended ocean carbon as well; exceptions to this improvement can be found in the high latitudes
and subtropical regions (Fig. 5d). The average reconstruction reduces the RMSE by between 60% to 90% on average, relative
260 to the climatology first-guess RMSE (Fig. 5b) and reduces the detrended RMSE by 30% to 80% (Fig. 5e). These relative
improvements are equivalent to the method capturing over 60% of the detrended variability in most regions of the ocean. The
high (close to 1) relative RMSE reductions found in the ensemble maximum statistics (Figs. 5c and e) suggest that there are
models which are substantially similar to one another and thus able to reduce the RMSE of the full signal by over 80% in most
regions and the RMSE of the detrended signal by over 60%.

265 There are noticeable regional variations throughout all of the improvement statistics. The western low-latitude Pacific and
subtropical Indian ocean show consistent local maxima in all of the improvement statistics. These regions of maximum relative
RMSE reduction are characterised by their strong correlation between salinity and DIC, suggesting that the most constrained
responses within the CMIP6 ensemble may be related to the control of alkalinity on DIC. Conversely, the small regions that
show potential degradation in the reconstructions due to the errors in the covariances are characterised by weak correlations
270 against atmospheric pCO_2 changes (for the full carbon signal reconstruction, Fig. 2a-c) and weakly positive covariances be-
tween T_a and DIC_a (for the detrended carbon system, Fig. 3b).

Thus, when considering pointwise observations of temperature and salinity, alongside global average CO_2 concentrations,
our sensitivity experiments indicate that a substantial amount of upper-ocean carbon variability can be reconstructed. These es-
timates can provide an upper bound on the reconstruction potential as global coverage of temperature and salinity observations
275 is theoretically ideal but difficult to accomplish, even with widespread autonomous observing tools. We therefore continue by
exploring how irregular observations of temperature and salinity impact reconstructions of upper-ocean carbon.

5 Reconstructing carbon using irregular Argo observations

The pointwise reconstructions of upper-ocean carbon from atmospheric pCO_2, temperature, and salinity reveal that the phys-
ical controls of solubility and alkalinity can explain a substantial amount of interannual variability in the upper-ocean carbon
280 system. Although temperature and salinity profiles have a larger global coverage than ocean carbon profiles, coverage remains
irregular and incomplete. Thus, we continue by exploring how the irregular coverage in Argo profiles impacts carbon recon-
structions by conducting further synthetic reconstructions with observations consistent with year 2002 and year 2015 Argo
coverage (Fig. 1).

When considering irregular sampling, reconstructions with global coverage use background covariances to propagate infor-
285 mation from Argo profile locations to the rest of the ocean. There is the potential for information from multiple Argo sites to be
used to reconstruct ocean carbon at a given location. If the covariance fields are correct, additional information should improve
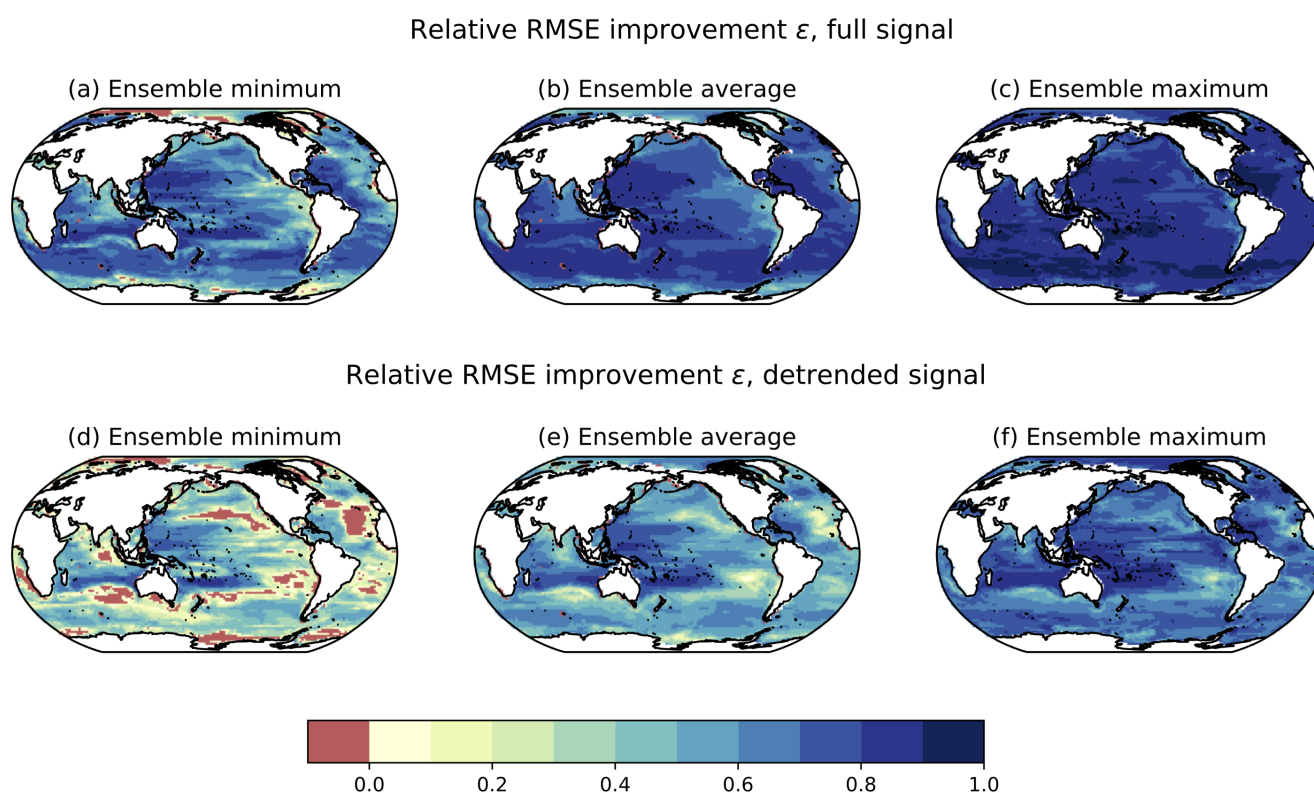
Figure 5. Ensemble statistics for the relative RMSE reductions ε using co-located temperature and salinity: (a) ensemble minimum, (b) ensemble average, and (c) ensemble maximum. Ensemble statistics for the relative RMSE reductions, but considering only the detrended carbon signal: (d) ensemble minimum, (e) ensemble average, and (f) ensemble maximum. Red areas indicate regions where the sensitivity tests show a RMSE increase relative to the assumption of climatology.

the reconstruction and increase the relative error reduction; however, any errors in the covariance fields can propagate and decrease the relative error reduction. Thus, the radius of influence for an observation should ideally balance the extra information and errors it provides in the reconstruction.

290   With the smaller radii of 1° and 2°, the method creates an incomplete reconstruction even under current-day Argo profile distributions. With a larger radius of 5°, most of the global ocean has nearby temperature and salinity profiles that can be used for the reconstruction (see Fig. 6 coverage). Overall, the benefit of increased coverage outweighs the increases in reconstruction errors (Appendix C). Thus, in this section we focus on the 5° radius results for simplicity .

With the 5° search for observations, the carbon reconstruction at each point will use between 0 and 81 temperature and
295   salinity inputs. Including the globally averaged atmospheric CO_2 inputs, DIC at each point can be reconstructed from at least 1 and up to 82 input variables (up to 81 temperature and salinity profiles + globally averaged atmospheric pCO_2).

Relative RMSE improvement ε, full signal

Relative RMSE improvement ε, detrended signal
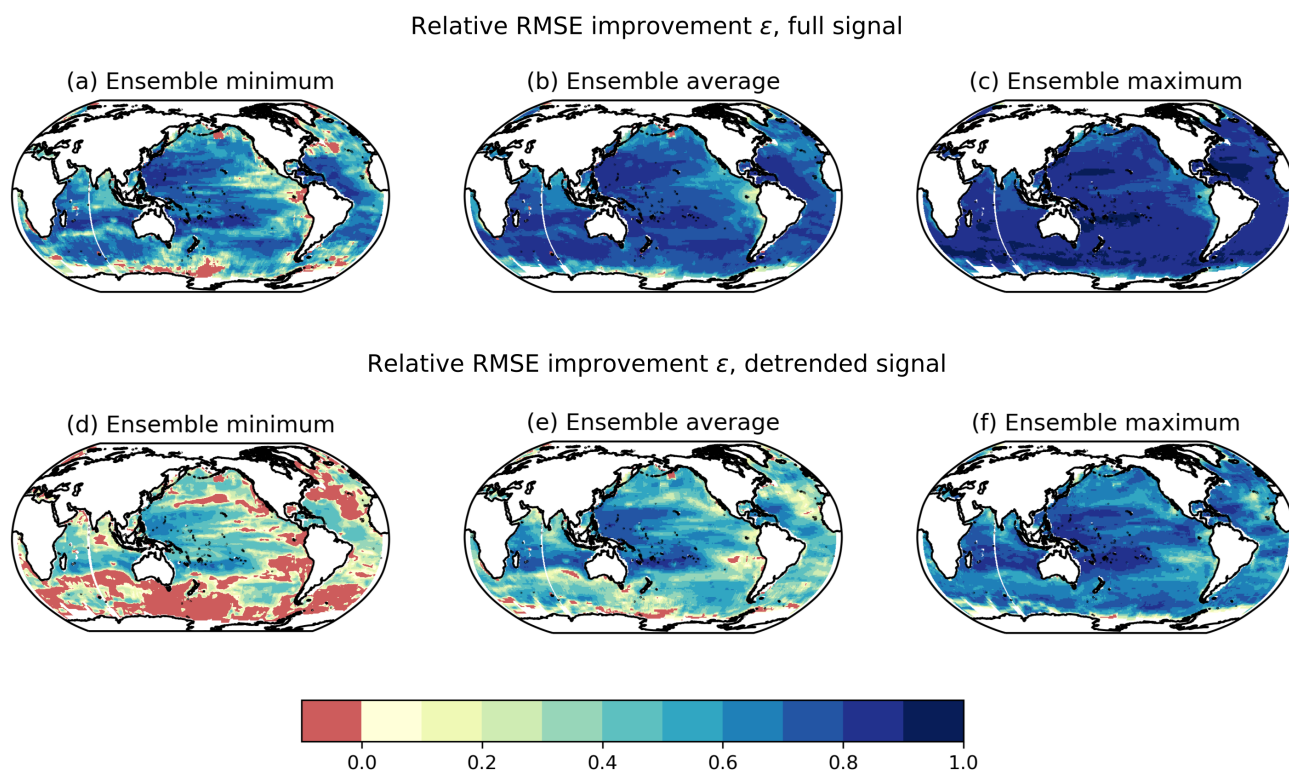
Figure 6. As for Fig. 5, but for the reconstruction scheme that uses Argo 2015-type temperature and salinity observations within $5°$ of the reconstruction point.

Errors in spatial covariances result in some regions with lower RMSE improvements compared to reconstructions using co-located observations (Fig. 6a, d). Across the sensitivity tests, most regions show an relative RMSE reduction of at least 50%, and when considering the detrended carbon signal most regions show an RMSE reduction of at least 40%. Notably,

300   the ensemble minimum RMSE improvement for the detrend signal shows substantial regions where one of the sensitivity tests produces poor reconstructions. This sensitivity test is that for MPI-ESM1.2-LR, which exhibits more fine-scale structure in its annual DIC, temperature, and salinity. The errors arise because the correlation length scales for MPI-ESM1.2-LR are smaller than those for the other models, resulting in poor fits.

For most of the upper ocean the RMSE improvements are similar to those found in the reconstruction with co-located

305   temperature and salinity measurements, and the regional variation in the RMSE improvements is similar regardless of whether co-located or removed observations are used. Thus the additional information provided by nearby measurements counteracts the error propagation inherent in the scheme, suggesting that the use of irregular Argo profiles provides sufficient information to reproduce a substantial amount of upper-ocean carbon variability.
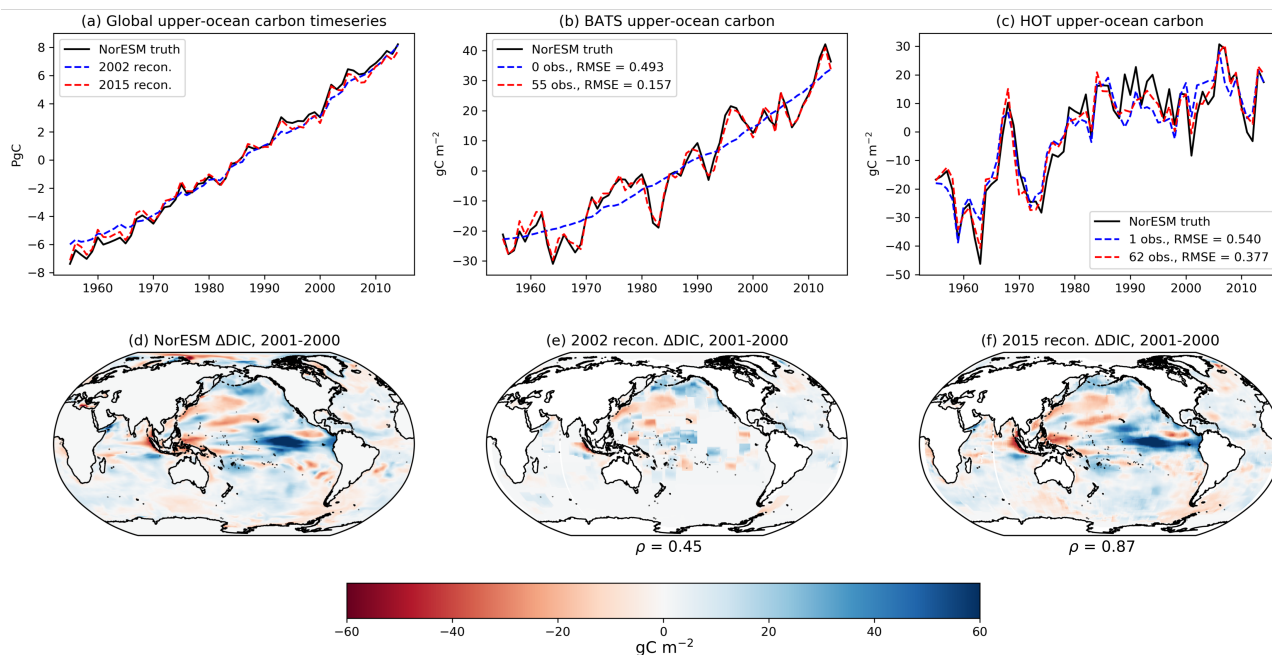
Figure 7. Reconstructions of NorESM upper-ocean carbon (black lines) using year 2002 Argo profile distributions (blue dashed lines) and year 2015 Argo profile distributions (red dashed lines): (a) globally integrated carbon, in units PgC, (b) Integrated carbon at BATS, in units gC, and (c) Integrated carbon at HOT, in units gC. Integrated carbon is calculated as a residual from the 1955-2014 average. Changes in upper-ocean carbon in the NorESM, in units gC m^{-2}, between years 2001 and 2000: (d) modelled truth, (e) reconstructed changes using year 2002 Argo profile distributions, and (f) reconstructed changes using year 2015 Argo profile distributions. Correlations below (e) and (f) are spatial correlations calculated against the model truth (d).

6   Potential of the method to provide global and regional carbon timeseries

310   To illustrate the potential for reconstructing upper-ocean carbon using temperature and salinity measurements, we reconstruct output from the Norwegian Earth System Model (NorESM2, Seland et al. (2020)) using the full ensemble covariance fields and temperature and salinity measurements at locations similar to Argo observation locations at year 2002 and year 2015. NorESM2 was chosen as an independent check as its depth coordinates follow isopycnals rather than distance from the sea surface like z-level models; as all the models in the CMIP6 ensemble are z-level models, this choice allows us to begin

315   to consider some potential errors from physical uncertainties and the restrictions of model assumptions. The global upper ocean carbon inventory is dominated by a positive, near-linear trend (Fig. 7a, black line). Both Argo-style reconstructions reproduce this trend; however, due to having fewer observations, the year 2002 reconstruction (dashed blue line) reproduces less interannual variability than the year 2015 reconstruction (dashed red line).

To show how the reconstructions operate on more regional scales, we provide the same reconstructions but for the modelled

320   carbon inventory at the Bermuda Atlantic Time Series (BATS, 31°50'N 64°10'W) and the Hawaii Ocean Time series (HOT,

22°45′N 158°00′W). These locations were chosen because they are the sites of multidecadal ocean DIC time series; thus, for future reconstructions the observations at these sites will can be used for validation. By providing reconstructions at these locations we can estimate the upper-bound in how well the method works, ignoring any errors in the observations themselves. Modelled DIC at these two locations show varying levels of CO_2-driven increases and interannual variability to provide context
325   as to how the method reconstructs both aspects.

At BATS, the signal is again driven by a long-term positive trend (Fig. 7b, black line). As the year 2002-style Argo distribution has no observations within 5°, the reconstruction uses atmospheric CO_2 measurements to reproduce the trend. For the reconstruction using a year 2015-style Argo distribution, the increase in observed locations results in a reconstruction that is able to reproduce the variability in the carbon inventory. The reconstruction using the year 2015 observation locations for
330   temperature and salinity inputs results in an RMSE that is 68% lower than that for the reconstruction that is based off of atmospheric pCO_2 changes alone.

At HOT, upper-ocean carbon exhibits more defined interannual and decadal variability (Fig. 7c, black line). The reconstruction using year 2002 Argo locations (blue dashed line) is able to reproduce the broad behaviour in ocean carbon, whereas the reconstruction using year 2015 Argo locations (red dashed line) shows an improvement in representation of the interannual
335   variability. By using multiple observations at year 2015 locations, the RMSE is reduced by 30% relative to the RMSE for the reconstruction using one observation. Thus the results at HOT provide further evidence that the inclusion of nearby Argo measurements provides skill in the reconstruction scheme, and emphasises the improvement that is already possible when using one nearby location for temperature and salinity observations.

Lastly, for a global view of the reconstruction we provide maps of the Norwegian ESM DIC change between years 2001
340   and 2000 (Fig. 7d) with the reconstructed difference using year 2002 Argo locations (Fig. 7e) and year 2015 Argo locations (Fig. 7f). The truth field for DIC changes show regional variability, with increases in the western North Pacific and strong decreases in the equatorial upwelling zone. The reconstructed change using 2002 Argo locations is restricted by the number of observations: most regions show little change between the two years. The reconstruction has a spatial correlation of 0.45 using the relatively sparse temperature and salinity sampling procedure. For the reconstruction using 2015 Argo locations, most of
345   DIC changes are reconstructed, and the reconstruction has a spatial correlation of 0.87 compared to the model truth.

7   Discussion and conclusions

In this study we present a new method for reconstructing upper-ocean carbon using observations of ocean temperatures and salinities. While the ocean plays a large role in determining the partitioning of carbon in the Earth system, sparse observations inhibit a full characterisation of ocean interior carbon. Using synthetic profiles and creating mapped reconstructions of
350   model truths, we have explored how wide-spread observations of temperature and salinity from recent autonomous sampling programmes can provide global information about ocean carbon.

When considering the carbon system in the near-surface ocean, the correlation fields between atmospheric pCO_2, ocean temperature, ocean salinity, and ocean carbon reflect first-order controls (Fig. 2). Increases in atmospheric pCO_2 are correlated

with increases in ocean heat and carbon due to the chemical and thermal disequilibria created by emissions. Increases in ocean

355   salinity are broadly correlated with increases in ocean carbon due to the impact of alkalinity on solubility. Regional variations

in the correlations between temperature and ocean carbon can be decomposed as a sum of a response related to added carbon

from emissions and a residual; the residual correlation reflects the impact of temperature on CO_2 solubility in seawater. Within

all these correlation fields there are regional variations. The structure in the correlations between temperature and DIC and

salinity and DIC attain their strongest values in different regions, and the cross-correlations between temperature and DIC are

360   generally more moderate, indicating that the information provided by these observations are complementary.

The strength and structure in the covariance fields results in optimal weights for pCO_2, temperature, and salinity that reflect

global first-order controls of solubility and alkalinity (Fig. 4). For this linear model, using co-located temperature and salinity

observations along with globally averaged atmospheric pCO_2 concentrations reduces the RMSE of the reconstructed DIC

content by over 75% on average, relative to a reconstruction based purely on a time-invariant climatology (Fig. 5). In addition

365   to replicating the long-term trend, the reconstructions are able to reproduce local interannual and decadal variability, capturing

over 60% of the detrended upper-ocean carbon signal. When considering the detrended carbon inventory, the sensitivity tests

show an average relative RMSE reduction of 50% in most regions. The lower skill in reconstructing interannual and decadal

variability is expected as the carbon system is nonlinear and controlled by other aspects such as circulation changes and ocean

biology, which are not explicitly included in the variables used to calculate the covariance fields.

370   We have also explored whether the use of irregular temperature and salinity observations impacts the capacity to reconstruct

ocean carbon. In theory, observations do not need to be co-located with the desired reconstruction, as the Ensemble Optimal

Interpolation scheme can include spatial covariances; however, there is the possibility of extra errors in the covariance fields

from poorly represented spatial variability. In our sensitivity tests, we find that using irregular observations consistent with cur-

rent Argo coverage can be combined with the CMIP6 covariance fields and provide similar levels of skill to the reconstruction.

375   When taking the average RMSE reduction across our sensitivity tests, we find a reduction of over 60% in most regions when

considering the full carbon signal and a reduction of over 40% in most regions when considering the detrended carbon signal

(Fig. 6b,d). However, we find that there are regions such as the Southern Ocean and the North Atlantic which show substantial

error increases relative to the climatological first guess in some of the sensitivity experiments (Fig. 6a,c). These increases in

errors are found particularly when considering the detrended signal (i.e., the signal from interannual and decadal variability).

380   The regions that show the largest error increases are those with high mesoscale ocean eddy activity, and are mainly in the

sensitivity test that reconstructs ocean carbon within MPI-ESM1.2-LR. The horizontal resolution in MPI-ESM1.2-LR model is

higher than that for the other models in our ensemble, suggesting that the covariances' representation of mesoscale-scale ocean

dynamics impacts the ability to translate nearby Argo observations into reconstructed carbon. Our sensitivity tests indicate that

the correlation lengthscales will be an important feature of any real-world carbon reconstruction. For future work, there are a

385   variety of methods regarding both the ensemble selection and search radius selection that may be used to account for the model

differences and model/truth discrepancy in spatial correlation lengthscales.

While we have opted to use only temperature and salinity as our ocean observations for this scheme, the ensemble optimal

interpolation method is flexible and can take additional oceanographic variables such as oxygen, pH, nutrients, or chlorophyll.

With increased coverage of these variables thanks to campaigns such as the Southern Ocean Carbon and Climate Observations
390   and Modeling programme (SOCCOM, Johnson et al. 2017) and Bio-Argo (Claustre et al., 2020), there is the potential for
non-trivial improvements by including these variables. With each new variable it is possible to quantify the amount of added
information and test the ensemble optimal interpolation method at various stages by examining the covariance fields and
uncertainties, and conducting sensitivity tests in a similar way as has been done in this work. As more complex reconstruction
schemes such as the machine learning algorithms of Landschützer et al. (2016) use these biogeochemical observations as well,
395   including them in the Ensemble Optimal Interpolation scheme could enable a direct comparison between linear and non-linear
mapping methods and can help to quantify the merits of linear versus non-linear assumptions when reconstructing ocean
interior carbon.

The Ensemble Optimal Interpolation scheme relies on some important assumptions. Firstly, the covariance fields assume that
the processes relating ocean carbon to other variables are stationary. We have used historical CMIP6 experiments to calculate
400   the covariances fields, so these covariances should be able to represent ocean carbon behaviour under current-day carbon
emission forcing; however, under low or negative emissions the relationships between atmospheric pCO_2, ocean temperature,
and ocean carbon changes will likely change due to hysteresis from continued heat uptake. Additionally, we have chosen to
focus on annual average and depth-integrated carbon content in order to focus on the physical carbon response. It is also
possible to consider seasonal or monthly variability by calculating covariance fields for each month or season, as is done for
405   surface pCO_2 and heat content (Smith and Murphy, 2007; Jones et al., 2015). In the current format, months or seasons that
are poorly observed will have solutions that return towards climatological inventories. However, the system can be set up in
such a way that lagged covariance fields propagate temperature and salinity information from observed to unobserved months
as well as locations. The lack of wintertime carbon observations has created significant biases in machine learning products,
as enhanced winter outgassing in polar waters is not present in the training data (Bushinsky et al., 2019). As the response to
410   gaps in data are sensitive to the reconstruction method, the Ensemble Optimal Interpolation reconstructions with seasonal data
could provide added insight as to how information on the carbon system is best propagated to unobserved regions and seasons.

The use of a model ensemble rather than a single model allows for quantification of some of the uncertainties within the
covariance fields. Our model ensemble has been constructed to consider both inter-model and intra-model uncertainty when
calculating the background error covariances. We have been able to explore the sensitivity of this method to the ensemble
415   makeup by reconstructing each model separately using out of sample covariances from the other models, and assessing the
minimum, mean and maximum improvements. The improvements in the reconstructions indicate both that the models have
regions with well-defined and strong correlations that lead to high improvements, such as the western tropical Pacific, as well as
regions where the models are relatively uncertain in their relationships, such as the Southern Ocean when considering irregular
spatial coverage (Figs. 5, 6). While these sensitivity experiments and synthetic reconstructions allow for useful insight into the
420   capacity of our reconstruction method, some level of uncertainty remains from the use of coarse-resolution CMIP6 models.
Ocean models underestimate decadal variability in anthropogenic carbon (DeVries et al., 2019) and CMIP6 models exhibit
biases in the Revelle factor (Terhaar et al., 2022). More iterative Ensemble Optimal Interpolation techniques can be used to

reduce the errors from the model ensemble by recalculating the covariance fields after observations have been included in the analysis (Smith and Murphy, 2007; Smith et al., 2015; Allison et al., 2019).

425   As these synthetic tests have suggested that a large amount of ocean carbon variability can be reconstructed using only hydrographic observations, future work will include expanding the method to account for observational errors and applying the method to real-world Argo profile data. The reconstructions from Argo data can be cross-validated against independent time series sites and GLODAP data, and we will explore the added information from non-DIC biogeochemical observations. We will expand the method to consider how the covariances behave on various depth levels: as Argo profiles measure temperature

430   and salinity down to 2000m depth, the vast majority of carbon changes can potentially be captured by the method, although for deeper reconstructions there are few independent carbon observations for cross-validation.

    In conclusion, we have presented a method that draws upon first-principle relationships between temperature, salinity, and carbon to reconstruct ocean carbon changes from Argo style observations. We find that, for the upper ocean, the CMIP6 models show consistent first principle constraints of solubility, alkalinity, and undersaturation, leading to accurate reproductions of

435   trends and variability in the model world. When considering current-day distributions of Argo profiles, reconstructions that use multiple nearby observations are able to still reconstruct the trend and variability in carbon content, and sensitivity tests using out-of-sample temperature and salinity observations show strong spatial correlations of around 0.8 with the model truth and capture over 60% of the detrended variance. The results of our work provide a strong theoretical basis for using state-of-the-art autonomous hydrographic observations to supplement sparse interior carbon coverage. This proof-of-concept work shows

440   that there is strong potential in using these measurements to create a new, independent estimate that can be used alongside reconstruction methods to better understand ocean carbon variability and its controls.

Code availability. The code to regrid, concatenate, and analyse CMIP6 model output is available in a Zenodo reposity (DOI: 10.5281/zen-odo.6669897).

Appendix A: Breakdown of covariance fields into pCO$_2$ and non-pCO$_2$ terms

445   It is possible to decompose the changes in ocean temperature, salinity, and DIC from their time-averaged states (T', S', DIC' from Section 2) into terms proportional to atmospheric pCO$_2$ and a residual:

$$T'(x,y,t) = \alpha(x,y)\,\mathrm{pCO}_2'(t) + T_a(x,y,t) \tag{A1}$$

$$S'(x,y,t) = \beta(x,y)\,\mathrm{pCO}_2'(t) + S_a(x,y,t) \tag{A2}$$

$$DIC'(x,y,t) = \gamma(x,y)\,\mathrm{pCO}_2'(t) + DIC_a(x,y,t), \tag{A3}$$

450   where α, β, γ are the least-squares coefficients against changes in globally uniform atmospheric pCO$_2$ changes and T_a, S_a, DIC_a indicate the residuals for these decompositions.

   While the decomposition into pCO$_2$ and non-pCO$_2$ terms is not inherently physically based, this breakdown may help us understand how the complex structure in the correlation fields in Fig. 2 arise. Using covariance identities, the covariance between, for instance, T' and S', is thus the sum of the covariances between the addends:

455    $\mathrm{cov}(T',S') = \mathrm{cov}(\alpha\,\mathrm{pCO}_2(t), \beta\,\mathrm{pCO}_2(t)) + \mathrm{cov}(\alpha\,\mathrm{pCO}_2(t), S'') + \mathrm{cov}(T'', \beta\,\mathrm{pCO}_2(t)) + \mathrm{cov}(T'', S''). \tag{A4}$

   The first term is a function of the variance of pCO$_2$, scaled by the product $\alpha\beta$. If the term proportional to atmospheric pCO$_2$ were orthogonal to the residual term for each of these decompositions, the covariances $cov(pCO_2, X_a) = 0$ for any variable X. An ordinary least squares solution does not lead to orthogonal components, but the covariances between atmospheric pCO$_2$ and DIC_a, T_a, or S_a are small (shown as correlations in Fig. A1, panels d h and i) and can therefore be ignored to first order.
460   Thus for these variables we ignore these terms as second-order components and can thus approximate the covariance as:

$$\mathrm{cov}(T',S') \approx \alpha\beta\,\mathrm{var}(\mathrm{pCO}_2) + \mathrm{cov}(T_a, S_a). \tag{A5}$$

   With this decomposition of the covariance fields into pCO$_2$ and non-pCO$_2$ terms, we have the added benefit of approximating the covariance fields between pCO$_2$ and one of T', S', or DIC' through a global scaling of the coefficient terms α, β, or γ, respectively.

465   Scaling the decomposition by the product of the standard deviations allows us to explore how the correlation fields arise in Fig. 2. We compare the original correlation fields from Fig. 2 with their decompositions according to (A4) in Fig. A1. The approximation errors show how well the approximation in (A5) reflects the full correlation field.

   For the correlations between T' and S' and S' and DIC', the signal is dominated by the term associated with $cov(T_a, S_a)$ and $cov(S_a, DIC_a)$, respectively. We therefore understand these covariance and correlation fields to be set by the variability
470   not forced by carbon emissions. Conversely, the correlation field between T' and DIC' appears to arise almost equally from the pCO$_2$ scaling and the $cov(T_a, DIC_a)$ terms. Thus we provide both terms in Section 3 to gain insight as to how the covariance fields reflect both the combined response to added carbon in the Earth system as well as the response to climate variability.
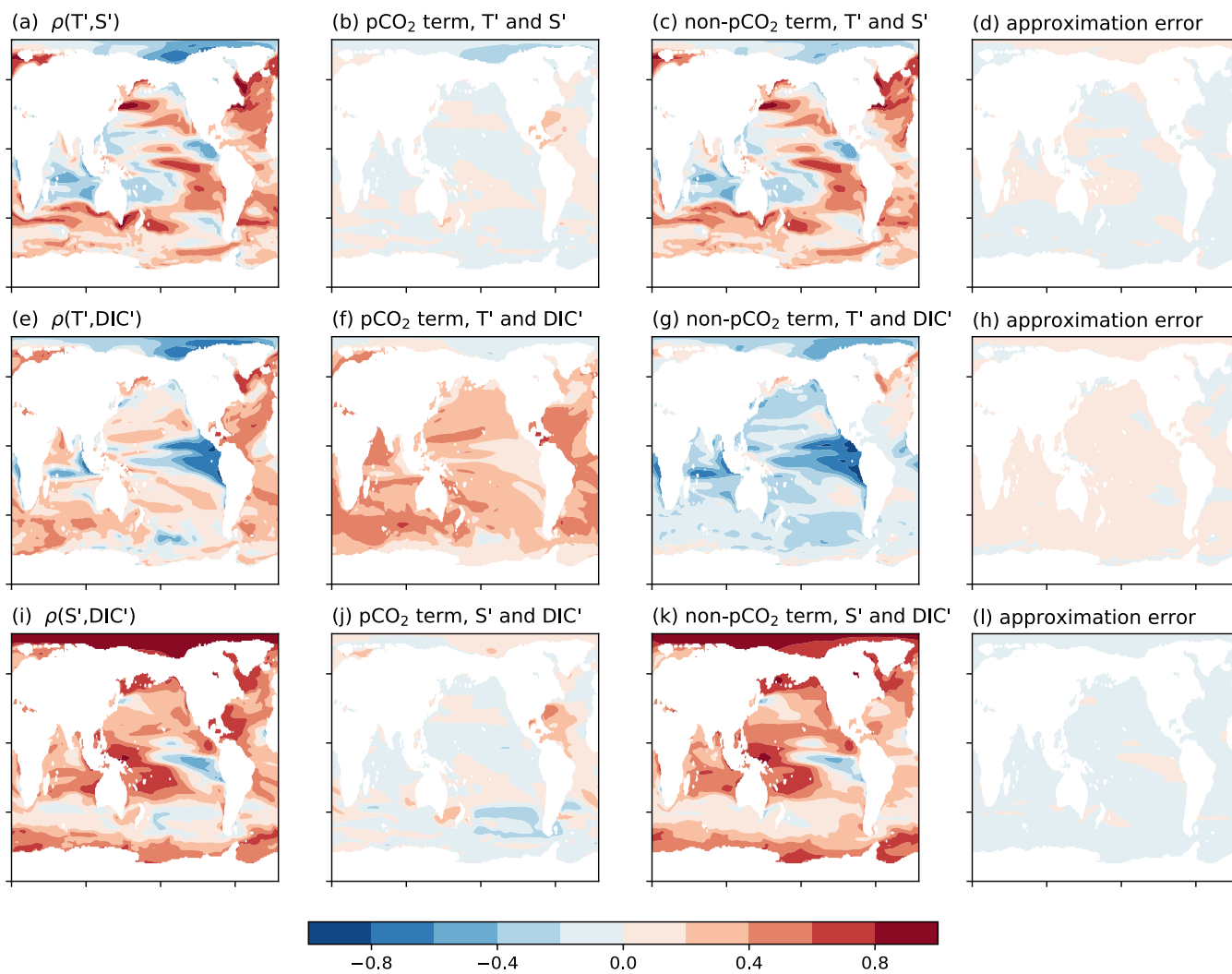
Figure A1. Correlation fields between non-pCO$_2$ variables $T', S',$ and DIC', with their terms from (A5) and approximation errors to measure the level of mismatch between (A4) and (A5).

Appendix B: Coefficients for sensitivity tests

Here we present the coefficients for reconstructing upper-ocean carbon using atmospheric CO_2 and co-located temperature and
475   salinity measurements.
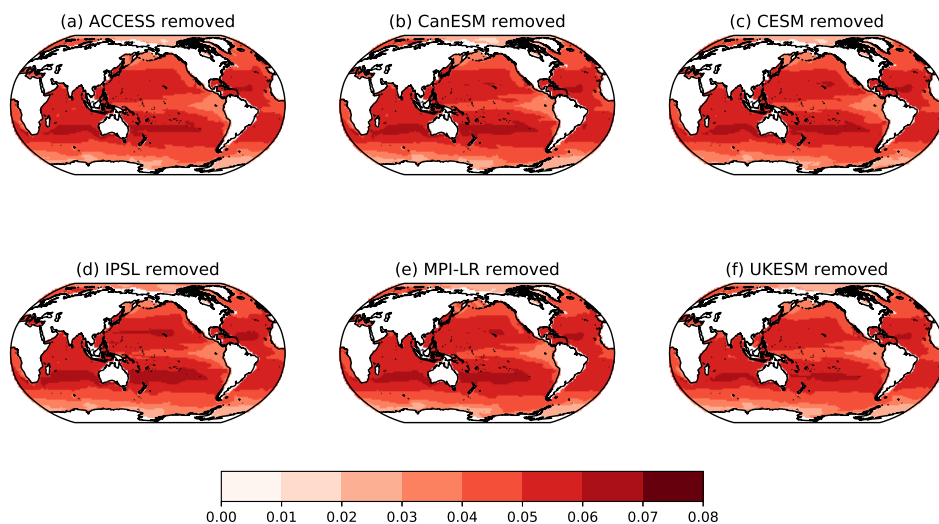
Figure B1. Optimal pCO_2 weights, in units mol C $(ppm\ CO_2\ m^2)^{-1}$, for solving for 0-100m integrated ocean DIC using a subset of models within the CMIP6 ensemble. The subsets are created by removing the runs of each model: (a) ACCESS, (b) CanESM, (c) CESM, (d) IPSL, (e) MPI-LR, and (f) UKESM
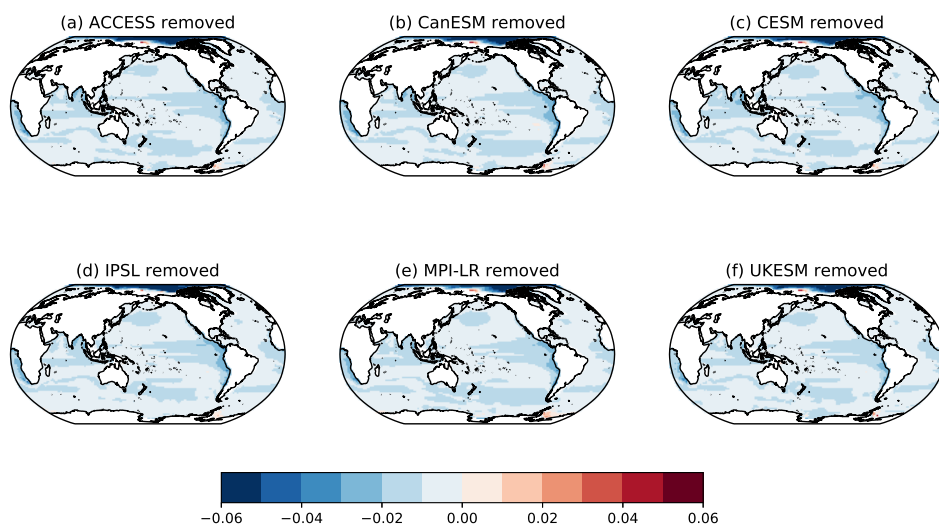
Figure B2. Optimal T weights, in units mol C $(^{\circ}\text{C m}^3)^{-1}$, for solving for 0-100m integrated ocean DIC using a subset of models within the CMIP6 ensemble. The subsets are created by removing the runs of each model: (a) ACCESS, (b) CanESM, (c) CESM, (d) IPSL, (e) MPI-LR, and (f) UKESM
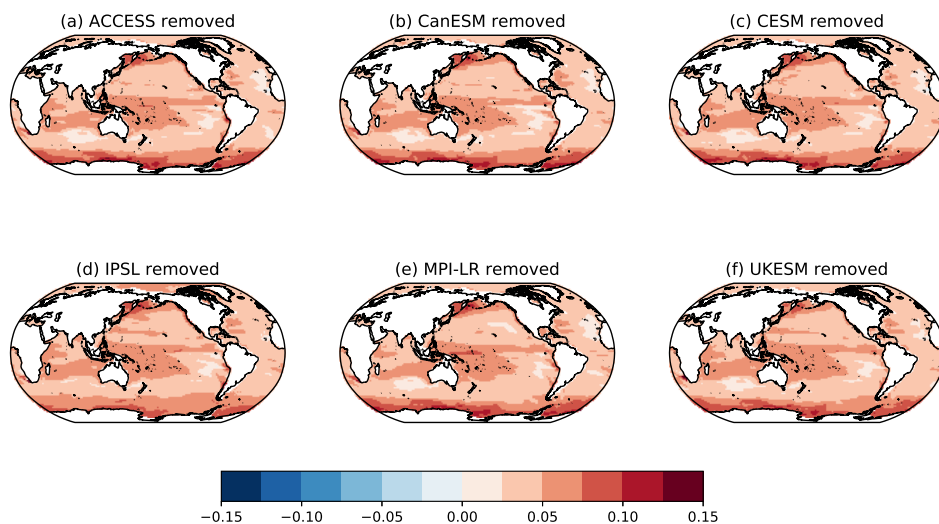
Figure B3. Optimal S weights, in units mol C $(\text{psu m}^3)^{-1}$, for solving for 0-100m integrated ocean DIC using a subset of models within the CMIP6 ensemble. The subsets are created by removing the runs of each model: (a) ACCESS, (b) CanESM, (c) CESM, (d) IPSL, (e) MPI-LR, and (f) UKESM

Appendix C: Comparison of RMSE reductions with Argo profiles and various search radii

Here we present some reconstruction RMSE reductions for the Norwegian ESM model using the CMIP6 covariance fields, Argo profiles consistent with year 2015 sampling, and various search radii. We use the Norwegian ESM as it uses isopycnic depth coordinates as opposed to the z-level coordinates used in the models within the ensemble. We find that the coverage in 480 the reconstructions increases to near-global coverage when allowing profiles within 5° to be used in the reconstruction and improves the detrended carbon reproduction; however, for regions that are reproduced with high skill using more restrictive search radii, the increase in search radius is accompanied by slight weakenings in the RMSE reductions.
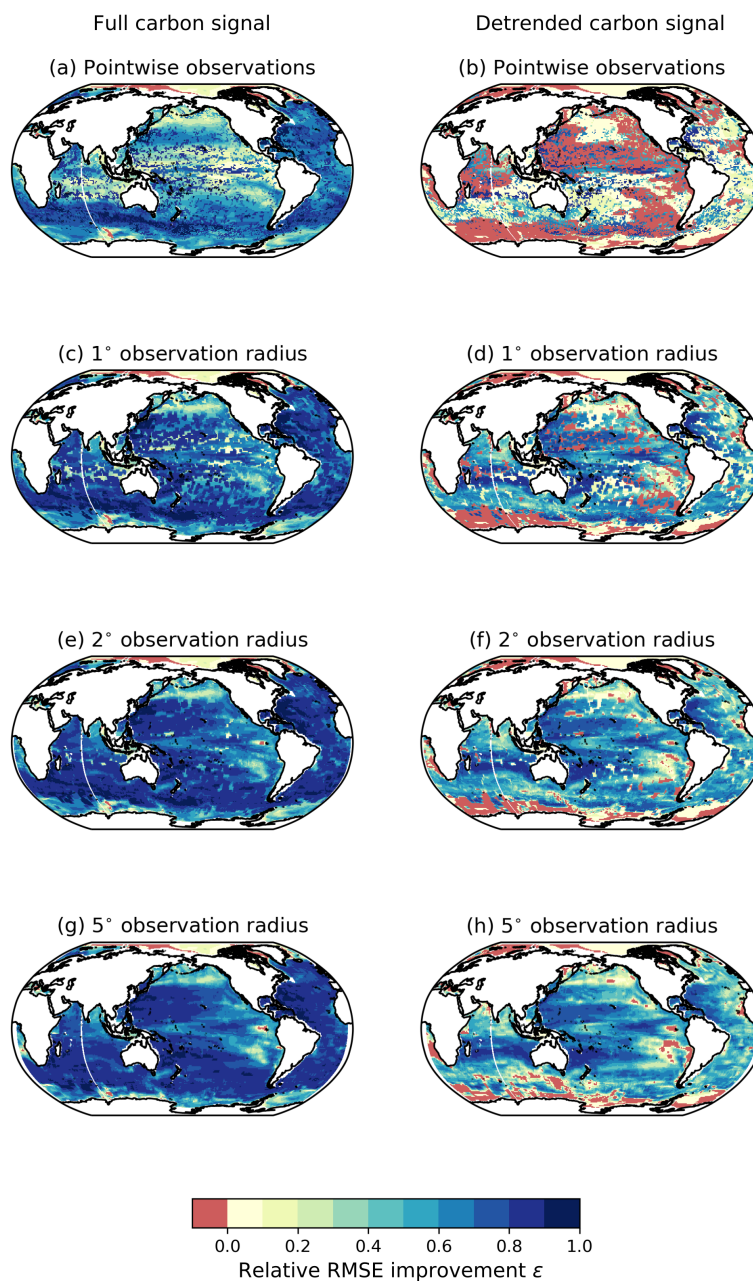
Figure C1. Relative RMSE reduction for the NorESM full carbon (left column) and detrended carbon (right column) signals using year 2015 Argo distributions and varying radii of influence: (a,b) using only co-located temperature and salinity profiles, (c,d) using profiles within $1°$ latitude or longitude, (e,f) using profiles within $2°$ latitude or longitude, and (g,h) using profiles within $5°$ latitude or longitude. Regions of red indicate regions where the reconstruction increases the RMSE relative to a first-order climatology assumption.

Author contributions. All authors contributed towards the conceptualisation of the work. KET created the method and analysis with input from RGW, AK, and DMS. KT wrote the manuscript draft, and all authors reviewed and edited the manuscript. All authors signed off on the manuscript for submission.

485

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. KET received support from the Leverhulme Trust via the Leverhulme Research Centre for Functional Materials Design. Authors KET, RGW, and AK received research support from the U.K. Natural Environmental Research Council (Grant number NE/T007788/1). DMS was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra.

References

Allison, L. C., Roberts, C. D., Palmer, M. D., Hermanson, L., Killick, R. E., Rayner, N. A., Smith, D. M., and Andrews, M. B.: Towards quantifying uncertainty in ocean heat content changes using synthetic profiles, Environmental Research Letters, 14, 084 037, https://doi.org/10.1088/1748-9326/ab2b0b, 2019.

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.

Bates, N. R., Astor, Y. M., Church, M. J., Currie, K., Dore, J. E., González-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santana-Casiano, J. M.: A Time-Series View of Changing Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO$_2$ and Ocean Acidification, Oceanography, 27, https://doi.org/10.5670/oceanog.2014.16, 2014.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J., Guenet, B., Guez, E., L., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, Journal of Advances in Modeling Earth Systems, 12, https://doi.org/10.1029/2019MS002010, 2020.

Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., Resplandy, L., Johnson, K. S., and Sarmiento, J. L.: Reassessing Southern Ocean Air-Sea CO$_2$ Flux Estimates With the Addition of Biogeochemical Float Observations, Global Biogeochemical Cycles, 33, 1370–1388, https://doi.org/https://doi.org/10.1029/2019GB006176, 2019.

Carroll, D., Menemenlis, D., Adkins, J. F., Bowman, K. W., Brix, H., Dutkiewicz, S., Fenty, I., Gierach, M. M., Hill, C., Jahn, O., Landschützer, P., Lauderdale, J. M., Liu, J., Manizza, M., Naviaux, J. D., Rödenbeck, C., Schimel, D. S., Van der Stocken, T., and Zhang, H.: The ECCO-Darwin Data-Assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multidecadal Surface Ocean pCO$_2$ and Air-Sea CO$_2$ Flux, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 888, https://doi.org/https://doi.org/10.1029/2019MS001888, 2020.

Cheng, L. and Zhu, J.: Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variations, Journal of Climate, 29, 5393–5416, 2016.

Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the global ocean with biogeochemical-Argo, Annual review of marine science, pp. 23–48, 2020.

Daley, R.: Atmospheric data analysis, Cambridge University Press, Cambridge, 1991.

Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S.,

Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, https://doi.org/10.1029/2019MS001916, 2020.

DeVries, T., Le Quéré, C., Andrews, O., Berthet, S., Hauck, J., Ilyina, T., Landschützer, P., Lenton, A., Lima, I. D., Nowicki, M., et al.: Decadal trends in the ocean carbon sink, Proceedings of the National Academy of Sciences, 116, 11 646–11 651, 2019.

530 Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang, L. M., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, O., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Luijkx, I. T., Jain, A., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, 535 A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G. R., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J.: Global Carbon Budget 2021, Earth System Science Data, 14, 1917–2005, https://doi.org/10.5194/essd-14-1917-540 2022, 2022.

Gloege, L., Yan, M., Zheng, T., and McKinley, G. A.: Improved Quantification of Ocean Carbon Uptake by Using Machine Learning to Merge Global Models and pCO$_2$ Data, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 620, https://doi.org/https://doi.org/10.1029/2021MS002620, e2021MS002620 2021MS002620, 2022.

Gruber, N., Landschützer, P., and Lovenduski, N. S.: The variable Southern Ocean carbon sink, Annual review of marine science, 11, 2019.

545 Gruber, N., Boyd, P. W., Frölicher, T. L., and Vogt, M.: Biogeochemical extremes and compound events in the ocean, Nature, 600, 395–407, 2021.

Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., Swift, D. D., Williams, N. L., Boss, E., Haëntjens, N., et al.: Biogeochemical sensor performance in the SOCCOM profiling float array, Journal of Geophysical Research: Oceans, 122, 6416–6436, 2017.

550 Jones, S. D., Le Quéré, C., Rödenbeck, C., Manning, A. C., and Olsen, A.: A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data, Journal of Advances in Modeling Earth Systems, 7, 1554–1575, https://doi.org/https://doi.org/10.1002/2014MS000416, 2015.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochemical Cycles, 30, 1396–1417, https://doi.org/https://doi.org/10.1002/2015GB005359, 2016.

555 Landschützer, P., Ilyina, T., and Lovenduski, N. S.: Detecting Regional Modes of Variability in Observation-Based Surface Ocean pCO$_2$, Geophysical Research Letters, 46, 2670–2679, https://doi.org/https://doi.org/10.1029/2018GL081756, 2019.

Lauderdale, J. M., Dutkiewicz, S., Williams, R. G., and Follows, M. J.: Quantifying the drivers of ocean-atmosphere CO$_2$ fluxes, Global Biogeochemical Cycles, 30, 983–999, https://doi.org/https://doi.org/10.1002/2016GB005400, 2016.

Lauvset, S. K., Lange, N., Tanhua, T., Bittig, H. C., Olsen, A., Kozyr, A., Álvarez, M., Becker, S., Brown, P. J., Carter, B. R., Cotrim da 560 Cunha, L., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jutterström, S., Jones, S. D., Karlsen, M. K., Lo Monaco, C., Michaelis, P., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Velo, A., Wanninkhof, R., Woosley, R. J., and Key, R. M.: An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2021, Earth System Science Data, 13, 5565–5589, https://doi.org/10.5194/essd-13-5565-2021, 2021.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler,
565   S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns,
       T., Jimenéz-de-la-Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L.,
       Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D.,
       Nyawira, S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H.,
       Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., Storch, J.,
570   Tian, F., Voigt, A., Vrese, P., Wieners, K., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System
       Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO$_2$, Journal of Advances in Modeling Earth Systems, 11, 998–1038,
       https://doi.org/10.1029/2018MS001400, 2019.

McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., and Lovenduski, N. S.: Timescales for detection of trends in the ocean
       carbon sink, Nature, 530, 469–472, 2016.

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

McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L., and Lovenduski, N. S.: External Forcing Explains Recent Decadal Variability of
       the Ocean Carbon Sink, AGU Advances, 1, e2019AV000 149, https://doi.org/10.1029/2019AV000149, 2020.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y.-C., Kirkevåg,
580   A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren,
       O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: Overview
       of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations,
       Geoscientific Model Development, 13, 6165–6200, https://doi.org/10.5194/gmd-13-6165-2020, 2020.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward,
585   S., Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews,
       T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths,
       P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker,
       R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1:
       Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, p. 46, 2019.

590   Sloyan, B. M., Wanninkhof, R., Kramp, M., Johnson, G. C., Talley, L. D., Tanhua, T., McDonagh, E., Cusack, C., O'Rourke, E., McGovern,
       E., Katsumata, K., Diggs, S., Hummon, J., Ishii, M., Azetsu-Scott, K., Boss, E., Ansorge, I., Perez, F. F., Mercier, H., Williams, M. J. M.,
       Anderson, L., Lee, J. H., Murata, A., Kouketsu, S., Jeansson, E., Hoppema, M., and Campos, E.: The Global Ocean Ship-Based Hydro-
       graphic Investigations Program (GO-SHIP): A Platform for Integrated Multidisciplinary Ocean Science, Frontiers in Marine Science, 6,
       https://doi.org/10.3389/fmars.2019.00445, 2019.

595   Smith, D. M. and Murphy, J. M.: An objective ocean temperature and salinity analysis using covariances from a global climate model, Journal
       of Geophysical Research: Oceans, 112, https://doi.org/https://doi.org/10.1029/2005JC003172, 2007.

Smith, D. M., Allan, R. P., Coward, A. C., Eade, R., Hyder, P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, C. D., and Scaife,
       A. A.: Earth's energy imbalance since 1960 in observations and CMIP5 models, Geophysical Research Letters, 42, 1205–1213,
       https://doi.org/https://doi.org/10.1002/2014GL062669, 2015.

600 Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., p. 51, 2019.

Terhaar, J., Frölicher, T. L., and Joos, F.: Observation-constrained estimates of the global ocean carbon sink from Earth System Models, Biogeosciences Discussions, 2022, 1–49, https://doi.org/10.5194/bg-2022-134, 2022.

605 Thomas, J., Waugh, D., and Gnanadesikan, A.: Relationship between Ocean Carbon and Heat Multidecadal Variability, Journal of Climate, 31, 1467 – 1482, https://doi.org/10.1175/JCLI-D-17-0134.1, 2018.

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

Williams, R. G. and Follows, M. J.: Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms, Cambridge University Press, 2011.

610 Williams, R. G., Katavouta, A., and Roussenov, V.: Regional Asymmetries in Ocean Heat and Carbon Storage due to Dynamic Redistribution in Climate Model Projections, Journal of Climate, 34, 3907 – 3925, https://doi.org/10.1175/JCLI-D-20-0519.1, 2021.

Wong, A. P. S., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., Scanderbeg, M., Bhaskar, T. V. S. U., Buck, J. J. H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André, X., Poffa, N., Yashayaev, I., Barker, P. M., Guinehut, S., Belbéoch, M., Ignaszewski, M., Baringer, M. O., Schmid, C., Lyman, J. M., McTaggart, K. E., Purkey, S. G.,

615 Zilberman, N., Alkire, M. B., Swift, D., Owens, W. B., Jayne, S. R., Hersh, C., Robbins, P., West-Mack, D., Bahr, F., Yoshida, S., Sutton, P. J. H., Cancouët, R., Coatanoan, C., Dobbler, D., Juan, A. G., Gourrion, J., Kolodziejczyk, N., Bernard, V., Bourlès, B., Claustre, H., D'Ortenzio, F., Le Reste, S., Le Traon, P.-Y., Rannou, J.-P., Saout-Grit, C., Speich, S., Thierry, V., Verbrugge, N., Angel-Benavides, I. M., Klein, B., Notarstefano, G., Poulain, P.-M., Vélez-Belchí, P., Suga, T., Ando, K., Iwasaska, N., Kobayashi, T., Masuda, S., Oka, E., Sato, K., Nakamura, T., Sato, K., Takatsuki, Y., Yoshida, T., Cowley, R., Lovell, J. L., Oke, P. R., van Wijk, E. M., Carse, F., Donnelly, M.,

620 Gould, W. J., Gowers, K., King, B. A., Loch, S. G., Mowat, M., Turton, J., Rama Rao, E. P., Ravichandran, M., Freeland, H. J., Gaboury, I., Gilbert, D., Greenan, B. J. W., Ouellet, M., Ross, T., Tran, A., Dong, M., Liu, Z., Xu, J., Kang, K., Jo, H., Kim, S.-D., and Park, H.-M.: Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats, Frontiers in Marine Science, 7, 700, https://doi.org/10.3389/fmars.2020.00700, 2020.

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y.-P., and Srbinovsky,

625 J.: The Australian Earth System Model: ACCESS-ESM1.5, Journal of Southern Hemisphere Earth Systems Science, 70, 193, https://doi.org/10.1071/ES19035, 2020.