Global Biogeochemical Cycles

RESEARCH ARTICLE
10.1029/2019GB006186

Key Points:
- Robust uncertainties for the recent change in the North Atlantic surface fCO_2 are determined by using observational-based and model products
- The increasing North Atlantic surface fCO_2 is overestimated by ESMs during 1992–2014, and not captured by models' internal variability
- Simulation initialised with biogeochemical observations correct for the models' bias in the trend in surface CO_2 trends

Supporting Information:
- Supporting Information S1

Correspondence to:
A. D. Lebehot,
alice.lebehot@uct.ac.za

Citation:
Lebehot, A. D., Halloran, P. R., Watson, A. J., McNeall, D. J., Ford, D. A., Landschützer, P., et al. (2019). Reconciling observation and model trends in North Atlantic surface CO_2. *Global Biogeochemical Cycles*, *33*. https://doi.org/10.1029/2019GB006186

Received 11 FEB 2019
Accepted 10 AUG 2019
Accepted article online 18 AUG 2019

©2019. American Geophysical Union. All Rights Reserved.

Reconciling Observation and Model Trends in North Atlantic Surface CO_2

Alice D. Lebehot[1,2,3], Paul R. Halloran[1], Andrew J. Watson[1], Doug McNeall[4], David A. Ford[4], Peter Landschützer[5], Siv K. Lauvset[6,7], and Ute Schuster[1]

[1]Geography, College of Life and Environmental Studies, University of Exeter Laver Building, Exeter, UK, [2]Marine Research Institute, Department of Oceanography, University of Cape Town, Cape Town, South Africa, [3]Southern Ocean Carbon-Climate Observatory (SOCCO), CSIR, Cape Town, South Africa, [4]Met Office, Exeter, UK, [5]Max Planck Institute for Meteorology, Hamburg, Germany, [6]NORCE Norwegian Research Center, Bjerknes Centre for Climate Research, Bergen, Norway, [7]Geophysical Institute, Bjerknes Centre for Climate Research, University of Bergen, Bergen, Norway

Abstract The North Atlantic Ocean is a region of intense uptake of atmospheric CO_2. To assess how this CO_2 sink has evolved over recent decades, various approaches have been used to estimate basin-wide uptake from the irregularly sampled in situ CO_2 observations. Until now, the lack of robust uncertainties associated with observation-based gap-filling methods required to produce these estimates has limited the capacity to validate climate model simulated surface ocean CO_2 concentrations. After robustly quantifying basin-wide and annually varying interpolation uncertainties using both observational and model data, we show that the North Atlantic surface ocean fugacity of CO_2 ($fCO_{2-ocean}$) increased at a significantly slower rate than that simulated by the latest generation of Earth System Models during the period 1992–2014. We further show, with initialized model simulations, that the inability of these models to capture the observed trend in surface $fCO_{2-ocean}$ is primarily due to biases in the models' ocean biogeochemistry. Our results imply that current projections may underestimate the contribution of the North Atlantic to mitigating increasing future atmospheric CO_2 concentrations.

1. Introduction

About 50% of the carbon dioxide (CO_2) emitted from fossil fuel and land-use change activities is absorbed each year by the natural terrestrial and marine sinks, in similar proportions (Le Quéré et al., 2018). While the land and ocean carbon sinks play a fundamental role in controlling the levels of atmospheric CO_2, and hence mitigating climate change, the limited number and spatial coverage of the CO_2 observations makes the quantification of their intensity and variability at both the global and regional scale challenging. Due to the land spatial heterogeneity, which is difficult to capture by point measurements (especially compared to the relatively well-mixed ocean), the land sink is difficult to quantify and is commonly estimated as the residual from total emissions, atmospheric CO_2 growth, and ocean sink (Le Quéré et al., 2018). As such, robust estimates of the marine carbon sink's response to natural variability and climate change must be produced, so that (1) change in the CO_2 airborne fraction can be quantified and (2) the relationship between anthropogenic CO_2 emissions and atmospheric CO_2 concentrations represented in models can be confidently evaluated.

The North Atlantic Ocean is a region of strong natural and anthropogenic CO_2 uptake (Khatiwala et al., 2013; Mikaloff Fletcher et al., 2007; Sabine et al., 2004). While the large-scale processes controlling atmospheric CO_2 uptake by the North Atlantic are well understood, estimates of the time-varying sink over recent decades range from a decline (Schuster & Watson, 2007), through variability (Bates et al., 2014; Gruber et al., 2009), to an increase (Landschützer et al., 2016; Schuster et al., 2013; Ullman et al., 2009). These differences are likely to arise from the choice of study interval (Fay & McKinley, 2013; McKinley et al., 2011), internal climate variability (McKinley et al., 2011, 2016), and limited quantification of the uncertainty arising from the interpolation of CO_2 observational coverage (Jones et al., 2015). Indeed, despite international efforts to provide a well-distributed CO_2 observational coverage (both spatially and temporally), substantial gaps remain, including in the North Atlantic (Figures 1 and S1 in the supporting information). While various techniques (e.g., Landschützer et al., 2013; Rödenbeck et al., 2015; Schuster et al., 2013) have been developed to estimate the surface ocean fugacity of CO_2 ($fCO_{2-ocean}$) away from measurement locations and have been widely used to determine air-sea CO_2 fluxes across the oceans, basin-wide and time-varying interpolation uncertainties either have not been calculated (e.g. Rödenbeck et al., 2015) or are too broad to allow the

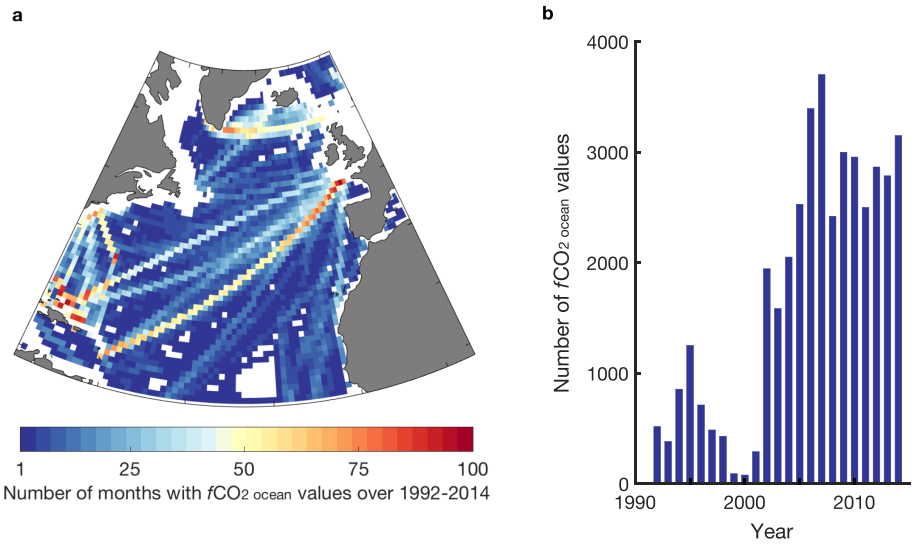

Figure 1. (a) Spatial and (b) temporal description of the Surface Ocean CO_2 ATlas (SOCAT) monthly gridded product Version 4 (Bakker et al., 2016) in the North Atlantic, for the period 1992–2014. Waters shallower than 1,000-m depth have been removed using the ETOPO1 bedrock product (Amante & Eakins, 2015). The term "$fCO_{2-ocean}$ values" refers to the monthly gridded values in SOCAT Version 4, calculated from the $fCO_{2-ocean}$ observations that were submitted to the SOCAT database. See Figure S1 for a representation of the annual spatial coverage of SOCATv4 for 1992–2014.

detection of significant temporal trends in surface $fCO_{2-ocean}$ (Jones et al., 2015, 2019). This lack of well-delimited interpolation uncertainty on observational-based products limits our ability to understand the marine carbon response to increasing atmospheric CO_2 and climate change, and also inhibits our evaluation of surface CO_2 concentrations simulated by climate models. By making use of the strengths of both observational and model data, here we robustly quantify basin-wide interpolation uncertainties of North Atlantic surface $fCO_{2-ocean}$ from 1992 to 2014, which allows us to (1) determine whether the change in the surface ocean CO_2 concentrations is significant, and (2) robustly compare the observation-based results with those simulated by the current generation of Earth System Models (ESMs) and identify potential shortcomings in those models.

We first present the observation-based interpolation technique used to provide basin-wide $fCO_{2-ocean}$ estimates and the novel interpolation uncertainty assessment (section 2). Once an appropriate method for robustly quantifying the uncertainties of the annually varying $fCO_{2-ocean}$ and its trend is identified, we then determine the recent change in the North Atlantic surface $fCO_{2-ocean}$ and evaluate the corresponding simulated change in state-of-the-art ESMs from CMIP5, the Coupled Model Intercomparison Project Phase 5 (section 3). Finally, we investigate the reasons behind the discrepancy between the models and the observation-based results by (1) using the models' preindustrial control runs to discuss the role of internal variability and (2) generating a set of ocean model simulations initialized with observations to study the role of biogeochemical initial conditions in driving the North Atlantic surface $fCO_{2-ocean}$ trend (section 4).

2. Methods

2.1. The MLR Interpolation Technique

Multiple linear regression (MLR) approaches rely on the mechanisms that link the predicted variable, which is described here by the spatially and temporally discontinuous observations of surface $fCO_{2-ocean}$, to a set of explanatory variables, which could be any available basin-wide variables that are involved in the $fCO_{2-ocean}$ response to anthropogenic changes, as well as physical and biogeochemical oceanic properties. To account for specific relationships between the predicted and explanatory variables across different biogeochemical regimes (e.g., a surface $fCO_{2-ocean}$ signal mostly temperature driven in the subtropical North Atlantic and principally temperature and biologically driven in the subpolar region; Schuster & Watson, 2007), MLR-based studies commonly build a linear fit within each geographical regime, whose separated $fCO_{2-ocean}$ results are finally merged to reconstruct the basin of interest (e.g., Iida et al., 2015; Watson et al., 2009).

Most studies assess the interpolation uncertainty through the comparison of the predicted results against the observations used to train the MLR and/or against observations that were not included in the MLR (independent products), therefore limiting the uncertainty assessment to the irregular observational coverage (e.g., Landschützer et al., 2014; Schuster et al., 2013). Substantial assumptions are therefore made when applying the "localized" uncertainty to the basin-wide and continuous predicted field; one assumption being that the interpolation method does not add any bias on the surface $fCO_{2-ocean}$ trend. By contrast, Watson et al. (2009) (hereinafter W09) used, for the first time, a biogeochemical model to evaluate the basin-wide uncertainty and test for potential biases introduced by the interpolation. Specifically, the model's surface $fCO_{2-ocean}$ field was subsampled at the same locations and days of the year of study as the observational coverage and was then treated as "real observations" in a separate MLR. By comparing the MLR-predicted $fCO_{2-ocean}$ to the model-truth values, W09 were able to generate a basin-wide uncertainty. As such, the MLR approach and uncertainty assessment developed by W09 is here extended and improved through a set of observation-based (section 2.1.1) and model-based MLRs (section 2.1.2).

2.1.1. The Observation-Based MLR

W09 built, for each 2-month interval of the year 2005 and for subdivisions of the North Atlantic of 30°, 20°, and 10° latitude, a linear relationship between the surface $fCO_{2-ocean}$ and three explanatory variables: the sea surface temperature (SST), mixed layer depth (MLD), and longitude. By being directly linked to changes of CO_2 concentration in seawater, the SST and MLD are characterized as mechanistically driven explanatory variables. Indeed, the solubility of CO_2 increases as the temperature of the surface waters decreases, and the deepening of the MLD through density-influenced and/or wind-mixing events can (1) enhance the dilution within the mixed layer of additional CO_2 taken up by the surface ocean and hence stimulate further uptake; (2) bring nutrient-enriched waters to the euphotic layer, potentially enhancing photosynthesis and the CO_2 uptake; and (3) bring dissolved inorganic carbon (DIC)-enriched waters to the surface, leading to local CO_2 outgassing (Sarmiento & Gruber, 2006). The longitude explanatory variable used here, and in W09's method, is included to account for east-west differences in water properties, for example, resulting from the contrasting temperature histories of water in eastern and western boundary currents, not accounted for by the other variables.

Here, we extend the W09 MLR method by (1) optimizing the approach by testing the use of latitude band widths of 60°, 30°, 20°, 10°, 5°, 2°, and 1° in order to account for the approximately latitudinally separated biogeochemical regimes resulting from the basin's circulation but also to investigate the impact of the spatial division on the predicted $fCO_{2-ocean}$; (2) widening the temporal extent of study from a single year to the period 1992–2014 (dates chosen to correspond to the starting year of the MLD product; Table S1); and (3) incorporating a further explanatory variable, the atmospheric CO_2 mixing ratio (xCO_2), in order to account for the time-varying impact of anthropogenic CO_2 emissions on the marine carbon system. The MLR analysis performed on observational products is hereinafter referred to as the "observation-based MLR."

The observation-based MLR was performed using surface $fCO_{2-ocean}$ from the Surface Ocean CO_2 ATlas (SOCAT) monthly gridded product Version 4 (hereinafter SOCATv4; Bakker et al., 2016), and using for the monthly SST, MLD, and xCO_2, the Optimum Interpolation Sea Surface Temperature (OISST) Version 2 product (Reynolds et al., 2007), the Estimating Circulation and Climate of the Ocean (ECCO2) version 2 product (Menemenlis et al., 2008), and the GLOBALVIEW-CO_2 reference matrix (GLOBALVIEW-CO2, 2013), respectively, from which (1) the period 1992–2014 and the North Atlantic (defined here as 10° N to 70° N and from 75° W to 5° E) were extracted, and (2) the regions shallower than 1,000-m water depth (as in W09) were removed using the ETOPO1 Bedrock product (Amante & Eakins, 2015) to minimize the freshwater inputs from rivers and coastal effects (cf. Table S1 for further description on the observational data processing). The observation-based MLR followed two steps:

1. The explanatory variables were subsampled at the locations and times at which $fCO_{2-ocean}$ values were available within SOCATv4 (in the North Atlantic over 1992-2014), referred as the "subsampled" data. The subsampled data were used to feed the MLR as follows:

$$fCO_{2\text{ -ocean, SOCATv4}} = \beta_0 + \beta_1 \cdot SST_{subsampled} + \beta_2 \cdot MLD_{subsampled} + \beta_3 \cdot xCO_{2_{subsampled}} + \beta_4 \cdot Lon_{subsampled}, \quad (1)$$

where β_0 is the intercept and $\beta_{1,2,3,4}$ are the regression coefficients returned by the statistical model.

2. The β coefficients were applied to the monthly basin-wide explanatory variables, to predict the monthly basin-wide $fCO_{2-ocean}$ from 1992 to 2014:

$$fCO_{2\text{ -ocean}} = \beta_0 + \beta_1 \cdot \text{SST} + \beta_2 \cdot \text{MLD} + \beta_3 \cdot xCO_2 + \beta_4 \cdot \text{Lon}. \quad (2)$$

The MLR computed over the 60° latitude bandwidth (corresponding to the entire North Atlantic basin width) therefore generated one set of β coefficients. To potentially improve the model's predictive skill (section 2.2), separate MLRs were trained over each latitude band of 30°, 20°, 10°, 5°, 2°, and 1° width, returning 2, 3, 6, 12, 30, and 60 sets of β coefficients (one set per subregion), respectively. The observation-based MLR analyses therefore included a total of 113 MLRs, leading to seven different monthly $fCO_{2-ocean}$ estimates for the North Atlantic from 1992 to 2014 (one for each spatial division method). The uncertainty assessment, achieved through the CMIP5-based MLR analyses (section 2.1.2), will allow us to identify the optimum spatial division method that provides bias-free (i.e., nonsignificant error) North Atlantic surface $fCO_{2-ocean}$ annual means and trend (section 2.2).

2.1.2. The CMIP5-Based MLRs

To quantify the annually varying interpolation uncertainty in the North Atlantic over 1992–2014, we use data from the 19 Coupled Model Intercomparison Project phase 5 (CMIP5) models (Table S2), obtained from the Earth System Grid Federation (ESGF) online portal. Outputs from the first ensemble member were downloaded at monthly frequency, for the historical experiment prior to 2006 and the Representative Climate Pathways (RCP) 8.5 from 2006 (Riahi et al., 2007; Taylor et al., 2012), and were regridded to the SOCAT 1° × 1° grid using the CDO package (http://www.mpimet.mpg.de/cdo; cf. Text S1 for further description on the CMIP5-data processing), from which North Atlantic shelf waters (<1,000-m water depth) were also removed using the ETOPO1 Bedrock product (Amante & Eakins, 2015).

Assuming the CMIP5 models to be perfectly known plausible alternative worlds, we use their outputs as investigation tools. This investigation approach offers a way to test how well the interpolation method performs, crucially, at places where no observational data are available. By providing a large model diversity, the CMIP5 framework specifically allows us to statistically investigate the effectiveness of the MLR interpolation technique used in the real world. Unlike W09, which assessed the interpolation uncertainty by training a MLR on a biogeochemical-modeled $fCO_{2-ocean}$ field with observational-based explanatory variables (a first attempt that unrealistically relied on the assumption that the model captures the real-world variability), the present analysis performs Observation Sampling Experiments (OSEs). By subsampling each CMIP5 model outputs (uniformly regridded) at the months and grid cells at which CO_2 observations were gathered within SOCATv4 over the period 1992–2014, OSEs allow to generate separate model-consistent MLRs, hereinafter referred as the "CMIP5-based MLRs."

For each CMIP5-based MLR, the steps described in the observation-based MLR (section 2.1.1) are followed using the subsampled $fCO_{2-ocean}$ model field from one of the 19 CMIP5 models and the SST, MLD, atmospheric xCO_2 fields from that CMIP5 model (and longitude, which is not a model-specific variable). Note that since the CMIP5 model simulations are conditioned with global annual atmospheric xCO_2 values, and not with seasonally and spatially varying values as in the real world, the CMIP5-based MLRs are built on model-relevant variables for consistency purposes (i.e., choosing annual atmospheric xCO_2 for CMIP5-based MLRs, while consistently selecting the seasonally and spatially varying xCO_2 for the observation-based MLR). As for the observation-based MLR, the CMIP5-based MLRs were also generated across the latitudinally divided North Atlantic. As such, the CMIP5-based MLRs generated over latitude band of 60°, 30°, 20°, 10°, 5°, 2°, and 1° width, returned a total of 19, 38, 57, 114, 228, 570, and 1,140 sets of β coefficients. For each spatial division method, the 19 reconstructed basin-wide surface $fCO_{2-ocean}$ products generated by the CMIP5-based MLRs were subtracted by their corresponding model-truth $fCO_{2-ocean}$ values (calculated interactively within each of the CMIP5 models), defining the $fCO_{2\text{-residuals}}$ (i.e., $fCO_{2\text{-residuals}} = fCO_{2\text{-ocean, MLR-predicted}} - fCO_{2\text{-ocean, model-truth}}$). The study of the $fCO_{2\text{-residuals}}$ enables us to investigate which of the seven spatial divisions used in the MLR method provides the best predictive skills for the North Atlantic surface $fCO_{2-ocean}$ annual means and the corresponding 1992–2014 trend (section 2.2).

2.2. Identifying the Appropriate MLR Setup

2.2.1. Annually Varying Uncertainty

To study whether the MLR method introduces a significant bias on the predicted annual $fCO_{2-ocean}$ in the North Atlantic over the period 1992–2014, the annual time series of the $fCO_{2\text{-residuals}}$ produced by each

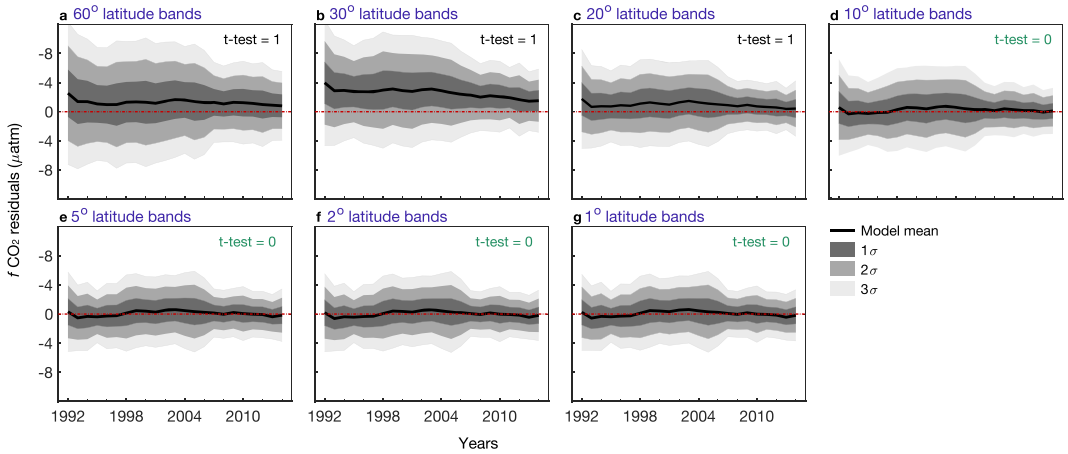

Figure 2. Annually varying uncertainty from the Coupled Model Intercomparison Project Phase 5-based multiple linear regression analyses. Each panel corresponds to the results from one of the seven latitudinally divided multiple linear regression methods. The black line corresponds to the multimodel mean of annual average $fCO_{2\text{-residuals}}$ (equation (3)). The dark, medium, and light gray shadings correspond, respectively, to the 1σ, 2σ, and 3σ across the 19 annual averages of $fCO_{2\text{-residuals}}$ (equation (4)). The dashed red line indicates the zero level. The t test result is indicated in each panel, where 0 means that there is no statistically significant mean bias in the residuals at the 5% significance level.

CMIP5-based MLR analysis is first calculated from the basin area-weighted monthly means. Then, the annually varying model mean (thick black line in Figure 2) and standard deviation σ (dark shading in Figure 2) errors are calculated, respectively, by

$$\mathrm{ME}_y = \sum_{m=1}^{M} \frac{\overline{fCO_{2\text{ -residual,y,m}}}}{M}, \tag{3}$$

$$\sigma(\mathrm{ME}_y) = \sqrt{\sum_{m=1}^{M} \frac{\left(\overline{fCO_{2\text{ -residual,y,m}}} - \mathrm{ME}_y\right)^2}{\mathrm{M}\text{ -1}}}, \tag{4}$$

where the horizontal line represents the North Atlantic area-weighted monthly means, which are averaged to the year y (from 1992 to 2014) and m to the CMIP5-based MLR analysis with M the total number of models (19).

Analysis of the annual bias in the North Atlantic averaged $fCO_{2-\mathrm{ocean}}$ field derived by the MLR method (Figure 2) shows that (1) the smaller the width of the latitude band over which the MLR is trained, the smaller the overall width of the uncertainty (the gray shadings in Figure 2 are much wider for the 60° study than for the 5° or 1° studies, with indiscernible visual improvements between the 5°, 2°, and 1° studies); (2) the uncertainty generally gets smaller over time, in line with the idea that the increase in number of $fCO_{2-\mathrm{ocean}}$ values (Figure 1b) improves the MLR's predictive skill; and (3) the MLR analyses built on latitude bands of 5°, 2°, and 1° width present similar results, suggesting that those three statistical methods provide a similar predictive skill on annual means. A t test is used to identify whether the uncertainty over the period 1992–2014 is significantly different from a distribution of mean zero at the 5% significance level. The MLR analyses computed successively over latitude bands of 10°, 5°, 2°, and 1° width reproduce the model-averaged $fCO_{2-\mathrm{ocean}}$ time series without the addition of a significant bias (the black thick line is statistically indistinguishable from the zero red line), at the 5% significance level (Figures 2d–2g). However, for the analyses using the larger latitude bands, the MLR overestimates the model-true annual $fCO_{2->\mathrm{ocean}}$ values at the 5% significance level (Figures 2a–2c). We conclude from this first assessment that the 10°, 5°, 2°, and 1° MLR analyses are potentially appropriate for calculating the annually varying surface $fCO_{2-\mathrm{ocean}}$.

2.2.2. Temporal Trend Uncertainty

The error introduced by the MLR interpolation technique on the North Atlantic surface $fCO_{2-\mathrm{ocean}}$ trend is studied using the CMIP5-based MLR analyses. For each CMIP5-based MLR analysis, linear trends in North

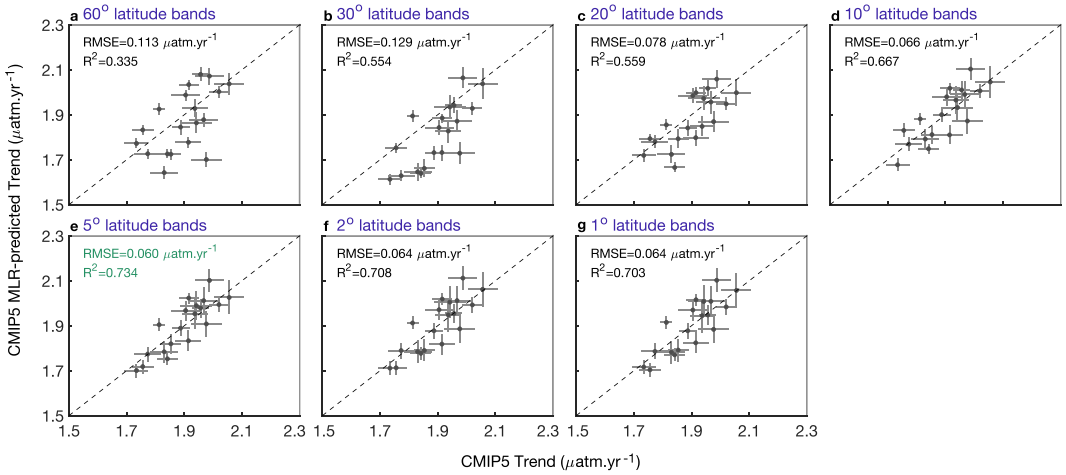

Figure 3. Trend uncertainty from the Coupled Model Intercomparison Project Phase 5 (CMIP5)-based multiple linear regression (MLR) analyses. Each panel corresponds to the results from one of the seven spatially divided MLR method. North Atlantic $fCO_{2-ocean}$ linear trends calculated from annual means over the period 1992–2014 from the CMIP5-based MLR product ($\Gamma_{MLR\text{-predicted}}$) versus the CMIP5 model-truth value ($\Gamma_{model\text{-true}}$; circles). The corresponding standard error of each linear trend (returned by the linear fit applied to basin-wide North Atlantic annual means) is indicated by the vertical and horizontal lines. Each panel includes the R^2 value and the root mean square error (RMSE) between $\Gamma_{MLR\text{-predicted}}$ and $\Gamma_{model\text{-true}}$. The dashed line indicates the one-to-one line. Points above the one-to-one line indicate that, for the corresponding CMIP5 models, the MLR overestimates the model-truth $fCO_{2-ocean}$ trend.

Atlantic surface $fCO_{2-ocean}$ were calculated using the MLR-predicted ($\Gamma_{MLR\text{-predicted}}$) and CMIP5 model-truth ($\Gamma_{model\text{-truth}}$) fields. Specifically, each basin-wide surface $fCO_{2-ocean}$ product was first averaged into North Atlantic area-weighted monthly means over 1992–2014 and then averaged into annual means, from which the linear surface $fCO_{2-ocean}$ trend and its associated standard error (returned by the linear fit) were obtained. By studying the differences between $\Gamma_{MLR\text{-predicted}}$ and $\Gamma_{model\text{-truth}}$ across the 19 CMIP5-based MLR analyses, we can therefore quantify the error on the North Atlantic surface $fCO_{2-ocean}$ trend over 1992–2014 introduced by the MLR (Figure 3).

For each of the seven different spatially divided MLR analyses, the R^2 and the root mean square error (RMSE) between $\Gamma_{MLR\text{-predicted}}$ and $\Gamma_{model\text{-truth}}$ were calculated (Figure 3). Across the methods, the 5° latitude bands analysis provides the highest R^2 value (0.734) but most importantly the smallest RMSE (0.060 μatm/year; Figure 3). The fact that the MLR methods on smaller latitude bands (over 2° and 1° width) provide slightly higher RMSE than the method on 5° latitude bands (Figures 3f and 3g) suggests that overfitting might be occurring. Indeed, the smaller the region over which the MLR is trained within a given period, the fewer observations are available, which could lead the MLR to become too specific to the trained data set and provide slightly poorer predictive skill for the overall population than when training the MLR over a wider region (Hastie et al., 2016).

In summary, the MLR method computed over 5° latitude bands across the North Atlantic provides unbiased (i.e., nonsignificant mean error) and robust results for both the annually varying surface $fCO_{2-ocean}$ and its corresponding trend over 1922–2014. As such, the observation-based MLR results generated from the 5° latitude bands method are adopted for the remaining analysis. Observation-based results are hereinafter presented with the associated interpolation uncertainty determined from the CMIP5-based MLR analyses (the ones generated from the 5° latitude band width setup). Specifically, the North Atlantic annually varying surface $fCO_{2-ocean}$ is constrained with the annually varying 1σ, 2σ, and 3σ (equation (4)), displayed in Figure 2e. The uncertainty on the observation-based trend in surface $fCO_{2-ocean}$ over the period 1992–2014 is given by the standard deviation of the difference between $\Gamma_{MLR\text{-predicted}}$ and $\Gamma_{model\text{-truth}}$ calculated across the 19 CMIP5 models (Figure 3e), which equals 0.060 μatm/year. Note that the annually varying uncertainties are in agreement with the uncertainty assessment carried out in W09 for 2005. Indeed, W09 quantified a 1σ error on the $fCO_{2-ocean}$ annual mean between 0.8 and 1.8 μatm, similar to the 1σ errors determined by our analysis (which ranges between 1.0 and 1.8 μatm across 1992–2014).

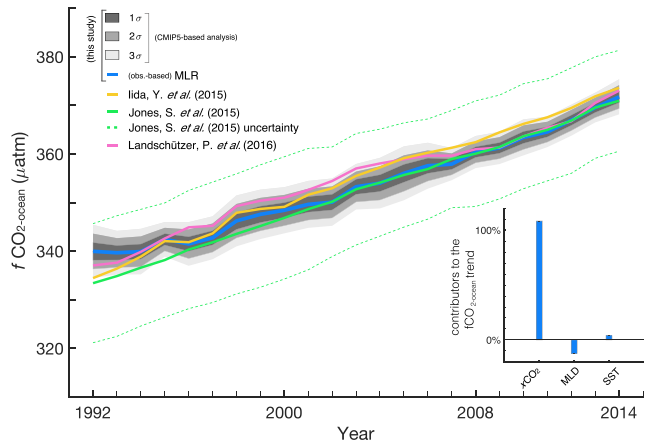

Figure 4. North Atlantic area-weighted annual averaged surface $fCO_{2-ocean}$, taken from the observation-based multiple linear regression (MLR; blue). The dark, medium, and light gray shadings represent the $\pm 1\sigma$, 2σ, and 3σ annual uncertainties obtained from the Coupled Model Intercomparison Project Phase 5-based MLR analysis, respectively (Figure 2e). The other lines present results from previous independent observation-based studies, using different techniques to reconstruct the trends, but similar data: The pink is from the neural network method of Landschützer et al. (2016, 2017), the brown is from a MLR method from Iida et al. (2015), and the green is from the Jones et al. (2019) statistical gap-filling method, which also provided an estimate for basin-wide uncertainties shown by the dashed lines. Note that Landschützer et al. (2016, 2017) and Iida et al. (2015) specifically provided $pCO_{2\text{-ocean}}$, but in terms of the illustrative analysis presented here, the difference between $pCO_{2\text{-ocean}}$ and $fCO_{2-ocean}$ is negligible. The bar plot shows the contribution of the xCO_2, mixed layer depth (MLD) and sea surface semperature (SST) variables to the predicted $fCO_{2-ocean}$ trend (section 3.2).

Further justification for using the CMIP5-based MLR uncertainty results to constrain the observation-based interpolated estimates is provided in a complementary analysis (Text S2; Figure S2). We show there that the two MLR studies (the one using the observational-based products and the one using the CMIP5 outputs) react similarly to their corresponding proxy variables. At the locations and times at which observations were made, the CMIP5-based MLR uncertainties are in agreement with the observation-based MLR uncertainties (on annual means; Text S2 and Figure S2). Since the annually varying uncertainty assessment is valid for the current CO_2 observational coverage, one can assume that the basin-wide annually varying uncertainty assessment is a robust estimate to delimit the annually varying observation-based $fCO_{2-ocean}$ estimates.

3. Results

3.1. Recent Change in the North Atlantic Surface $fCO_{2-ocean}$

In this section we explore the implications of the uncertainties in annual variation and trend on the North Atlantic $fCO_{2-ocean}$ using the MLR technique applied to 5° latitude bands as determined above (Figures 2e and 3e). While the observation-based $fCO_{2-ocean}$ predicted by the MLR is available at a monthly frequency and over 5° latitude bands, a study of the long-term change (23 years) at those resolutions would require estimates that are bias-free (i.e., the mean interpolation error not statistically different from zero) and robust within each subregion in each season, requirements that are not met for the month of August (Figure S3) and for some subregions of the North Atlantic (Figures S4 and S5). For instance, the MLR over 5° latitude band widths shows that while the method is robust in the subtropical North Atlantic, the range of uncertainties increases as we move northward (Figure S4), suggesting that the proxy variables used within the MLR might miss some information regarding the long-term change in the subpolar region and/or that not enough observations exist at high latitudes to detect a well-delimited signal. As such, we deliberately focus on basin-wide annual means for which we can demonstrate that the MLR results are unbiased and robust—a requirement for the comparison with the CMIP5 models (section 3.3).

The North Atlantic surface $fCO_{2-ocean}$ obtained from the observation-based MLR is first area-weighted into monthly means and then averaged into annual means (thick blue line in Figure 4). Over the period 1992–2014, the North Atlantic surface $fCO_{2-ocean}$ increased approximately linearly at a rate of 1.47 ± 0.06 μatm/year (Figure 4). The standard error on the trend due to the linear fit is 0.04 μatm/year, which is encompassed by the 1σ interpolation uncertainty. The relatively small year-to-year variability in surface $fCO_{2-ocean}$ (as characterized by the relatively small standard error on the linear fit) is expected because (1) interpolation methods systematically tend to smooth high-frequency variability and hence ultimately smooth the interannual signal and (2) the interpolation method used here is a linear regression model, which by construction we expect to lead to a robust estimate of the first-order trend. The increase of surface $fCO_{2-ocean}$ is considerably less than that of $fCO_{2\text{-atmosphere}}$ over the same interval (1.88 ± 0.02 μatm/year; Text S3; dashed line in Figure 4; GLOBALVIEW-CO2, 2013; Kalnay et al., 1996; Reynolds et al., 2007), resulting in an increased atmosphere-ocean fCO_2 gradient (Figure S6). With all else being equal, this would result in increased oceanic CO_2 uptake.

The idea that atmospheric and surface ocean CO_2 concentrations can be seen to diverge in response to anthropogenic CO_2 emissions, while ultimately an inevitable consequence of the Revelle factor (Revelle & Suess, 1957), could be considered counterintuitive. Indeed, McKinley et al. (2011) found ocean and atmosphere CO_2 trends to converge in the North Atlantic on multidecadal time scales but also highlighted that in the latter period of study (1993–2005), the largest region examined in their study, the permanently stratified Subtropical Atlantic, had a lower ocean than atmosphere CO_2 trend. While McKinley et al. (2011) suggest that the permanently stratified Subtropical Atlantic trend may be the result of multidecadal variability in the Atlantic, studies other than that presented here, which further extend the ocean CO_2 time series (Iida

et al., 2015; Jones et al., 2019; Landschützer et al., 2017) and from which can be calculated the area-weighted North Atlantic mean (in opposition to McKinley et al., 2011, which provide trends from nonbasin-interpolated values), support our results that at a basin scale the North Atlantic is maintaining a lower CO_2 trend than that observed in the atmosphere (Figures 4 and S7b). Indeed, our multidecadal $fCO_{2-\mathrm{ocean}}$ surface results are consistent with three methodologically independent and complementary techniques (Figure 4) based on similar data: a Neural Network approach (Landschützer et al., 2016, 2017), an alternative MLR approach Iida et al. (2015), and the Jones et al. (2019) statistical gap-filling method, which also provided a corresponding basin-wide uncertainty (Figure 4; Table S1). For each of the three independent techniques, the interpolated monthly surface $fCO_{2-\mathrm{ocean}}$ (or $pCO_{2\text{-}ocean}$) was extracted for the North Atlantic open ocean and for the period 1992–2014 (Table S1), spatially averaged using area-weighted means and averaged to annual means, from which the trend was finally calculated. The methods from Landschützer et al. (2016, 2017), Iida et al. (2015), and Jones et al. (2019) obtained a $fCO_{2-\mathrm{ocean}}$ (or $pCO_{2\text{-ocean}}$) trend of 1.49 ± 0.05, 1.71 ± 0.03, and 1.69 ± 1.03 μatm/year over 1992-2014, respectively (where the uncertainty corresponds to the standard error returned by the linear fit for the first two methods and corresponds for Jones et al. (2019) to the range of possible trends given their annually varying uncertainty), which is comparable to our trend estimate of 1.47 ± 0.06 μatm/year. While the results from Iida et al. (2015) are reaching the limit of agreement with the present work (given the 3σ range), the lack of time-varying uncertainties, which would likely be greater than the uncertainty on the linear fit (i.e., 0.03 μatm/year), might alter the interstudy comparisons. The uncertainty provided by Jones et al. (2019) covers a wide range of possible $fCO_{2-\mathrm{ocean}}$ trends within that interval, limiting what can be concluded about North Atlantic CO_2 trends from this approach. Finally, while our method was specifically designed to provide robust basin-wide $fCO_{2-\mathrm{ocean}}$ annual estimates, we show in an additional analysis that the localized surface $fCO_{2-\mathrm{ocean}}$ is in good agreement with the monthly CO_2 measurements obtained within Bermuda Atlantic Time-series Study (BATS) vicinity over our period of study (Figure S8).

3.2. Identifying the Drivers of the Recent Change in the North Atlantic Surface $fCO_{2-ocean}$

To examine the drivers of the North Atlantic surface $fCO_{2-\mathrm{ocean}}$ trend, the role of each explanatory variable in the interpolation technique is studied. The set of β coefficients from the observation-based MLRs over 5° latitude band width (equation (1)) was applied, within each band, to the separately varying subregional xCO_2, MLD, and SST following equations (5), (6), and (7), respectively (note that this approach was not used for longitude as it is a nontemporally varying variable). For example, the contribution of xCO_2 to the rate of change of the North Atlantic surface $fCO_{2-\mathrm{ocean}}$ was studied by calculating a new (unrealistic) $fCO_{2-\mathrm{ocean}}$ field within each 5° latitude band, using the corresponding β coefficients to the mean value of each explanatory variable, except for xCO_2, which was allowed to vary in time and space (equation (5)). The $12\,fCO_{2\text{ -ocean}<-xCO_2}$ monthly varying fields (one per each 5° latitude band) were then merged to reconstruct the basin-wide North Atlantic, monthly averaged with an area-weighted mean and annually averaged, from which the mechanistic trend was finally calculated. As such, by studying the sign and amplitude of the mechanistic trends in $fCO_{2\text{ -ocean} < \text{-}xCO_2}$, $fCO_{2\text{-ocean}<\text{-MLD}}$, and $fCO_{2\text{-ocean}<\text{-SST}}$, the dominant driver (among the given explanatory variables) and the mechanisms involved in the rate of change in surface $fCO_{2-\mathrm{ocean}}$ can be identified (bar plot in Figure 4).

$$fCO_{2\text{ -ocean}<-xCO_2} = \beta_0 + \beta_1\cdot\overline{\mathrm{SST}} + \beta_2\cdot\overline{\mathrm{MLD}} + \beta_3\cdot xCO_2 + \beta_4\cdot\overline{\mathrm{Lon}}, \tag{5}$$

$$fCO_{2\text{ -ocean} < \text{- MLD}} = \beta_0 + \beta_1\cdot\overline{\mathrm{SST}} + \beta_2\cdot\mathrm{MLD} + \beta_3\cdot\overline{xCO_2} + \beta_4\cdot\overline{\mathrm{Lon}}, \tag{6}$$

$$fCO_{2\text{ -ocean} < \text{- SST}} = \beta_0 + \beta_1\cdot\mathrm{SST} + \beta_2\cdot\overline{\mathrm{MLD}} + \beta_3\cdot\overline{xCO_2} + \beta_4\cdot\overline{\mathrm{Lon}}, \tag{7}$$

where β_0 is the intercept and $\beta_{1,2,3,4}$ are the regression coefficients (for a given 5° latitude band) returned by the linear model (equation (1)) and horizontal lines indicate that the data are averaged to a constant value.

The $fCO_{2-\mathrm{ocean}}$ trends resulting from separately varying xCO_2, MLD, and SST are 1.60, −0.18, and 0.057 μatm/year, corresponding to 108.5%, −12.4%, and 3.9% of the original trend (i.e., 1.47 μatm/year),

Table 1
Trends in Surface $fCO_{2-ocean}$ for Each CMIP5 Model and From the Observation-Based Estimates

Model name	$fCO_{2-ocean}$ trend (μatm/year)
CESM1-BGC	1.92 ± 0.03
CanESM2	2.02 ± 0.03
GFDL-ESM2G	1.73 ± 0.04
GFDL-ESM2M	1.77 ± 0.05
GISS-E2-H-CC	1.76 ± 0.04
GISS-E2-R-CC	1.84 ± 0.04
HadGEM2-CC	2.06 ± 0.05
HadGEM2-ES	1.99 ± 0.05
IPSL-CM5A-LR	1.94 ± 0.04
IPSL-CM5A-MR	1.83 ± 0.05
IPSL-CM5B-LR	1.94 ± 0.06
MIROC-ESM-CHEM	1.98 ± 0.05
MIROC-ESM	1.97 ± 0.05
MPI-ESM-LR	1.96 ± 0.03
MPI-ESM-MR	1.90 ± 0.05
MRI-ESM1	1.91 ± 0.05
NorESM-ME	1.81 ± 0.02
bcc-csm1-1-m	1.85 ± 0.04
bcc-csm1-1	1.89 ± 0.03
Model-mean	**1.90 ± 0.09**
Observation-based	**1.47 ± 0.06**

Note. CMIP5 = Coupled Model Intercomparison Project Phase 5. All trends were calculated over the North Atlantic and over the period 1992–2014. Significant air-sea CO_2. The uncertainty on the CMIP5 models estimates corresponds to the standard error of the linear fit, except for the model-mean values, which correspond to the intermodel variability.

respectively (bar plot in Figure 4). By having the largest trend amplitude, the atmospheric xCO_2 appears as the predominant driver explaining the increase in surface $fCO_{2-ocean}$ over the period 1992–2014, while expected provides evidence that the anthropogenic signal dominates the annual surface $fCO_{2-ocean}$ in the North Atlantic over this period. This result anticipates the results of McKinley et al. (2011), which found thata 25-year-long interval was required for the long-term signal in surface $fCO_{2-ocean}$ to emerge from the North Atlantic decadal variability (note however that their study was based on a different interval, 1981–2009). Over the period 1992–2014, the MLD and SST play a less important role than the atmospheric xCO_2 in controlling the surface $fCO_{2-ocean}$, and their signs are consistent with our understanding of the mechanisms between those two variables and the surface $fCO_{2-ocean}$. The negative sign in the $fCO_{2\text{-ocean}<\text{-MLD}}$ trend indicates that an overall increase in the MLD would lead to a decrease in the surface $fCO_{2-ocean}$, suggesting that the enhancement of the dilution of CO_2 in the mixed layer and/or the stimulation of the biological activity from the input of nutrient-enriched deep waters to the surface are the main MLD-related mechanisms involved in the surface $fCO_{2-ocean}$ for the basin-wide North Atlantic. As such, the impact of carbon-enriched deep waters to the surface as MLD deepens appears to be minimal over the period 1992–2014 and is unlikely to explain the rate of change of surface $fCO_{2-ocean}$. The positive sign in the $fCO_{2\text{-ocean}<\text{-SST}}$ trend indicates that, over the period 1992–2014, an increase in the SST leads to an increase in the North Atlantic surface $fCO_{2-ocean}$, which is consistent with the decrease in the solubility of CO_2 into seawater as surface waters warm. Nonetheless, we acknowledge that the drivers in the recent change in the North Atlantic surface $fCO_{2-ocean}$ would vary across the basin (with a remaining dominance from atmospheric xCO_2), for instance in the subpolar region, which experienced localized deep water formation events at the beginning of the period of study (Yashayaev et al., 2007) and which could therefore result in a MLD-driven mechanism (Figure S9).

3.3. Evaluating the Recent Change in the North Atlantic Surface CO_2 Concentrations in the CMIP5 Models

Recent change in the surface $fCO_{2-ocean}$ simulated by 19 CMIP5 models (Table 1) is evaluated through comparison with our new error-bounded observational-based time series and trend (Figure 5). We find that (1)

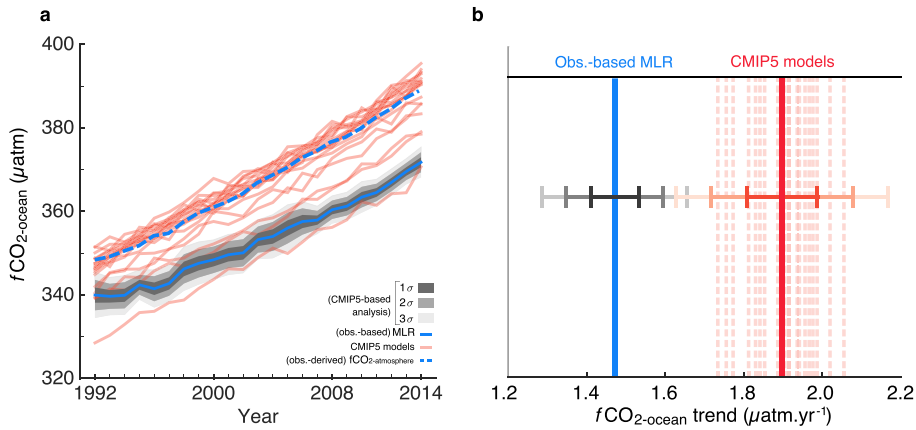

Figure 5. Evaluation of the recent change in surface $fCO_{2-ocean}$ in the Coupled Model Intercomparison Project Phase 5 (CMIP5) models against observation-based estimates. (a) Annually area-weighted averaged North Atlantic $fCO_{2-ocean}$ over the period 1992–2014 (except for the blue dashed line, which corresponds to the annually varying observation-derived atmospheric fCO_2; GLOBALVIEW-CO2, 2013; Kalnay et al., 1996; Reynolds et al., 2007; Text S3; which is closely followed by the CMIP5 models' atmospheric fCO_2, as shown in Figure S7b) and (b) resulting linear trends over the 1992–2014 interval. Blue corresponds to observation-based multiple linear regression (MLR) results and orange to each of the available CMIP5 models (Table 1). The dark, medium, and light gray shadings, respectively, represent the $\pm 1\sigma$, 2σ, and 3σ annual and trend interpolation uncertainties. In (b), the thick orange line corresponds to the CMIP5 model-mean trend value, and the $\pm 1\sigma$, 2σ, and 3σ across the models by the associated error bars.

the CMIP5 models overall behave similarly to each other in response to the increase in atmospheric CO_2 concentrations and (2) the CMIP5 models differ from the observation-based estimates in terms of the amplitude of the surface $fCO_{2-ocean}$, and importantly its trend (Figures 5a and 5b). Over the period 1992–2014, the North Atlantic $fCO_{2-ocean}$ trend in the CMIP5 models is on average 1.90 ± 0.09 μatm/year (where the $\pm 1\sigma$ value corresponds to the intermodel variability; Table 1). The models' surface ocean concentration closely follows the rise of atmospheric CO_2, which they experienced (the model-mean $fCO_{2\text{-atmosphere}}$ trend, i.e., the calculation of $fCO_{2\text{-atmosphere}}$ from the prescribed historical and RCP8.5 xCO_2 using modeled temperature and pressure, over the period 1992–2014 is 1.92 ± 0.01 μatm/year; Figure S7b), consequently limiting the air-sea CO_2 gradient and therefore the removal of CO_2 from the atmosphere. The $fCO_{2-ocean}$ trends in the CMIP5 models (1.90 ± 0.09 μatm/year) are significantly larger than the observation-based $fCO_{2-ocean}$ trend (i.e., 1.47 μatm/year), at the 5% significance level (right-tailed t test statistics).

The fact that the CMIP5 models have a larger $fCO_{2-ocean}$ trend than the observations means that the difference between $fCO_{2\text{-atmosphere}}$ and $fCO_{2-ocean}$ (i.e., ΔfCO_2) increases at a slower rate in the models than in the real world (for a negligible difference in the $fCO_{2\text{-atmosphere}}$ between the real-world and the models, as shown in Figure S7b). Therefore, the air-sea CO_2 flux, which is proportional to ΔfCO_2, would increase at a slower rate in the models than in the real world (with potential CO_2 outgassing), in the absence of significant trend in the gas transfer velocity and solubility.

4. Discussion

The statistically significant discrepancy in the surface $fCO_{2-ocean}$ trends between the CMIP5 models and the observation-based estimate is likely to result from one or more of four factors: (1) the CMIP5 models and the real world could be forced with slightly different atmospheric CO_2 concentrations, which could have impacted the surface $fCO_{2-ocean}$ and led to slightly different trends between the two systems; (2) the specific number and/or combination of the chosen CMIP5-based MLR analyses used to calculate the trend uncertainty may result in an anomalous $fCO_{2-ocean}$ trend uncertainty, impacting the outcome of the model-observation comparison; (3) the real world could be experiencing, over the period of study, a phase of natural variability not captured by any of the CMIP5 models; and/or (4) the CMIP5 models could poorly represent or miss some key characteristics of the marine CO_2 system necessary to capture the observed surface $fCO_{2-ocean}$ trend. These four possibilities are explored here.

4.1. A Model Bias due to Different Atmospheric xCO_2?

To verify that the discrepancy in the surface $fCO_{2-ocean}$ does not arise from differences in the respective atmospheric xCO_2 products, the annual atmospheric xCO_2 values in the real world and in the model world are compared (Figure S7a). Overall, the xCO_2 annual time series from GLOABALVIEW-CO2 (extracted for the North Atlantic and area-weighted into annual means) and from the "historical+RCP8.5" scenarios are similar, with indistinguishable resulting trends over 1992–2014 (i.e., 1.95 ± 0.02 ppm/year and 1.98 ± 0.02 ppm/year, respectively; Figure S7a). For the period up until 2005 (year from which the CMIP5 models are forced with RCP8.5 scenario xCO_2 values), the atmospheric xCO_2 trends from the observation-based product and the RCP8.5 projection values are also indistinguishable (1.83 ± 0.04 ppm/year and 1.80 ± 0.03 ppm/year, respectively; Figure S7a). As such, the trend discrepancy in surface $fCO_{2-ocean}$ between the real world and the models cannot be explained by the minimal differences in atmospheric xCO_2 between the two systems.

4.2. Model-Observation Discrepancy Explained by a Sensitive Interpolation Uncertainty?

The trend uncertainty assessment, which was used to constrain the observation-based trend estimate, relied on the CMIP5-based MLR analyses and was determined from the distribution across the 19 CMIP5 models of the difference between $\Gamma_{\text{MLR-predicted}}$ and $\Gamma_{\text{model-true}}$ (see section 2.2). As such, we performed a sensitivity analysis on the trend uncertainty to test whether the difference between the observation-based and the CMIP5 $fCO_{2-ocean}$ trends remains significant if we were to determine the uncertainty using fewer and different combinations of CMIP5-based MLR analyses. We calculated the trend uncertainty using a number of models k (a subset of the available models) that varies from 2 to 18 and considers the different possible combinations C_k^M of those models:

$$C_k^M = \frac{M!}{k!(M-k)!}, \tag{8}$$

where M is the total number of models available (19) and k the number of selected models (from 2 to 18) in a subset used to calculate the uncertainty. The standard deviation (i.e., trend uncertainty) calculations are repeated for each model subset size and combination.

The observation-based $fCO_{2-ocean}$ trend with the uncertainties resulting from the sensitivity analysis (i.e., 1.47 ± (1, 2, 3) $\sigma_{C_k^M}$ μatm/year are compared to the CMIP5 model ensemble of $fCO_{2-ocean}$ trends with an unpaired two-sample left-tailed t test, for each of the possible combinations (Figure S10). A t test statistics show that in all scenarios, the $fCO_{2-ocean}$ observation-based trend is always significantly smaller than the CMIP5 trends, at the 5% significance level (Figure S10). The sensitivity analysis therefore indicates that the CMIP5 models robustly overestimate the observation-based $fCO_{2-ocean}$ trend.

4.3. Internal Variability

One of the major challenges when interpreting the time-varying behavior in a model ensemble against observations on decadal to multidecadal time scales is that the model ensemble and the real world could be experiencing different phases of internal variability (e.g., the North Atlantic Oscillation or Atlantic Multidecadal Variability, both of which have been implicated in North Atlantic CO_2 uptake variability; McKinley et al., 2011; Schuster et al., 2009; Thomas et al., 2008). The model ensemble may therefore be significantly different from the observations because it does not capture the component of natural variability sampled by the real system. To assess this possibility, the internal variability in the model ensemble and in the real world should be quantified. While studies (e.g., DeVries et al., 2019; Landschützer et al., 2016; Rödenbeck et al., 2015) suggest substantial decadal variability within observation-based estimates linked to climate variability (e.g., Landschützer et al., 2019), the relatively short length of the $fCO_{2-ocean}$ observational record limits our confidence in the representation of the true internal variability and its interaction with CO_2 uptake at decadal and longer time scales (e.g., McKinley et al., 2011). A further complication is that this variability is superimposed on the anthropogenically forced climate change and the two can only be disentangled by assuming the validity of internal variability generated by models (e.g., Hegerl & Zwiers, 2011).

To investigate the role of internal variability within the North Atlantic surface $fCO_{2-ocean}$ trend, an initial analysis was conducted using the forced model fields. We specifically calculated the surface $fCO_{2-ocean}$

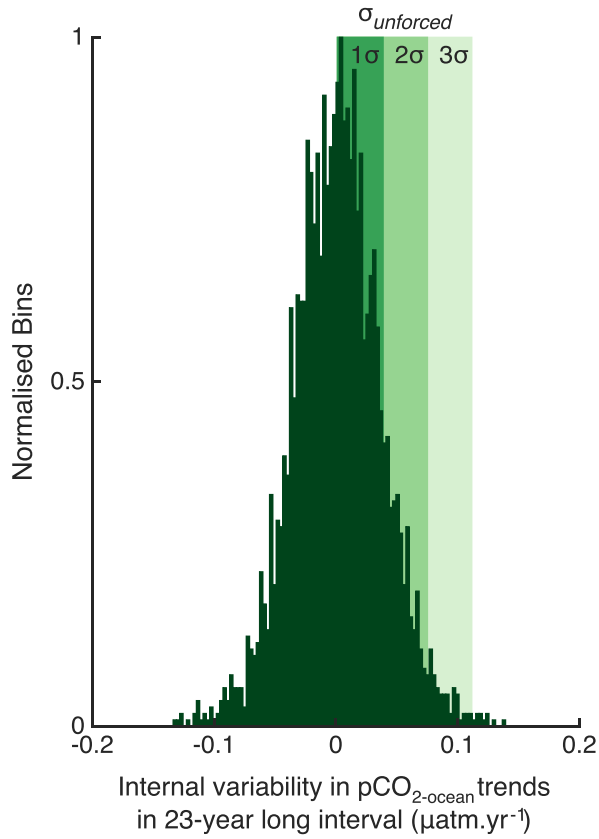

Figure 6. Quantifying the North Atlantic internal variability in the modeled trend. Histogram of the $pCO_{2\text{-ocean}}$ linear trends calculated over 23-year running intervals, using the first 240 years of the preindustrial experiment outputs, available for 14 Coupled Model Intercomparison Project Phase 5 models (Table S2). The dark to light green bands correspond to $1\sigma_{unforced}$, $2\sigma_{unforced}$, and $3\sigma_{unforced}$ of those $pCO_{2\text{-ocean}}$ trends, respectively.

trends simulated by the CMIP5 models within relatively shorter intervals than the period of study 1992–2014 (precisely over the periods 1992–2013, 1993–2014, 1992–2012, and 1994–2014; Figure S11). We show that the surface $fCO_{2-\text{ocean}}$ trends within 21- and 22-year-long intervals, as simulated by the CMIP5 models, are within the same range as the trends over the period 1992–2014 (Figure S11), suggesting that the signal captured by the CMIP5 models is consistent with overall forced system, as opposed to internal variability.

Nevertheless, models offer idealized platforms to fully quantify the internal climate variability using their unforced control simulations. To attempt to quantify internal variability in surface ocean CO_2 concentrations in a more robust manner (relative to the analysis conducted in Figure S11), we use the CMIP5 model preindustrial control simulations. The North Atlantic internal variability (i.e., the unforced variability) in the models' surface $pCO_{2\text{-ocean}}$ contained in 23-year-long intervals is quantified by using the models' preindustrial experimental runs, which describe the climate system without anthropogenic forcing. Out of the 19 CMIP5 models used in the model trend evaluation study, 14 provided $pCO_{2\text{-ocean}}$ data from their preindustrial control simulation (Table S2). In the context of this study, the difference between $pCO_{2\text{-ocean}}$ and $fCO_{2-\text{ocean}}$ is assumed to be negligible. Since the 14 CMIP5 models' control simulations were run over different lengths of time (from 240 to 1,000 years), the first 240 years in each of the control simulations was used for consistency. Over 240 years, 217 possible 23-year-long continuous intervals are defined, leading to the generation of 217 $pCO_{2\text{-ocean}}$ linear trends for each of the 14 CMIP5 control simulations (Figure 6). The standard deviation $\sigma_{unforced}$ of the $pCO_{2\text{-ocean}}$ trends calculated across all the possible 23-year-long intervals and across the available control simulations (a total of 217 × 14 = 3,038 trends) equals 0.036 µatm/year (corresponding to the dark green shading in Figure 6) should represent the amplitude of the unforced variability in the models. The internal variability in the surface $pCO_{2\text{-ocean}}$ trend over a 23-year-long

interval (up to $3\sigma_{unforced} = 0.108$ μatm/year), as simulated by the CMIP5 models, is about four times smaller than the difference between the CMIP5 model mean and the observation-based $fCO_{2-ocean}$ trend (i.e., $1.90 - 1.47 = 0.43$ μatm/year), indicating that the systematic trend overestimation in the CMIP5 models cannot be explained by the models' internal variability. To further support this statement, additional analysis (Text S3 and Figure S12) suggests that an ensemble of 19 forced model runs (with the model diversity as provided by the CMIP5 framework) is large enough to sample most of the models' unforced variability captured within an interval of 23 years. Due to the increase in atmospheric CO_2 concentrations since the beginning of the industrial era, the surface $fCO_{2-ocean}$ in the present day is higher than during preindustrial times, leading to the surface ocean being less buffered and potentially more variable than prior to the existence of anthropogenic forcing. As such, the dispersion of the trends simulated by the preindustrial runs may not necessarily be directly comparable with the dispersion of the trends simulated in an anthropogenically forced climate. However, the fact that none of the models used to investigate the trends over 1992–2014 capture the observation-based trend, while they all experienced the anthropogenic atmospheric CO_2 rise and consequent reduction in buffering, further suggests that the ensemble size is large enough to account for model internal variability (Text S3 and Figure S12).

While the $fCO_{2-ocean}$ trends in the 19 ensemble members seems to predominantly capture the forced change rather than the models' unforced variability, it cannot be ruled out that the observation-based $fCO_{2-ocean}$ trend estimate (calculated over a 23-year-long interval) may describe a phase of multidecadal internal variability, which is not generated by the models (McKinley et al., 2017; Schuster et al., 2009; Thomas et al., 2008). Indeed, the observational-based study of McKinley et al. (2011) identified that in the North Atlantic, the long-term trend took 25 years to emerge from the variability occurring on decadal time scales. Multiple lines of evidence—observational analysis indicating that xCO_2 was the dominant driver of our identified trend, model unforced control run analysis, and use of a 19 member model ensemble—have pointed to the trend difference being anthropogenically forced and further evidence that the discrepancy between the CMIP5 model and the observed trends is anthropogenic in origin comes from a mechanistic exploration of the difference in trends below.

4.4. Investigating the Mechanisms Explaining the CMIP5 Model Bias

While the CMIP5 models are the most advanced tools available at this time to explore the Earth's climate response to anthropogenic forcing, and the coordination of their simulations to provide multimodel ensembles helps us avoid drawing model-specific conclusions, such multimodel analyses lead to challenges. Different modeling groups will have made different assumptions in building and setting up the model, for example, with different initial conditions and spinup procedures (Séférian et al., 2016). As such, to understand and identify the potential mechanisms responsible for the systematic bias in the $fCO_{2-ocean}$ trends in the CMIP5 models, we performed five ocean-only ensembles using a single model. The five ensembles are identically initialized with observation-based products but are forced with surface conditions from five different sources spanning the interval of interest, which allow us to investigate the impacts of the initialization and of the surface forcing fields on the simulated $fCO_{2-ocean}$ trends.

Simulations were performed using a 1°, 75 vertical level, global ocean-only physical-biogeochemical model, the GO5.0 (Megann et al., 2014) configuration of the Nucleus for European Modelling of the Ocean (NEMO) hydrodynamic model (Madec, 2008) coupled with v4.1 of the CICE model (Hunke et al., 2010), and the Hadley Centre Ocean Carbon Cycle (HadOCC; Palmer & Totterdell, 2001) biogeochemical model. All NEMO fields were initialized to zero, except for temperature and salinity, which were taken from the EN4 objective analysis v4.1.1 (Good et al., 2013; Gouretski & Reseghetti, 2010). For HadOCC, initial conditions for nutrients were taken from the World Ocean Atlas climatology (Garcia et al., 2010), for DIC and Total Alkalinity (TA) from the Global Ocean Data Analysis Project (GLODAP) climatology (Key et al., 2004; whose version is more representative of the 1980s, period over which the simulation initialization is made, than the second updated version), and for phytoplankton, zooplankton, and detritus from the end of a previous simulation (Ford et al., 2012). The atmospheric pCO_2 values were prescribed using globally and monthly averaged surface data based on observations (Dlugokencky & Tans, 2016).

Considering initially the first ensemble member, five simulations were run from 3 January 1979 to 31 December 2014 and were forced by prescribing the surface conditions from five different sources (one

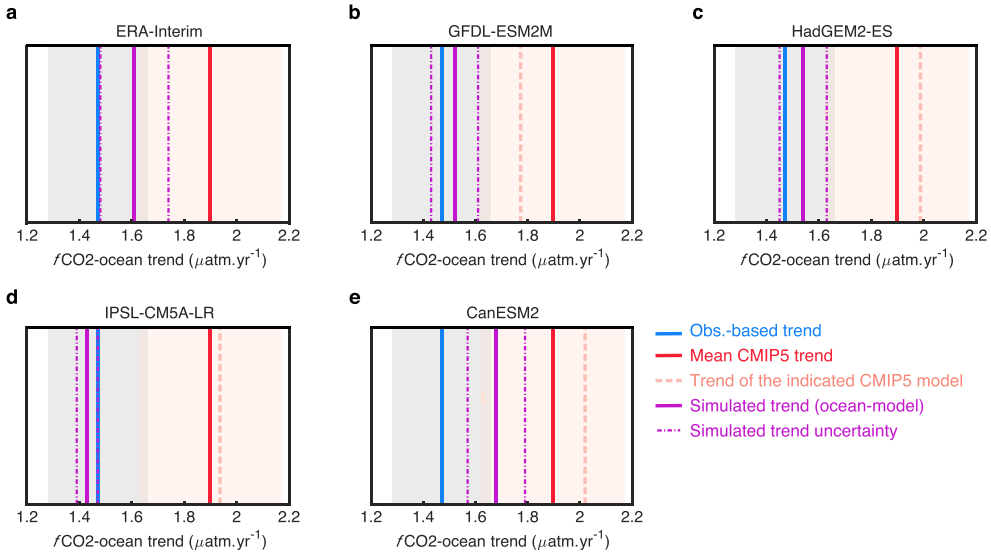

Figure 7. North Atlantic $fCO_{2-ocean}$ trends (thick magenta line) and uncertainty due to interannual variability in the atmospheric forcing (dashed magenta lines) returned from the ocean-only simulations that were forced with (a) ERA-Interim, (b) GFDL-ESM2M, (c) HadGEM2-ES, (d) IPSL-CM5A-LR, and (e) CanESM2 daily surface conditions. The orange dashed line in (b)–(e) shows the trend value for the corresponding Coupled Model Intercomparison Project Phase 5 (CMIP5) model (e.g., in b, the dashed orange line is the $fCO_{2-ocean}$ trend calculated from the GFDL-ESM2M model, as in Table 1). All simulated trends are here model-drift corrected (see Text S4; Figures S13 and S14). The blue and red lines correspond to the observation-based $fCO_{2-ocean}$ trend and the CMIP5 model-mean, respectively. The gray and orange shadows correspond to the 3σ returned by the CMIP5-based multiple linear regression analysis and by the intermodel spread.

source for each experiment). The prescribed surface forcing includes air temperature, snowfall, specific humidity, vector winds, precipitation, downwelling shortwave, and longwave radiation. The motivation behind running five simulations forced with five different surface conditions was to test whether the $fCO_{2-ocean}$ simulated trend was highly impacted by the atmospheric forcing (heat, moisture, and momentum fluxes) or by the initialization of the ocean variables, which was kept the same across the simulations. The first simulation used daily surface conditions from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis (Dee et al., 2011) and is hereinafter referred as the "ERA-Interim forced simulation." The remaining four simulations used daily surface conditions from four CMIP5 model outputs, specifically from the GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, and CanESM2 (Text S1) and are hereinafter referred as the "CMIP5-forced simulations." The four CMIP5 models were chosen to approximately span the $fCO_{2-ocean}$ CMIP5 model behaviors (Table 1).

For each of the five simulations of the first ensemble member, the surface $fCO_{2-ocean}$ outputs were (1) regridded into a $1° \times 1°$ regular grid using the bilinear interpolation within the CDO package (http://www.mpimet.mpg.de/cdo), (2) extracted for the open waters of the North Atlantic (the shelf waters above 1,000-m depth were removed using ETOPO1; Amante & Eakins, 2015), (3) monthly averaged (as the outputs were saved at a daily frequency), and (4) averaged into annual area-weighted means; leading for each simulation to a $fCO_{2-ocean}$ annual time series from 1979 to 2014. For each of the five annual $fCO_{2-ocean}$ annual time series, the linear trend was calculated over the period of interest 1992–2014. As such, the first 13 years of the simulations (i.e., from 1979 to 1991) correspond to the spinup period. While a 13-year spinup is inevitably of insufficient length for the model to reach equilibrium, an analysis of the model fields suggested that the large initial drifts had settled down during this period (the initial perturbation settles after ~3–4 years, example in Figure S13), certainly for the surface processes which we are most interested in over the time scales of this study, and which are most driven by the atmospheric forcing. As such, we are satisfied that the hindcast is sufficiently able to reproduce the mean state and observed variability of the ocean (Ford & Barciela, 2017). Furthermore, we require a short spinup to avoid the DIC and TA fields drifting too far from their initialized observed state and therefore to test the carbon cycle behavior when the model's DIC and TA are close to the observed (initialized) state on the simulated surface $fCO_{2-ocean}$ trends. Three additional

members were run with constant atmospheric CO_2 mixing ratio for each of the five different atmospheric forcing experiments to (1) quantify the model drift in each forcing experiment and (2) confirm that the drift in the trend is not overly sensitive to interannual variability in the forcing (Text S4; Figure S13).

After quantifying and removing the model drift on the simulated $fCO_{2-ocean}$ trends using the runs from the constant atmospheric CO_2 mixing ratio members (Text S4; Figure S14), we find that the simulated North Atlantic $fCO_{2-ocean}$ trends show good agreement with the observation-based trend and among themselves (potentially less agreement when considering the CanESM2-forced simulation; Figure 7). By being forced with an observation-based product, the ERA-Interim run provides an improvement in simulating the North Atlantic $fCO_{2-ocean}$ trend compared to the averaged CMIP5 model (Figure 7a). Nevertheless, a realistic surface forcing field is not a necessary condition for simulating an appropriate North Atlantic $fCO_{2-ocean}$ trend. Crucially, however, the initialization with observation-based biogeochemical (DIC and TA) and physical (T and S) fields is necessary for the trend to match the observations. Indeed, the CMIP5-forced simulations are performed with nonrealistic atmospheric conditions, and yet they provide realistic North Atlantic $fCO_{2-ocean}$ trends (Figures 7b-7e). Over the period 1992–2014, the resulting surface $fCO_{2-ocean}$ trends calculated from the four CMIP5-forced simulations (i.e., using the surface fluxes from GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, and CanESM2) are, respectively, 1.52 ± 0.09 μatm/year, 1.54 ± 0.09 μatm/year, 1.43 ± 0.04 μatm/year, and 1.68 ± 0.11 μatm/year, corresponding to a change that is respectively 83%, 86%, 92%, and 62% closer to the observation-based $fCO_{2-ocean}$ trend than their corresponding CMIP5 model-truth trends were (Figures 7b–7e). Besides being ocean-only simulations, the major difference between our new simulations and the CMIP5 simulations is that ours are initialized from observed ocean physics and biogeochemistry shortly before the results are produced, whereas the CMIP5 models have been spun up for periods up to 10,000 years to allow them to approach their unique preindustrial equilibrium (Séférian et al., 2016), from which the CMIP5 transient simulations are initiated. Over a long spinup period important carbon cycle fields, such as alkalinity, will drift from the observed state toward the models own equilibrium state (Figures S15–S18). Our simulations are therefore attempting to quantify the surface ocean CO_2 response to the real-world carbon cycle state, whereas the CMIP5 models describe a surface ocean CO_2 response with their model state.

5. Conclusions

While the latest generation of ESMs are widely used to underpin policy making, and of particular relevance here, to determine the allowable CO_2 emissions to remain below agreed levels of global warming, their evaluation typically focuses on how well they represent the climatological state. Since questions asked by policy makers relate to how the system is changing, which are answered by using ESMs, it is important to assess how these models simulate change. Evaluation of the models' ability to simulate change requires observational information about how the real world has changed; a challenge in a field such as ocean biogeochemistry, where observational records are short and often sparse. By carefully assessing and refining a method to interpolate sparse surface $fCO_{2-ocean}$ observations across the North Atlantic over the period 1992–2014 and quantifying robust annually varying uncertainties associated with the resultant time series and trend, and by investigating the impact of internal variability on the simulated trends, we have been able to identify that the CMIP5 model ensemble significantly overestimated the trend in the North Atlantic surface CO_2 concentration over the period 1992–2014. The possible mechanisms impacting the simulated North Atlantic surface $fCO_{2-ocean}$ were explored with an ocean-only model, which suggested that the discrepancy between the surface $fCO_{2-ocean}$ trend in the real world and in the CMIP5 models arises from inadequacies in their simulation of the background biogeochemical state. Inadequacies in CMIP5 physics and/or biology lead to substantial biases in TA that may be the cause for the $fCO_{2-ocean}$ trend overestimation in the models through its impacts on the buffer capacity (Figure S18). The present analysis therefore challenges our ability to predict the decadal-to-centennial evolution of the North Atlantic sink for CO_2 in the future using the current state-of-the-art ESMs. The model biases identified here and the extension of continuous ocean CO_2 measurements must be addressed if the climate-science community is to provide the best possible guidance on what anthropogenic CO_2 emissions are consistent with agreed atmospheric CO_2 targets. Nevertheless, our study further shows that models provide an essential platform to investigate the uncertainty and error associated with an observation-based interpolation technique. We therefore encourage the community to use OSEs to provide

further critical understanding of the robustness of interpolated products and their corresponding basin-wide and time-varying uncertainties.

Acronyms

BATS	Bermuda Atlantic Time-series Study
CMIP5	Coupled Model Intercomparison Project Phase 5
DIC	Dissolved Inorganic Carbon
ECMWF	European Centre for Medium-Range Weather Forecasts
ESGF	Earth System Grid Federation
ESM	Earth System Model
GLODAP	Global Ocean Data Analysis Project
HadOCC	Hadley Centre Ocean Carbon Cycle
MLD	Mixed Layer Depth
MLR	Multiple Linear Regression
NEMO	Nucleus for European Modelling of the Ocean
RCP	Representative Concentration Pathways
RMSE	Root Mean Square Error
SOCAT	Surface Ocean CO_2 Atlas
SST	Sea Surface Temperature
TA	Total Alkalinity

Acknowledgments
A. D. L., P. R. H., and U. S. were supported by the Natural Environment Research Council (NERC), RAGNARoCC (NE/K002473/1) and Atlantic Biogeochemical Fluxes (NE/M005070/1) projects. A. D. L. and P. R. H. were additionally supported by NERC Grant NE/P015142/1. A. J. W. was supported by the Royal Society. A. D. L. and D. M. were supported by the Joint BEIS/Defra Met Office Hadley Centre Climate Programme (GA1101). The authors thank Peter Cox, Pierre Friedlingstein, Laurent Bopp, and Steve Jones for discussion. Products generated from the observation-based MLRs and CMIP5-based MLRs, and from the ocean-only simulations, are available from doi:10.5285/8c1568f8-941e-512a-e053-6c86abc0d1e5 (https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/8c1568f8-941e-512a-e053-6c86abc0d1e5/).

References

Amante, C., & Eakins, B. (2015). ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis: National Oceanic Atmospheric Administration. https://doi.org/10.7289/V5C8276M

Bakker, D. C., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., et al. (2016). A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT). *Earth System Science Data*, *8*(2), 383–413. https://doi.org/10.5194/essd-8-383-2016

Bates, N. R., Astor, Y., Church, M., Currie, K., Lorenzoni, L, Muller-Karger, F. E., et al. (2014). A time-series view of changing surface ocean chemistry due to ocean uptake of anthropogenic CO_2 and ocean acidification. *Oceanography*, *27*(1), 126–141. https://doi.org/10.5670/oceanog.2014.16

DeVries, T., Le Quéré, C., Andrews, O., Berthet, S., Hauck, J., Ilyina, T., et al. (2019). Decadal trends in the ocean carbon sink. *Proceedings of the National Academy of Sciences*, *116*(24), 11,646–11,651. https://doi.org/10.1073/pnas.1900371116

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

Dlugokencky, E., & Tans, P. (2016). NOAA/ESRL.

Fay, A. R., & McKinley, G. A. (2013). Global trends in surface ocean pCO_2 from in situ data. *Global Biogeochemical Cycles*, *27*, 541–557. https://doi.org/10.1002/gbc.20051

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

Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., & Demaria, J. (2012). Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model. *Ocean Science*, *8*(5), 751–771. https://doi.org/10.5194/os-8-751-2012

GLOBALVIEW-CO2 (2013). Multi-laboratory compilation of synchronized and gap-filled atmospheric carbon dioxide records for the period 1979-2012. Boulder, Colorado, U.S.A: NOAA Global Monitoring Division. https://doi.org/10.3334/OBSPACK/1002

Garcia, H. E., Locarnini, R. A., Boyer, T. P., & Antonov, J. I. (2010). World Ocean Atlas 2009. Nutrients (Phosphate, Nitrate, Silicate).

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*, 6704–671. https://doi.org/10.1002/2013JC009067

Gouretski, V., & Reseghetti, F. (2010). On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep Sea Research Part I: Oceanographic Research Papers*, *57*(6), 812–833.

Gruber, N., Gloor, M., Mikaloff, F., Doney, ScottC., Dutkiewicz Stephanie, Follows, M. J., et al. (2009). Oceanic sources, sinks, and transport of atmospheric CO_2. *Global Biogeochemical Cycles*, *23*, GB1005. https://doi.org/10.1029/2008GB003349

Hastie, T., Tibshirani, R., & Friedman, J. (2016). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Verlag New York: Springer.

Hegerl, G., & Zwiers, F. (2011). Use of models in detection and attribution of climate change. *Wiley Interdisciplinary Reviews-Climate Change*, *2*(4), 570–591. https://doi.org/10.1002/wcc.121

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., & Elliott, S. (2010). CICE: the Los Alamos Sea Ice Model Documentation and Software User's Manual Version 4. 1.

Iida, Y., Kojima, A., Takatani, Y., Nakano, T., Sugimoto, H., Midorikawa, T., & Ishii, M. (2015). Trends in pCO_2 and sea-air CO_2 flux over the global open oceans for the last two decades. *Journal of Oceanography*, *71*(6), 637–661. https://doi.org/10.1007/s10872-015-0306-4

Jones, S. D., Le Quéré, C., Rödenbeck, C., Manning, A. C., & Olsen, A. (2015). 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/10.1002/2014MS000416

Jones, S. D., Le Quéré, C., Rödenbeck, C., Manning, A. C., & Olsen, A (2019). *Update to the Data and Code archive for the interpolation of surface ocean carbon dioxide*: PANGAEA. https://doi.org/10.1594/PANGAEA.898196

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al.(1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., et al. (2004). A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles*, *18*, GB4031. https://doi.org/10.1029/2004GB002247

Khatiwala, S., Tanhua, T., Mikaloff Fletcher, S., Gerber, M., Doney, S. C., Graven, H. D., et al. (2013). Global ocean storage of anthropogenic carbon. *Biogeosciences*, *10*(4), 2169–2191. https://doi.org/10.5194/bg-10-2169-2013

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

Landschützer, P., Gruber, N., & Bakker, D. C. E. (2017). An updated observation-based global monthly gridded sea surface pCO2 and air-sea CO2 flux product from 1982 through 2015 and its monthly climatology (NCEI Accession 0160558). Version 2. 2.

Landschützer, P., Gruber, N., Bakker, D. C. E., & Schuster, U. (2014). Recent variability of the global ocean carbon sink. *Global Biogeochemical Cycles*, *28*, 927–949. https://doi.org/10.1002/2014GB004853

Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., et al. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences*, *10*(11), 7793–7815. https://doi.org/10.5194/bg-10-7793-2013

Landschützer, P., Ilyina, T., & Lovenduski, N. S. (2019). Detecting regional modes of variability in observation-based surface ocean pCO2. *Geophysical Research Letters*, *46*, 12670–2679. https://doi.org/10.1029/2018GL081756

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., et al. (2018). Global Carbon Budget 2017. *Earth System Science Data*, *10*(1), 405–448. https://doi.org/10.5194/essd-10-405-2018

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

McKinley, G. A., Fay, A. R., Lovenduski, N. S., & Pilcher, D. J. (2017). 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

McKinley, G. A., Fay, A. R., Takahashi, T., & Metzl, N. (2011). Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales. *Nature Geoscience*, *4*(9), 606–610. https://doi.org/10.1038/ngeo1193

McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., & Lovenduski, N. S. (2016). Timescales for detection of trends in the ocean carbon sink. *Nature*, *530*(7591), 469–472. https://doi.org/10.1038/nature16958

Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., et al. (2014). GO5. 0: The joint NERC-Met Office NEMO global ocean model for use in coupled and forced applications. *Geoscientific Model Development*, *7*(3), 1069–1092.

Menemenlis, D., Campin, J., Heimbach, P., Hill, C., Lee, T., Nguyen, A., et al. (2008). ECCO2: High resolution global ocean and sea ice data synthesis. *Mercator Ocean Quarterly Newsletter*, *31*(October), 13–21.

Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Gloor, M., Doney, S. C., Dutkiewicz, S., et al. (2007). Inverse estimates of the oceanic sources and sinks of natural CO_2 and the implied oceanic carbon transport. *Global Biogeochemical Cycles*, *21*, GB1010. https://doi.org/10.1029/2006GB002751

Palmer, J. R., & Totterdell, I. J. (2001). Production and export in a global ecosystem model. *Deep-Sea Research Part I: Oceanographic Research Papers*, *48*(5), 1169–1198.

Revelle, R., & Suess, H. (1957). Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO2 during the past decades. *Tellus*, *9*(1), 18–27. https://doi.org/10.1111/j.2153-3490.1957.tb01849.x

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D., Casey, K. S., & Schlax, M. G.(2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. https://doi.org/10.1175/2007JCLI1824.1

Riahi, K., Grübler, A., & Nakicenovic, N. (2007). Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change*, *74*(7), 887–935. https://doi.org/10.1016/j.techfore.2006.05.026

Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., et al. (2015). Data-based estimates of the ocean carbon sink variability first - results of the Surface Ocean pCO(2) Mapping intercomparison (SOCOM). *Biogeosciences*, *12*(23), 7251–7278. https://doi.org/10.5194/bg-12-7251-2015

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., et al. (2004). The oceanic sink for anthropogenic CO_2. *Science*, *305*(5682), 367–371. https://doi.org/10.1126/science.1097403

Sarmiento, J. L., & Gruber, N. (2006). *Ocean Biogeochemical Dynamics*. Princeton, Woodstock: Princeton University Press.

Schuster, U., McKinley, G. A., Bates, N., Chevallier, F., Doney, S. C., Fay, A. R., et al. (2013). An assessment of the Atlantic and Arctic sea-air CO_2 fluxes, 1990-2009. *Biogeosciences*, *10*(1), 607–627. https://doi.org/10.5194/bg-10-607-2013

Schuster, U., & Watson, A. J. (2007). A variable and decreasing sink for atmospheric CO_2 in the North Atlantic. *Journal of Geophysical Research*, *112*, C11006. https://doi.org/10.1029/2006JC003941

Schuster, U., Watson, A. J., Bates, N. R., Corbiere, A., Gonzalez-Davila, M., Metzl, N., et al. (2009). Trends in North Atlantic sea-surface fCO_2 from 1990 to 2006. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *56*(8-10), 620–629. https://doi.org/10.1016/j.dsr2.2008.12.011

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., et al. (2016). Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, *9*(5), 1827–1851. https://doi.org/10.5194/gmd-9-1827-2016

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *American Meteorological Society*, *93*, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Thomas, H., Prowe, A. E. F., Lima, I. D., Doney, S. C., Wanninkhof, R., Greatbatch, R. J., et al. (2008). Changes in the North Atlantic Oscillation influence CO_2 uptake in the North Atlantic over the past 2 decades. *Global Biogeochemical Cycles*, *22*, GB4027. https://doi.org/10.1029/2007GB003167

Ullman, D. J., McKinley, G. A., Bennington, V., & Dutkiewicz, S. (2009). Trends in the North Atlantic carbon sink: 1992–2006. *Global Biogeochemical Cycles*, *23*, GB4011. https://doi.org/10.1029/2008GB003383

Watson, A. J., Schuster, U., Bakker, D. C. E., Bates, N. R., Corbière, A., González-Dávila, M., et al. (2009). Tracking the variable North Atlantic sink for atmospheric CO_2. *Science*, *326*(5958), 1391–1393. https://doi.org/10.1126/science.1177394

Yashayaev, I., Bersch, M., & van Aken, H. M. (2007). Spreading of the Labrador Sea Water to the Irminger and Iceland basins. *Geophysical Research Letters*, *34*(L10602). https://doi.org/10.1029/2006GL028999

References From Supporting Information

Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T. Y., et al. (2013). Basic performance of a new Earth system model of the Meteorological Research Institute (MRI-ESM1). *Papers in Meteorology and Geophysics*, *64*, 1–19. https://doi.org/10.2467/mripapers.64.1

Aumont, O., & Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, *20*, GB2017. https://doi.org/10.1029/2005GB002591

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., et al. (2013). The Norwegian Earth System Model, NorESM1-M, Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, *6*(3), 687–720. https://doi.org/10.5194/gmd-6-687-2013

Chylek, P., Li, J., Dubey, M. K., Wang, M., & Lesins, G. (2011). Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2. *Atmospheric Chemistry and Physics Discussions*, *11*(8), 22,893–22,907. https://doi.org/10.5194/acpd-11-22893-2011

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., et al. (2011). Development and evaluation of an Earth-System model HadGEM2. *Geoscientific Model Development*, *4*(4), 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011

Cooper, D. J., Watson, A. J., & Ling, R. D. (1998). Variation of p(CO2) along a North Atlantic shipping route (U.K. to the Caribbean): A year of automated observations. *Marine Chemistry*, *60*(1-2), 147–164. https://doi.org/10.1016/S0304-4203(97)00082-0

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dynamics*, *40*, 2123–2165. https://doi.org/10.1007/s00382-012-1636-1

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2012). GFDLs ESM2 global coupled climate-carbon Earth System Models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, *21*, 6646–6665.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2013). GFDLs ESM2 global coupled climate-carbon Earth System Models. Part II: Carbon system formulation and baseline simulation characteristics. *Journal of Climate*, *26*, 2247–2267. https://doi.org/10.1175/JCLI-D-12-00150.1

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., et al. (2011). The Community Climate System Model version 4. *Journal of Climate*, *24*(19), 4947–4991. https://doi.org/10.1175/2011JCLI4083.1

Giorgetta, M. A., Jungclaus, J. H., Reick, C. H., Legutke, S., Brovkin, V., Crueger, T., et al. (2013). Climate change from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project 5. *Journal of Advances in Modeling Earth Systems*, *5*, 572–597. https://doi.org/10.1002/jame.20038

Hansen, J., Sato, M., Ruedy, R., Kharecha, P., Lacis, A., Miller, R., et al. (2007). Climate simulations for 1880-2003 with GISS model. *Climate Dynamics*, *29*(7-8), 661–696. https://doi.org/10.1007/s00382-007-0255-8

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Nunez-Riboni, I. (2013). The global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth System Model in different CMIP5 experimental realizations. *Geophysical Research Letters*, *40*, 5909–5914. https://doi.org/10.1002/jame.20017

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., et al. (2013). Characteristics of the ocean simulations in the Max Planck InstituteOcean Model (MPIOM) the ocean component of the MPI Earth System Model. *Journal of Advances in Modeling Earth Systems*, *5*, 422–446. https://doi.org/10.1002/jame.20023

Kara, A. B., Rochford, P. A., & Hurlburt, H. E. (2000). An optimal definition for ocean mixed layer depth. *Journal of Geophysical Research*, *105*(C7), 16803. https://doi.org/10.1029/2000JC900072

Kawamiya, M., Kishi, M. J., & Suginohara, N. (2000). An ecosystem model for the North Pacific embedded in a general circulation model Part II: Mechanisms forming seasonal variations of chlorophyll. *Journal of Marine Systems*, *25*, 159–178.

Kortzinger, A. (1999). Determination of carbon dioxide partial pressure (p(CO2)), in Methods of Seawater Analysis. https://doi.org/10.1002/9783527613984

Landschützer, P., Gruber, N., & Bakker, D. C. E. (2015). A 30 years observation-based global monthly gridded sea surface pCO_2 product from 1982 through 2011. https://doi.org/10.3334/CDIAC/OTG.SPCO2_1982_2011_ETH_SOM-FFN

Lauvset, S. K., Key, R. M., Olsen, A., Van Heuven, S., Velo, A., Lin, X., et al. (2016). A new global interior ocean mapped climatology: The 1°× 1° GLODAP version 2. *Earth System Science Data*, *8*(2), 325–340. https://doi.org/10.5194/essd-8-325-2016

Lewis, E., & Wallace, D. W. R. (1998). *Program developed for CO2 system calculations. ORNL/CDIAC-105. Carbon Dioxide Information Analysis Center*. Oak Ridge, Tennessee: Oak Ridge National Laboratory, U.S Department of Energy.

Long, M., Lindsay, K., & Peacock, S. (2011). Twentieth-century oceanic carbon uptake and storage in CESM1(BGC)*. *Journal of Climate*, *26*(3), 6775–6800. https://doi.org/10.1175/JCLI-D-12-00184.1

McDougall, T. J., & Barker, P. M. (2011). Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox, SCOR/IAPSO WG127.

Pierrot, D., Neill, C., Sullivan, K., Castle, R., Wanninkhof, R., Luüger, H., et al. (2009). Recommendations for autonomous underway pCO_2 measuring systems and data-reduction routines. *Deep-Sea Research II*, *56*, 512–522. https://doi.org/10.1016/j.dsr2.2008.12.005

Romanou, A., Romanski, J., & Gregg, W. W. (2014). Natural ocean carbon cycle sensitivity to parameterizations of the recycling in a climate model. *Biogeosciences*, *11*, 1137–1154. https://doi.org/10.5194/bg-11-1137-2014

Sabine, C. L., Hankin, S., Koyuk, H., Bakker, D. C., Pfeil, B., Olsen, A., et al. (2013). Surface Ocean CO_2 Atlas (SOCAT) gridded data products. *Earth System Science Data*, *5*(1), 145–153. https://doi.org/10.5194/essd-5-145-2013

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., et al. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, *6*, 141–184. https://doi.org/10.1002/2013MS000265

Sudo, K., Takahashi, M., Kurokawa, J., & Akimoto, H. (2002). CHASER: A global chemical model of the troposphere 1. Model description. *Journal of Geophysical Research*, *107*(D17), 4339. https://doi.org/10.1029/2001JD001113

Sun, S., & Bleck, R. (2006). Multi-century simulations with the coupled GISS-HYCOM climate model: Control experiments. *Climate Dynamics*, *26*(4), 407–428. https://doi.org/10.1007/s00382-005-0091-7

Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., et al. (2013). Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). *Geoscientific Model Development Discussions*, *5*(4), 3035–3087. https://doi.org/10.5194/gmdd-5-3035-2012

Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, *12*(6), 351–362. https://doi.org/10.4319/lom.2014.12.351

Wanninkhof, R., Park, G. H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., et al. (2013). Global ocean carbon uptake: Magnitude, variability and trends. *Biogeosciences*, *10*(3), 1983–2000. https://doi.org/10.5194/bg-10-1983-2013

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development Discussions*, *4*(2), 1063–1128. https://doi.org/10.5194/gmd-4-845-2011

Weiss, R. (1974). Carbon dioxide in water and seawater: The solubility of a non-ideal gas. *Marine Chemistry*, *2*, 203–215.

Wu, T., Li, W., Ji, J., Xin, X., Li, L., Wang, Z., et al. (2013). Global carbon budgets simulated by the beijing climate center climate system model for the last century. *Journal of Geophysical Research: Atmospheres*, *118*, 4326–4347. https://doi.org/10.1002/jgrd.50320

Wu, T., Song, L., Li, W., Wang, Z., Zhang, H., Xin, X., et al. (2008). An overview of BCC climate system model development and application for climate change studies. *Journal of Meteorological Research*, *28*, 034–056. https://doi.org/10.1007/s13351-014-3041-7

Yukimoto, S., Yoshimura, H., Hosaka, M., Sakami, T., Tsujino, H., Hirabara, M., et al. (2011). Meteorological Research Institute-Earth System Model Version 1 (MRI-ESM1), Model Description. Japan: Meteorological Research Institute.

Zahariev, K., Christian, J., & Denman, K. (2008). Preindustrial, historical, and fertilization simulations using a global ocean carbon model with new parameterizations of iron limitation, calcification, and N_2 fixation. *Progress in Oceanography*, *77*(1), 56–82. https://doi.org/10.1016/j.pocean.2008.01.007