# Supporting Information for "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], Ute Schuster[1]

[1]Geography, College of Life and Environmental Studies, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, UK

[2]Marine Research Institute, Department of Oceanography, University of Cape Town, South Africa

[3]Southern Ocean Carbon-Climate Observatory (SOCCO), CSIR, Cape Town, South Africa

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

[5]Max Planck Institute for Meteorology, Bundesstrasse 53, 20146, Hamburg, Germany

[6]NORCE Norwegian Research Center, Bjerknes Centre for Climate Research, Jahnebakken 5, 5007 Bergen, Norway

[7]Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Allègaten 70, 5007 Bergen, Norway

## Contents of this file



## Text S1.

---

*Current affiliation

### Observation-based products

Here, the term "observational-based products" covers a wide range of datasets that use observations that have been processed in different manners: high-frequency measurements that have been merged into a regular spatial and temporal grid, localised measurements have been interpolated using various statistical tools, data derived from satellite measurements or from reanalysis methods. The general information on all observational-based products used in the study, and the various data-processing steps are described in Table S1.

### CMIP5 model data

Data download: Output from the CMIP5 models' (Taylor, Stouffer, and Meehl (2012); Table S2) was obtained from the ESGF online portal (`https://esgf-node.llnl.gov/projects/cmip5/`). From the available CMIP5 models (Taylor et al. (2012); Table S2), the following fields were downloaded for the $1^{st}$ ensemble member (labelled *r1i1p1* on ESGF) and for the historical experiment prior 2006, and the RCP8.5 scenario (Riahi et al., 2007) from 2006 (text in brackets indicates the variable name labelled on ESGF):

• **daily** near-surface specific humidity (*huss*, in $kg.kg^{-1}$), near-surface air temperature (*tas*, in K), surface longwave (*rlds*) and shortwave (*rsds*) downwelling radiations (in $W.m^{-2}$), eastward (*uas*) and northward (*vas*) near-surface winds (in $m.s^{-1}$), precipitation (*pr*) and snowfall (*prsn*) fluxes (in $kg.m^{-2}.s^{-1}$, equivalent to $mm.s^{-1}$ for a density of 1,000 $kg.m^{-3}$).

• **monthly** sea surface partial pressure (*spco2*, in Pa), temperature (*tos*, in K) and pressure (*psl*, in Pa), potential temperature (*thetao*, in K) and salinity (*so*, in psu)

• **yearly** dissolved inorganic carbon (DIC; *dissic*) and total alkalinity (TA; *talk*).

All variables are surface fields (i.e. 3-dimensional fields, with time×lat×lon), except for potential temperature, salinity, DIC and TA which are also available at depth (i.e. 4-dimensional fields, with time×depth×lat×lon). Note that the DIC and TA fields were also available at monthly frequency but surface only, which would have limited study at depth (Figure S15). Additionally, the $1^{st}$ ensemble member of the control run (picontrol on ESGF) for the monthly sea surface *spco2* (in Pa) was downloaded for 14 CMIP5 models (Table S2).

Conversion steps: To provide identical units across the models' outputs, and similar units/variables as observational products, and hence facilitate inter-model and model-observation comparisons, few adjustments were made:

• The IPSL models (i.e. IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR) did not provide on the ESGF portal the variable *spco2* but provided *dpco2* in $\mu$atm, from which *spco2* can be deduced. To do so, *dpco2* was first converted into Pa by multiplying by 0.101325 (i.e. 1 atm = 101325 Pa). The partial pressure of $CO_2$ in the atmosphere (i.e. $pCO_{2-\mathrm{atmosphere}}$) was then calculated from the atmospheric $xCO_2$ forcing data (Riahi et al., 2007) and the respective modelled SST and pressure fields (Taylor et al., 2012; Pierrot et al., 2009):

$$pCO_{2-\mathrm{atmosphere}} = xCO_{2-\mathrm{atmosphere}} \times 10^{-6} \times (P - pH_2O) \quad (1)$$

$$pH_2O = 0.981 \times \exp(14.32602 - 5306.03/T) \times 101325 \quad (2)$$

where, $P$ is the sea level pressure (Pa), $T$ the temperature (K), and $B$ and $\delta$ are the temperature dependent coefficients ($m^3 \cdot mol^{-1}$) (Weiss, 1974; Kortzinger, 1999), $xCO_{2-atmosphere}$ the $CO_2$ mixing ratio and $pH_2O$ the partial pressure of saturated water vapour (Pa) (Cooper et al., 1998).

Finally, $pCO_{2-atmosphere}$ was subtracted by *dpco2*. The deduced *spco2* was therefore in Pa, as the rest of the models.

• To match with the SOCAT observational product, surface *spco2* from each model was converted into fugacity and from Pa to $\mu$atm, using the modelled SST and pressure fields (Equation 3; Weiss (1974)) and by dividing by 0.101325.

$$fCO_2 = pCO_2 \times \exp\left(P\frac{B + 2\delta}{RT}\right) \quad (3)$$

where $R$ the ideal gas constant (8.314 $J \cdot K^{-1} \cdot mol^{-1}$).

Note that for simplicity, and due to the different ways results from different experiments have been used, this conversion was applied to the historical and RCP85 experiments but not to the picontrol experiment.

• Non-realistic salinity values for the CESM1-BGC model indicated that the units were in kg/kg, instead of psu (i.e. g/kg). The salinity field for this model was therefore multiplied by 1,000.

• The land mask for the *thetao* and *so* fields in the MRI-ESM1 model was set with zeros instead of "NaN". To provide clear land mask and avoid unrealistic MLD values at those grid points, all zero points were set to "NaN".

Regridding: The variety of model resolution and vertical/horizontal type of grids lead to difficulties when comparing models with each-other and/or with observational products

that are generally based on regular grids. To allow direct comparison of fields, one data processing step was therefore to regrid all model outputs into a regular grid.

All monthly variables (except *thetao* and *so*) were horizontally regridded to a transitional 180×360 grid using the bilinear interpolation within the CDO package (`http://www.mpimet.mpg.de/cdo`) and were then adjusted to the SOCATv4 grid (also a 1°×1° grid, but with longitude going from -179.5°E to 179.5°E) using the bilinear interpolation function within the Python Iris package (`http://scitools.org.uk/iris/docs/v1.7/index.html`).

The 4-dimensional *dissic* and *talk* fields were also vertically regridded according to the GLODAPv2 depth levels (Lauvset et al., 2016) using the CDO package (`http://www.mpimet.mpg.de/cdo`). However, since the first depth level in GLODAPv2 is 0 m and in the models is generally shallower (e.g. 5 m), the interpolation could not be computed between 0 m and the first level in the models (except for the NorESM1-ME model whose first depth level is 0 m). As such, the regridding provided "NaN" values at the surface and therefore led surface analyses of DIC and TA to be achieved at 10 m depth.

All daily surface fields, which were used to force an ocean model (c.f. Section 4.3. in main text), were temporally and spatially adjusted to models' set-up and grid. First, the time component of all variables was adapted to a Gregorian calendar. Models that had a 365-day calendar and hence no leap years (GFDL-ESM2M, CanESM2 and IPSL-CM5A-LR) has the extra day, February the $29^{th}$, added on leap years by repeating the fields for February $28^{th}$. The model with a 360-day calendar and hence 12 months of each 30-day long (HadGEM2-ES) had an extra day at the end of the months of January, July, August, October and November on non-leap years and also March on leap years, by repeating

the fields from day 30. Then all fields were spatially regridded to a transitional 180×360 standard-grid using the CDO package (`http://www.mpimet.mpg.de/cdo`), and then re-gridded to ORCA-1 (i.e. the grid of NEMO ocean model) using the Surface Interpolation Environment SOSIE (`https://github.com/brodeau/sosie`).

Mixed Layer Depth calculations: While the ESGF portal provided the variable *mlotst* (ocean mixed layer thickness defined by sigma-t) for some models, it was preferred, for clarity and uniformity reasons, to calculate the MLD for all models using the same method as in the observational-based product (Kara et al., 2000). Nevertheless, determining the MLD at each grid cell, for each month and for each model was a computationally expensive task. As such, calculations were achieved for the available period of MLD data from the observational-based product (from 1992; Table S1), and using the first 40 depth levels of each model, if available. Indeed, each model has its specific vertical resolution, going from 31 to 70 depth levels (Table S3). While the depth level 40 corresponds to relatively different depths across the models (from 985 m to 5720 m; Table S3), the MLD is expected to be found somewhere between the surface and depth level 40, as the observed global MLD mean is 89 ± 268 m (Menemenlis et al., 2008). For each model, the MLD was determined as follows:

1. To facilitate the calculations steps, the salinity and potential temperature fields were horizontally regridded into a 1°×1° latitude×longitude grid (not the SOCATv4 grid).

2. Salinity and potential temperature (in that order) were respectively converted into Absolute Salinity (AS) and Conservative Temperature (CT) using the Gibbs Seawater Oceanographic Matlab toolbox v3.04 (`http://www.TEOS-10.org`) (McDougall & Barker, 2011).

3. Density was calculated from AS and CT, also using the Gibbs Seawater toolbox.

4. MLD was calculated using CT and AS at the reference level (i.e. first depth level), from which the density criterion corresponding to a temperature change of 0.8° was determined, and also using the density profiles (Kara et al., 2000). The method first determines the bottom of the uniform layer and then identifies the depth interval which contains the density criterion (this interval typically corresponds to the zone of transition between the well-mixed surface and intermediate waters). If this depth interval is found, the MLD lies within that interval and is determined by linear interpolation (Kara et al., 2000). If the transitioning layer is not found, the steps are repeated using instead the second depth level as the reference level (Kara et al., 2000). If still not found, the MLD is set as NaN for this grid-cell.

5. The 3-dimensional (time×lat×lon) MLD field was regridded to the SOCATv4 grid using the Python Iris package. The resulting calculations gave overall realistic values (Table S3).

Atmospheric $xCO_2$: A monthly 3-dimensional grid (time×latitude×longitude) storing annual values of $xCO_2$ was created for the CMIP5-based MLR analyses. For each year, the same annual $xCO_2$ value was applied to each month, at each 1°×1° grid cell (following the SOCATv4 grid). Those $xCO_2$ values, obtained from the RCP Database Version 2.0.5 (Riahi et al., 2007), contain historical values up to 2005 and values following the RCP8.5 scenario from 2006.

**Text S2.**

When applying the annually-varying uncertainty determined from the CMIP5-based MLR analyses (results based on the 5° method) to the annually-varying $fCO_{2-ocean}$ deduced from the observation-based MLR, one assumes that the two MLR studies react similarly to their corresponding explanatory variables. For example, the model-mean $fCO_{2-residuals}$ time series is statistically not different from zero (Figure 2e), which does not guarantee that the basin-wide annually-varying $fCO_{2-residuals}$ from the observation-based MLR would behave in the same way as this, with a bias-free annually-varying $fCO_{2-residuals}$ (which cannot be assessed). To justify the application of the annually-varying uncertainty obtained from the CMIP5-based MLR to annually-varying results from the observation-based MLR, a comparative analysis on their corresponding $fCO_{2-residuals}$ is undertaken. Since the observation-based $fCO_{2-residuals}$ are only defined at places where observations were made, such comparison can only be achieved after subsampling the CMIP5-based $fCO_{2-residuals}$ at the locations, months and years of the observations. The observation-based $fCO_{2-residuals}$ were compared to the subsampled CMIP5-based $fCO_{2-residuals}$ as annually averaged area-weighted means, and more specifically to the model-mean and spread (Equations 3 and 4 in main text; corresponding respectively to the black line and grey shadings on Figure S2).

Overall, the patterns in the annual subsampled $fCO_{2-residuals}$ averages, from both the CMIP5-based and the observation-based MLR analyses, are impacted by the number of data points available for the different years (Figure S2), particularly in 2000 which the year with the lowest number of $fCO_{2-ocean}$ values in the North Atlantic during 1992-2014 (Figure 1). With only 80 grid cells with a $fCO_{2-ocean}$ value available in 2000 in the North

Atlantic, the residual mean is very likely to be impacted by potential grid cells where the MLR failed at capturing the model-true value (e.g. in the Canary upwelling system; Figure S1). During the 1992-2014 period, about 70% of the observation-based annual $fCO_{2-\text{residuals}}$ (i.e. 16 years out of 23) are within $2\sigma$ uncertainty range from the CMIP5-based MLR (i.e. 70% of the blue crosses are in the dark and medium grey shadings on Figure S2). Over 1992-2014, there are no significant trends in the annual $fCO_{2-\text{residuals}}$ calculated from the observation-based MLR, and from each of the subsampled CMIP5-based MLR residuals, at the 5% significance level. These results show that, over the period 1992-2014, the time-varying uncertainty calculated from the CMIP5-based MLR analysis (Figure 2e) is a robust estimate of the interpolation technique associated with the time-varying observation-based $fCO_{2-\text{ocean}}$.

**Text S3.**

This section provides additional evidence to addresses the question: Is the ensemble of 19 CMIP5 models large enough to capture the models' internal variability and therefore provide a model evaluation that reflect the overall forced signal rather than the unforced signal? For instance, if the model ensemble evaluation was based on the model-mean and standard deviation calculated from only 3 CMIP5 models instead of 19, the results might substantially vary depending on which CMIP5 models were selected. For clarity purposes, the standard deviation across the $fCO_{2-\text{ocean}}$ trends calculated from the forced model simulations over the period 1992-2014 (i.e. Table 1) is hereinafter referred as $\sigma_{forced}$, and the standard deviations across the $pCO_{2-\text{ocean}}$ trends calculated from the control runs over the possible 23-year intervals are hereinafter referred as $\sigma_{unforced}$.

If the standard deviation $\sigma_{forced}$ across an ensemble of forced $f\mathrm{CO}_{2-\mathrm{ocean}}$ modelled trends is larger than the $2\sigma_{unforced}$ of internal variability's range (i.e. $2\times0.036 = 0.07$ $\mu$atm·yr$^{-1}$), it suggests that the ensemble is wide enough to captures 95% of internal variability's range (for normally distributed data) and that the evaluation of the CMIP5 models constituting that ensemble most likely describe a mean state that is outside internal variability. With the amplitude of internal variability in trends over 23-year long intervals quantified (Figure 6), the minimum number of models that are required to create an ensemble that captures 95% of modelled internal variability (i.e. $2\sigma_{unforced} = 0.07$ $\mu$atm·yr$^{-1}$ for normally distributed data) can be determined. For each possible ensemble, whose size varies from 2 to 18 CMIP5 models (19 being the ensemble used in Figure 5b) considering all possible model combinations (Equation 9 in main text), the new standard deviation $\sigma_{forced}$ is calculated following:

$$\sigma_{forced} = \sqrt{\sum_{m=1}^{k} \frac{\left(\Gamma_m - \frac{1}{k}\sum_{m=1}^{k}\Gamma_m\right)^2}{k-1}} \tag{4}$$

where $\Gamma_m$ corresponds to the $f\mathrm{CO}_{2-\mathrm{ocean}}$ trends calculated over the period 1992-2014 for the CMIP5 model $m$ selected within the combination of models forming the new ensemble of size $k$, following Equation 9 in main text. For example, for an ensemble of three CMIP5 models (i.e. $k$=3), there are 969 possible combinations of 3 models among 19 models (Equation 9 in main text), leading to 969 values of $\sigma_{forced}$ (reported as a box plot on Figure S10).

When the ensemble contains less than 17 CMIP5 models, the amplitude of $\sigma_{forced}$ (i.e. y-axis in Figure S10) can be of a same amplitude as $2\sigma_{unforced}$, depending on the model configurations. For example, for an ensemble size of 16 CMIP5 models, few outlier $\sigma_{forced}$

values (represented by the grey dots in Figure S10) are within the $2\sigma_{forced}$ shading. As such, the evaluation of an ensemble of less than 17 CMIP5 models might, depending on the selected model combination, not adequately capture interval variability (e.g. McKinley et al. (2016)). The smaller the ensemble size, the more likely it can under-sample the unforced variability. When the ensemble contains all the available CMIP5 models (i.e. $k$=19), the amplitude of $\sigma_{forced}$, which equals 0.09 $\mu$atm·yr$^{-1}$, is larger than $2\sigma_{unforced}$ (i.e. 0.07 $\mu$atm·yr$^{-1}$), which suggests that the ensemble of 19 forced model runs (as used in this study) is large enough to sample the models' unforced variability captured within an interval of 23 years.

**Text S4.**

To identify whether the $fCO_{2-\mathrm{ocean}}$ trends simulated by the ocean-only model are a result of potential model drift and thus to provide a meaningful comparison between the simulated, the observation-based and the CMIP5 models $fCO_{2-\mathrm{ocean}}$ trends, additional simulations for which the atmospheric $CO_2$ concentration was held constant were run. Under constant atmospheric $CO_2$ concentration conditions, if the surface $fCO_{2-\mathrm{ocean}}$ remains approximately constant, the model is considered as non-drifting. Here, the method (1) quantifies the model-drift in each of the five simulations, and (2) accordingly removes the model drift from the simulated $fCO_{2-\mathrm{ocean}}$ trends.

To quantify the model drift in each of the five ocean-only simulations, fifteen additional simulations with prescribed constant atmospheric $CO_2$ mixing ratio, were run; hereinafter referred as the "model-drift simulations". The model-drift simulations all experienced the same constant atmospheric $CO_2$ mixing ratio of 336.85 ppm, which is the mixing ratio

value for the year 1979 (i.e. the starting year of the changing $CO_2$ runs) (Dlugokencky & Tans, 2016). The fifteen model-drift simulations were divided into five groups, each of them was respectively forced with the daily surface atmosphere conditions from ERA-Interim, GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and CanESM2. Each group is therefore made of three model-drift simulations (constituting three ensemble members), each of which were initialised at three different dates and hence run over three different periods of 36 years: from February 1979 to January 2015, from January 1984 to January 2020, and from January 1988 to January 2024. By shifting the atmospheric forcing years (i.e. from 1979 to 1984 and from 1979 to 1988), the set of model-drift simulations per set of surface forcings allows us to separate the model drift that is purely due to the model adjusting to the initial conditions from the trend that results from the year to year variability in the atmospheric forcing conditions. Since the ERA-Interim reanalysis was available until 2015, and since two ERA-Interim-forced model-drift simulations required data until 2020 and 2024, the ERA-Interim surface conditions from the period January 1983 to January 1988 were used to cover the period from January 2015 to January 2020, and the data from the period January 1983 to January 1992 were used to cover the period from January 2015 to January 2020. Since the CMIP5 models' surface conditions are known in the future, the daily models' surface conditions from the scenario RCP8.5 were used up to 2024 for the CMIP5-forced model drift simulations.

For each of the fifteen model-drift simulations, the $fCO_{2-ocean}$ outputs followed the same data processing steps as for the five changing atmospheric $CO_2$ runs. From the fifteen annual time series of the North Atlantic area-weighted monthly means of surface $fCO_{2-ocean}$ (Figure S13), the linear trends were calculated from the year 14 (i.e. the end of the spinup

phase) to the year 36 (i.e the end of the simulation), corresponding to the periods 1992-2014, 1997-2019 and 2001-2023 and for the simulations started in 1979, 1984 and 1988, respectively (Table S4). The amplitude of the model drift in the surface $fCO_{2-ocean}$ varies across the differently forced and initiated simulations, suggesting that the different mixing and solubility have an impact of the resulting $fCO_{2-ocean}$ (Table S4). Overall, the ocean-only model simulations experience significant negative surface $fCO_{2-ocean}$ trends, except when forced with the HadGEM2-ES surface forced conditions starting in 1984 and 1988, at the 5% significant level (Table S4). As such, in the case where the surface $fCO_{2-ocean}$ is increasing over the period 1992-2014 in the changing atmospheric $CO_2$ runs, the identified model-drift would actually mean that the surface $fCO_{2-ocean}$ is actually experiencing a stronger rate of increase compare to the non-corrected model-drift results.

To correct the model drift in each of the five North Atlantic $fCO_{2-ocean}$ trends simulated by the atmospheric varying $xCO_2$ runs, a series of three main steps were carried out. Here is an example for the ERA-Interim forced simulations (Figure S12):

1. For the model-drift simulation started in 1979, a linear regression was fitted between the years 14 and 36 of the simulation (thick blue line in Figure S14a). The linear trend and standard error $\sigma$ values returned by the linear fit (-0.29 $\pm$ 0.06 $\mu$atm·yr$^{-1}$; Table S4) were used to construct six additional linear time series (dashed blue lines in Figure S14a). The first time series was constructed so that it had a trend equalling the +1 $\sigma$ limit from the main fit: -0.29 + 0.06 $\mu$atm·yr$^{-1}$ = -0.23 $\mu$atm·yr$^{-1}$. The second time series was similarly constructed but from the -1 $\sigma$ limit from the main fit, which therefore had a trend of -0.35 $\mu$atm·yr$^{-1}$. The third to the sixth time series were finally constructed from the $\pm$ 2, 3 $\sigma$ limits from the main fit, leading to four time series with the trends of -0.47,

-0.41, -0.17 and -0.11 $\mu$atm·yr$^{-1}$. As such, there are a total of seven fitted time series for the model-drift simulation starting in 1979. These steps were repeated for the model-drift simulation starting in 1984 and in 1988 (Figure S14b-c, respectively), leading to a total of 21 model drift estimates for the ERA-Interim forced model drift simulations (all the lines in Figure S14a-c)

2. The 21 model drift estimates were subtracted one at a time from the $fCO_{2-ocean}$ time series over the period 1992-2014 for the changing atmospheric $CO_2$ run (i.e. dashed line in Figure S14d, whose linear trend is 1.29 $\pm$ 0.06 $\mu$atm·yr$^{-1}$, where 0.06 $\mu$atm·yr$^{-1}$ corresponds to the linear fit's standard error), leading to 21 possible corrected $fCO_{2-ocean}$ time series (i.e. coloured lines in Figure S14d). Each time series is presented as an anomaly plot relative to its corresponding $fCO_{2-ocean}$ value in 1991.

3. A linear fit was applied to each of the 21 corrected $fCO_{2-ocean}$ time series (i.e. coloured lines in Figure S14e), whose mean and standard deviation define to the model-drift corrected final North Atlantic $fCO_{2-ocean}$ trend value simulated by the ERA-Interim forced simulation (i.e. solid black line and error bars Figure S14e): 1.61 $\pm$ 0.13 $\mu$atm·yr$^{-1}$ (Table S5).

The above steps were repeated across each of the CMIP5-forced simulations and their North Atlantic surface $fCO_{2-ocean}$ trends prior and after the model drift correction are provided in Table S5.

## References

Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T. Y., ... Kitoh, A. (2013). Basic performance of a new earth system model of the Meteorological

Research Institute (MRI-ESM1). *Papers in Meteorology and Geophysics*, *64*, 1–19. doi: 10.2467/mripapers.64.1

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

Aumont, O., & Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, *20*(2), 1–15. doi: 10.1029/2005GB002591

Bakker, D. C., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., ... Xu, S. (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. doi: 10.5194/essd-8-383-2016

Bates, N. R., Astor, Y. M., Church, M. J., Currie, K., Dore, J. E., González-Dávila, M., ... Santana-Casiano, J. M. (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. Retrieved from `http://dx.doi.org/10.5670/oceanog.2014.16.` doi: 10.5670/oceanog.2014.16⟨http://dx.doi.org/10.5670/oceanog.2014.16⟩).

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevå g, A., Seland, O., ... Kristjánsson, J. E. (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. Retrieved from `http://www.geosci-model-dev.net/6/687/2013/` doi: 10.5194/gmd-6-687-2013

Chylek, P., Li, J., Dubey, M. K., Wang, M., & Lesins, G. (2011). Observed and model sim-

ulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2. *Atmospheric Chemistry and Physics Discussions*, *11*(8), 22893–22907. doi: 10.5194/acpd-11-22893-2011

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., ... Woodward, S. (2011). Development and evaluation of an Earth-System model – HadGEM2. *Geoscientific Model Development*, *4*(4), 1051–1075. doi: 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. doi: 10.1016/S0304-4203(97)00082-0

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorological Society*, *137*, 553–597. doi: 10.1002/qj.828

Dlugokencky, E., & Tans, P. (2016). *NOAA/ESRL.* Retrieved from `www.esrl.noaA.gov/gmd/ccgg/trends/`

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

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., ... Zadeh, N. (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., ... Zadeh, N. (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. doi: 10.1175/JCLI-D-12-00150.1

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

Giorgetta, M. A., Jungclaus, J. H., Reick, C. H., Legutke, S., Brovkin, V., Crueger, T., ... Stevens, B. (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*(3), 572–597. doi: 10.1002/jame.20038

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

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

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. doi: 10.1007/s10872-015-0306-4

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. doi: 10 .1002/jame.20017

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.* doi: https://doi.pangaea.de/10.1594/PANGAEA.898196

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., ... von Storch, J. S. (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*(21), 422–446. doi: 10 .1002/jame.20023

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. Retrieved from `http://www.esrl.noaA .gov/psd/` doi: 10.1175/1520-0477(1996)077⟨0437:TNYRP⟩2.0.CO;2

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. doi: 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*. XX. doi: 10.1002/9783527613984

Landschützer, P., Gruber, N., & Bakker, D. C. E. (2016). Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, *30*(10), 1396–1417. doi: 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.*

Lauvset, S. K., Key, R. M., Olsen, A., Van Heuven, S., Velo, A., Lin, X., ... Watelet, S. (2016). A new global interior ocean mapped climatology: The 1°× 1° GLODAP version 2. *Earth System Science Data*, *8*(2), 325–340. doi: 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 National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.* Retrieved from `http://cdiac.ornl.gov/oceans/co2rprt.html`

Long, M., Lindsay, K., & Peacock, S. (2011). Twentieth-Century Oceanic Carbon Uptake and Storage in CESM1(BGC)*. *Journal of Climate*, *26*(3), 6775–6800. doi: 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 [Computer software manual].

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. doi: 10.1038/nature16958

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

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.

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

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. doi: 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. doi: 10.1016/j.techfore.2006.05.026

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. doi: 10.5194/bg-11-1137-2014

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

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

University Press.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., ... Zhang, J. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.*, *6*, 141–184. doi: 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*, 4339. doi: 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. doi: 10.1007/s00382-005-0091-7

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

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

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., ... Kawamiya, M. (2011). MIROC-ESM: model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development Discussions*, *4*(2), 1063–1128. doi: 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., ... Zhang, J. (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. doi: 10 .1002/jgrd.50320,2013

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

Yukimoto, S., Yoshimura, H., Hosaka, M., Sakami, T., Tsujino, H., Hirabara, M., ... Kitoh, A. (2011). *Meteorological Research Institute-Earth System Model Version 1 (MRI-ESM1), Model Description* (Tech. Rep.). Meteorological Research Institute, Japan.

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. doi: 10.1016/j.pocean.2008.01.007

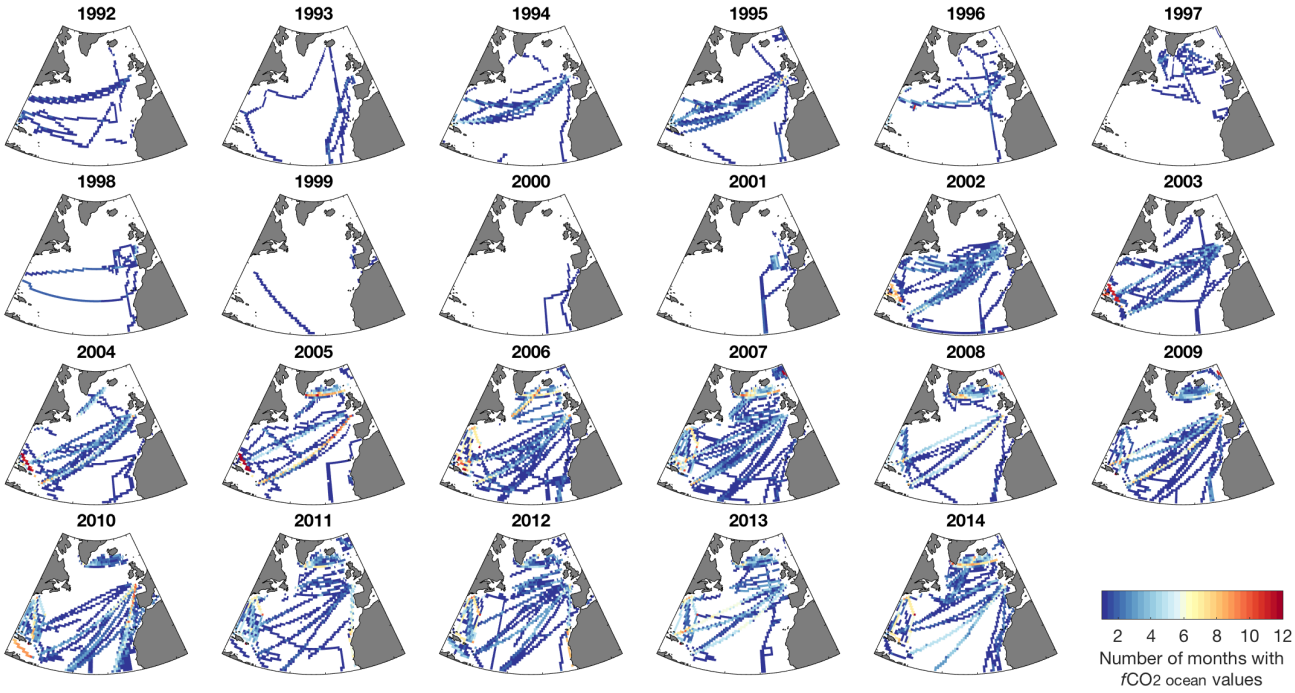


**Figure S1.** Spatial description of the SOCAT v4 monthly gridded product (Bakker et al., 2016) in the open-waters of the North Atlantic (i.e. waters shallower than 1,000 m depth have been removed using the ETOPO1 bedrock product, (Amante & Eakins, 2015)) per year for the period 1992-2014. The term "$fCO_{2-ocean}$ values" refers to the monthly gridded values in SOCATv4, calculated from the $fCO_{2-ocean}$ observations that were submitted to the SOCAT database.

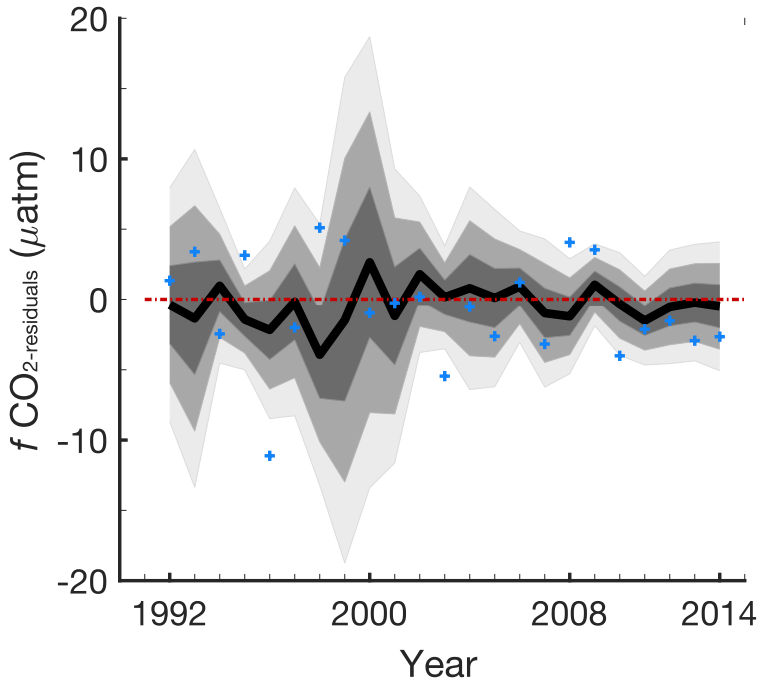


**Figure S2.** Comparison of the residuals between the CMIP5-based and observation-based MLRs. Model-mean of the annually-averaged subsampled $fCO_{2-\text{residuals}}$ (black line; Equation 3 in main text), with associated 1, 2 and 3 $\sigma$ across the models (dark, medium and light grey, respectively; Equation 4 in main text). The term "subsampled" refers to the extraction of model data (here the $fCO_{2-\text{residuals}}$) at position and time where observations are available, based on SOCATv4. The blue crosses are the annual averages of the $fCO_{2-\text{residuals}}$ from the observation-based MLR analysis. Results correspond to the MLR generated over 5° latitude band width over the North Atlantic.

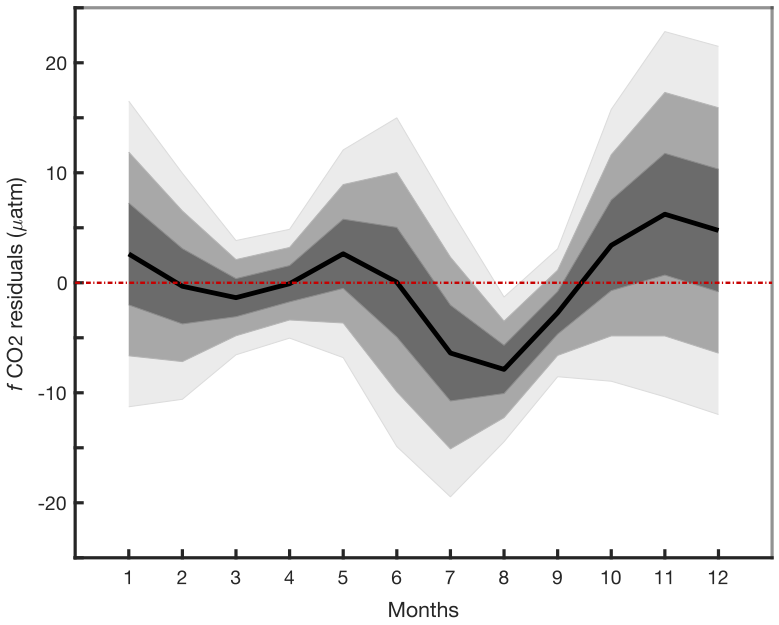


**Figure S3.** Seasonally-varying uncertainty from the CMIP5-based MLR analyses, based on the 5° spatially-dividing method and calculated using the monthly $fCO_{2-\text{residuals}}$ over 1992-2014. The black line corresponds to the multi-model mean of monthly average $fCO_{2-\text{residuals}}$. The dark, medium and light grey shadings correspond, respectively, to the 1, 2 and $3\sigma$ of the residuals across the 19 monthly averages of $fCO_{2-\text{residuals}}$. The dashed red line indicates the zero level. The MLR adds a significant negative bias in the mean seasonal signal of August, suggesting that the MLR method developed here is not suitable to provide robust monthly uncertainties.

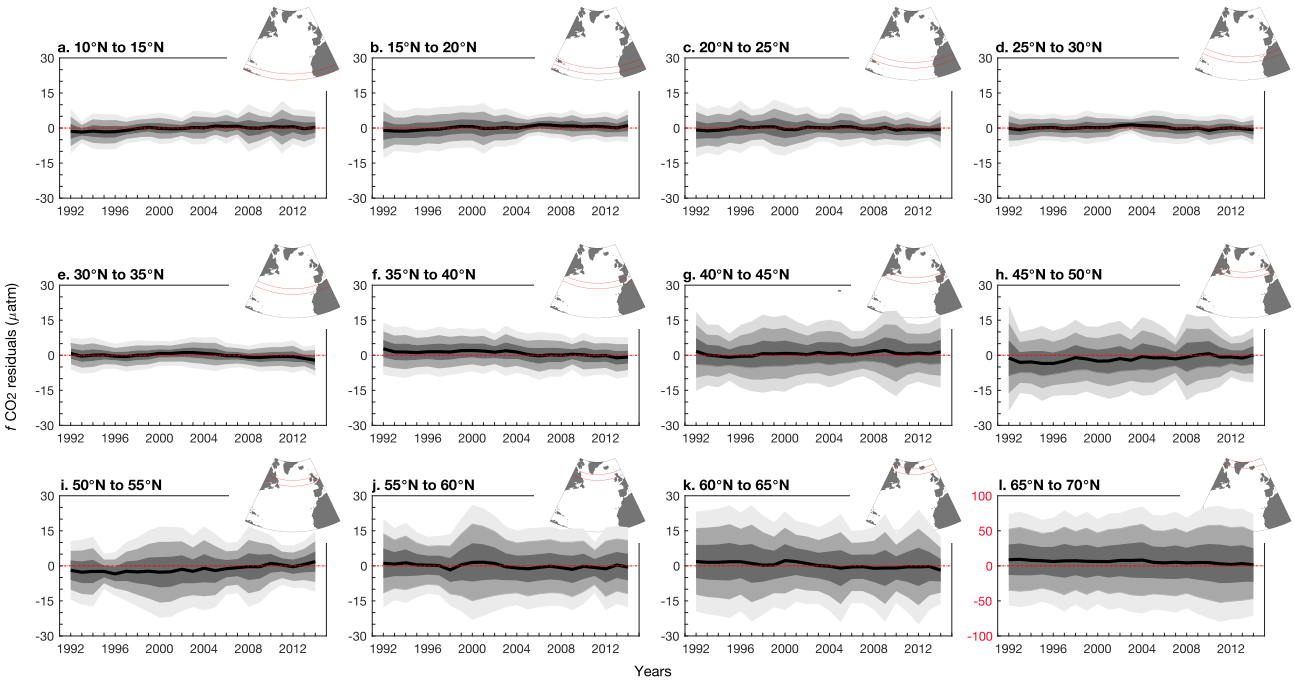


**Figure S4.** Spatial investigation of the annually-varying uncertainty from the CMIP5-based MLR analyses, from the 5° latitudinally-dividing method. Each panel corresponds to the results for each sub-region of the North Atlantic, which is indicated by the red lines on the corresponding map. The black line is the multi-model mean of annual average $fCO_{2-\text{residuals}}$ (Equation 3 in the main text). The dark, medium and light grey shadings correspond, respectively, to the 1, 2 and $3\sigma$ across the 19 annual averages of $fCO_{2-\text{residuals}}$ (Equation 4 in the main text). The dashed red line indicates the zero level. Note that the last panel (l. from 65°N to 70°N) displays a y-axis range that does not follow the range of other panels.

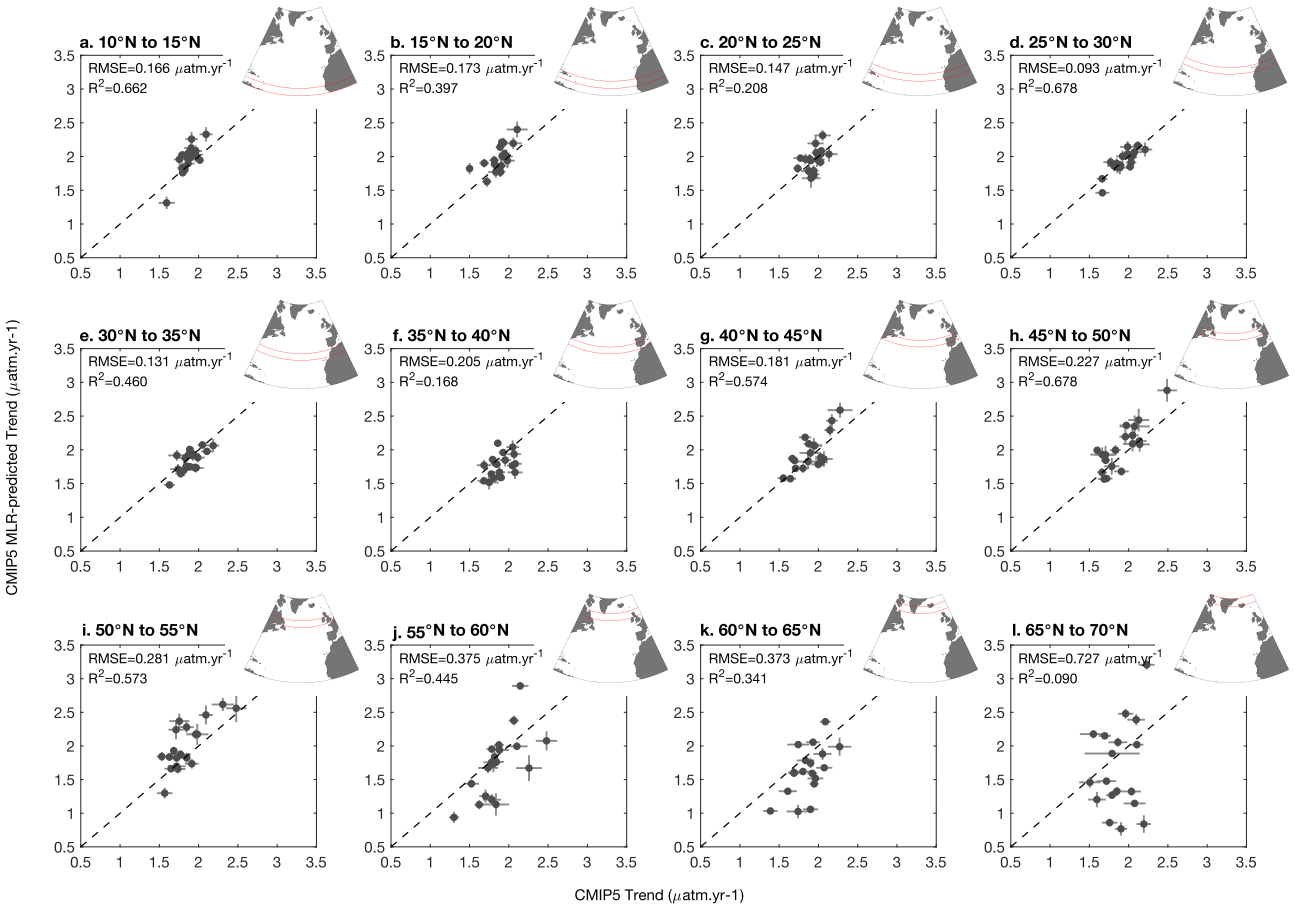


**Figure S5.** Spatial investigation of the trend uncertainty from the CMIP5-based MLR analyses, from the 5° latitudinally-dividing method. Each panel corresponds to the results for each sub-region of the North Atlantic, which is indicated by the red lines on the corresponding map. North Atlantic $fCO_{2-ocean}$ linear trends calculated from annual means over the period 1992-2014 from the CMIP5-based MLR product ($\Gamma_{MLR-predicted}$) versus the CMIP5 model-truth value ($\Gamma_{model-true}$) (circles). The corresponding standard error of each linear trend (returned by the linear fit applied to the annual means for the sub-region of study) is indicated by the vertical and horizontal lines. Each panel includes the $R^2$ value and the RMSE between $\Gamma_{MLR-predicted}$ and $\Gamma_{model-true}$. The dashed line indicates the 1-to-1 line. Points above the 1-to-1 line indicate that, for the corresponding CMIP5 models, the MLR overestimates the model-truth $fCO_{2-ocean}$ trend in the subregion of study.

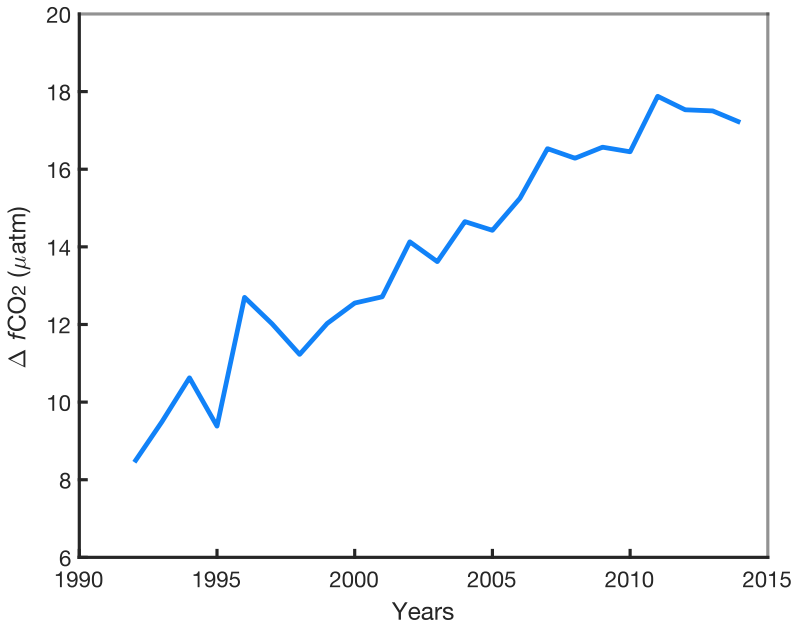


**Figure S6.** Annually-varying $\Delta fCO_2$ in the North Atlantic over 1992-2014, calculated from difference between the annually-varying $fCO_{2-\text{atmopshere}}$ and the annually-varying $fCO_{2-\text{ocean}}$ predicted by the 5° spatially-dividing observation-based MLR method (all annual averages are deduced from monthly area-weighted means for the basin-wide North Atlantic).

**Table S1.** Description and processing steps of observational-based products. The term "surface fluxes" corresponds to: temperature at 2 m, specific humidity at 2 m, surface shortwave and longwave radiation downwards, total precipitation, snowfall, x-direction and y-direction wind components at 10 m.

| Variable | Product | Description | Processing | Reference |
|---|---|---|---|---|
| $fCO_{2-ocean}$ | SOCATv4 | 1°×1°<br>Monthly<br>1970-2015 | Waters < 1,000 m depth removed | (Bakker et al., 2016)<br>(Sabine et al., 2013) |
| | Statistical gap-filling | 1°×1°<br>Monthly<br>1985-2017 | Waters < 1,000 m depth removed | (Jones et al., 2019) |
| $pCO_{2-ocean}$ | Neural-Network | 1°×1°<br>Monthly<br>1982-2015 | Waters < 1,000 m depth removed | (Landschützer et al., 2016, 2017) |
| | A MLR method | 1°×1°<br>Monthly<br>1990-2015 | Waters < 1,000 m depth removed | (Iida et al., 2015) |
| SST | OISST v2 | 1°x1°<br>Monthly<br>1981-present | | (Reynolds et al., 2007) |
| MLD | ECCO2 | 0.25°×0.25°<br>Daily<br>1992-present | Monthly averaged<br>Interpolated into SOCATv4 grid | (Menemenlis et al., 2008) |
| $xCO_2$ | GLOBALVIEW-$CO_2$ | Function of latitude<br>8-day frequency<br>1979-2014 | Interpolated into daily<br>Averaged into monthly<br>Gridded to SOCATv4 grid | (GLOBALVIEW-CO2, 2013) |
| SLP | NCEP/NCAR Reanalysis 1 | 2.5°×2.5°<br>Monthly<br>1948-present | Interpolated to SOCATv4 grid<br>Averaged into monthly<br>Adjusted to SOCATv4 grid | (Kalnay et al., 1996) |
| Wind Speed W | NCEP/NCAR Reanalysis 1 | 0.25°×0.25°<br>6-hourly<br>1948-present<br>$0.995\sigma$ level | Calculated $W^2$<br>$W^2$ monthly average<br>Adjusted to SOCATv4 grid | (Kalnay et al., 1996) |
| Bathymetry | ETOPO1 | 1' arc resolution | Adjusted to SOCATv4 grid<br>Adjusted to the "gap-filling" grid | (Amante & Eakins, 2015) |
| DIC, TA Climatologies | GLODAPv2 | 1°x1°<br>33 depth levels | Adjusted to SOCATv4 grid | (Lauvset et al., 2016) |
| Surface fluxes | ERA-Interim | 0.75°×0.75°<br>Daily<br>1979-2015 | Adjusted to ORCA1 grid | (Dee et al., 2011) |

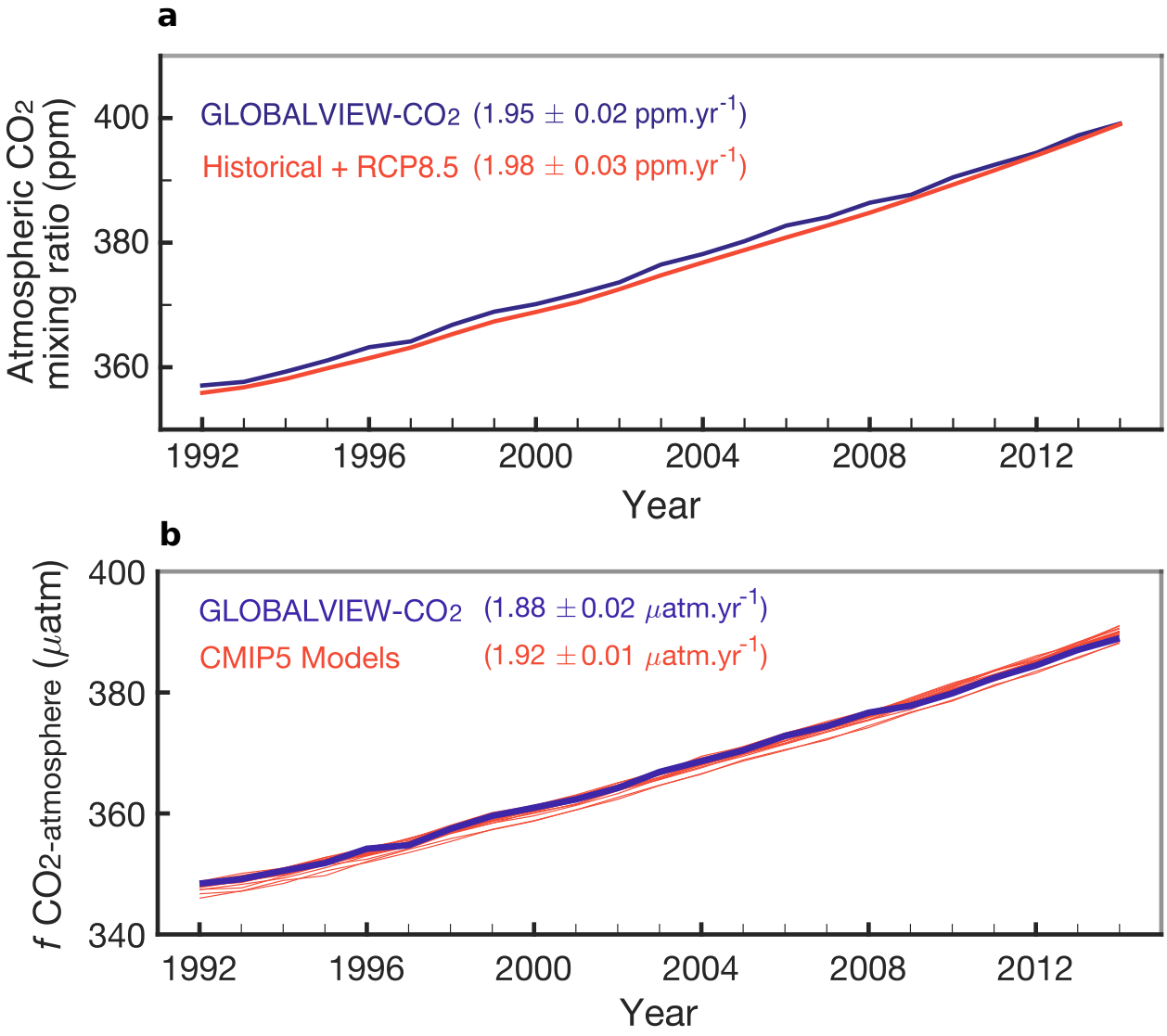


**Figure S7.** North Atlantic annually-varying atmospheric (a) $xCO_2$ and (b) $fCO_2$ from GLOBALVIEW-CO2 (GLOBALVIEW-CO2, 2013) and the "historical + RCP8.5" (Riahi et al., 2007; Taylor et al., 2012). Trends indicated in the plots are calculated over the period 1992-2014, and where the uncertainty corresponds to the standard error returned by the linear fit (except for the CMIP5 models on panel (b) where the uncertainty corresponds to the standard deviation across the models' trends). Note that for the period 1992-2005 the trend in atmospheric $xCO_2$ is 1.83 ± 0.04 ppm.yr$^{-1}$ and 1.80 ± 0.03 ppm.yr$^{-1}$ for GLOBALVIEW-CO2 and the CMIP5 historical datasets, respectively (where the uncertainty also corresponds to the standard error returned by the linear fit).

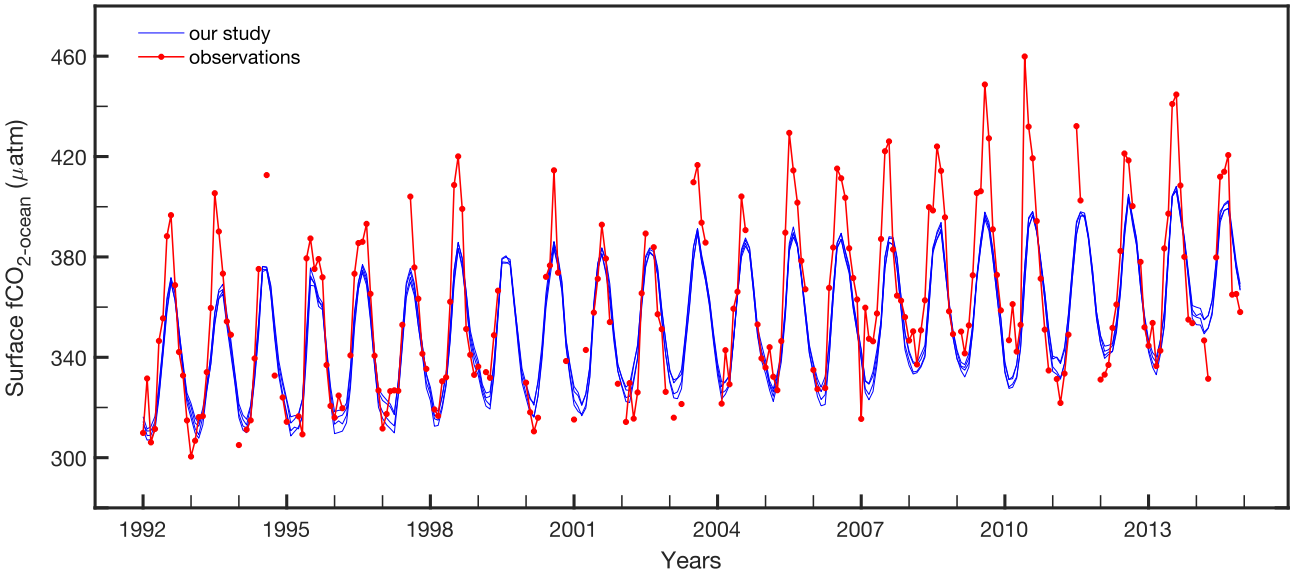


**Figure S8.** Monthly surface $fCO_{2-ocean}$ at the BATS station from the observations (red) (Bates et al., 2014) and produced by the 5° spatially-dividing observation-based MLR method (this study; blue). Specifically, the observed $fCO_{2-ocean}$ was generated using the CO2SYS Matlab toolbox (Lewis & Wallace, 1998) and the MLR-predicted values were taken for four grid cells neighbouring the BATS stations (box delimited by 31.5°N and 32.5°N and -64.5°W and -63.5°W).

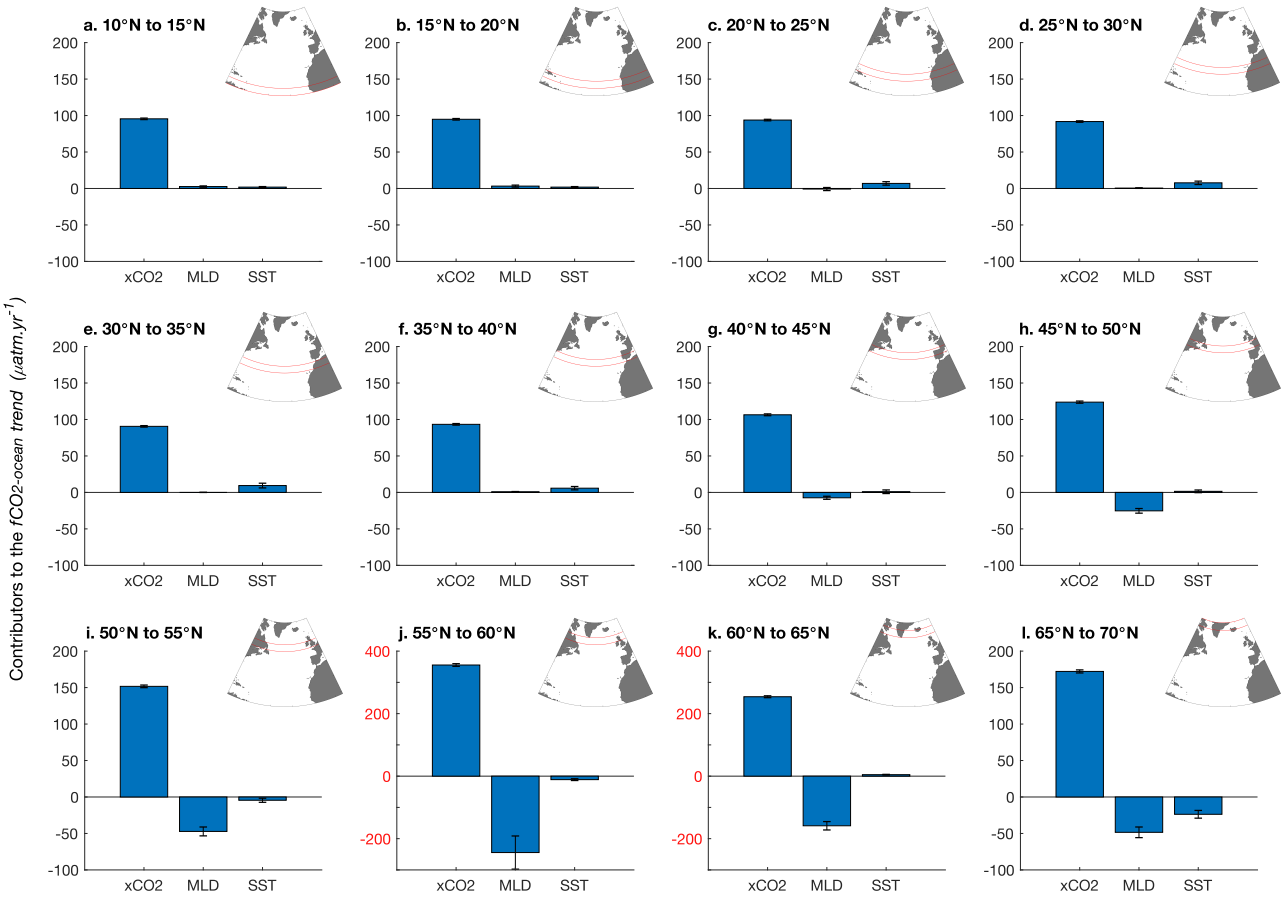


**Figure S9.** Spatial investigation of the contributors (atmospheric $xCO_2$, MLD, SST) of the observation-based surface $fCO_{2-\mathrm{ocean}}$ regional trend, from the 5° latitudinally-dividing method. Each panel corresponds to the results for each sub-region of the North Atlantic, which is indicated by the red lines on the corresponding map. Note that panels (j) and (k) have different y-axis limits. See Section 3.2 for details on determining the drivers of the trend.

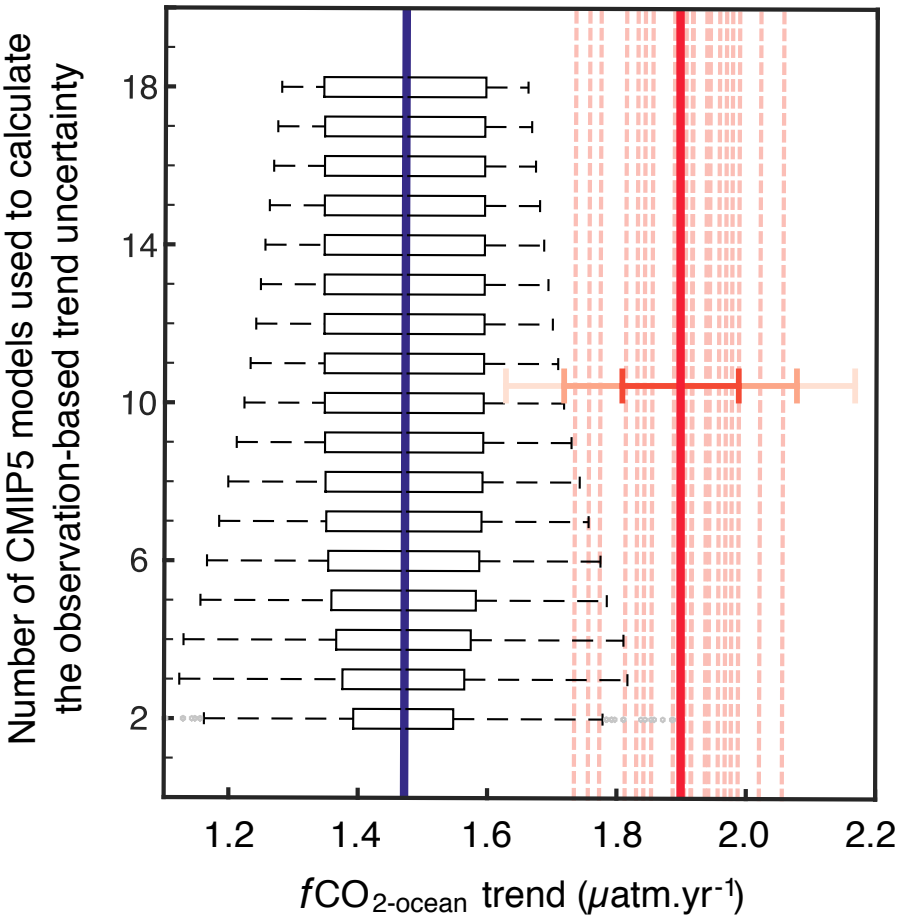


**Figure S10.** A sensitivity analysis on the $fCO_2$ trend uncertainty. $fCO_{2-ocean}$ trends for the 19 CMIP5 models (orange) and from the observation-based MLR results (blue). Each box plot contains the $\pm$ (1, 2, 3) $\sigma_{C_k^M}$ trend uncertainties (Equation 9 in main text), relative to the observation-based trend, using all the possible combination of results from the CMIP5-based MLR analysis. The y-axis indicates the total number of models that were included in the trend uncertainty analysis. Trends are calculated for the North Atlantic and for the period 1992-2014.

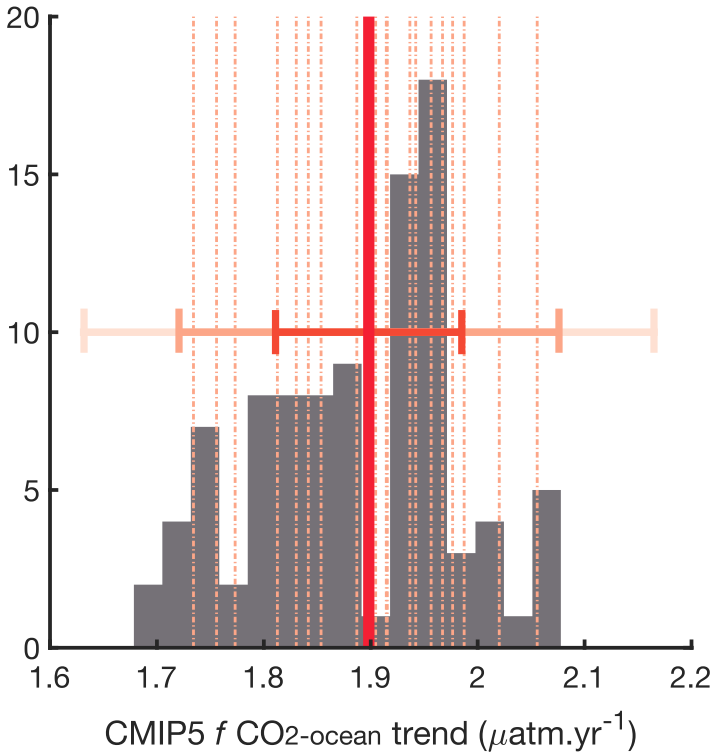


**Figure S11.** North Atlantic $fCO_{2-ocean}$ trends in the CMIP5 models over the period 1992-2014 (dashed lines), with model-mean (thick orange) and inter-model 1, 2 and 3 standard deviation (error bars). Trends calculated over 22 and 21 year-long intervals (1992-2013, 1993-2014, 1994-2014, 1992-2012) across the 19 CMIP5 models are displayed by the grey histogram.

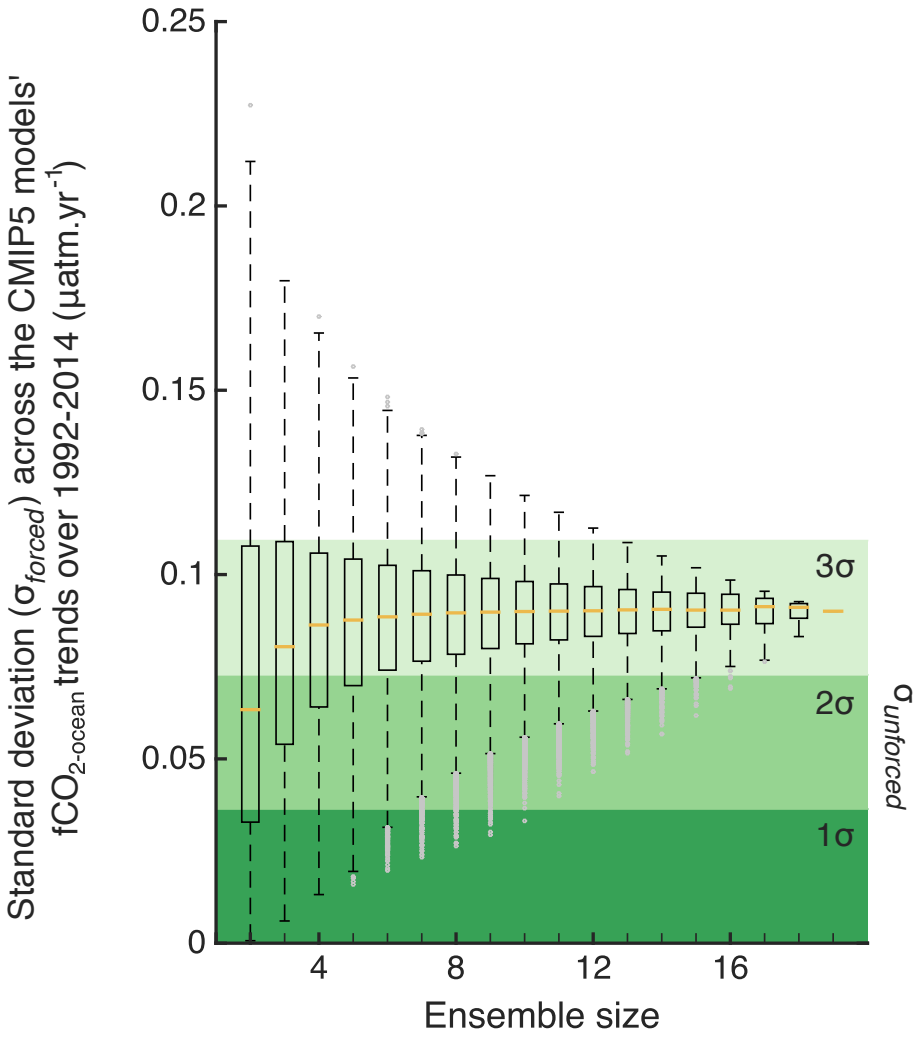


**Figure S12.** Box plot of the standard deviations $\sigma_{forced}$ across the $fCO_{2-\text{ocean}}$ linear trends calculated in 1992-2014 (using the historical and RCP8.5 experiments) when considering different ensemble sizes (i.e. the total number of models included in the standard deviation calculations) and all the possible combinations of models given an ensemble size. The dark to light green bands correspond to 1, 2 and $3\sigma_{unforced}$ of those $pCO_{2-\text{ocean}}$ trends, respectively from Figure 6. See Text S4 for further discussion.

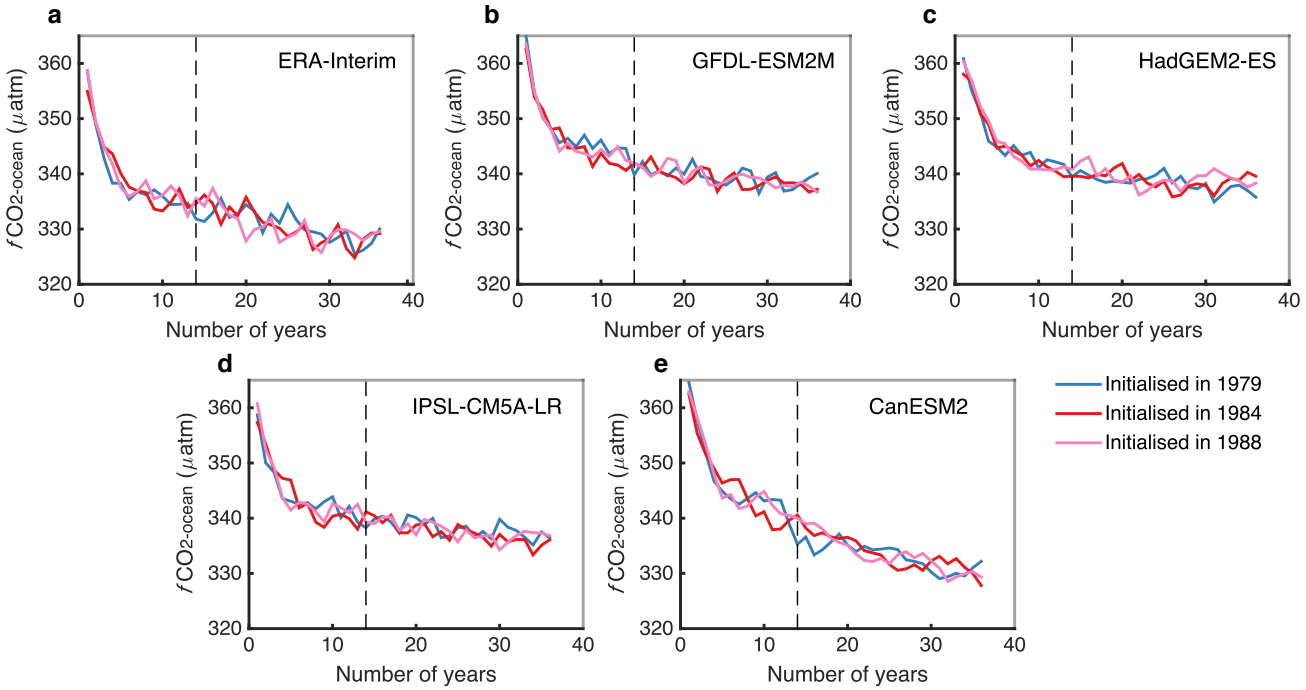


**Figure S13.** Quantification of the model-drift in the North Atlantic surface $fCO_{2-ocean}$. North Atlantic (area-weighted) $fCO_{2-\mathrm{ocean}}$ annual time series calculated from the simulations for which the atmospheric $CO_2$ mixing ratio was held to 336.85 ppm and were forced with (a), ERA-Interim, (b), GFDL-ESM2M, (c), HadGEM2-ES, (d), IPSL-CM5A-LR and (e), CanESM2 surface conditions. For each of the five experiments, three simulations (ensemble members) were made with three different starting and therefore initialisation date: January 1979 (blue), January 1984 (red) and January 1988 (pink). The first 13 years (left to the dashed line) correspond to the spinup phase, which are excluded from the model-drift assessment. The model-drift is quantified using the data on the right of the dashed line. For each simulation, the corresponding model-drift (i.e. linear trend and corresponding standard error) is indicated in Table S4.

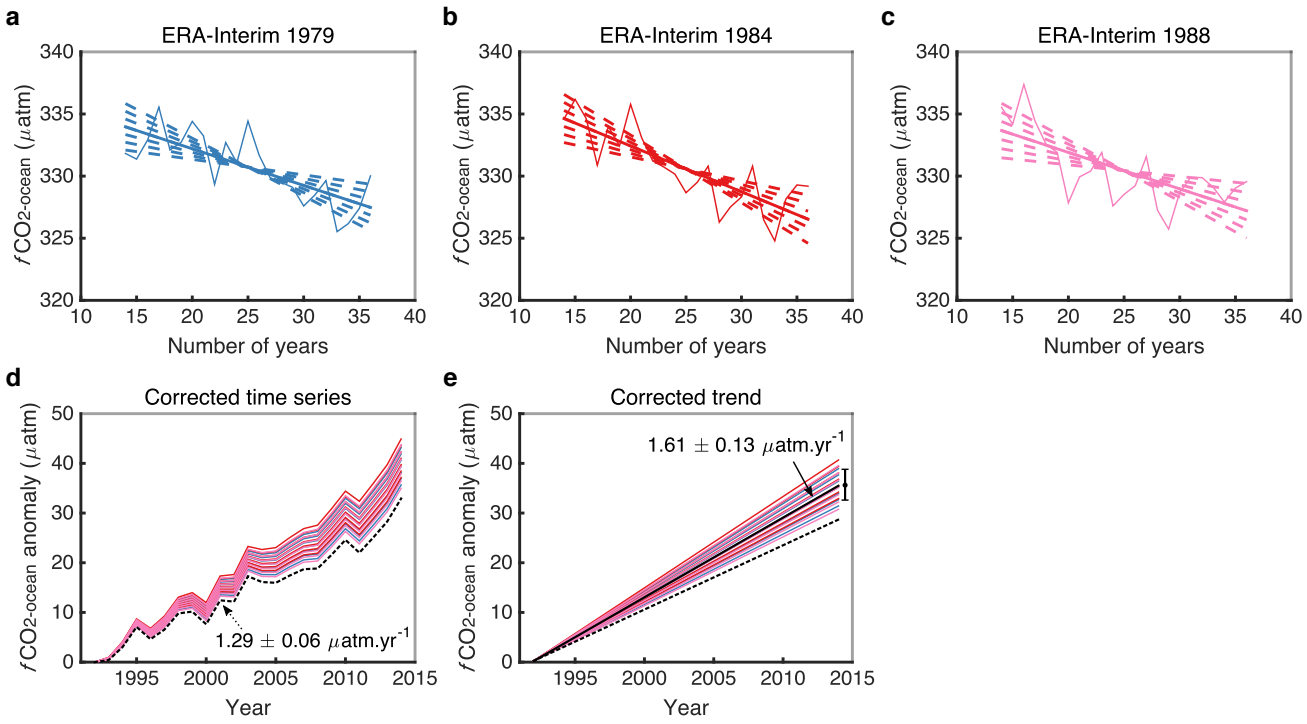


**Figure S14.** Example of the model drift correction steps for the ERA-Interim forced simulation. The figure description is detailed in Text S4. The coloured lines in (d) and (e) refer to the different starting year of the model drift simulations, as indicated in (a), (b) and (c). All annual values were computed for the area-weighted North Atlantic, without shelf waters.

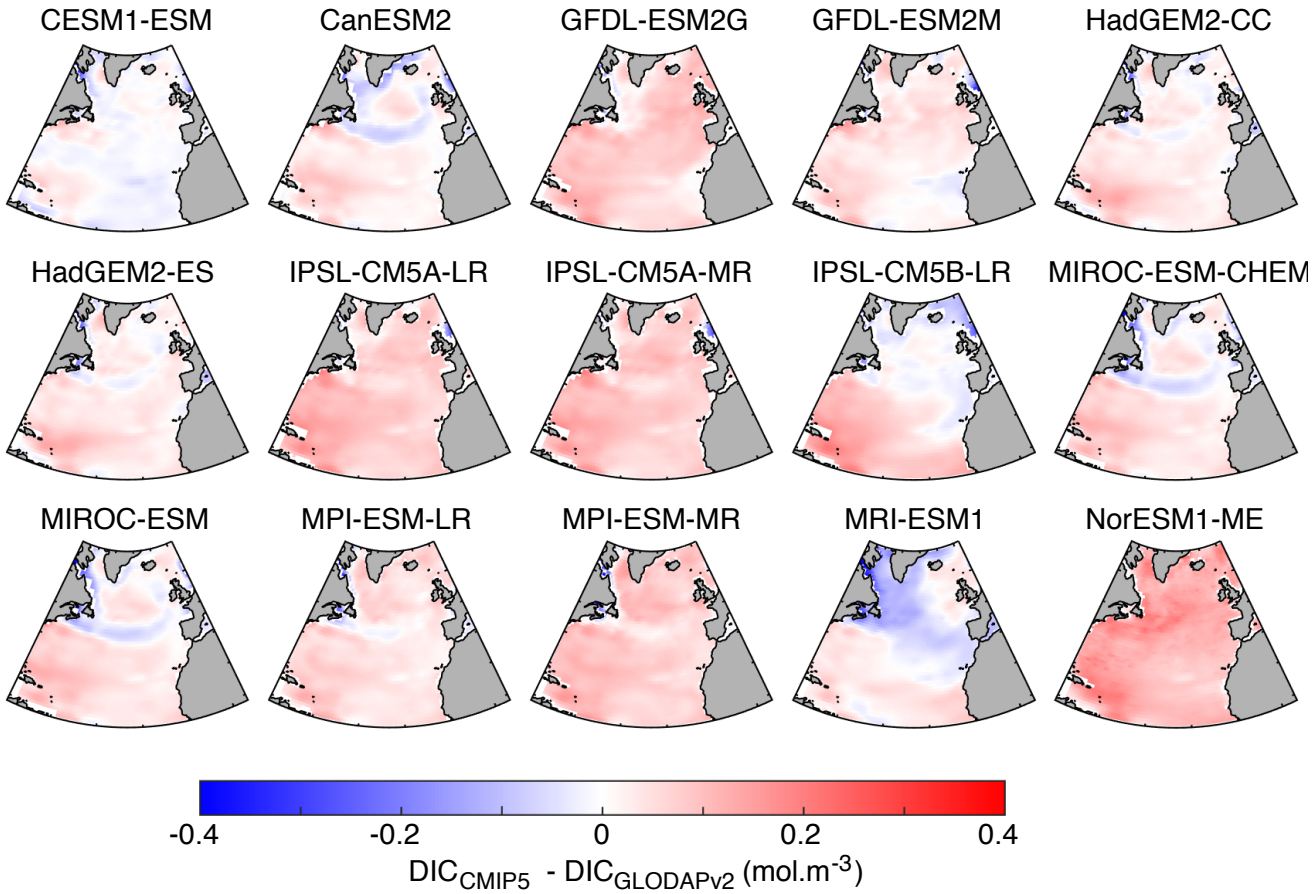


**Figure S15.** Figure S13: Difference between each CMIP5 model and GLODAPv2 DIC climatologies (Lauvset et al., 2016) at 10 m depth in the North Atlantic. The GLODAPv2 climatology for surface DIC used available DIC measurements collected over 1972-2013 for mapping and was then normalised to 2002 (Lauvset et al., 2016). The interval 1996-2008 was chosen to calculate the DIC climatology in the CMIP5 models (Table S2), as it is centred in 2002 and is a 13-year long interval, which is a similar length as the other climatologies calculated within GLODAPv2 for the variables that are potentially being affected by the anthropogenic change.

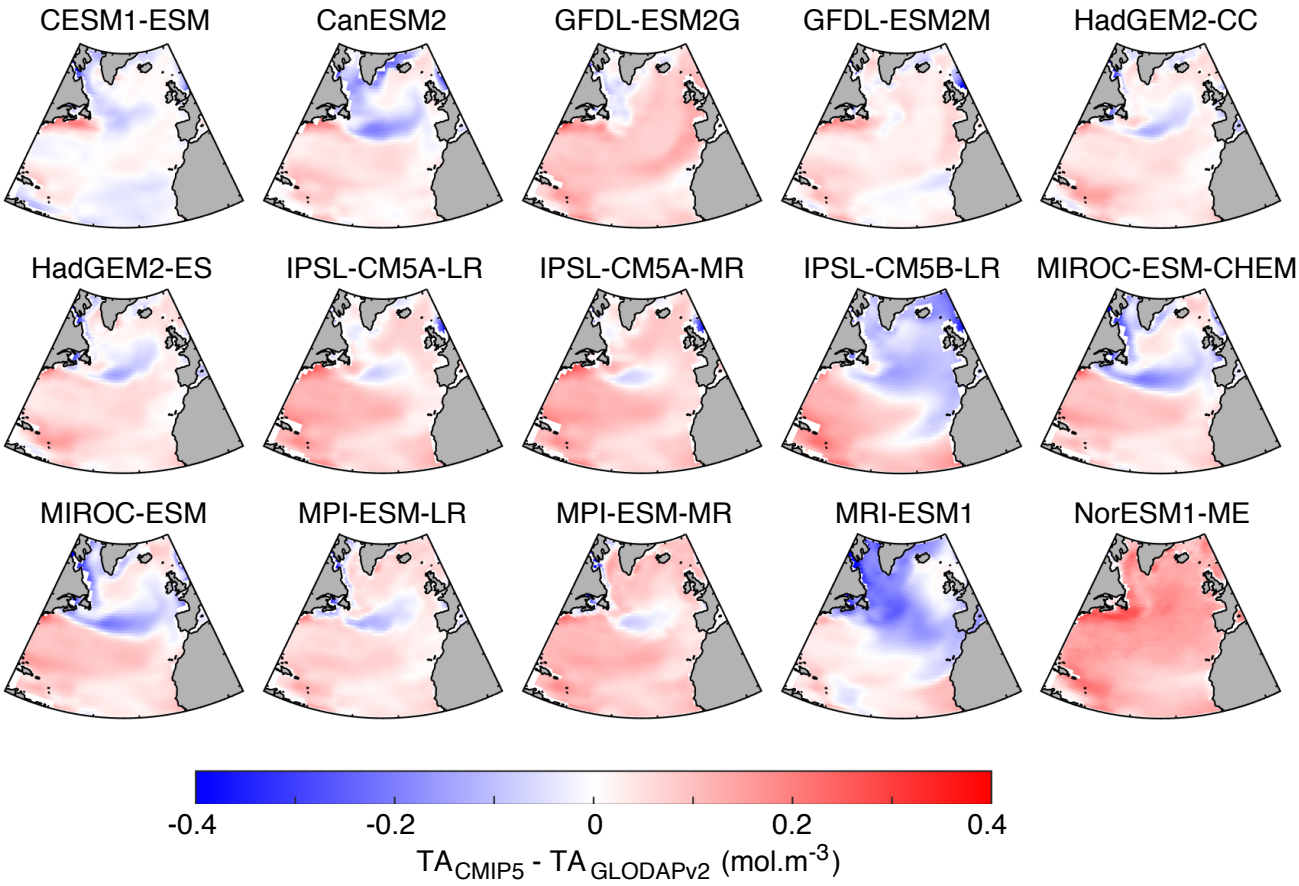


**Figure S16.** Difference between each CMIP5 model and GLODAPv2 TA climatologies (Lauvset et al., 2016) at 10 m depth in the North Atlantic. The GLODAPv2 climatology for surface TA used available TA measurements collected over 1972-2013 for mapping (Lauvset et al., 2016). The interval 1972-2013 was chosen to calculate the TA climatology in the CMIP5 models.

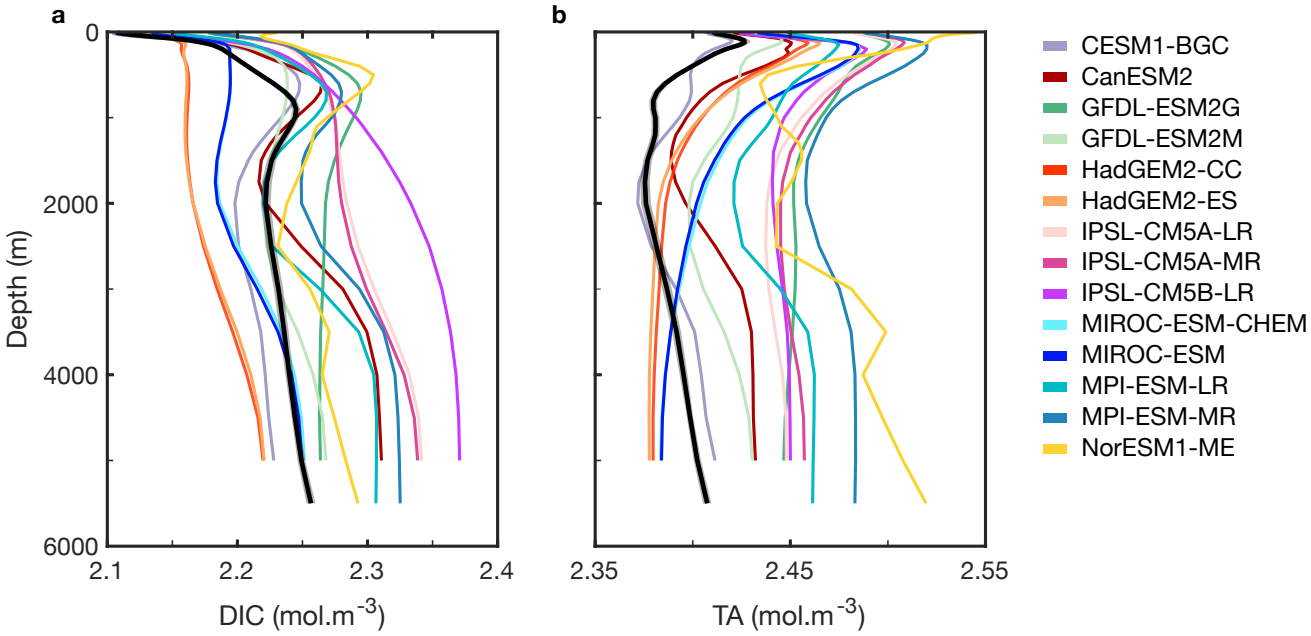


**Figure S17.** (a), DIC and (b), TA area-weighted mean profile in the North Atlantic from the CMIP5 models, and from GLODAPv2 (thick black) (Lauvset et al., 2016). Note that the area-weighted mean interpolation error is displayed around the observation-based climatology profiles but is indiscernible from the mean profile. From the 15 CMIP5 models that provided yearly (4D) DIC and TA fields in the online portal, 14 are displayed here. Indeed, the MRI-ESM1 presented unrealistic profiles, which were linked to issues with the regridding step, particularly because the model used a land mask that was filled with zeros instead of a common "not-a-number" mask (impacting the coastal DIC and TA values, and hence the basin-wide mean profile).

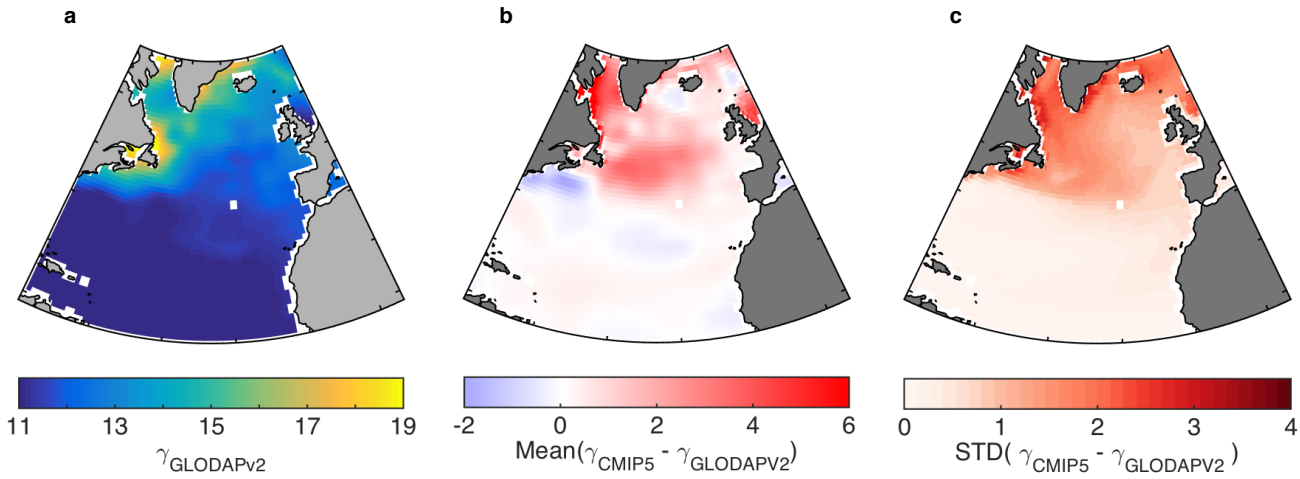


**Figure S18.** (a), Revelle factor based on the GLODAPv2 DIC and TA climatological binned values (Lauvset et al., 2016), (b), Mean difference between the 15 CMIP5 models and the observation-based Revelle factor (Table S2), with the inter-model difference variability shown in (c). The Revelle factor was calculated using the Sarmiento and Gruber (2006) definition.

**Table S2.** List of the CMIP5 models used in this study. Details about the data processing of CMIP5 models' outputs (e.g. download, conversions, regridding) are provided in Text S1, under "CMIP5 model data".

| Model Name | Biogeochemical Model | Available variables | Reference |
|---|---|---|---|
| | | $pCO_{2-ocean}$ picontrol / DIC-TA | |
| **CESM1-BGC** | **BEC** | ✓/ ✓ | (Gent et al., 2011) |
| National Center for Atmospheric Research | | | (Long et al., 2011) |
| **CanESM2** | **CMOC** | ✓/ ✓ | (Zahariev et al., 2008) |
| Canadian Centre for Climate Modelling and Analysis | | | (Chylek et al., 2011) |
| **GFDL-ESM2G** | **TOPAZ2** | - / ✓ | (Dunne et al., 2012) |
| Geophysical Fluid Dynamics Laboratory | | | (Dunne et al., 2013) |
| **GFDL-ESM2M** | **TOPAZ2** | ✓/ ✓ | (Dunne et al., 2012) |
| Geophysical Fluid Dynamics Laboratory | | | (Dunne et al., 2013) |
| **GISS-E2-H-CC** | **NOBM** | ✓/ - | (Sun & Bleck, 2006) |
| NASA Goddard Institute for Space Studies | | | (Schmidt et al., 2014) |
| | | | (Romanou et al., 2014) |
| **GISS-E2-R-CC** | **NOBM** | ✓/ - | (Hansen et al., 2007) |
| NASA Goddard Institute for Space Studies | | | (Schmidt et al., 2014) |
| | | | (Romanou et al., 2014) |
| **HadGEM2-CC** | **diat-HadOCC** | ✓/ ✓ | (Palmer & Totterdell, 2001) |
| Met Office Hadley Centre | | | (Collins et al., 2011) |
| **HadGEM2-ES** | **diat-HadOCC** | ✓/ ✓ | (Palmer & Totterdell, 2001) |
| Met Office Hadley Centre | | | (Collins et al., 2011) |
| **IPSL-CM5A-LR** | **PISCES** | - / ✓ | (Aumont & Bopp, 2006) |
| Institut Pierre-Simon Laplace | | | (Dufresne et al., 2013) |
| **IPSL-CM5A-MR** | **PISCES** | - / ✓ | (Aumont & Bopp, 2006) |
| Institut Pierre-Simon Laplace | | | (Dufresne et al., 2013) |
| **IPSL-CM5B-LR** | **PISCES** | - / ✓ | (Aumont & Bopp, 2006) |
| Institut Pierre-Simon Laplace | | | (Dufresne et al., 2013) |
| **MIROC-ESM-CHEM** | **NPZD** | ✓/ ✓ | (Kawamiya et al., 2000) |
| Japan Agency for Marine-Earth Science and Technology | | | (Sudo et al., 2002) |
| | | | (Watanabe et al., 2011) |
| **MIROC-ESM** | **NPZD** | ✓/ ✓ | (Kawamiya et al., 2000) |
| Japan Agency for Marine-Earth Science and Technology | | | (Watanabe et al., 2011) |
| **MPI-ESM-LR** | **HAMOCC5.2** | ✓/ ✓ | (Giorgetta et al., 2013) |
| Max Planck Institute for Meteorology | | | (Ilyina et al., 2013) |
| | | | (Jungclaus et al., 2013) |
| **MPI-ESM-MR** | **HAMOCC5.2** | ✓/ ✓ | (Giorgetta et al., 2013) |
| Max Planck Institute for Meteorology | | | (Ilyina et al., 2013) |
| | | | (Jungclaus et al., 2013) |
| **MRI-ESM1** | **NPDZ** | - / ✓ | (Yukimoto et al., 2011) |
| Meteorological Research Institute | | | (Adachi et al., 2013) |
| **NorESM-ME** | **HAMOCC5.1** | ✓/ ✓ | (Bentsen et al., 2013) |
| Norwegian Climate Centre | | | (Tjiputra et al., 2013) |
| **bcc-csm1-1-m** | **MOM4 FMS** | ✓/ - | (Wu et al., 2008) |
| Beijing Climate Centre | | | (Wu et al., 2013) |
| **bcc-csm1-1** | **MOM4 FMS** | ✓/ - | (Wu et al., 2008) |
| Beijing Climate Centre | | | (Wu et al., 2013) |

**Table S3.** Mixed Layer Depth processing. ($2^{nd}$ column), Total number of depth levels within each model. ($3^{rd}$ column), Approximated depth value at level 40 or at the last (deepest) available level. ($4^{th}$ column), area-weighted mean of the global MLD from 1992-2014 (same period as the used observational-based product), with standard deviation from the area-weighted mean.

| **Model** | **Vertical resolution** | **Depth at level 40 or last available (m)** | **Mean MLD (m)** |
|---|---|---|---|
| CESM1-BGC | 60 | 985 | 78 ± 67 |
| CanESM2 | 40 | 5233 | 83 ± 93 |
| GFDL-ESM2G | 50 | 2049 | 80 ± 111 |
| GFDL-ESM2M | 50 | 2049 | 84 ± 119 |
| GISS-E2-H-CC | 33 | 5500 | 78 ± 121 |
| GISS-E2-R-CC | 32 | 4887 | 92 ± 170 |
| HadGEM2-CC | 40 | 5328 | 78 ± 73 |
| HadGEM2-ES | 40 | 5328 | 78 ± 71 |
| IPSL-CM5A-LR | 31 | 5250 | 77 ± 102 |
| IPSL-CM5A-MR | 31 | 5250 | 78 ± 102 |
| IPSL-CM5B-LR | 31 | 5250 | 76 ± 91 |
| MIROC-ESM-CHEM | 44 | 4525 | 90 ± 99 |
| MIROC-ESM | 44 | 4525 | 91 ± 97 |
| MPI-ESM-LR | 40 | 5720 | 84 ± 111 |
| MPI-ESM-MR | 40 | 5720 | 81 ± 108 |
| MRI-ESM1 | 51 | 2500 | 66 ± 94 |
| NorESM1-ME | 70 | 1150 | 125 ± 112 |
| bcc-csm1-1-m | 40 | 5334 | 99 ± 178 |
| bcc-csm1-1 | 40 | 5334 | 90 ± 126 |

**Table S4.** Linear trends ($\mu$atm·yr$^{-1}$) of the model drift in the North Atlantic $fCO_{2-ocean}$ calculated from year number 14 to 36 (Figure S11) for the fifteen model drift simulations, a period that corresponds to a 23-year period long similar to the interval of study 1992-2014 (for the changing atmospheric $CO_2$ runs; i.e. the $1^{st}$ ensemble member). Columns refer to the surface forcing field experiments and lines to the starting year of the simulation (Figure S11). The uncertainties correspond to the standard error returned by each linear fit. Significant trends are in bold, at the 5% significance level (a trend is significant when its p-value $< 0.05$, from F-statistics).

| | **ERA-Interim** | **GFDL-ESM2M** | **HadGEM2-ES** | **IPSL-CM5A-LR** | **CanESM2** |
|---|---|---|---|---|---|
| 1979 | **-0.29 ± 0.06** | **-0.11 ± 0.04** | **-0.15 ± 0.03** | **-0.14 ± 0.04** | **-0.29 ± 0.04** |
| 1984 | **-0.37 ± 0.06** | **-0.18 ± 0.04** | -0.08 ± 0.05 | **-0.24 ± 0.03** | **-0.42 ± 0.03** |
| 1988 | **-0.29 ± 0.07** | **-0.19 ± 0.04** | -0.10 ± 0.05 | **-0.15 ± 0.04** | **-0.44 ± 0.04** |

**Table S5.** Simulated trends in the North Atlantic surface $fCO_{2-ocean}$. For the ERA-Interim forced and the CMIP5-forced experiments, $fCO_{2-\mathrm{ocean}}$ trends calculated over the period 1992-2014 prior and after model drift corrections. For the non drift-corrected trends, the uncertainty corresponds to the standard error returned by the linear fit, while for the drifted-corrected trends, the uncertainty corresponds to the standard deviation across the 21 trends identified from the model-drift analysis (c.f. Text S4 for details on the model-drift correction steps).

| **Forcing field** | **Non drift corrected trends** ($\mu$**atm·yr**$^{-1}$) | **Drift-corrected trends** ($\mu$**atm·yr**$^{-1}$) |
|---|---|---|
| ERA-Interim | 1.29 ± 0.06 | 1.61 ± 0.13 |
| GFDL-ESM2M | 1.36 ± 0.05 | 1.52 ± 0.09 |
| HadGEM2-ES | 1.43 ± 0.04 | 1.54 ± 0.09 |
| IPSL-CM5A-LR | 1.25 ± 0.05 | 1.43 ± 0.04 |
| CanESM2 | 1.29 ± 0.05 | 1.68 ± 0.11 |