



**Reconstruction and Multiyear Prediction of Global Air-sea CO₂ Flux using NOAA GFDL's
SPEAR Ensemble**

Xiao Liu^{*,1}, John P. Dunne¹, Liwei Jia¹, Charles A. Stock¹, Xiaosong Yang¹, Matthew J. Harrison¹,
Liping Zhang^{1,2}, Anthony Rosati³

¹The Geophysical Fluid Dynamics Laboratory (GFDL), National Oceanic and Atmospheric
Administration (NOAA); Princeton, NJ, USA

²University Corporation for Atmospheric Research, Boulder, Colorado, U.S

³Independent Researcher

*Correspondence to: Xiao Liu (xiao.liu@noaa.gov)

In preparation for *Earth System Dynamics*

Special Issue: "Advances in annual to decadal prediction, projection, and attribution"



Abstract. The ocean acts as a critical carbon sink, but its efficiency in absorbing anthropogenic CO₂ varies significantly over multiyear to decadal timescales. Accurately predicting this variability is essential for anticipating atmospheric CO₂ growth and establishing the unperturbed baselines necessary for verifying climate mitigation efforts, such as marine Carbon Dioxide Removal (mCDR) activities. This study introduces a fully coupled physical-biogeochemical prediction framework, which integrates the NOAA GFDL Seamless System for Prediction and Earth System Research (SPEAR) with the COBALTv3 ocean biogeochemical model. We conducted ensembles of uninitialized historical simulations, data-assimilative reconstructions, and retrospective initialized predictions of global air-sea CO₂ flux, evaluating their skill against observation-constrained products for a recent 40-year period (1984-2023).

The uninitialized ensemble skillfully predicts the amplitude of the observed historical increases, but struggles to resolve multiyear and decadal variability. We show that initialization significantly improves prediction skill. Globally, skill is enhanced for lead times up to two years, extending up to five years in specific higher-latitude regions. Through Average Predictability Time (APT) analysis, we isolated distinct physical drivers of the dominant predictability. We find that skillful predictions up to two years are primarily governed by the El Niño–Southern Oscillation (ENSO) and its impact on tropical upwelling. Moreover, we identified a multidecadal, potentially predictable signal linked to long-term changes in Eastern Boundary Current upwelling, as well as Southern Ocean mixed layer depth and sea surface temperature. However, verifying this long-term potential predictability remains fundamentally constrained by the sparsity of multidecadal observational records in these remote or nearshore regions. This underscores the critical need for sustained, optimized ocean observations and an improved understanding of the uncertainties associated with existing observational data.

Short summary. This study uses NOAA GFDL’s SPEAR model to improve predictions of how the ocean absorbs human emitted CO₂. While standard models capture long-term trends, they struggle with yearly fluctuations. Researchers found that "initializing" the model with observed data significantly boosts their prediction accuracy for up to two years. This framework helps separate the predictions of natural variability from human-driven carbon changes, a vital step for verifying future marine carbon removal efforts.



1. Introduction

The world's ocean and terrestrial systems together absorb roughly 50% of the anthropogenic CO₂ released into the atmosphere (Friedlingstein et al., 2025a; Gruber et al., 2019). The efficiency of these natural carbon sinks, however, is highly variable, fluctuating on timescales ranging from months to decades (Landschützer et al., 2015; McKinley et al., 2020). This variability dictates how much emitted CO₂ remains in the atmosphere with the associated uncertainty limiting understanding of acidification and radiative warming. As emissions continue to grow, understanding historical patterns and predicting near-future changes in carbon sinks are increasingly vital for disentangling natural variability from human-driven processes, as well as for anticipating ecosystem tipping points to guide management decisions (Ilyina et al., 2021; Lovenduski et al., 2019).

Since 2007, an enormous international effort has been undertaken to integrate observational datasets, reanalysis products, machine-learning algorithms, and numerical models to produce the Global Carbon Budget (GCB; see Friedlingstein et al., 2025a for GCBv2024), which provides historical estimates of ocean and terrestrial carbon sinks. Despite numerous new insights enabled by the GCB, large uncertainties in the carbon budget remain. A scarcity of direct measurements, challenges with the quality controls of autonomous platform proxies, and systematic discrepancies between data and model products call for continuous efforts to constrain uncertainties in carbon sink reconstructions (Claustre et al., 2020; Fay and McKinley, 2021; Gloege et al., 2021; Hauck et al., 2020).

Recent GCB estimates indicate that the ocean has become the largest net sink of CO₂ over the past decade as the efficiency of the terrestrial sink has declined (Friedlingstein et al., 2025b). While much of the year-to-year variation in both sinks is driven by El Niño Southern Oscillation (ENSO), the ocean sink also fluctuates on longer timescales (Li and Ilyina, 2018; Séférian et al., 2018). This underlying predictability is largely due to the ocean's thermal and chemical inertia, making multiyear to decadal predictions both a viable scientific goal and a necessary step toward predicting atmospheric CO₂ growth to guide effective policymaking.

Furthermore, the growing interest in marine Carbon Dioxide Removal (mCDR) strategies, such as alkalinity enhancement or iron fertilization to artificially boost the efficiency of the ocean sink (Doney et al., 2025), brings new urgency to constraining natural ocean carbon sinks. Verifying the additionality and durability of these proposed efforts requires a thorough understanding of natural



versus perturbation-induced variability (Fennel, 2026). As such, establishing a reliable baseline that provides an unperturbed multiyear outlook of carbon states and fluxes is essential to support mCDR verification frameworks.

In this study, we conducted and analyzed a novel set of retrospective prediction simulations using the NOAA Geophysical Fluid Dynamics Laboratory (GFDL) Seamless System for Prediction and EArth System Research (SPEAR) model (Delworth et al., 2020) to assess the prediction skill of global air-sea CO₂ flux on multiyear timescales. These retrospective predictions were verified against independent reconstructions included in recent GCB estimates (Friedlingstein et al., 2025a) to evaluate prediction skill, as well as against the model's own reconstruction to assess the upper bounds of predictability. Our study identifies ENSO as the primary source of near-term (i.e., approximately two years) prediction skill, with significant longer-term predictability linked to multiyear to decadal variability in upwelling, vertical mixing, and sea surface temperature in specific ocean regions.

2. Methods

2.1 SPEAR modeling and prediction system

The Seamless System for Prediction and EArth System Research (SPEAR) is a modeling framework developed at NOAA GFDL for seasonal to multidecadal predictions and projections. The SPEAR configuration employed herein, SPEAR_MED, comprises an atmospheric component (GFDL AM4; Zhao et al., 2018) with a resolution of approximately 50 km and 33 vertical levels, a land component (GFDL LM4; Zhao et al., 2018) at a 50 km resolution, and an ocean and sea-ice component (GFDL MOM6-SIS2; Adcroft et al., 2019) at a 100 km resolution, featuring meridional refinement down to ~30 km in the tropics. Comprehensive details regarding the core system can be found in Delworth et al. (2020) and references therein.

In this study, we integrate the latest version of GFDL's ocean biogeochemical model, COBALTv3 (Stock et al., 2020, 2025), into the SPEAR framework. COBALTv3 employs 40 tracers to simulate the dynamic evolution of four phytoplankton and three zooplankton functional types, as well as free-living bacteria. Carbonate chemistry and thermodynamic calculations within COBALTv3 are configured for GFDL Earth System Modeling (e.g., ESM4.1; Dunne et al., 2020). Ultimately, the integration of



COBALTv3 equips the SPEAR system to perform fully coupled atmosphere-ocean physical and biogeochemical reconstructions and predictions.

A 500-year preindustrial control simulation was conducted to equilibrate the ocean biogeochemistry, particularly the carbonate states and air-sea CO₂ flux. The atmospheric, land, and physical ocean components were initialized from a prior 300-year preindustrial control simulation (Delworth et al., 2020). Ocean biogeochemical states were initialized using climatological means for macronutrients and dissolved oxygen from the 2018 World Ocean Atlas (WOA18; Garcia et al., 2019), as well as preindustrial dissolved inorganic carbon (DIC) and alkalinity (ALK) fields from the Global Ocean Data Analysis Project Version 2 (GLODAPv2; Lauvset et al., 2016).

A 10-member ensemble historical simulation was then commenced by sampling the control run every 30 years between year 191 and year 461 and integrated for 174 years, covering the period from 1850 to 2023. Radiative forcing agents for this historical simulation, which include greenhouse gases, volcanic aerosols, ozone species, and land-use changes, are based on observations and prescribed to the model for years prior to 2015 following Eyring et al. (2016). For subsequent years, projections from the Shared Socioeconomic Pathway 2-4.5 (SSP2-4.5; O'Neill et al., 2016) were applied. To constrain air-sea CO₂ fluxes, we restored the simulated atmospheric CO₂ concentrations toward observed values on an annual timescale. Hereafter, this ensemble is referred to as the *uninitialized* simulation, indicating that the deterministic trajectories of these simulations arise solely from the evolving radiative forcing (i.e., there is no attempt to align the simulated climate variability with the observed state before making a prediction).

2.2 Air-sea CO₂ flux reconstruction and retrospective predictions

A reconstructive historical ensemble of simulations (hereafter, the *SPEAR reconstruction*; 1958-2023) and an extensive set of retrospective predictions were conducted following the methodology of Yang et al. (2021). This approach does not directly assimilate ocean interior information; instead, it focuses on the restoration of observed winds, as well as atmospheric and surface ocean temperatures, which drive the ocean interior. It is nonetheless able to recreate variations in ocean heat content and large-scale circulation while maintaining a dynamically balanced interior solution.



Ensemble members for the reconstruction were branched from their respective uninitialized runs on January 1, 1958. From this point forward, 3-dimensional atmospheric temperature and winds within the model were restored toward values from the Japanese 55-year Reanalysis (JRA55; Kobayashi et al., 2015) on a 12-hourly time scale. Unlike Yang et al. (2021), sea surface temperature (SST) in the ocean model was not directly restored to observational data. This modification was necessary to account for the high thermal sensitivity of the partial pressure of CO₂ in seawater ($p\text{CO}_{2,\text{sw}}$). Instead, SST and its subsequent impact on $p\text{CO}_{2,\text{sw}}$ were indirectly constrained via the restored atmospheric temperature and winds. We show that the modified configuration, despite not directly restoring SST, captures much of the observed interannual variability in SST with a correlation coefficient greater than 0.8 over 70% of the ocean area (Fig. S1). Poor correlation is limited to the Southern Ocean, consistent with Yang et al. (2021), which applied dampened restoring in the high latitudes. Retrospective predictions, hereafter referred to as *initialized* predictions, were then launched from each member of the reconstruction annually on January 1, spanning from 1980 through 2023. Each prediction was integrated for five years. This design enables a skill assessment over a 40-year period (1984-2023) at lead times of one to five years.

2.3 Skill assessment

The simulated air-sea CO₂ flux was evaluated against the ensemble mean of eight observation-constrained products included in Global Carbon Budget version 2024 (Friedlingstein et al., 2025a), which have coverage for the assessment period: VLIZ-SOMFFN (Landschützer et al., 2016), Jena-MLS (Rödenbeck et al., 2022), UExP-FNN-U (Ford et al., 2024), NIES-ML3 (Zeng et al., 2022), JMA-MLR (Iida et al., 2021), OceanSODA-ETHZv2 (Gregor et al., 2024), LDEO-HPD (Bennington et al., 2022), and CSIR-ML6 (Gregor et al., 2019). These products utilize long-term observations of the ocean surface fugacity of CO₂ ($f\text{CO}_2$; e.g., from the Surface Ocean CO₂ Atlas, or SOCAT; Bakker et al., 2016) converted to the partial pressure of CO₂ ($p\text{CO}_{2,\text{sw}}$). They also incorporate observed atmospheric CO₂ concentrations, various reanalysis wind products, and modern statistical techniques, such as machine learning, to fill spatial and temporal gaps within the sparse $f\text{CO}_2$ observational record based on derived relationships with other ocean variables. Note that these products, although utilizing similar $f\text{CO}_2$ observations, often employ different wind products, offsets, and gap-filling algorithms, leading to substantial uncertainties, especially in the high latitudes (Fig. S2).



To calculate the ensemble mean of these products (Fig. S3), we first averaged at the gridpoint level and then integrated over the global ocean. This mean product is hereafter referred to as the *GCB reconstruction*. Because both the GCB and SPEAR reconstructions are subject to biases and uncertainties, we evaluated the initialized predictions against both baselines. This dual evaluation identifies both the model's prediction skill, verified by comparing the SPEAR predictions against the independent GCB reconstruction, and the model's potential prediction skill, or predictability, determined by comparing the SPEAR predictions against the SPEAR reconstruction. Prediction skill or predictability is quantified using the Anomaly Correlation Coefficient (ACC) between the model predictions and either reconstruction, with the corresponding ensemble climatological mean at each lead time removed to account for drift. To isolate the model's skill in capturing interannual and multiyear variability in air-sea CO₂ fluxes, we also calculated a detrended ACC, removing the linear trend from each reconstruction and prediction prior to skill assessment.

To assess the statistical significance of skill improvement resulting from initialization, we employed a Monte Carlo resampling strategy. In each iteration, a new time series was created to calculate a new ACC or detrended ACC against its reference. This was done by randomly sampling one of the 10 ensemble members to provide a flux estimate for a given year instead of using the ensemble mean (e.g., the estimate for 1983 may come from member #6, while that for 1984 may come from member #3). A new ACC or detrended ACC was then calculated for each resampled time series against its reference. The reference was based on the ensemble mean of either GCB or SPEAR reconstruction. This process was repeated 1,000 times in the same manner for both the uninitialized and initialized ensembles. A Student's two-sample *t*-test was used to determine whether the mean ACC ($n = 1000$) of the initialized predictions was significantly higher than that of the uninitialized predictions at the 95% confidence level.

Finally, we consider the skill from a persistence forecast as an additional baseline. This assumes that current conditions will continue unchanged into the future; thus, a prediction at a lead time of n years can be generated by simply shifting the original time series (SPEAR or GCB reconstruction) backward by n years, which is then correlated with the original (e.g., persistent skill ACC for a lead time of one year was calculated from correlating a reconstruction of 1984-2023 with the same reconstruction of 1983-2022). Overall, our approach closely follows a recent multi-model comparison study by Ilyina et al. (2021) that evaluated the prediction skill of ocean and land carbon sinks across eight Earth System Model prediction systems.



2.4 Average Predictability Time analysis

We apply Average Predictability Time (APT) analysis to the SPEAR predictions to investigate the dominant predictable timescales and sources of predictability for global air-sea CO₂ flux. Briefly, APT analysis is a statistical method that helps reveal the maximum predictable timescales of a forecast system by isolating predictable signals from noise (DelSole and Tippett, 2009a,b). This framework has been successfully applied to SPEAR and other prediction systems to identify predictable components and diagnose their sources of predictability over relevant timescales (e.g., Clark et al., 2025; DelSole et al., 2013; Jia et al., 2015, 2022; Yang et al., 2024). While full mathematical details of the optimization problem are available in DelSole and Tippett (2009a,b), APT can be briefly defined here as the sum of a predictability measure, namely, the predictable fraction of total variance, over all forecast lead times. For a given time series of a state (e.g., ocean temperature, air-sea CO₂ flux), this is often expressed as:

$$APT = 2 \sum_{t=1}^{\infty} \frac{\sigma_{signal}^2(t)}{\sigma_{total}^2} \Delta t \quad (1)$$

where $\sigma_{signal}^2(t)$ represents the signal variance at fixed lead time t , and σ_{total}^2 is the total variance. Δt is one year in this study. Because only discrete lead time is considered here, the APT is defined as twice the sum of the predictability measure. By solving an eigenvalue problem to find the linear combinations of variables that maximize this sum, APT isolates the components of the global air-sea CO₂ flux that remain predictable for the longest duration, naturally filtering out unpredictable, high-frequency noise.

APT analysis was first applied to ensemble forecast problems in Yang et al. (2013), where the signal and total variance were approximated by the corresponding ensemble covariance matrices at fixed lead time t . Following DelSole and Tippett (2009a,b) and Yang et al. (2013), we seek the projection vector $\mathbf{q}$ from the generalized eigenvalue equation below to maximize APT:

$$2 \sum_{t=1}^L \widehat{\Sigma}_{sig}(t) \mathbf{q} = \lambda \widehat{\Sigma}_{tot} \mathbf{q} \quad (2)$$

where $\widehat{\Sigma}_{sig}(t)$ is the covariance matrix of the forecast ensemble mean at fixed lead time t , $\widehat{\Sigma}_{tot}$ is the covariance matrix of the total ensemble, and L is the maximum lead time, which is five years. Because both covariance matrices are symmetric, the eigenvectors of the equation above produce components that are uncorrelated. The eigenvalue, λ , then corresponds to the value of APT associated with each component. This approach is analogous to empirical orthogonal function (EOF) analysis, which



decomposes a multivariate dataset into spatial patterns and corresponding time series (Principal Components, or PCs) that optimize variance. Here, we optimize predictability instead of variance, such that the first component maximizes APT and is the most predictable, the second component maximizes APT subject to being uncorrelated with the first and is the second most predictable, and so on. The spatial pattern $\mathbf{p}$ associated with each component can be obtained by:

$$\mathbf{p} = \widehat{\Sigma}_{\text{tot}} \mathbf{q} \quad (3)$$

The predicted air-sea CO₂ flux field is represented on a $1^\circ \times 1^\circ$ grid, corresponding to 64,800 spatial grid points. The dimension of the covariance matrices is $64,800 \times 64,800$, too large to seek a numerical solution for $\mathbf{q}$. To reduce dimensions in the original dataset, we first solved the eigenvalue equation in PC space by projecting the area-weighted, anomalous air-sea CO₂ field onto 10 leading PCs. APT was then maximized using only these 10 PCs, which explains 46% of the total variance in the dataset. Sensitivity tests using six to 20 PCs suggest that our results using 10 PCs are robust. We then transformed the derived spatial patterns from PC space back into physical space using:

$$\mathbf{p} = \Sigma' \mathbf{p}' \quad (1)$$

where Σ' is a $64,800 \times 10$ matrix containing the 10 PC patterns across 64,800 spatial grid points, while $\mathbf{p}'$ represents the predictable spatial patterns in PC space.

3. Results

3.1. SPEAR reconstruction of global air-sea CO₂ flux

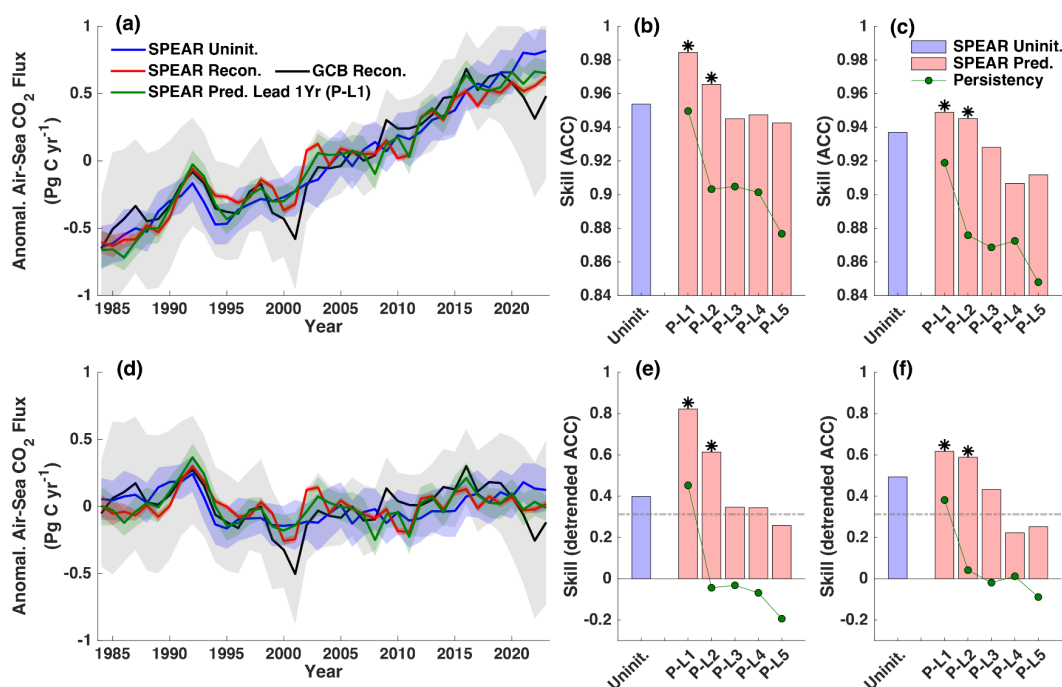
The SPEAR reconstruction of the global air-sea CO₂ flux effectively captures both the amplitude of the long-term trend and much of the multiyear variability present in the observation-constrained GCB reconstruction (ACC = 0.94; detrended ACC = 0.59; Fig. 1a,d). Multiyear variations are most prominent following the volcanic eruption of Mount Pinatubo in the early 1990s and near the turn of the century, coinciding with strong El Niño-Southern Oscillation (ENSO) transitions. Discrepancies between the two reconstructions become most evident between 2018 and 2022, where the GCB reconstruction exhibits an abrupt decline while the SPEAR reconstruction plateaus. This recent divergence may be attributed to uncertainties introduced into both baselines. On the SPEAR side, the model transitioned to using radiative forcing agents from the SSP2-4.5 projections in 2015. While the projected atmospheric CO₂ trend tracked observations well over the past decade, observed CO₂ and sulfur dioxide (SO₂, acting as a global cooling agent) emissions displayed faster and more abrupt



interannual changes than the SSP2-4.5 models anticipated. Examples include the capped SO₂ emissions introduced by the IMO mandate in 2020 (Jordan and Henry, 2024) and the reduced CO₂ emissions during the global pandemic from 2020 to 2021 (Jones et al., 2021). These abrupt changes, along with the observed “Triple-Dip La Niña” from 2020 to 2023, substantially modulated the global heat budget and the ocean’s responses, which were not fully captured in SPEAR under the SSP2-4.5 forcing constraints. On the other hand, the number of *f*CO₂ observations utilized in the GCB reconstruction grew steadily from 1984 but has dropped by nearly 50% since 2017 (Friedlingstein et al., 2025b). Furthermore, the recent addition of rare observations from poorly-sampled areas, such as the Southern Ocean, has contributed to large uncertainties within the data-constrained reconstructions over the 2018-2022 period (Fay et al., 2025). Notably, an ensemble of ocean-only simulations forced by reanalysis atmospheric forcing failed to capture the 2018-2022 decline in the GCB reconstruction (Friedlingstein et al., 2025a), compounding aforementioned evidence to reduce confidence in GCB estimates.



Figure 1. a) Yearly time series of the anomalous globally integrated air-sea CO₂ flux calculated from observation-based products (GCB, $n = 8$), SPEAR reconstruction ($n = 10$), and SPEAR initialized predictions at a one-year lead time ($n = 10$). Anomalies are relative to each time series' own climatological mean. b) Barplot showing the Anomaly Correlation Coefficient (ACC) of the global air-sea CO₂ flux in the SPEAR initialized predictions at lead times of one to five years relative to the SPEAR reconstruction. Asterisks atop the bars indicate significant difference ($p < 0.05$ with paired t -test, $n = 1000$) between the initialized prediction at that lead time and the uninitialized simulation based on Monte Carlo resampling. Green filled circles show ACCs from persistent predictions using the SPEAR reconstruction. c) Same as (b), but the ACCs are calculated relative to the GCB reconstruction instead of the SPEAR reconstruction. Similarly, ACCs for the persistent predictions are calculated using the GCB reconstruction here. d, e, and f) Same as panels (a-c), but each time series of the globally integrated air-sea CO₂ flux is first detrended linearly before an ACC is calculated. Dash lines in (e) and (f) indicate the critical correlation threshold of 0.312 for a two-tailed Pearson t -test at the 95% confidence level and for $df = 38$. They do not show in (b) and (c) because 0.312 is out of the y-axis range, meaning all ACCs in these panels are significant.



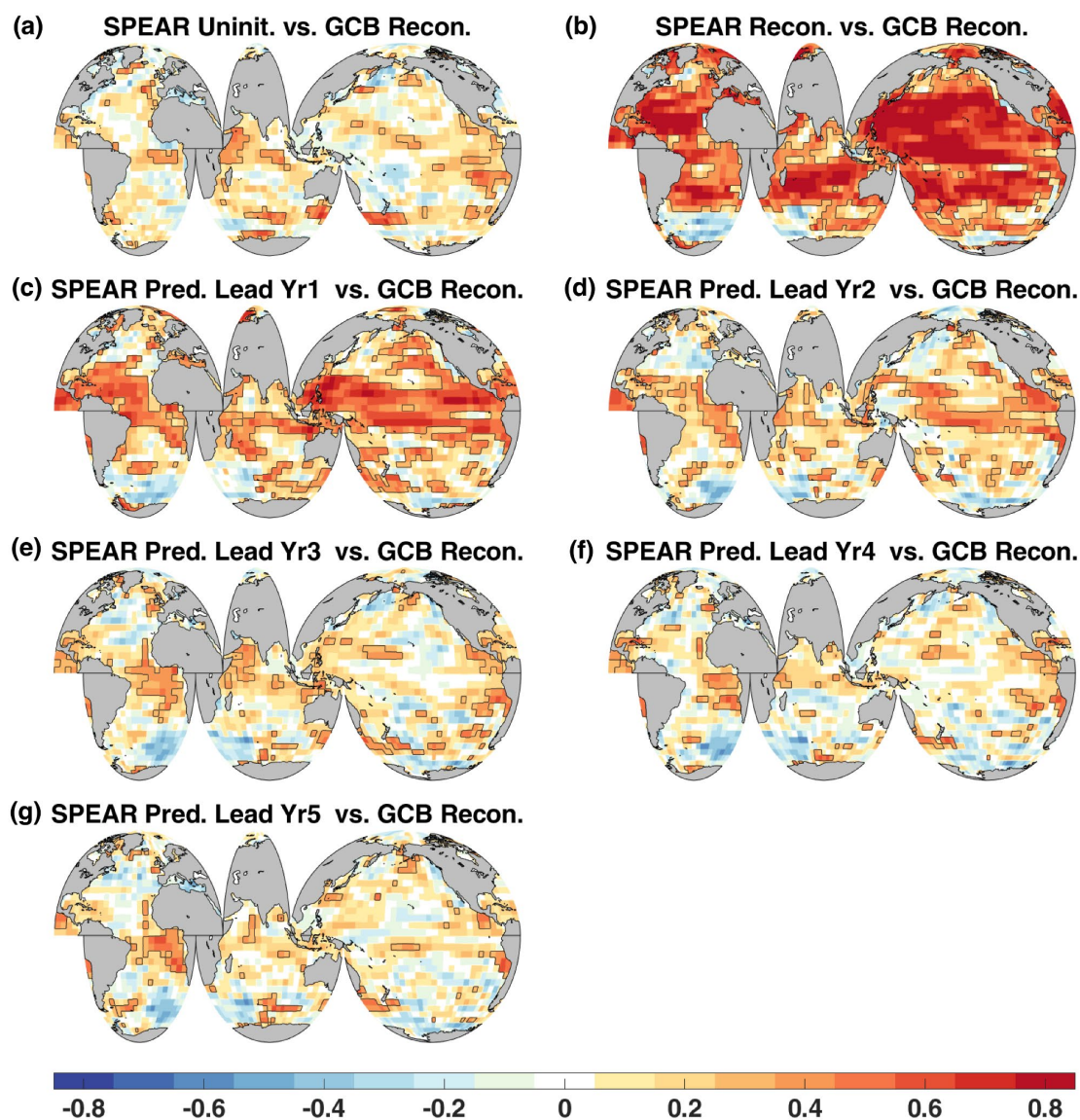


The uninitialized historical ensemble demonstrates skill in predicting the reconstructed long-term trend similar to that of the initialized predictions (Fig. 1a-c), consistent with prior findings establishing that Earth System Models reliably project long-term changes in the ocean carbon sink (McKinley et al., 2016; Tjiputra et al., 2014). However, detrended multiyear variations are largely muted in the uninitialized ensemble mean because individual ensemble members simulate out-of-phase instances of internal climate variability. The residual variations in the uninitialized ensemble mean primarily stem from well-defined external drivers, such as the aerosol forcing signal from the 1991 eruption of Mount Pinatubo (Bilbao et al., 2024; Stenchikov et al., 1998). These forcing-driven variations in the uninitialized simulation contribute to a moderate detrended ACC of 0.40 and 0.49 with the SPEAR and GCB reconstruction, respectively (Fig. 1d-f).

A grid-cell level analysis of the detrended ACC reveals substantially improved skill in the SPEAR reconstruction compared to the uninitialized runs, which results from the additional atmospheric constraints provided by temperature and wind restoring (Fig. 2a,b). These enhancements are most prominent in the subtropics and mid-latitudes, where relatively young surface waters respond rapidly to atmospheric forcing. Conversely, we observe minimal or even negative changes in skill in the northern high latitudes and across large portions of the Southern Ocean, particularly south of the Antarctic Circumpolar Current. In these high-latitude regions, the surface response is impacted by complex variations in convective circulation, sea-ice formation, and melting. Skill degradation is also evident in the nearshore grid cells associated with Eastern Boundary Currents (EBCs) and the monsoon-influenced Indian Ocean. We attribute these regional challenges to inherent limitations in both reconstructions: Observations can be exceptionally sparse in remote and nearshore environments, often reducing confidence in GCB estimates (Fay and McKinley, 2021; Gloege et al., 2021). Concurrently, the SPEAR reconstruction is limited by the model's ability to realistically represent deep-ocean interior variations under a relatively short period of atmospheric constraints (note that temperature and wind restoring commenced in 1958), as well as by the ocean model's coarse grid resolution ($0.33\text{-}1^\circ$), which struggles to resolve nearshore circulations and their variability (Liu et al., 2019).



Figure 2. Grid-cell level Anomaly Correlation Coefficient (detrended ACC) of air-sea CO₂ flux from the SPEAR a) uninitialized simulation, b) historical reconstruction, and c-g) initialized predictions at lead times of one to five years, respectively, relative to the GCB reconstruction. Detrended ACCs were calculated based on ensemble means. Contours show the critical correlation threshold of 0.312 for a two-tailed Pearson *t*-test at the 95% confidence level and for *df* = 38.





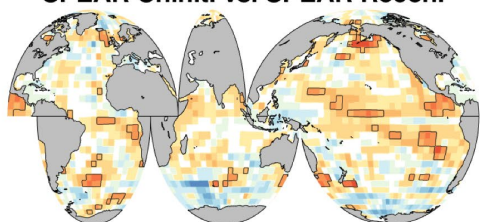
3.2 Initialized predictions of global air-sea CO₂ flux

When compared against the SPEAR reconstruction, the initialized predictions show significant skill improvement at lead times of one and two years, with a global detrended ACC of 0.82 and 0.61, respectively, relative to 0.40 for the uninitialized simulation (Fig. 1b). Regionally, skill improvement extends to a lead time of five years (Fig. 3). Here, we define the predictability horizon at a specific grid cell as the maximum lead year at which predictions still benefit from initialization (i.e., the initialization skill is significantly higher than the uninitialized skill, based on Monte Carlo resampling). Our analysis shows that 38% of the ocean area, primarily in the high latitudes, exhibit a predictability horizon of five years, capped by the five-year duration of our predictions, and potentially longer (Figs. 4i and 5). However, we recognize that this skill improvement is often subtle despite statistical significance, and does not necessarily translate into a practically skillful prediction. In fact, at a lead time of five years, only 18% of the ocean area, such as the subpolar North Pacific and the Southern Ocean, shows significant skill with a detrended ACC greater than the critical threshold (Fig. 3f). When evaluated against the independent GCB reconstruction, the prediction skill is further reduced: only 16% of the ocean can be skillfully predicted two years in advance, with skill improvement seen mainly in the low latitudes (Figs. 2d, 4d). Skill improvement diminishes entirely at lead times of four and five years (Fig. 4h,j). Reconciling high-latitude variations in the SPEAR predictions with the GCB products remains a primary challenge, which we attribute to biases and uncertainties in both the observations and the model as discussed in Sect. 3.1.

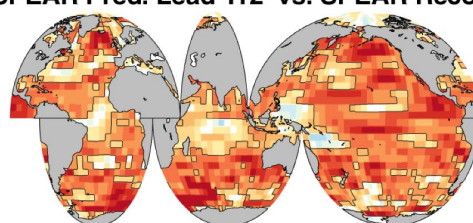
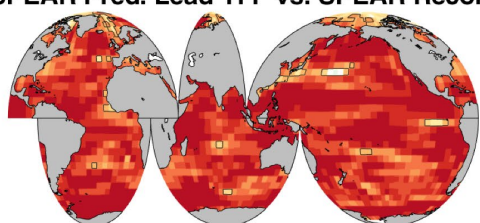


Figure 3. Grid-cell level Anomaly Correlation Coefficient (detrended ACC) of air-sea CO₂ flux from the SPEAR a) uninitialized simulation and c,e,g) initialized predictions at a lead time of one, two, and three years, respective, relative to the SPEAR reconstruction. b,d,f,h) Same as (a,c,e,g), but relative to the GCB reconstruction. Detrended ACC values are calculated based on ensemble means. Black dots indicate grid cells with a detrended ACC greater than the critical correlation threshold of 0.312 for a two-tailed Pearson *t*-test at the 95% confidence level and for *df* = 38.

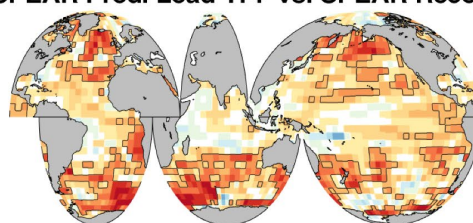
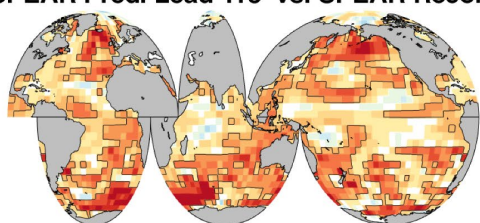
(a) **SPEAR Uninit. vs. SPEAR Recon.**



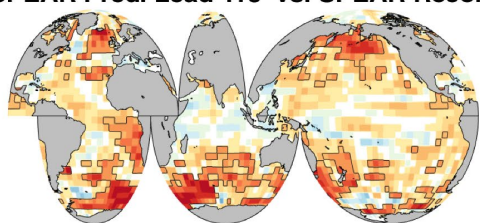
(b) **SPEAR Pred. Lead Yr1 vs. SPEAR Recon.** (c) **SPEAR Pred. Lead Yr2 vs. SPEAR Recon.**



(d) **SPEAR Pred. Lead Yr3 vs. SPEAR Recon.** (e) **SPEAR Pred. Lead Yr4 vs. SPEAR Recon.**



(f) **SPEAR Pred. Lead Yr5 vs. SPEAR Recon.**



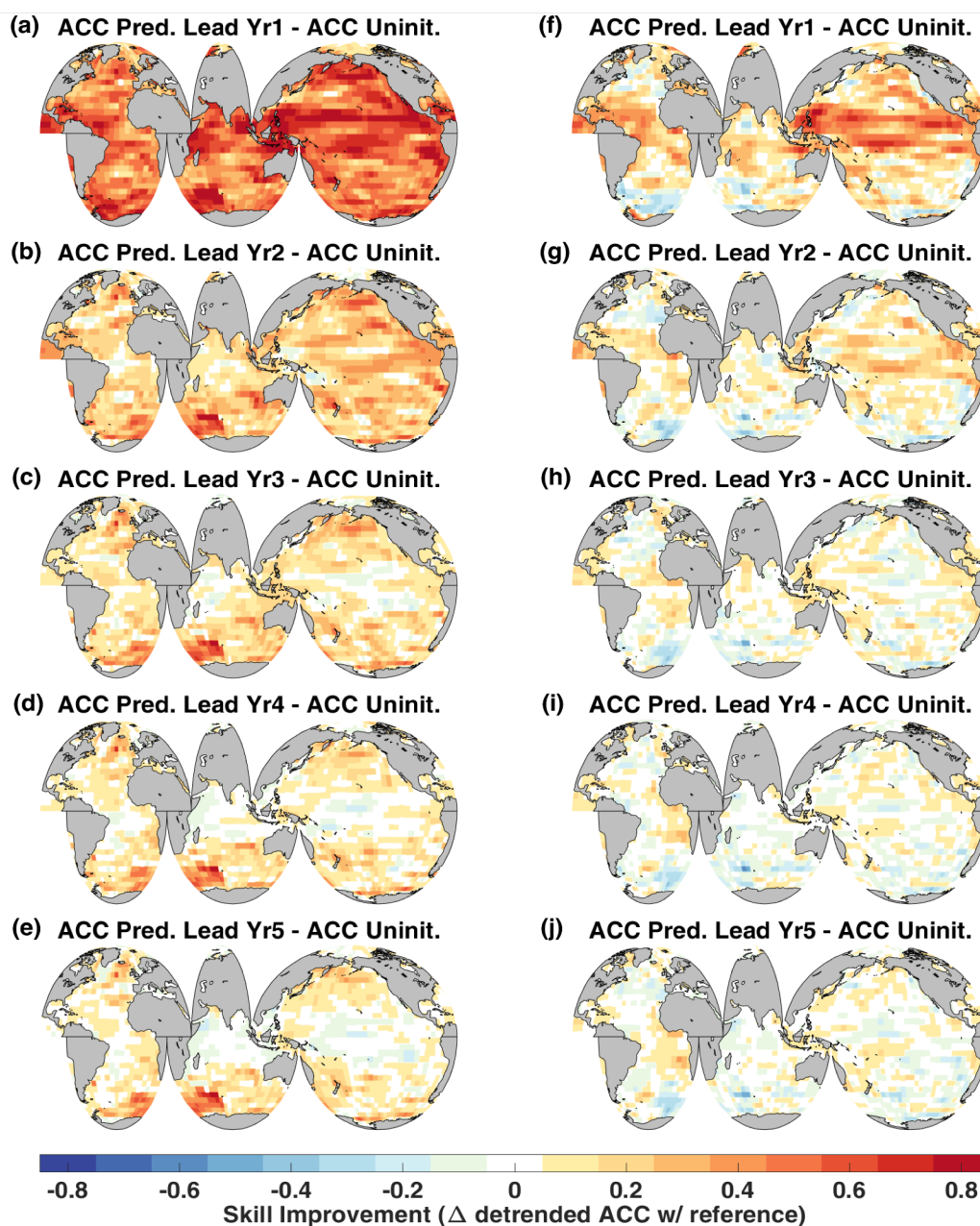
391

392

393



Figure 4. Skill (detrended ACC) improvement in the SPEAR initialized predictions at lead times of one to five years relative to the uninitialized simulations. Panels (a-e) and (e-j) show results based on differences of detrended ACC with the SPEAR and GCB reconstruction, respectively. Detrended ACC values are calculated based on ensemble means.

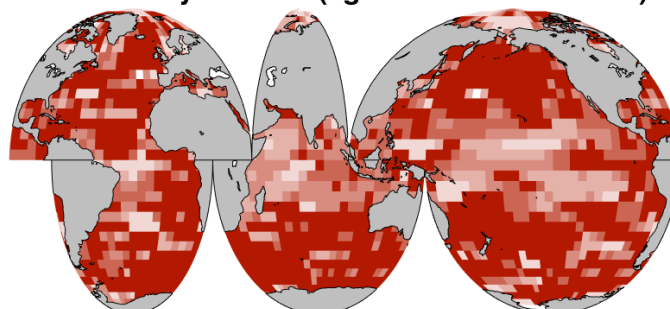


399
400

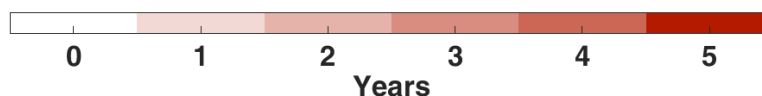
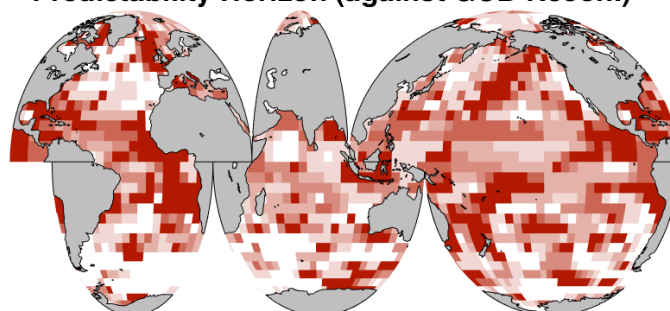


401 Figure 5. Predictability horizons of detrended air-sea CO₂ flux when compared against a) the SPEAR
402 reconstruction and b) the GCB reconstruction. This horizon is defined as the maximum lead year at
403 which the initialized prediction still has a significantly higher detrended ACC than the uninitialized
404 simulation when compared against either reference.
405

(a) Predictability Horizon (against SPEAR Recon.)



(b) Predictability Horizon (against GCB Recon.)



406

407



3.3 Multidecadal variability and long-term potential predictability

We applied Average Predictability Time (APT; see Sect. 2.4) analysis to the SPEAR predictions across all lead times to better understand the dominant predictable timescales and sources of predictability for global air-sea CO₂ flux (Fig. 6). The leading predictable component (PrC1), which explains 25.6% of the total variance in the ensemble predictions, exhibits a widespread positive trend over the 40-year analysis period (Fig. 6a,b). This pattern is strongest in the Southern Ocean, the eastern equatorial Pacific, and the subpolar Atlantic. Projecting the SPEAR and GCB reconstructions onto the PrC1 spatial pattern yields similar trend-like time series. The PrC1 time series of the predictions compares well with either reconstruction, yielding high correlations ($ACC > 0.9$) across all lead times (Fig. 6c). We attribute this to a shared long-term trend, driven by rising atmospheric CO₂ concentrations and other imposed forcing, such as the 1991-1993 peak associated with the Mount Pinatubo eruption, in both reconstructions. PrC1 is highly predictable at least five years in advance when validated against either reconstruction. Crucially, the uninitialized simulations show similarly high ACCs with both reconstructions, suggesting the predictability of PrC1 primarily stems from the model's response to external forcing (i.e., radiative forcing, atmospheric CO₂ restoring) rather than initialization.

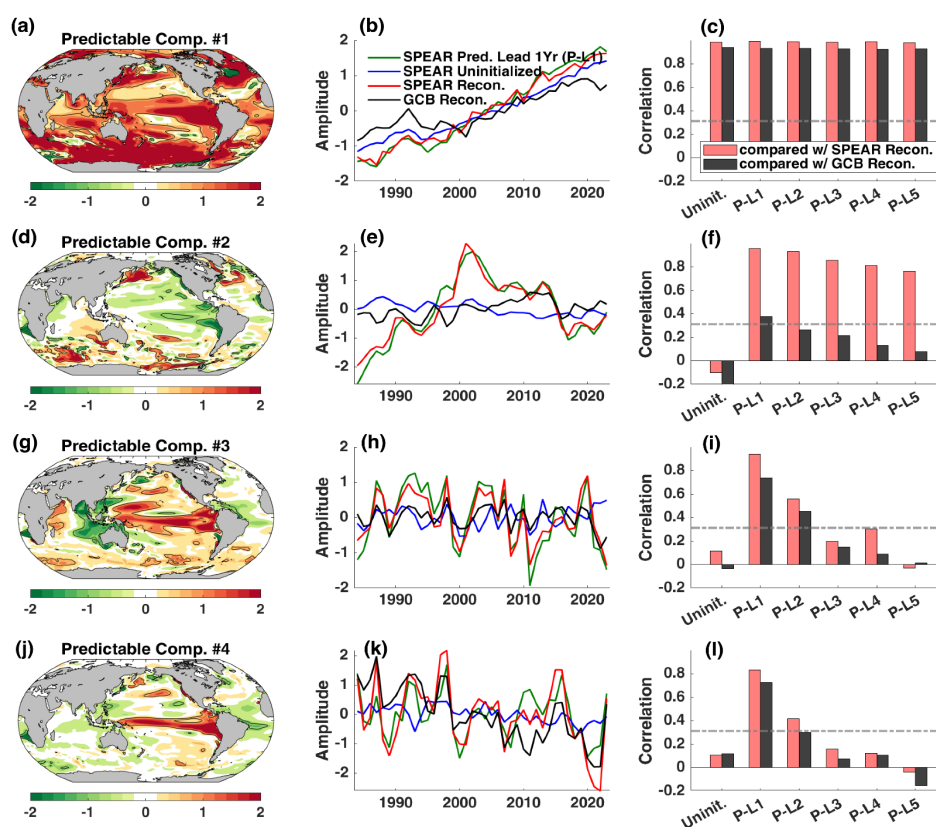
The time series associated with the second predictable component (PrC2, explaining 3.6% of the variance) exhibits strong multidecadal variability, characterized by a rapid increase from 1984 to 2001 followed by a moderate decline (Fig. 6e). Unlike PrC1, PrC2 displays a contrasting spatial pattern: negative anomalies (anti-correlated with the time series) dominate the Eastern Boundary Currents, while positive anomalies (correlated with the time series) are found in high-latitude regions, such as the western North Pacific Subpolar Gyre and along the Antarctic ice shelves. When compared against the SPEAR reconstruction, the correlation between the two PrC2 time series is 0.94 at a lead time of one year and remains as high as 0.76 at a lead time of five years, indicating long-term potential predictability (five years and potentially longer) for PrC2 (Fig. 6f). However, these multidecadal variations are absent in the much flatter time series of the GCB reconstruction when projected onto the PrC2 pattern. This suggests that the observational data underlying the GCB reconstruction may be insufficient to capture this multidecadal variability, particularly since PrC2 is driven by variations in the data-limited, high-latitude areas and the nearshore EBCs. These variations also happen to be the strongest during the earlier decades of the record, when observations are particularly sparse. Importantly, the PrC2 time series from the uninitialized simulation correlates with neither



440 reconstruction, confirming that the long-term potential predictability of PrC2 distinctly benefits from
441 initialization.
442



Figure 6. The four leading predictable components of air-sea CO₂ flux from Average Predictability Time (APT) analysis. Each predictable component can be further decomposed into a spatial pattern (a,d,g,j) and a unique time series (b,e,h,k) associated with each simulation, e.g., blue lines show the uninitialized simulation, while green lines show the initialized prediction at a one-year lead time. Similarly, projecting the GCB (black lines) and SPEAR reconstruction (red lines) onto the spatial patterns also yields their own time series. c,f,i,l) Anomaly Correlation Coefficient (detrended ACC) calculated for uninitialized and initialized predictions when compared against the SPEAR reconstruction (black bars) and GCB reconstruction (pink bars). The dashed gray line represents the critical correlation threshold of 0.312 based on a two-tailed Pearson *t*-test at 95% confidence level and for *df* = 38.





Previous studies suggest that interior ocean memory can lead to multiyear predictability of air-sea CO₂ flux in specific ocean basins (Ilyina et al., 2021; Li et al., 2016, 2019; Lovenduski et al., 2019). We further investigate the source of this potential predictability for PrC2 by correlating changes in relevant ocean variables with the PrC2 time series of the SPEAR reconstruction. We focused on the EBCs and the Southern Ocean (50-70°S), where the PrC2 pattern is most pronounced. The variables analyzed included both surface variables contributing to air-sea CO₂ flux variability, namely the gradient of atmospheric and oceanic $p\text{CO}_2$ ($\Delta p\text{CO}_2$), sea surface temperature (SST), salinity (SSS), and height (SSH), wind speed and vectors, sea ice concentration (SIC), mixed layer depth (MLD), and net primary productivity (NPP), as well as 3-dimensional variables, such as ocean age, temperature (T), and salinity (S), and dissolved inorganic carbon (DIC) at depths, which serve as indicators for interior ocean processes and the memory they carry. We report only those variables highly correlated to the PrC2 time series as potential explanatory sources of predictability.

Within the nearshore grids of the EBCs, the area-weighted averages of $\Delta p\text{CO}_2$ (atmospheric minus seawater) and SSH anomaly tightly track the inverted PrC2 time series for global air-sea CO₂ flux, maintaining a correlation of 0.8 to 0.9 across all lead times (Fig. 7a-d). A negative SSH anomaly in these upwelling zones generally indicates enhanced upwelling of older, high $p\text{CO}_{2,\text{sw}}$ waters to the surface, which leads to lower $\Delta p\text{CO}_2$ and enhanced oceanic outgassing or reduced uptake of atmospheric CO₂. Consequently, enhanced upwelling prior to 2001 likely contributed to a negative trend in air-sea CO₂ flux, which is the opposite of the PrC2 time series and corresponds to the negative amplitude in the spatial pattern for the EBCs. This trend subsequently reversed, with positive SSH and $\Delta p\text{CO}_2$ anomaly emerging in recent years. An analysis of ocean age also shows that older (then younger) waters appeared at 200 m when air-sea CO₂ flux declined (then increased), confirming the upwelling driven shifts. We found no significant relationships between these variables and ocean temperature, both at the surface and at depth, suggesting that changes in density were not a primary contributor to the SSH shifts.

The Southern Ocean (SO; 50-70°S) compensates for these multidecadal changes in EBC air-sea CO₂ fluxes. Specifically, the reconstructed SO mixed layer experienced a rapid shoaling between 1984 and 2000, coinciding with an enhanced $\Delta p\text{CO}_2$ and a surface cooling of about 0.13°C (Fig. 8a,c). The MLD then remained unchanged until 2015, while the SST continued a moderate decline. The most recent decade was characterized by an accelerated MLD deepening along with rapid surface warming in the SO. Multidecadal variability in the reconstructed SST generally agrees with the NOAA OISSTv2 data



product. We found strong correlations between the reconstructed MLD and SST in the SO and the PrC2 time series for global air-sea CO₂ flux (Fig. 8b,d). The increased CO₂ uptake in the earlier decades is likely due to both mixed layer shoaling, which suppresses the surfacing of deep, high $p\text{CO}_{2,\text{sw}}$ waters, and enhanced CO₂ solubility driven by cooling. The reverse processes likely explain the recent decline in CO₂ uptake.

We note that the reconstructed SST displays a turning point around 2011, which lags behind that of the OISSTv2 product and the PrC2 time series by at least several years. This delay is likely due to a lagged response of the model interior ocean to boundary restoring, which substantially modulates SST. This decoupling between SST and PrC2 suggests that MLD in the SO and EBC upwelling may act together to drive the PrC2 turning point around 2001. The anti-correlation between the SO-averaged $\Delta p\text{CO}_2$, MLD, and SST and the PrC2 time series for air-sea CO₂ flux remains very high (positive amplitude for $\Delta p\text{CO}_2$ and negative amplitude for MLD and SST) across all lead times, with an ACC between 0.76 and 0.86 (Fig. 8b,d,f).

While our analysis has focused on identifying local drivers of this multidecadal variability and predictability, previous studies show that ocean interior memory can also influence climate patterns far beyond local systems, particularly in the EBCs and Southern Ocean. For example, Jeong et al. (2026) uses fully coupled, idealized experiments to show that Antarctic sea-ice loss alone can act as a cross-hemispheric driver for Pacific Decadal Oscillation phase shifts. In their study, an extended period of Antarctic ice losses causes Southern Ocean warming, enhancing air-sea heat fluxes and regional circulation. These signals then propagate through the ocean interior, modifying tropical cloud-radiative feedbacks and rainfall, and triggering PDO-like anomalies across the tropical and northern Pacific over multiyear to decadal timescales. We expect that the local predictability in atmospheric and ocean surface drivers discussed here similarly reflect transient responses to ocean interior memory and source of longer predictability in remote areas.



Figure 7. Detrended, yearly time series of SPEAR reconstructed a) $\Delta p\text{CO}_2$ (atmospheric minus seawater), b) Sea Surface Height (SSH), and c) ocean age at 200 m in the major coastal upwelling systems (dark gray with symbols), and the PrC2 time series of the SPEAR reconstruction (red), uninitialized simulation (blue), and initialized prediction at a lead time of five years (green) for air-sea CO_2 flux are shown. b,d,f) Barplots showing the Anomaly Correlation Coefficient (ACC) between the PrC2 time series of initialized predictions for air-sea CO_2 flux and the reconstructed $\Delta p\text{CO}_2$, SSH, and ocean age, respectively, across all lead times. Linear detrending and a five-year moving average was applied to the $\Delta p\text{CO}_2$, SSH, and ocean age time series. The dashed gray line represents the critical correlation threshold of ± 0.312 based on a two-tailed Pearson t -test at 95% confidence level and for $df = 38$.

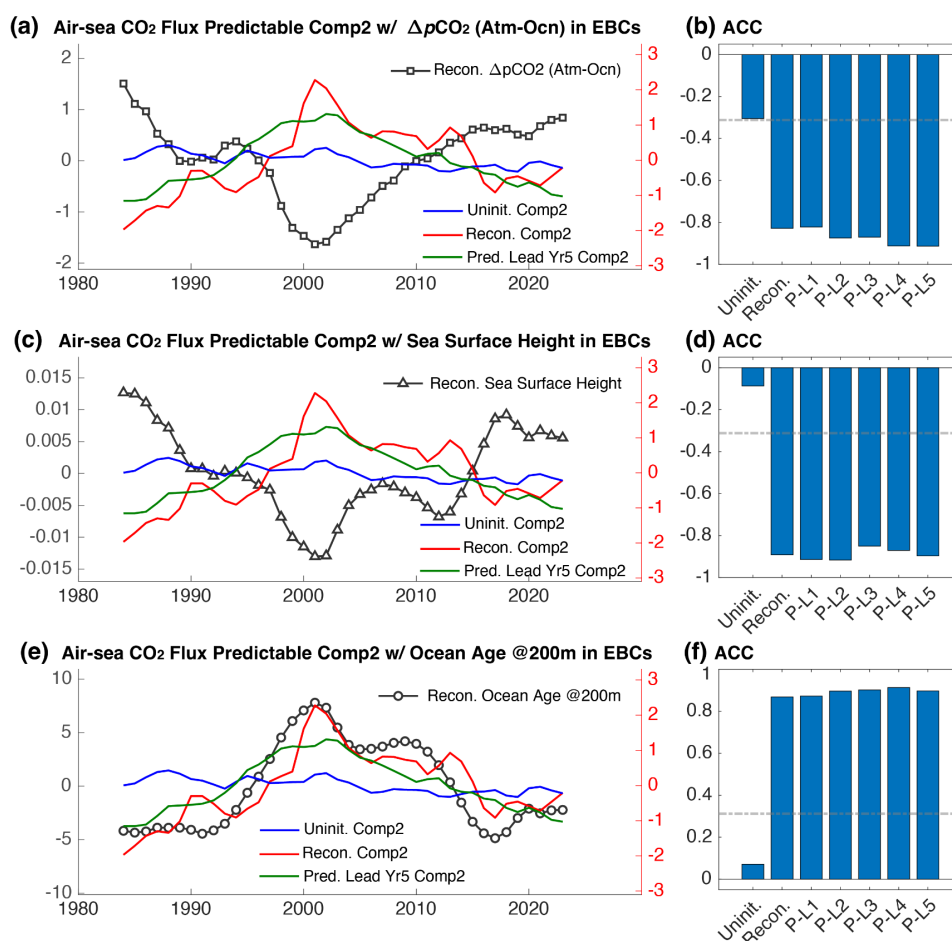
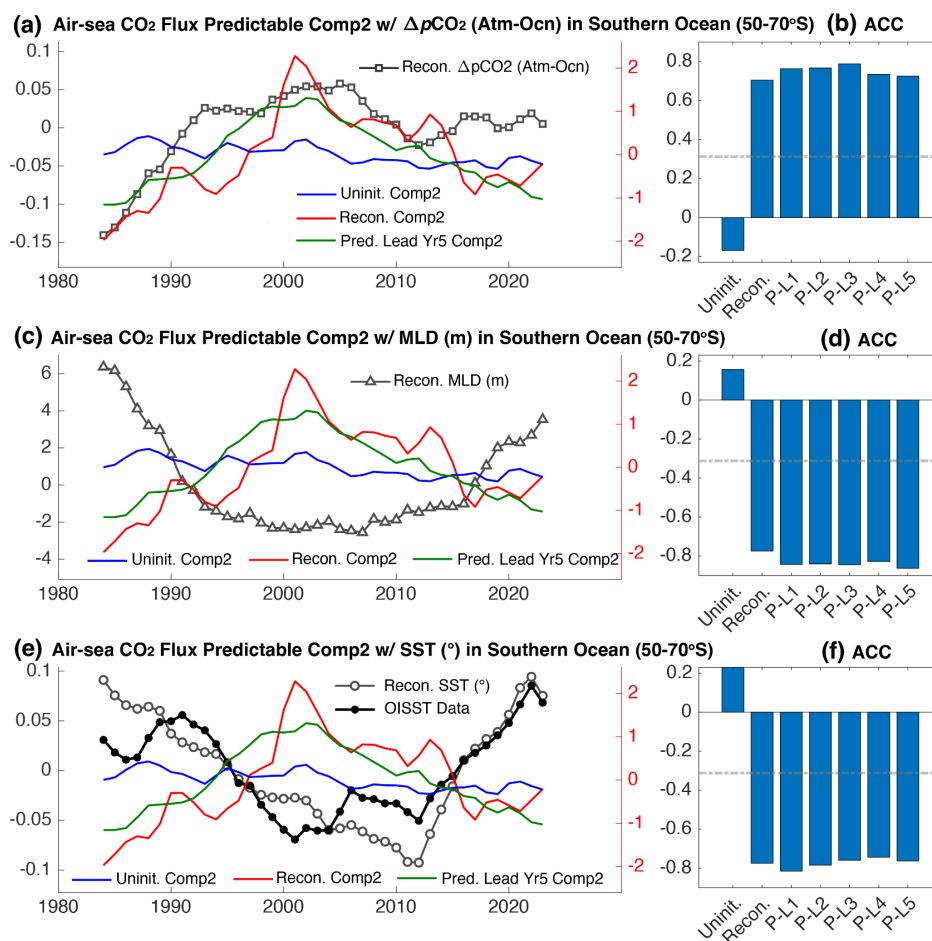




Figure 8. Detrended, yearly time series of SPEAR reconstructed a) $\Delta p\text{CO}_2$ (atmospheric minus seawater), b) Mixed Layer Depth (MLD), and c) Sea Surface Temperature (SST) in the Southern Ocean (50-70°; dark gray or black with symbols), and the PrC2 time series of the SPEAR reconstruction (red), uninitialized simulation (blue), and initialized prediction at a lead time of five years (green) for air-sea CO_2 flux. b,d,f) Barplots showing the Anomaly Correlation Coefficient (ACC) between the PrC2 time series of initialized predictions for air-sea CO_2 flux and the reconstructed $\Delta p\text{CO}_2$, MLD, and SST, respectively, across all lead times. Linear detrending and a five-year moving average was applied to the $\Delta p\text{CO}_2$, MLD, and SST time series. The dashed gray line represents the critical correlation threshold of ± 0.312 based on a two-tailed Pearson t -test at 95% confidence level and for $df = 38$.





3.4 Short-term (one to three years) variability and skillful predictions

The third and fourth predictable components (PrC3 and PrC4, together explaining 7.2% of the variance) exhibit a clear and strong association with the El Niño-Southern Oscillation (ENSO), characterized by large short-term (one to three years) variations across the tropical Pacific and a compensating pattern over Australia and Indonesia (Fig. 6h,i). Note that this ENSO association explains more variance than PrC2, despite shorter APTs. A lagged correlation analysis against the NOAA Physical Sciences Laboratory's observation-based Multivariate ENSO Index (MEIv2) reveals that both the PrC3 and PrC4 time series are highly correlated with ENSO evolution, albeit with a lag of up to one year for PrC3 and no lag for PrC4 (Fig. 9). Together, these two components demonstrate the temporal evolution of air-sea CO₂ flux in association with ENSO, with PrC4 and PrC3 corresponding to a prompt and delayed response, respectively.

During El Niño events, the tropical thermocline flattens, suppressing the upwelling of high $p\text{CO}_{2,\text{sw}}$ waters to the surface in the east (Feely et al., 1999). This reduction in outgassing manifests as a positive anomaly in CO₂ uptake; the opposite occurs during La Niña events. Previous studies indicate that ocean biogeochemical responses, including air-sea CO₂ flux, often lag behind SST peaks by several months (Chatterjee et al., 2017; McKinley et al., 2004); consequently, these signals may be aggregated into the following year's annual average, explaining the lagged and spatially dispersed correlation for PrC3.

For PrC3, a high correlation with MEIv2 ($r \approx 0.7$) is observed in both the GCB and SPEAR reconstructions, as well as in the initialized predictions at a lead time of one year (Fig. 6b). However, this correlation skill decays quickly to 0.4 and 0.38 at lead times of two and three years, respectively, with no predictability beyond this horizon. A similar skill is found for PrC4, confirming the inherent challenge of predicting ENSO events more than two years in advance.

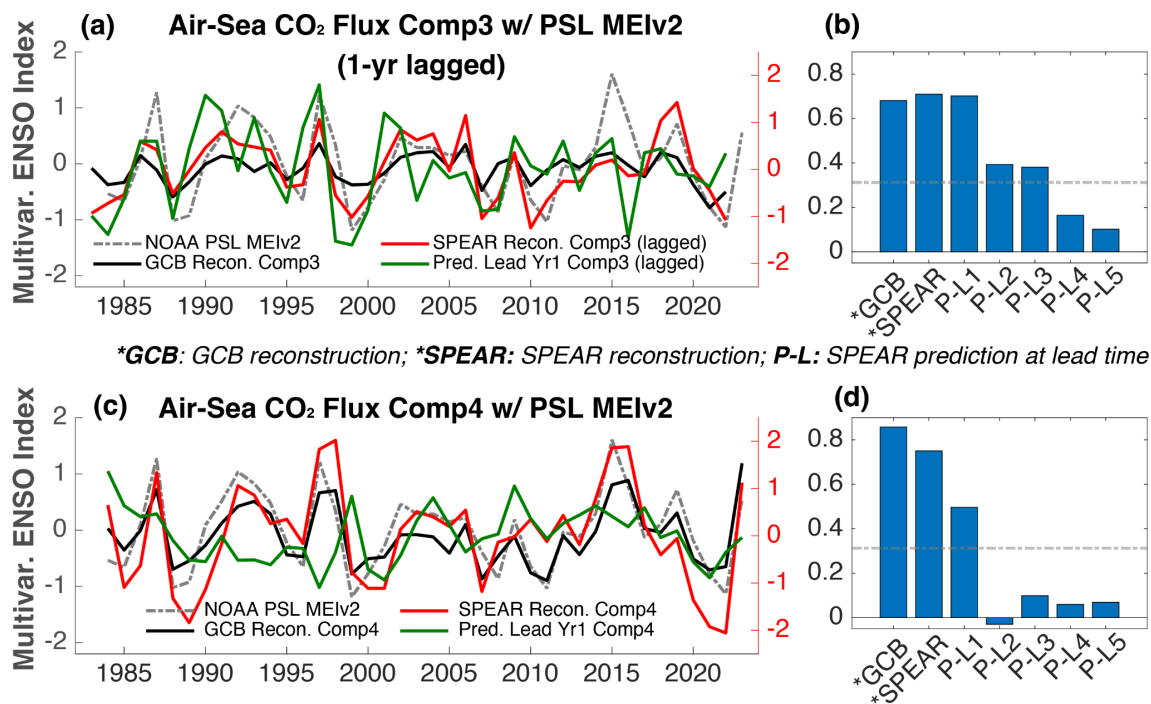
We conclude that short-term variability in tropical air-sea CO₂ flux is primarily linked to ENSO and can be skillfully predicted one to two years in advance. Furthermore, ENSO serves as the primary short-term signal and source of prediction skill for global air-sea CO₂ flux. Beyond a two-year horizon,



572 the practical prediction skill of global air-sea CO₂ flux is constrained both by the model's limited
573 capability to forecast ENSO at longer lead times and by the lack of multidecadal, high-quality
574 observational records necessary to verify potential long-term predictability in the EBCs and the
575 Southern Ocean, which we only observed by comparing predictions against the model reconstruction.



576 Figure 9. a) Yearly time series of NOAA PSL’s Multivariate ENSO Index (MEIv2) and the one-year
577 lagged PrC3 time series of SPEAR reconstruction (red) and initialized prediction at a lead time of one
578 year (green) for air-sea CO₂ flux. b) Barplot showing the Anomaly Correlation Coefficient (ACC)
579 between the PrC3 time series of both reconstructions and the initialized prediction of air-sea CO₂ flux
580 against MEIv2. c,d) same as panels (a,b) except showing for PrC4 without lagging.
581



582
583



4. Discussion

Earth System predictions on multiyear to decadal timescales (two to 10 years ahead) represent a major frontier in climate research. Traditionally, initialized decadal prediction frameworks focused on physical variables and phenomena, such as sea surface temperature, ocean heat content, and fluctuations in Meridional Overturning Circulations (MOCs; Keenlyside et al., 2008; Yeager et al., 2012). Over the last decade, leading modeling centers expanded these frameworks into Earth System prediction systems by coupling initialized physical ocean-atmosphere models with prognostic marine biogeochemistry and terrestrial carbon cycle modules, allowing for multiyear to decadal predictions of oceanic and terrestrial carbon sinks (Li et al., 2016, 2019, 2023; Park et al., 2018; Lovenduski et al., 2019; Sospedra-Alfonso et al., 2021; Bilbao et al., 2021; Bethke et al., 2021).

For the air-sea CO₂ flux, most models established a global predictability horizon of two to four years, although the approaches of skill assessment have been inconsistent across these studies. Ilyina et al. (2021) provided the first multi-modal synthesis by assessing the prediction skill at a lead time of two years against GCB version 2019 estimates. The SPEAR assessment largely follows methods used in this synthesis. With a detrended ACC of 0.57, the SPEAR framework sits above the 0.45 average of the assessed models, representing a major improvement over GFDL's earlier-generation Earth System prediction system, GFDL-ESM2. Here we discuss the features, limitations, and future directions of SPEAR development in the context of existing Earth System prediction frameworks across leading modeling centers.

4.1 Necessity and limitations of using only atmospheric assimilation in CO₂ reconstruction

Our model reconstruction adopted the data assimilation strategies utilized in SPEAR's decadal prediction system, where atmospheric temperature and winds are restored toward the JRA55 reanalysis (Yang et al., 2021). Crucially, Sea Surface Temperature (SST) restoring employed by Yang et al., (2021) was disabled, allowing ocean physics and biogeochemistry to evolve freely in response to these atmospheric constraints. This also strongly contrasts with SPEAR's seasonal prediction framework, which integrates stronger ocean observational constraints via a three-dimensional ocean data assimilation scheme based on an Ensemble Kalman Filter (EnKF ODA) and an Ocean Tendency Adjustment (OTA) bias correction strategy (Lu et al., 2020).



The decision to only employ atmospheric restoring in this study was driven by two primary factors. The first issue involved observed performance degradation under SST restoring due to the artificial decoupling of heat and carbon fluxes. We found that the air-sea CO₂ flux is highly sensitive to SST restoring. In the real ocean, heat and carbon fluxes are physically coupled. However, SST restoring mathematically injects or extracts heat without a corresponding physical atmospheric flux. In regions with large negative SST biases, for example, restoring acts as an unbounded heat source, artificially warming the water and potentially driving continuous, spurious CO₂ outgassing. Our sensitivity experiments indicated that SST restoring deteriorates grid-cell level reconstruction by either exaggerating or dampening long-term CO₂ flux trends.

The second issue arose from ocean interior assimilation, which generated spurious vertical velocities. Assimilating interior ocean data (via ODA or OTA) tends to generate spurious vertical velocities in our model, particularly in the tropics. Our experiments with interior assimilation exhibited stronger trade wind biases, overactive larger-scale equatorial upwelling, and localized vertical velocities due to momentum imbalances. Without direct biogeochemical data assimilation, the biogeochemical model readily responds to this spurious upwelling by simulating excessive surface nutrient delivery and primary productivity, which degrades the realistic representation of natural biogeochemical impacts on the air-sea CO₂ flux. Similar effects have been documented in previous ocean biogeochemical reconstructions utilizing ODA (Sanchez-Gomez et al., 2016; Park et al., 2018). To mitigate these ODA-induced effects, Park et al. (2018), which employed an earlier GFDL coupled model with ODA, suggested that tightening atmospheric wind constraints while relaxing equatorial oceanic constraints helps improve chlorophyll reconstruction. However, testing this approach in our model failed to improve air-sea CO₂ reconstruction due to worsened trade wind and upwelling biases near the equator.

Consequently, relying on atmospheric restoring alone loosens direct ocean observational constraints on buoyancy-driven circulations such as the Atlantic Meridional Overturning Circulation and Southern Ocean deep convection, which strongly influence multiyear predictions of key physical variables, such as North Atlantic sea level rise (Zhang et al., 2023, 2024), Antarctic sea ice extent (Zhang et al., 2022; Morioka et al., 2023), and Antarctic bottom water formation (Zhang et al., 2017). The causes and consequences of their variations, when not well constrained, can potentially deteriorate regional skill of air-sea CO₂ flux. Additionally, relying on atmospheric restoring alone meant our model required an extended spin-up period to synchronize its simulated climate phase with observations following the onset of restoring. In regions like the Southern Ocean, this synchronization period lasted



approximately 25-30 years, ultimately limiting our capacity for historical validation longer than the 40-year period chosen here.

Despite these challenges, other studies have successfully integrated ocean interior data into air-sea CO₂ flux reconstructions. For example, Li et al., (2016, 2019, 2023) utilized a relaxed “anomaly assimilation” strategy, while Bernardello et al. (2024) and Sospedra-Alfonso et al. (2021) employed relaxed 3-dimensional nudging toward smoothed ocean reanalysis products while intentionally masking out the tropics. The adjusted relaxation timescales adopted in these strategies (e.g., 10 days in the upper layer increased to one year at depth - much longer than those typically used in operational ocean data assimilation) suggest that the reconstructions are nudged away from model climatologies, but not necessarily toward direct observations.

While these pragmatic solutions may be highly model-specific, a critical future direction for our prediction framework is the exploration of novel assimilation strategies. Because interior ocean observations are essential for maximizing prediction skills for phenomena such as ENSO (Alves et al., 2004; Balmaseda and Anderson, 2009; Fujii et al., 2015; Lu et al., 2020), finding a method to robustly integrate them without simultaneously degrading the coupled biogeochemical dynamics remains a key priority for advancing Earth System predictions. Along those lines, integrating interior ocean biogeochemical observations to further constrain the evolution of the biogeochemical model, and its response to the assimilation of physical ocean observations, would be a particularly beneficial next step.

4.2 Interactive atmospheric, ocean, and land carbon cycle

While our model activates the interactive carbon cycle between the atmosphere, ocean, and land, we chose to restore the atmospheric CO₂ concentrations toward observation values on an annual timescale. This approach was selected to constrain the uncertainties associated with the air-land CO₂ flux, which exhibits higher volatility and shorter predictability than the air-sea CO₂ flux (Ilyina et al., 2021). This methodology is generally consistent with conventional decadal prediction systems that focus primarily on climate predictions, where atmospheric CO₂ concentrations are prescribed. However, by restoring or prescribing atmospheric CO₂, we inherently bypass the critical feedback of the land carbon cycle to the atmosphere and the ocean, precluding the evaluation of prediction skill for atmospheric CO₂ growth. Recently, Li et al. (2023) introduced a global carbon prediction system driven by prescribed



CO₂ emissions rather than concentrations. This framework operates crucially on a closed global carbon budget with a fully interactive atmospheric-ocean-land carbon cycle. Several recent Earth System prediction systems have since adopted this emission-driven strategy and are synthesized in Li et al. (2026). While our study intentionally employed the conventional, concentration-driven strategy as a necessary step to isolate and evaluate the prediction skill for air-sea CO₂ flux, a vital future direction is to rigorously evaluate the representation of the land carbon cycle in our model. Ultimately, this will drive the development of a fully emission-driven prediction system capable of forecasting atmospheric CO₂ growth to better inform climate mitigation policymaking.

5. Conclusions

This study demonstrates the capability of the NOAA GFDL SPEAR framework to provide skillful multiyear predictions of global air-sea CO₂ flux. By utilizing an atmospheric-only restoration strategy for reconstructions, we successfully balanced the need to constrain physical variability with the imperative to avoid spurious biogeochemical responses that often arise from direct interior ocean data assimilation. Our results confirm that initialization significantly enhances prediction skill for air-sea CO₂ flux, establishing a global skill horizon of up to two years. This near-term predictability is primarily driven by ENSO, which modulates tropical upwelling and air-sea exchange. Beyond this near-term horizon, our analysis reveals potential multidecadal predictability linked to variations in Eastern Boundary Current upwelling and Southern Ocean mixed layer depth and sea surface temperature. While current observational products face challenges in verifying this long-term signal partly due to data sparsity in high-latitude and nearshore environments, our modeling experiments highlight the untapped potential of initialized Earth System predictions. As sustained and optimized ocean observations improve, these predictive frameworks will become essential tools for disentangling natural variability from anthropogenic impacts, informing our understanding of near-future carbon cycle variability, and establishing the baselines necessary to verify mCDR efforts.

Data and code availability NOAA PSL's Multivariate ENSO Index Version 2 (MEIv2) can be accessed at <https://psl.noaa.gov/enso/mei/>. NOAA PSL's Pacific Decadal Oscillation (PDO) index can be accessed at <https://psl.noaa.gov/pdo/>. The GCB2024 data products of air-sea CO₂ flux are available online at <https://zenodo.org/records/14639761>. The SPEAR reconstructed and predicted (at a lead time of one to five years) fields of air-sea CO₂ flux, sea surface partial pressure of CO₂, sea surface



715 temperature, and mixed layer depth, as well as MATLAB code and supporting datasets that can be
716 used to replicate figures here, are available online at <https://zenodo.org/records/21344715>.

717

718 **Author contributions** XL, JPD, and CAS designed the study; XL and XY developed model code; XL
719 performed model simulations and all data analysis; XL wrote the manuscript with contributions from
720 all co-authors.

721

722 **Acknowledgements.** This analysis was supported by the National Oceanic and Atmospheric
723 Administration, U.S. Department of Commerce. The statements, findings, conclusions, and
724 recommendations are those of the authors and do not necessarily reflect the views of the National
725 Oceanic and Atmospheric Administration, or the U.S. Department of Commerce. We thank G.
726 Negrete-Garcia and K. Noh who performed a GFDL internal review for an earlier draft of this
727 manuscript, and F. Lu who provided constructive insights and suggestions to this work.

Competing interests. The contact author has declared that none of the authors has any competing interests.



References

- Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al.: The GFDL global ocean and sea ice model OM4.0: model description and simulation features, *J. Adv. Model. Earth Syst.*, 11, 3167–3211, <https://doi.org/10.1029/2019MS001726>, 2019.
- Alves, O., Balmaseda, M. A., Anderson, D., and Stockdale, T.: Sensitivity of dynamical seasonal forecasts to ocean initial conditions, *Q. J. Roy. Meteor. Soc.*, 130, 649–667, 2004.
- Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., et al.: A multi-decade record of high-quality fCO₂ data in version 3 of the Surface Ocean CO₂ Atlas (SOCAT), *Earth Syst. Sci. Data*, 8, 383–413, 2016.
- Balmaseda, M. A., and Anderson, D.: Impact of initialization strategies and observations on seasonal forecast skill, *Geophys. Res. Lett.*, 36, L01701, 2009.
- Bennington, V., Gloege, L., and McKinley, G. A.: Variability in the global ocean carbon sink from 1959 to 2020 by correcting models with observations, *Geophys. Res. Lett.*, 49, e2022GL098632, <https://doi.org/10.1029/2022GL098632>, 2022.
- Bernardello, R., Sicardi, V., Lapin, V., Ortega, P., Ruprich-Robert, Y., Tourigny, E., and Ferrer, E.: Ocean biogeochemical reconstructions to estimate historical ocean CO₂ uptake, *Earth Syst. Dynam.*, 15, 1255–1275, 2024.
- Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., Passos, L., Bentsen, M., Guo, C., Gupta, A., Tjiputra, J., Kirkevåg, A., Olivié, D., Seland, Ø., Solsvik Vågane, J., Fan, Y., and Eldevik, T.: NorCPM1 and its contribution to CMIP6 DCP, *Geosci. Model Dev.*, 14, 7073–7116, <https://doi.org/10.5194/gmd-14-7073-2021>, 2021.
- Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Cruz-García, R., Cvijanovic, I., Doblas-Reyes, F. J., Donat, M., Dutra, E., Echevarría, P., Ho, A.-C., Loosveldt-Tomas, S., Moreno-Chamarro, E., Pérez-Zanon, N., Ramos, A., Ruprich-Robert, Y., Sicardi, V., Tourigny, E., and Vegas-Regidor, J.: Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth, *Earth Syst. Dynam.*, 12, 173–196, <https://doi.org/10.5194/esd-12-173-2021>, 2021.
- Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P. A., et al.: Impact of volcanic eruptions on CMIP6 decadal predictions: a multi-model analysis, *Earth Syst. Dynam.*, 15, 501–525, 2024.
- Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., et al.: Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks, *Proc. Natl. Acad. Sci. USA*, 104, 18866–18870, 2007.
- Chatterjee, A., Gierach, M. M., Sutton, A. J., Feely, R. A., Crisp, D., Eldering, A., et al.: Influence of El Niño on atmospheric CO₂ over the tropical Pacific Ocean: findings from NASA's OCO-2 mission, *Science*, 358, eaam5776, 2017.



- 767 Clark, J. P., Johnson, N. C., Park, M., Bernardez, M., and Delworth, T. L.: Predictable patterns of
768 seasonal atmospheric river variability over North America during winter, *Geophys. Res. Lett.*, 52,
769 e2024GL112411, 2025.
- 770 Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the global ocean with biogeochemical-
771 Argo, *Front. Mar. Sci.*, 7, 99, 2020.
- 772 DelSole, T., and Tippett, M. K.: Average predictability time: Part I. Theory, *J. Atmos. Sci.*, 66, 1172–
773 1187, 2009a.
- 774 DelSole, T., and Tippett, M. K.: Average predictability time: Part II. Seamless winter season
775 predictions, *J. Atmos. Sci.*, 66, 1188–1204, 2009b.
- 776 DelSole, T., Jia, L., and Tippett, M. K.: Predictability of decadal changes in the North Pacific, *J.*
777 *Climate*, 26, 4207–4220, 2013.
- 778 Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J. H., Dunne, K. A., et al.: SPEAR: The
779 next generation GFDL modeling system for seasonal to multidecadal prediction and projection, *J. Adv.*
780 *Model. Earth Syst.*, 12, e2019MS001895, 2020.
- 781 Doney, S. C., Busch, D. S., Cooley, S. R., and Kroeker, K. J.: The science, engineering, and validation
782 of marine carbon dioxide removal and storage, *Annu. Rev. Mar. Sci.*, 16, 217–241, 2025.
- 783 Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al.: The GFDL
784 Earth System Model version 4.1 (GFDL-ESM 4.1): overall coupled model description and simulation
785 characteristics, *J. Adv. Model. Earth Syst.*, 12, e2019MS002015, 2020.
- 786 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.:
787 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and
788 organization, *Geosci. Model Dev.*, 9, 1937–1958, 2016.
- 789 Fay, A. R., and McKinley, G. A.: Observed regional fluxes to constrain modeled estimates of the
790 ocean carbon sink, *Geophys. Res. Lett.*, 48, e2021GL095325, 2021.
- 791 Fay, A. R., Heimdal, T. H., Acquaviva, V., Shaum, A. P., and McKinley, G. A.: Sensitivity of ocean
792 carbon sink estimates to rare observations, *Geophys. Res. Lett.*, 52, e2025GL117961,
793 <https://doi.org/10.1029/2025GL117961>, 2025.
- 794 Feely, R. A., Wanninkhof, R., Takahashi, T., and Tans, P.: Influence of El Niño on the equatorial
795 Pacific contribution to atmospheric CO₂ accumulation, *Nature*, 398, 597–601, 1999.
- 796 Fennel, K.: The verification challenge of marine carbon dioxide removal, *Annu. Rev. Mar. Sci.*, 18,
797 141–164, 2026.
- 798 Ford, D. J., Blannin, J., Watts, J., Watson, A. J., Landschützer, P., Jersild, A., and Shutler, J. D.: A
799 comprehensive analysis of air-sea CO₂ flux uncertainties constructed from surface ocean data
800 products, *Global Biogeochem. Cy.*, 38, e2024GB008188, <https://doi.org/10.1029/2024GB008188>,
801 2024.



- 802 Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le
803 Quéré, C., Li, H., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C.,
804 Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., et al.: Global Carbon Budget 2024,
805 Earth Syst. Sci. Data, 17, 965–1039, 2025a.
- 806 Friedlingstein, P., Le Quéré, C., O'Sullivan, M., Hauck, J., Landschützer, P., Luijkx, I. T., Li, H., van
807 der Woude, A., Schwingshackl, C., Pongratz, J., Regnier, P., Andrew, R. M., Bakker, D. C. E.,
808 Canadell, J. G., Ciais, P., Gasser, T., Jones, M. W., Lan, X., Morgan, E., et al.: Emerging climate
809 impact on carbon sinks in a consolidated carbon budget, Nature, [https://doi.org/10.1038/s41586-025-](https://doi.org/10.1038/s41586-025-09802-5)
810 [09802-5](https://doi.org/10.1038/s41586-025-09802-5), 2025b.
- 811 Fujii, Y., Cummings, J., Xue, Y., Schiller, A., Lee, T., Balmaseda, M. A., et al.: Evaluation of the
812 Tropical Pacific Observing System from the ocean data assimilation perspective, Q. J. Roy. Meteor.
813 Soc., 141, 2481–2496, 2015.
- 814 Garcia, H. E., Boyer, T. P., Baranova, O. K., Locarnini, R. A., Mishonov, A. V., Grodsky, A., Paver,
815 C. R., Weathers, K. W., Smolyar, I. V., Reagan, J. R., Seidov, D., and Zweng, M. M.: World Ocean
816 Atlas 2018: product documentation (A. Mishonov, Tech. Ed.), National Centers for Environmental
817 Information, <https://repository.library.noaa.gov/view/noaa/70445>, 2019.
- 818 Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., et al.:
819 Quantifying errors in observationally based estimates of ocean carbon sink variability, Global
820 Biogeochem. Cy., 35, e2020GB006788, 2021.
- 821 Gregor, L., Lebehot, A. D., Kok, S., and Scheel Monteiro, P. M.: A comparative assessment of the
822 uncertainties of global surface ocean CO₂ estimates using a machine-learning ensemble (CSIR-ML6
823 version 2019a) – have we hit the wall?, Geosci. Model Dev., 12, 5113–5136,
824 <https://doi.org/10.5194/gmd-12-5113-2019>, 2019.
- 825 Gregor, L., Shutler, J., and Gruber, N.: High-resolution variability of the ocean carbon sink, Global
826 Biogeochem. Cy., 38, e2024GB008127, <https://doi.org/10.1029/2024GB008127>, 2024.
- 827 Gruber, N., Clement, D., Carter, B. R., Feely, R. A., Van Heuven, S., Hoppema, M., et al.: The oceanic
828 sink for anthropogenic CO₂ from 1994 to 2007, Science, 363, 1193–1199, 2019.
- 829 Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C. E., Bopp, L., et al.: Consistency and
830 challenges in the ocean carbon sink estimate for the Global Carbon Budget, Front. Mar. Sci., 7,
831 571720, 2020.
- 832 Iida, Y., Takatani, Y., Kojima, A., and Ishii, M.: Global trends of ocean CO₂ sink and ocean
833 acidification: an observation-based reconstruction of surface ocean inorganic carbon variables, J.
834 Oceanogr., 77, 323–358, <https://doi.org/10.1007/s10872-020-00571-5>, 2021.
- 835 Ilyina, T., Li, H., Spring, A., Müller, W. A., Bopp, L., Chikamoto, M. O., et al.: Predictable variations
836 of the carbon sinks and atmospheric CO₂ growth in a multi-model framework, Geophys. Res. Lett., 48,
837 e2020GL090695, 2021.
- 838 Jeong, H., Park, H. S., Yeh, S. W., et al.: Antarctic sea-ice loss shifts the Pacific Decadal Oscillation
839 toward a positive phase, Commun. Earth Environ., 7, 432, [https://doi.org/10.1038/s43247-026-03489-](https://doi.org/10.1038/s43247-026-03489-w)
840 [w](https://doi.org/10.1038/s43247-026-03489-w), 2026.



- 841 Jia, L., Yang, X., Vecchi, G., Gudgel, R., Delworth, T., Rosati, A., et al.: Improved seasonal prediction
842 of temperature and precipitation over land in a high-resolution GFDL climate model, *J. Climate*, 28,
843 2044–2062, 2015.
- 844 Jia, L., Delworth, T. L., Kapnick, S. B., Yang, X., Johnson, N. C., Cooke, W., Lu, F., Harrison, M. J.,
845 Rosati, A., Zeng, F., McHugh, C., Wittenberg, A. T., Zhang, L., Murakami, H., and Tseng, K.-C.:
846 Skillful seasonal prediction of North American summertime heat extremes, *J. Climate*, 35, 14331–
847 14345, 2022.
- 848 Jones, C. D., Hickman, J. E., Rumbold, S. T., et al.: The climate response to emissions reductions due
849 to COVID-19: initial results from CovidMIP, *Geophys. Res. Lett.*, 48, e2020GL091883, 2021.
- 850 Keenlyside, N., Latif, M., Jungclauss, J., et al.: Advancing decadal-scale climate prediction in the North
851 Atlantic sector, *Nature*, 453, 84–88, <https://doi.org/10.1038/nature06921>, 2008.
- 852 Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., Van Heuven, S., et
853 al.: The reinvigoration of the Southern Ocean carbon sink, *Science*, 349, 1221–1224, 2015.
- 854 Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean
855 carbon sink, *Global Biogeochem. Cy.*, 30, 1396–1417, <https://doi.org/10.1002/2015GB005359>, 2016.
- 856 Lauvset, S. K., Key, R. M., Olsen, A., Van Heuven, S., Velo, A., Lin, X., et al.: A new global interior
857 ocean mapped climatology: the 1°×1° GLODAP version 2, *Earth Syst. Sci. Data*, 8, 325–340, 2016.
- 858 Li, H., Ilyina, T., Müller, W. A., and Landschützer, P.: Decadal predictions of the North Atlantic CO₂
859 uptake, *Nat. Commun.*, 7, 11076, 2016.
- 860 Li, H., and Ilyina, T.: Current and future decadal trends in the oceanic carbon uptake are dominated by
861 internal variability, *Geophys. Res. Lett.*, 45, 916–925, 2018.
- 862 Li, H., Ilyina, T., Müller, W. A., and Sienz, F.: Multiyear predictability of ocean biogeochemistry in
863 the Southern Ocean, *J. Adv. Model. Earth Syst.*, 11, 1146–1163, 2019.
- 864 Li, H., Ilyina, T., Loughran, T., Spring, A., and Pongratz, J.: Reconstructions and predictions of the
865 global carbon budget with an emission-driven Earth system model, *Earth Syst. Dynam.*, 14, 101–119,
866 2023.
- 867 Li, H., Ilyina, T., Dunkl, I., Brune, S., Müller, W. A., Bernardello, R., et al.: Multi-model
868 reconstructions and predictions of the CO₂ fluxes and atmospheric CO₂ variations, *Bull. Amer.*
869 *Meteor. Soc.*, 2026.
- 870 Liu, X., Dunne, J. P., Stock, C. A., Harrison, M. J., Adcroft, A., and Zadeh, N.: Simulating water
871 residence time in the coastal ocean: a global perspective, *Geophys. Res. Lett.*, 46, 13910–13919, 2019.
- 872 Lovenduski, N. S., Yeager, S. G., Lindsay, K., and Lombardozzi, D. L.: Predicting near-term
873 variability in ocean carbon uptake, *Earth Syst. Dynam.*, 10, 45–56, 2019.
- 874 Lu, F., Harrison, M. J., Rosati, A., Delworth, T. L., Yang, X., Cooke, W. F., et al.: GFDL's SPEAR
875 seasonal prediction system: initialization and ocean tendency adjustment (OTA) for coupled model
876 predictions, *J. Adv. Model. Earth Syst.*, 12, e2020MS002149, 2020.



- 877 McKinley, G. A., Rödenbeck, C., Gloege, L., and Fay, A. R.: Pacific Ocean carbon dioxide and El
878 Niño, *Global Biogeochem. Cy.*, 18, 2004.
- 879 McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., and Lovenduski, N. S.:
880 Timescales for detection of trends in the ocean carbon sink, *Nature*, 530, 469–472, 2016.
- 881 McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L., and Lovenduski, N. S.: Ocean carbon
882 uptake under aggressive emission reductions, *Nature*, 584, 227–231, 2020.
- 883 Morioka, Y., Zhang, L., Delworth, T. L., Yang, X., Zeng, F., Nonaka, M., and Behera, S. K.:
884 Multidecadal variability and predictability of Antarctic sea ice in the GFDL SPEAR_LO model, *The*
885 *Cryosphere*, 17, 5219–5240, <https://doi.org/10.5194/tc-17-5219-2023>, 2023.
- 886 O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al.: The
887 scenario model intercomparison project (ScenarioMIP) for CMIP6, *Geosci. Model Dev.*, 9, 3461–
888 3482, 2016.
- 889 Park, J.-Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J. G., and Zhang, S.: Modeling
890 global ocean biogeochemistry with physical data assimilation: a pragmatic solution to the equatorial
891 instability, *J. Adv. Model. Earth Syst.*, 10, 891–906, 2018.
- 892 Rödenbeck, C., DeVries, T., Hauck, J., Le Quéré, C., and Keeling, R. F.: Data-based estimates of
893 interannual sea–air CO₂ flux variations 1957–2020 and their relation to environmental drivers,
894 *Biogeosciences*, 19, 2627–2652, <https://doi.org/10.5194/bg-19-2627-2022>, 2022.
- 895 Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., and Terray, L.: Drift dynamics in
896 a coupled model initialized for decadal forecasts, *Clim. Dynam.*, 46, 1819–1840, 2016.
- 897 Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al.: Multiyear
898 predictability of tropical marine productivity, *Proc. Natl. Acad. Sci. USA*, 115, E10–E18, 2018.
- 899 Sospedra-Alfonso, R., Merryfield, W. J., and Boer, G. J.: Trivial improvements in predictive skill due
900 to direct reconstruction of the global carbon cycle, *Earth Syst. Dynam.*, 12, 1139–1153, 2021.
- 901 Stenchikov, G. L., Kirchner, I., Robock, A., Graf, H. F., Antuña, J. C., Grainger, R. G., et al.:
902 Radiative forcing from the 1991 Mount Pinatubo volcanic eruption, *J. Geophys. Res.-Atmos.*, 103,
903 13837–13858, 1998.
- 904 Stock, C. A., Dunne, J. P., Luo, J. Y., Ross, A. C., VanOostende, N., Zadeh, N., et al.:
905 Photoacclimation and photoadaptation sensitivity in a global ocean ecosystem model, *J. Adv. Model.*
906 *Earth Syst.*, 17, e2024MS004701, <https://doi.org/10.1029/2024MS004701>, 2025.
- 907 Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., et al.: Long-
908 term changes in the ocean carbon cycle: a transient CMIP5 global carbon cycle projection, *J. Climate*,
909 27, 8787–8805, 2014.
- 910 Yang, X., Rosati, A., Zhang, S., Delworth, T. L., Gudgel, R. G., Zhang, W., et al.: Predictability of the
911 Atlantic meridional overturning circulation and North Atlantic surface temperature in a GFDL climate
912 model, *J. Climate*, 26, 3042–3062, 2013.



- 913 Yang, X., Jia, L., Delworth, T. L., Gudgel, R., Cooke, W., Harrison, M. J., et al.: Decadal predictability
914 of temperature and precipitation means and extremes in a GFDL coupled model, *J. Climate*, 34, 6205–
915 6228, 2021.
- 916 Yang, X., Delworth, T. L., Jia, L., et al.: Skillful seasonal prediction of wind energy resources in the
917 contiguous United States, *Commun. Earth Environ.*, 5, 313, [https://doi.org/10.1038/s43247-024-](https://doi.org/10.1038/s43247-024-01457-w)
918 [01457-w](https://doi.org/10.1038/s43247-024-01457-w), 2024.
- 919 Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., and Teng, H.: A decadal prediction case study:
920 Late twentieth-century North Atlantic Ocean heat content, *J. Climate*, 25, 5173–5189, 2012.
- 921 Zeng, J., Iida, Y., Matsunaga, T., and Shirai, T.: Surface ocean CO₂ concentration and air-sea flux
922 estimate by machine learning with modelled variable trends, *Front. Mar. Sci.*, 9, 989233,
923 <https://doi.org/10.3389/fmars.2022.989233>, 2022.
- 924 Zhang, L., Delworth, T. L., and Jia, L.: Diagnosis of decadal predictability of Southern Ocean sea
925 surface temperature in the GFDL CM2.1 model, *J. Climate*, 30, [https://doi.org/10.1175/JCLI-D-16-](https://doi.org/10.1175/JCLI-D-16-0537.1)
926 [0537.1](https://doi.org/10.1175/JCLI-D-16-0537.1), 2017.
- 927 Zhang, L., Delworth, T. L., Yang, X., Zeng, F., Lu, F., Morioka, Y., and Bushuk, M.: The relative role
928 of the subsurface Southern Ocean in driving negative Antarctic Sea ice extent anomalies in 2016–
929 2021, *Commun. Earth Environ.*, 3, 302, <https://doi.org/10.1038/s43247-022-00624-1>, 2022.
- 930 Zhang, L., Delworth, T. L., Yang, X., and Zeng, F.: Skillful multiyear to decadal predictions of sea
931 level in the North Atlantic Ocean and U.S. East Coast, *Commun. Earth Environ.*, 4, 420,
932 <https://doi.org/10.1038/s43247-023-01093-w>, 2023.
- 933 Zhang, L., Delworth, T. L., Yang, X., Zeng, F., Gu, Q., and Li, S.: Causes and multiyear predictability
934 of the rapid acceleration of U.S. Southeast Sea level rise after 2010, *npj Clim. Atmos. Sci.*, 7, 113,
935 2024.
- 936 Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., et al.: The GFDL global
937 atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning
938 strategies, *J. Adv. Model. Earth Syst.*, 10, 735–769, <https://doi.org/10.1002/2017MS001209>, 2018.
- 939