Predictable variations of the carbon sinks and atmospheric CO_2 growth in a multi-model framework

T. Ilyina[1], H. Li[1], A. Spring[1,2], W. A. Müller[1], L. Bopp[3], M. O. Chikamoto[4], G. Danabasoglu[5], M. Dobrynin[6], J. Dunne[7], F. Fransner[8], P. Friedlingstein[9], W. Lee[10], N. S. Lovenduski[11], W.J. Merryfield[10], J. Mignot[12], J.Y. Park[13], R. Séférian[14], R. Sospedra-Alfonso[10], M. Watanabe[15], S. Yeager[5]

[1]Max Planck Institute for Meteorology, Bundesstraße 53, 20146 Hamburg, Germany
[2]International Max-Planck Research School of Earth System Modelling, Bundesstraße 53, 20146, Hamburg, Germany
[3]LMD-IPSL, CNRS, Ecole Normale Supérieure / PSL Res. Univ, Ecole Polytechnique, Sorbonne Université, Paris, France
[4]Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas USA
[5]National Center for Atmospheric Research, Boulder, Colorado, USA
[6]Deutscher Wetterdienst (DWD), Hamburg, Germany
[7]NOAA/OAR Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08540 USA
[8]Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway
[9]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK
[10]Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada
[11]Department of Atmospheric and Oceanic Sciences and Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado, USA
[12]LOCEAN, Sorbonne Universites/IRD/CNRS/MNHN, Paris, France
[13]Department of Earth and Environmental Sciences, Jeonbuk National University, Jeollabuk-do 54896 Republic of Korea
[14]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[15]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 3173-25, Showa-machi, Kanazawa-ku, Yokohama, Kanagawa, 236-0001, Japan

Corresponding author: Tatiana Ilyina, `tatiana.ilyina@mpimet.mpg.de`

Key Points:

- Predictive skill of the global ocean carbon sink due to initialization is up to 6 years, with longer regional predictability in single models.
- Predictive skill due to initialization for the land carbon sink of up to 2 years is primarily maintained in the tropics and extra-tropics.
- Anomalies of atmospheric CO_2 growth rate are predictable up to 2 years and are limited by the land carbon sink predictability horizon.

Abstract

Inter-annual to decadal variability in the strength of the land and ocean carbon sinks impede accurate predictions of year-to-year atmospheric carbon dioxide (CO_2) growth rate. Such information is crucial to verify the effectiveness of fossil fuel emissions reduction measures. Using a multi-model framework comprising prediction systems based on Earth system models, we find a predictive skill for the global ocean carbon sink of up to 6 years. Longer regional predictability horizons and robust spatial patterns are found across single models. On land, a predictive skill of up to 2 years is primarily maintained in the tropics and extra-tropics enabled by the initialization of the physical climate variables towards observations. We further show that anomalies of atmospheric CO_2 growth rate inferred from natural variations of the land and ocean carbon sinks are predictable at lead time of 2 years and the skill is limited by the land carbon sink predictability horizon.

Plain Language Summary

Variations of the natural land and ocean carbon sinks in response to climate variability strongly regulate year-to-year variations in the growth rate of atmospheric carbon dioxide (CO_2). Information on the near-term evolution of the carbon sinks and CO_2 in the atmosphere is necessary to understand where the anthropogenic carbon would go in response to emission reduction efforts addressing global warming mitigation. Predictions of this near-term evolution would thus assist policy-relevant analysis. Here we use a set of prediction systems based on Earth system models to establish predictive skills of the ocean and land carbon sinks and to infer predictability of atmospheric CO_2 growth rate. We show predictability horizons of up to 6 years for the globally integrated ocean carbon sink in individual models with even higher predictive skill in some models and regions. Variations of the land carbon sink are predictable up to 2 years and limit predictability of changes in atmospheric CO_2 growth rate at lead time of 2 years. Our study demonstrates an emerging capacity of the initialized simulations for skillful predictions of the global carbon sink and atmospheric CO_2 variations.

1 Introduction

On interannual to decadal time-scales, atmospheric CO_2 growth rates exhibit pronounced anomalies driven by varying strengths of the land and ocean carbon sinks; these

anomalies are linked to climate variability (Peters et al., 2017; Friedlingstein et al., 2019; Landschützer et al., 2019). Variability in ocean carbon uptake is associated with major carbon uptake regions such as the Southern Ocean and the North Atlantic (Landschützer et al., 2019). Inter-annual variations of the land carbon sink are primarily driven by the terrestrial biosphere response to El Niño Southern Oscillation (ENSO) (Ropelewski & Halpert, 1987; Jones et al., 2001; Zeng et al., 2005; Kim et al., 2016). Year-to-year variations of the air-land carbon flux are about one order of magnitude higher than variations in the air-sea CO_2 fluxes (Doney et al., 2006). Hence, predicted El Niño variability has been used, in combination with an average CO_2 growth rate due to anthropogenic CO_2 emissions, to predict, from a simple linear regression, the atmospheric CO_2 growth at Mauna Loa for the subsequent year (Betts et al., 2016). Predicting changes in atmospheric CO_2 growth rate beyond this horizon remains a major challenge. Such information will be essential for the evaluation of mitigation efforts in real-time in the presence of internal climate variability in support of policy-relevant analysis for the UNFCCC global stocktakes (UNFCCC, 2015).

Recent initialized predictions of near-term future climate have proven successful (Marotzke et al., 2016; Smith et al., 2007) with predictive power of carbon sinks also emerging. Li et al. (2019) established a predictive skill of the globally aggregated air-sea CO_2 fluxes of up to 2 years assessed against an observational product. Longer predictability in regions like the North Atlantic and the Southern Ocean is suggested (Li et al., 2016; Lovenduski, Yeager, et al., 2019; Fransner et al., 2020). ESM-based initialized prediction systems also demonstrate predictability of other marine biogeochemical properties such as net primary production, export production, and seawater pH (Park et al., 2019; Séférian et al., 2014; Yeager et al., 2018; Brady et al., 2020; Fransner et al., 2020; Krumhardt et al., 2020). On the land side, a potential prediction skill of 2 years was established for terrestrial net ecosystem production (Lovenduski, Bonan, et al., 2019), but only of 9 months for tropical land-atmosphere carbon flux (Zeng et al., 2008). Perfect-model frameworks based on idealized simulations suggest analogous predictability horizons for the carbon sinks (Séférian et al., 2018; Spring & Ilyina, 2020). However, previous studies were either limited to internally consistent model environments of perfect models (Séférian et al., 2018; Spring & Ilyina, 2020; Frölicher et al., 2020) or single initialized models (Li et al., 2019, 2016; Lovenduski, Yeager, et al., 2019; Yeager et al., 2018; Fransner et al., 2020;

Krumhardt et al., 2020). Furthermore, they did not address predictability of variations in atmospheric CO_2 growth.

Here, we assess how well different ESM-based initialized prediction systems capture variations of the global land and ocean carbon sinks and their predictability. We make a step further and for the first time examine the resulting predictability of variations in the growth rate of atmospheric CO_2 that is driven by the response of carbon sinks to climate variability. As predictions of carbon sink evolution still remain a cutting-edge activity of only a few modeling groups, a common protocol is not yet available (Merryfield et al., 2020). Our multi-model framework comprises ESM-based prediction systems that contributed to the Decadal Climate Prediction Project (DCPP; Boer et al. (2016)) within the Coupled Model Intercomparison Project Phase 6 (CMIP6), as well as those which run with the CMIP5 forcing. This enables us to establish predictive skills in a larger number of models, whilst performance of CMIP5 and CMIP6 model versions with respect to different aspects of the carbon cycle has been addressed in recent studies (Arora et al., 2019; Séférian et al., 2020; Kwiatkowski et al., 2020). Prediction systems follow somewhat different initialization techniques and data assimilation methods based on the "best effort" of the different modeling centers. This approach arises from the overall DCPP philosophy of not specifying single details of the implementation and design of the multi-model predictions and thereby encompass aspects of the inherent uncertainty of climate predictions (Boer et al., 2016).

2 Materials and Methods

We use a multi-model framework comprising several ESM-based prediction systems, including CanESM5 (Swart et al., 2019), CESM-DPLE (Yeager et al., 2018), GFDL-ESM2 (Park et al., 2018), IPSL-CM6A-LR (Boucher et al., 2020), MIROC-ES2L (Watanabe et al., 2020), MPI-ESM-LR (Giorgetta et al., 2013), MPI-ESM1.2-HR (Mauritsen et al., 2019), and NorCPM1 (Counillon et al., 2016). Details of each prediction system are given in Supporting Information. Simulations with CanESM5, IPSL-CM6A-LR, MIROC-ES2L, MPI-ESM1.2-HR, and NorCPM1 contributed to CMIP6 DCPP following historical forcing until 2014 and climate change scenario SSP2-4.5. Simulations with CESM-DPLE, GFDL-ESM2, and MPI-ESM-LR were performed under CMIP5 historical forcing until the year 2005 and followed either RCP4.5 (GFDL-ESM2, MPI-ESM-LR) or RCP8.5 (CESM-DPLE) climate change scenario thereafter.

The ensemble size in single prediction systems ranges between at least 10 members for most of the models up to 40 for CESM-DPLE (Table S1), enabling us to demonstrate the added value of a larger ensemble. For NorCPM1, we merged the two decadal hindcast products with 10 members each, producing one ensemble of 20 members. In the MPI-ESM based systems, only the lower resolution MPI-ESM-LR included both the land and the ocean biogeochemistry components. MPI-ESM1.2-HR was configured with a higher resolution in the atmosphere and ocean, but did not integrate the land biogeochemistry component.

In all models the carbon cycle components are only indirectly initialized with the data assimilative physics. Hence, we assess observed variability in carbon sinks captured through initialization of prediction systems by the observed state of the physical climate. All simulations ran with prescribed evolution of atmospheric CO_2 concentrations and land use change.

We present three types of simulations. Reconstruction simulations include observed signals of climate variability introduced by assimilative observed and reanalysis products over a hindcast period. Uninitialized simulations are based on continuous historical simulations following CMIP6 or CMIP5 forcing (not the observed signals), i.e. the model physics evolves independently and the resulting climate variability does not necessarily match the observed one. Initialized simulations are retrospective prediction simulations that start from a respective reconstruction simulations and develop internal climate variability that may be out of phase with observed climate variability. We compare the initialized simulations against the uninitialized ones to assess predictive skill that is established due to initialization. This predictive skill is characterized by the anomaly correlation coefficients (ACC) between the model simulations and different reference data products. The anomalies are calculated by removing the climatological mean for the reconstruction and uninitialized simulations, and for the initialized simulations with additionally respect to the lead time. Note that we present the improved predictive skill due to initialization based on the comparison of ACC in the initialized predictions relative to that in the uninitialized simulations. We use a bootstrapping resample method to quantify the significance of the improved predictive skill (Li et al., 2019). The spatial map of predictive skill and the corresponding significance is generated by the central evaluation system MurCSS, which is a commonly used evaluation tool in decadal predictions (Illing et al., 2014). The focus time period of this study is from 1982-2013,

when the global carbon cycle experienced large internanual to decadal variations. The time series are all linearly detrended to emphasize the predictability in interannual to decadal variability. The global time series are integrated based on the original model grid. For the spatial pattern of ACC calculation, the variables are conservatively interpolated into 5 degree.

For land carbon uptake, direct observational estimates capturing the regional and global temporal variability are not available, hence we use the Global Carbon Budget 2019 (GCB; Friedlingstein et al. (2019)) carbon sinks estimate as a benchmark. Undergoing annual updates, GCB offers a comprehensive and temporally consistent time-series of stand-alone land and ocean carbon cycle model simulations forced with observed climate data or climate reanalysis and additional observational products (atmospheric CO_2, land cover change, etc.). For ocean carbon uptake, we additionally use the SOM-FFN (Landschützer et al., 2015) observationally based product. In addition, the HadISST data (Rayner et al., 2003) is used to compare with model simulations of sea surface temperature and to calculate the ENSO index.

3 Variations of ocean and land carbon sinks in initialized simulations

First we examine the ability of reconstructions and initialized predictions to simulate observed interannual variations in carbon sinks. Both reconstructions and initialized predictions at lead time of 2 years appropriately capture multi-year variations of the anomalous air-sea flux of CO_2 represented in the GCB and data-based SOM-FFN estimates (Fig.1 left). The uninitialized simulations mostly capture only ocean carbon sink increases in response to rising carbon emissions and thus follow a smoother temporal evolution. Furthermore, reconstructions suggest stronger multi-year variations in the ocean carbon sink and outperform the uninitialized simulations in GFDL-ESM2, MIROC-ES2L, MPI-ESM1.2-HR, and in NorCPM1 (only in comparison to SOM-FFN data). Lower correlations of reconstruction simulations as opposed to the uninitialized ones in CanESM5, IPSL-CM6A-LR, and MPI-ESM-LR can be related to two aspects of the design of our analysis. First, the assimilation techniques may not be optimally calibrated to represent ocean biogeochemistry in reconstruction (Park et al., 2018; Li et al., 2019). Second, GCB and SOM-FFN estimates chosen as the reference here are prone to their own uncertainties. GCB estimates are essentially an average of various stand-alone hindcast model simulations. The neural network approach of SOM-FFN is limited by spatial and tempo-

ral gaps in observations. While the different model outputs show a large spread in air-sea CO_2 flux, they overall fall within the range of the SOCOM data products (Rödenbeck et al., 2015). The weakening ocean carbon sink captured in the SOM-FFN data product in the 1990s is revealed by the stronger negative trends in the reconstruction and initialized simulations vs. the uninitialized ones, which are more pronounced in some predictions systems (in MPI-ESM-LR, MPI-ESM-HR, MIROC-ES2L, and partially in GFDL-ESM2). Other models (CanESM5, CESM-DPLE, NorCPM1, IPSL-CM6A-LR) capture a lower amplitude of the weakened ocean carbon sink, more consistent with the GCB estimate. Starting from the beginning of the 21st century, reconstruction simulations show an enhancement of the ocean carbon uptake with a stronger increase in the ocean carbon sink at the beginning of the 21st century as compared to the uninitialized ones. This decadal shift in evolution of the ocean carbon sink at the onset of the 21st century is attributable to climate modulated variability and is consistent with the SOM-FFN data estimate (Landschützer et al., 2015).

The fewer land carbon reconstruction simulations available to us all outperform the uninitialized simulations in capturing the major year-to-year variations as indicated by higher correlations with GCB (Fig.1 right). This correlation skill with the GCB estimates is maintained at lead year 2. Unsurprisingly, uninitialized simulations do not capture the timing of air-land CO_2 flux variations. Response to the warm and cold episodes of ENSO, the major driver of year-to-year variability of the air-land carbon fluxes, is clearly manifested in the GCB estimates and reconstructions (Fig. S1). It is notable that air-land CO_2 flux in CanESM5 has the highest correlation with GCB in reconstruction simulation, supported by the highest of all models correlation in the uninitialized simulation (Fig.1d). For NorCPM1 and MPI-ESM-LR assimilation data helps to establish correlation in reconstruction simulations. While there has been some progress in global models over the past decades (Bellenger et al., 2014), representing ENSO still remains a major challenge. Yet, a major improvement in the reconstruction simulations with respect to air-land CO_2 flux, gives us confidence that initialized prediction systems capture the important processes that link the land carbon cycle to ENSO. The reconstruction simulations produce a distinct weakening of the land carbon uptake in response to major El Niño events, followed by a strong increase in the land carbon sink during La Niña events. These variations are not captured in the uninitialized simulations as they are not in phase with the observed climate variability.

4 Predictability of carbon sinks and atmospheric CO_2 growth rate

We next examine effects of the global land ocean carbon sink variations on the inferred variability and predictability of atmospheric CO_2 growth rate (Fig.2). Note that all prediction systems available to us are forced with prescribed evolution of atmospheric concentrations of CO_2 (rather than with prescribed emissions of CO_2) and so the atmospheric compartments of those models do not respond to land or ocean CO_2 fluxes. Here, the detrended sum of the global land and ocean carbon fluxes serves as a diagnostic of variations in the temporal evolution of the atmospheric CO_2 growth driven by climate modulated variability of carbon sinks. These variations of a few PgC in the reconstruction simulations generally follow the evolution inferred from the GCB estimate (Fig.2a).

We find predictability of variations in atmospheric CO_2 growth at lead times of 2 years in most models, as indicated by higher correlations with GCB of the initialized simulations in comparison to the uninitialized ones (Fig.2 b, c). Given the higher amplitude of interannual air-land CO_2 flux variability, atmospheric carbon growth rate anomalies predominantly follow the land carbon sink evolution, and the ocean carbon sink acts to dampen the land modulated interannual variations of atmospheric CO_2 (Doney et al., 2006; Lee et al., 1998). Indeed, the improved correlation skill of air-land CO_2 fluxes with the GCB estimates is maintained at lead year 2 and outperforms the uninitialized simulations in all models except MIROC-ES2L (Fig.1f).

We further assess predictability horizons of the global ocean and land carbon sinks, as well as of the diagnosed changes in atmospheric CO_2 growth represented by the lead years with improved predictive skill due to initialization (Fig.3). Predictive skill of the ocean carbon sink significantly improves with initialization up to lead year 5 against the SOM-FFN data product in MPI-ESM1.2-HR and up to lead year 6 in CESM-DPLE and NorCPM1, respectively (Fig.3a). The predictive skill of CESM-DPLE is higher than reported in a previous study (Lovenduski, Yeager, et al., 2019) mainly because we focus on a different time period and use the SOM-FFN observationally based estimates rather than reconstruction here. A larger ensemble size of CESM-DPLE relative to the outputs from the other prediction systems maintains the predictive skill significance. Considering fewer ensemble members degrades its predictive skill significance (as indicated by the p-value dependence on ensemble size; Fig. S2). A previous study (Li & Ilyina, 2018) suggests that a large ensemble size is needed to capture decadal variations in the ocean car-

bon sink. Therefore, an increased ensemble prediction size could enhance the predictive skill of global carbon fluxes in other prediction systems, as well as in a multi-model ensemble.

Predictive skill due to initialization up to lead year 2 for land carbon sink verified against GCB estimates is found in CanESM5, IPSL-CM6A-LR, MPI-ESM, and NorCPM1 (Fig.3b). This skill, supported by higher coherence between GCB estimates and initialized simulations at lead time of 2 years in most models (Fig.1f), goes well beyond a seasonal skill attainable in previous studies. A slightly lower and insignificant skill was found for CESM-DPLE because of the initialization of atmosphere and land from a random ensemble member of CESM-LE (see Materials and Methods and Lovenduski, Bonan, et al. (2019)).

The atmospheric CO_2 growth rate changes induced by land and ocean carbon sink variations show predictive skill to lead year 2 (Fig.3) in the same models which have significant 2 year predictive horizons for the land carbon sink (i.e. in CanESM5, IPSL-CM6A-LR, MPI-ESM, and NorCPM1). Given the longer predictive horizons of the ocean carbon sink, our results indicate that predictability of the atmospheric CO_2 growth in these models is limited by the land carbon sink predictability. Analogously, a previous study, based on a perfect model framework (Spring & Ilyina, 2020), demonstrates that the predictive skill of atmospheric CO_2 concentration of 3 years is dampened by land.

5 Spatial patterns of predictability horizons of CO_2 fluxes

The prediction systems use different initialization techniques and data assimilation methods, but do they establish robust spatial patterns of predictability horizons in the carbon cycle? To address this question we examine predictability horizons due to added value of initialization, represented by the lead years when correlations of the initialized simulations are larger than those in the uninitialized ones. We find overall rather consistent CO_2 flux predictability horizons established due to initialization in the different prediction systems (Fig.4).

In some ocean regions, the improved skill is retained for up to 9-10 years, thereby going beyond the predictability horizons of the physical climate variables (Séférian et al., 2014; Li et al., 2016). We find regional improvements in air-sea CO_2 flux predictability due to initialization (as indicated by the difference between the initialized and unini-

tialized simulations and consistent spatial patterns of predictive skill at lead year 2 across models shown in Fig. S3-S4). These improved regions differ across the models and when assessing them vs. the GCB estimates or the SOM-FFN (Fig. S3) and the reconstructions (Fig. S4), highlighting the importance of high-quality reference for skillful predictions of the ocean carbon sink. The air-sea CO_2 flux dynamics is regulated by the temporal gradient of surface ocean pCO_2. Because of the fast equilibration of CO_2 between atmosphere and surface ocean in most areas, pCO_2 tracks atmospheric CO_2 evolution. This feature is fairly well captured in ocean biogeochemical models (Roy et al., 2011). Furthermore, our previous findings (Li et al., 2019) suggest that temperature variations largely control shorter-term (<3 years) predictability of the ocean carbon sink, while longer-term (>3 years) predictability is associated with nonthermal drivers. Coherent representation of the spatial patterns of the air-sea CO_2 flux in the different prediction systems may be driven by the robust representation of SST variations in the initialized predictions considered here (Fig. S5).

On the land side, statistically significant improvements due to initialization (in CanESM5, IPSL-CM6A-LR, MPI-EMS-LR, and NorCPM1) are suggested in regions of the tropics (e.g. Amazon, West Africa) and extra-tropics (e.g. Middle East, US Great Plains, Eastern Russia). These prediction systems represent land carbon fluxes improved due to initialization at lead time of 2 years. Less pronounced predictive skill of land carbon fluxes is found in CESM-DPLE due to the initialization of atmosphere and land from a random ensemble member; see Materials and Methods and Lovenduski, Bonan, et al. (2019)).

6 Conclusions

One major requirement related to the goal of the Paris Agreement of "limiting warming to well below 2°C, and pursuing efforts to 1.5°C", is to discern the pathways of anthropogenic carbon in the Earth system in order to verify the effectiveness of fossil fuel emissions reduction measures. A major scientific challenge in this context will be to predict the inter-annual and decadal variations of the natural carbon sinks and the related variations in the growth rate of atmospheric CO_2, as well as their susceptibility to ongoing climate change. Thus, predictability of variations of the global carbon cycle is a crucial emerging topic requiring fast advance as it relates to the global stocktaking requirements of the Paris Agreement.

Here we provide a first multi-model assessment of the initialized carbon cycle predictions, which is an important step towards skillful near-term predictions of the evolution of the land and ocean carbon sinks and the resulting variations in atmospheric CO_2 growth in response to climate variability and changes in anthropogenic carbon emissions. We find improved predictive skill due to initialization in both ocean and land carbon sinks. Predictive skill due to initialization for the global air-sea CO_2 flux is up to 6 years. There is indication of even higher regional skill in single models and regions. Representation of air-land CO_2 flux improved due to initialization in all models considered in this study. We demonstrate predictive horizons of up to 2 years in 4 out of the 6 models considered in this study. As year-to-year variations in atmospheric CO_2 are largely determined by variations of the land carbon sink, the predictability horizon of 2 years found for the atmospheric CO_2 growth rate is maintained by predictability of air-land CO_2 flux.

Ongoing challenges in predictions of the global carbon cycle include a lack of observationally based products suitable to initialize the ESMs and to verify prediction skill, the unavailability of standardized multi-model simulations that include prognostic carbon cycle components, and the insufficient prediction ensemble size that impairs significance assessment. Despite these challenges, our analysis provides clear indications that further advancement of the physical and biogeochemical components of prediction systems and larger ensembles could timely address some of these challenges as new prediction simulations and updated observational products become available. Our analysis demonstrates an emerging capacity of the initialized simulations for skillful predictions of the carbon cycle. Thus, such multi-model initialized predictions would offer a powerful tool in support of governmental and economical decisions related to verification and efficiency assessment of near-term carbon emission reduction pathways.

Acknowledgments

Funding: T.I., H.L., W.A.M. were supported by the Federal Ministry of Education and Research in Germany (BMBF) through the research program MiKlip (grant no. 01LP1517B). T.I., P.F., H.L., L.B., R.S. were supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 641816 (CRESCENDO). T.I., P.F., H.L. L.B. were supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 821003 (4C). R.S. was supported by the TRIATLAS project under grant agreement no. 817578. N.S.L. was supported by the U.S.

National Science Foundation (OCE-1752724). F.F. was supported by the Nansen Legacy Project (276730) and Det Kongelige Norske Videnskabers Selskap. M.C. was supported by National Science Foundation (OCE-1635465) and grant-in-aid for Scientific Research on Innovative Areas (15H05817). J.-Y.P. was supported by NOAA's marine ecosystem tipping points initiative and the National Research Foundation of Korea (NRF-2020R1A4A3079510). M.W. was supported by the Integrated Research Program for Advancing Climate Models (TOUGOU) Grant Number JPMXD0717935457 and JPMXD0717935715 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. The National Center for Atmospheric Research (NCAR) is a major facility sponsored by the US National Science Foundation (NSF) under Cooperative Agreement No. 1852977. Simulations with MPI-ESM were performed at the German Climate Computing Center (DKRZ). The CESM-DPLE was generated using computational resources provided by the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy under contract no. DEAC02-05CH11231, as well as by an Accelerated Scientific Discovery grant for Cheyenne (https://doi.org/10.5065/D6RX99HX) that was awarded by NCAR's Computational and Information Systems Laboratory. Primary data and scripts used in the analysis of this study are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de.

References

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., ... Ziehn, T. (2019). Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models. *Biogeosciences Discussions*, *2019*, 1–124. Retrieved from `https://bg.copernicus.org/preprints/bg-2019-473/` doi: 10.5194/bg-2019-473

Bellenger, H., Guilyardi, É., Leloup, J., Lengaigne, M., & Vialard, J. (2014). ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dynamics*, *42*(7-8), 1999–2018.

Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., & Kennedy, J. J. (2016). El Niño and a record co_2 rise. *Nature Climate Change*, *6*(9), 806–810.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., ... others (2016). The decadal climate prediction project (DCPP) contribution to CMIP6. *Geoscientific Model Development (Online)*, *9*(10).

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... others (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7).

Brady, R. X., Lovenduski, N. S., Yeager, S. G., Long, M. C., & Lindsay, K. (2020). Skillful multiyear predictions of ocean acidification in the california current system. *Nature Communications*, *11*(1), 1–9.

Counillon, F., Keenlyside, N., Bethke, I., Wang, Y., Billeau, S., Shen, M. L., & Bentsen, M. (2016). Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 32437. Retrieved from `https://doi.org/10.3402/tellusa.v68.32437` doi: 10.3402/tellusa.v68.32437

Doney, S. C., Lindsay, K., Fung, I., & John, J. (2006). Natural variability in a stable, 1000-yr global coupled climate–carbon cycle simulation. *Journal of climate*, *19*(13), 3033–3054.

Fransner, F., Counillon, F., Bethke, I., Tjiputra, J., Samuelsen, A., Nummelin, A., & Olsen, A. (2020). Ocean biogeochemical predictions—initialization and limits of predictability. *Frontiers in Marine Science*, *7*, 386.

Friedlingstein, P., Jones, M., O'sullivan, M., Andrew, R., Hauck, J., Peters, G., ... others (2019). Global carbon budget 2019. *Earth System Science Data*, *11*(4), 1783–1838.

Frölicher, T. L., Ramseyer, L., Raible, C. C., Rodgers, K. B., & Dunne, J. (2020). Potential predictability of marine ecosystem drivers. *Biogeosciences*, *17*, 2061-2083.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., ... others (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597.

Illing, S., Kadow, C., Oliver, K., & Cubasch, U. (2014). Murcss: A tool for standardized evaluation of decadal hindcast systems. *Journal of Open Research Software*, *2*(1).

Jones, C. D., Collins, M., Cox, P. M., & Spall, S. A. (2001). The carbon cycle response to ENSO: A coupled climate–carbon cycle model study. *Journal of Cli-*

mate, *14*(21), 4113–4129.

Kim, J.-S., Kug, J.-S., Yoon, J.-H., & Jeong, S.-J. (2016). Increased atmospheric co_2 growth rate during El Niño driven by reduced terrestrial productivity in the CMIP5 ESMs. *Journal of Climate*, *29*(24), 8783–8805.

Krumhardt, K. M., Lovenduski, N. S., Long, M. C., Luo, J. Y., Lindsay, K., Yeager, S., & Harrison, C. (2020). Potential predictability of net primary production in the ocean. *Global Biogeochemical Cycles*, *34*, e2020GB006531.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., ... others (2020). Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences*, *17*(13), 3439–3470.

Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C., Van Heuven, S., ... others (2015). The reinvigoration of the southern ocean carbon sink. *Science*, *349*(6253), 1221–1224.

Landschützer, P., Ilyina, T., & Lovenduski, N. S. (2019). Detecting regional modes of variability in observation-based surface ocean pco_2. *Geophysical Research Letters*, *46*(5), 2670–2679.

Lee, K., Wanninkhof, R., Takahashi, T., Doney, S. C., & Feely, R. A. (1998). Low interannual variability in recent oceanic uptake of atmospheric carbon dioxide. *Nature*, *396*(6707), 155–159.

Li, H., & Ilyina, T. (2018). Current and future decadal trends in the oceanic carbon uptake are dominated by internal variability. *Geophysical Research Letters*, *45*(2), 916–925.

Li, H., Ilyina, T., Müller, W. A., & Landschützer, P. (2019). Predicting the variable ocean carbon sink. *Science advances*, *5*(4), eaav6471.

Li, H., Ilyina, T., Müller, W. A., & Sienz, F. (2016). Decadal predictions of the north atlantic co_2 uptake. *Nature communications*, *7*(1), 1–7.

Lovenduski, N. S., Bonan, G. B., Yeager, S. G., Lindsay, K., & Lombardozzi, D. L. (2019). High predictability of terrestrial carbon fluxes from an initialized decadal prediction system. *Environmental Research Letters*, *14*(12), 124074.

Lovenduski, N. S., Yeager, S. G., Lindsay, K., & Long, M. C. (2019). Predicting near-term variability in ocean carbon uptake. *Earth System Dynamics (Online)*, *10*(1).

Marotzke, J., Müller, W. A., Vamborg, F. S., Becker, P., Cubasch, U., Feldmann, H., ... others (2016). MiKlip: A national research project on decadal climate prediction. *Bulletin of the American Meteorological Society*, *97*(12), 2379–2394.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... others (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing co_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038.

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A., ... others (2020). Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, *101*(6), E869–E896.

Park, J.-Y., Stock, C. A., Dunne, J. P., Yang, X., & Rosati, A. (2019). Seasonal to multiannual marine ecosystem prediction with a global Earth system model. *Science*, *365*(6450), 284–288.

Park, J.-Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J., & Zhang, S. (2018). Modeling global ocean biogeochemistry with physical data assimilation: a pragmatic solution to the equatorial instability. *Journal of Advances in Modeling Earth Systems*, *10*(3), 891–906.

Peters, G. P., Le Quéré, C., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Ilyina, T., ... others (2017). Towards real-time verification of CO2 emissions. *Nature Climate Change*, *7*(12), 848–850.

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., ... Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(D14).

Rödenbeck, C., Bakker, D. C., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., ... others (2015). Data-based estimates of the ocean carbon sink variability–first results of the surface ocean pco_2 mapping intercomparison (SOCOM). *Biogeosciences*, *12*, 7251–7278.

Ropelewski, C. F., & Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly weather review*, *115*(8), 1606–1626.

Roy, T., Bopp, L., Gehlen, M., Schneider, B., Cadule, P., Frölicher, T. L., ... Joos,

F. (2011). Regional impacts of climate change and atmospheric co_2 on future ocean carbon uptake: A multimodel linear feedback analysis. *Journal of Climate*, *24*(9), 2300–2318.

Séférian, R., Berthet, S., & Chevallier, M. (2018). Assessing the decadal predictability of land and ocean carbon uptake. *Geophysical Research Letters*, *45*(5), 2455–2466.

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., ... others (2020). Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. *Current Climate Change Reports*, *6*, 95–119.

Séférian, R., Bopp, L., Gehlen, M., Swingedouw, D., Mignot, J., Guilyardi, E., & Servonnat, J. (2014). Multiyear predictability of tropical marine productivity. *Proceedings of the National Academy of Sciences*, *111*(32), 11646–11651.

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007). Improved surface temperature prediction for the coming decade from a global climate model. *science*, *317*(5839), 796–799.

Spring, A., & Ilyina, T. (2020). Predictability horizons in the global carbon cycle inferred from a perfect-model framework. *Geophysical Research Letters*, *47*(9), e2019GL085311.

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... others (2019). The Canadian Earth System Model version 5 (CanESM5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873.

UNFCCC, V. (2015). Adoption of the paris agreement. i: proposal by the president (draft decision). *United Nations Office, Geneva (Switzerland)*.

Watanabe, M., Tatebe, H., Koyama, H., Hajima, T., Watanabe, M., & Kawamiya, M. (2020). Importance of El Niño reproducibility for reconstructing historical co_2 flux variations in the equatorial Pacific. *Ocean Science Discussions*, *2020*, 1–21. Retrieved from `https://os.copernicus.org/preprints/os-2020-32/` doi: 10.5194/os-2020-32

Yeager, S., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., ... others (2018). Predicting near-term changes in the earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bulletin of the American Meteorological Society*, *99*(9), 1867–1886.

Zeng, N., Mariotti, A., & Wetzel, P. (2005). Terrestrial mechanisms of interannual CO2 variability. *Global biogeochemical cycles*, *19*(1).

Zeng, N., Yoon, J.-H., Vintzileos, A., Collatz, G. J., Kalnay, E., Mariotti, A., ... Lord, S. (2008). Dynamical prediction of terrestrial ecosystems and the global carbon cycle: A 25-year hindcast experiment. *Global Biogeochemical Cycles*, *22*(4).

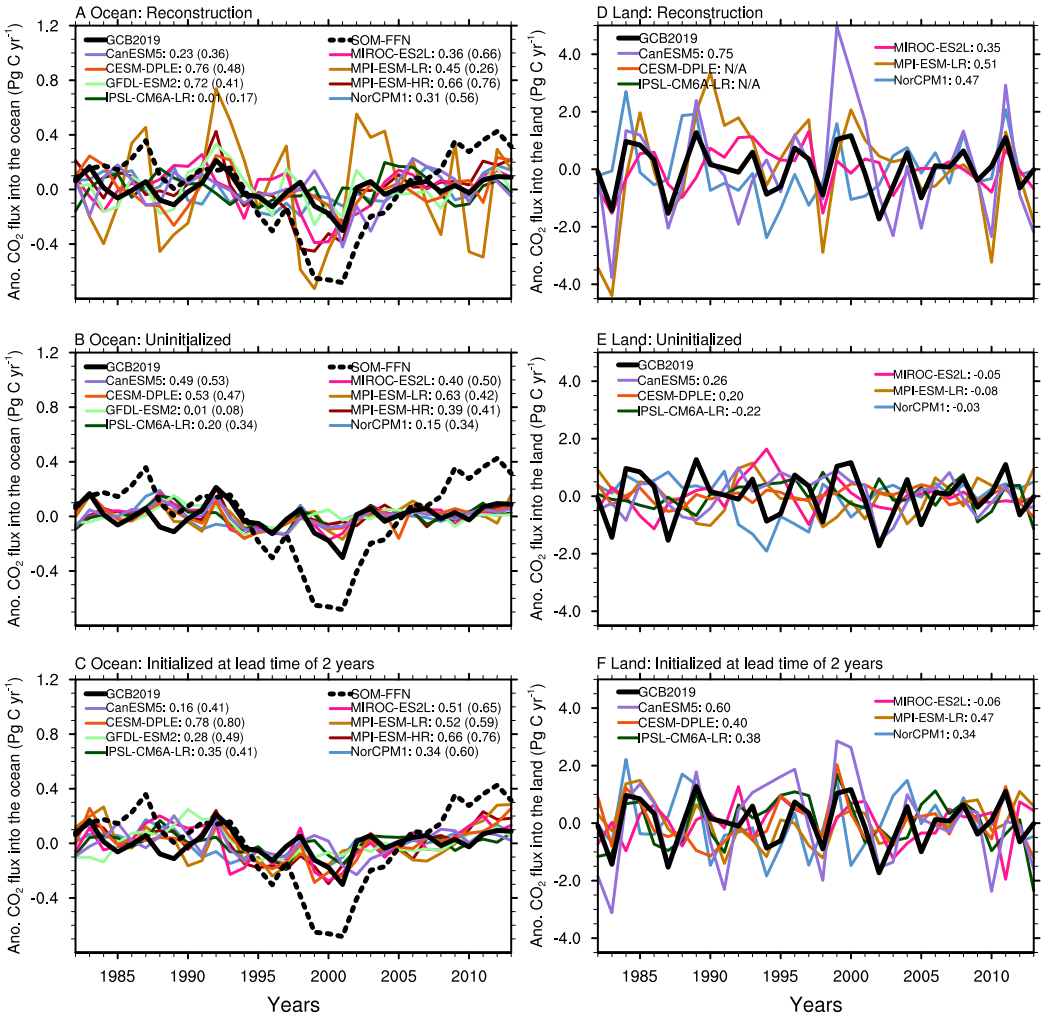

Figure 1. Time series of the global anomalous CO_2 flux relative to the climatological mean in each modeling system into the ocean (left) and land (right) from reconstruction (top), uninitialized simulation, (middle) and initialized retrospective prediction (bottom) simulations at lead time of 2 years. The long-term linear trends are removed for all the time-series. Left panels include available observation-based estimates from SOM-FFN. Numbers on the legends show the correlations relative to GCB and correlations relative to SOM-FFN data based estimates of the CO_2 flux into the ocean (shown in brackets). Outputs for air-land CO_2 fluxes from the reconstruction simulation were not available from IPSL-CM6A-LR and CESM-DPLE.

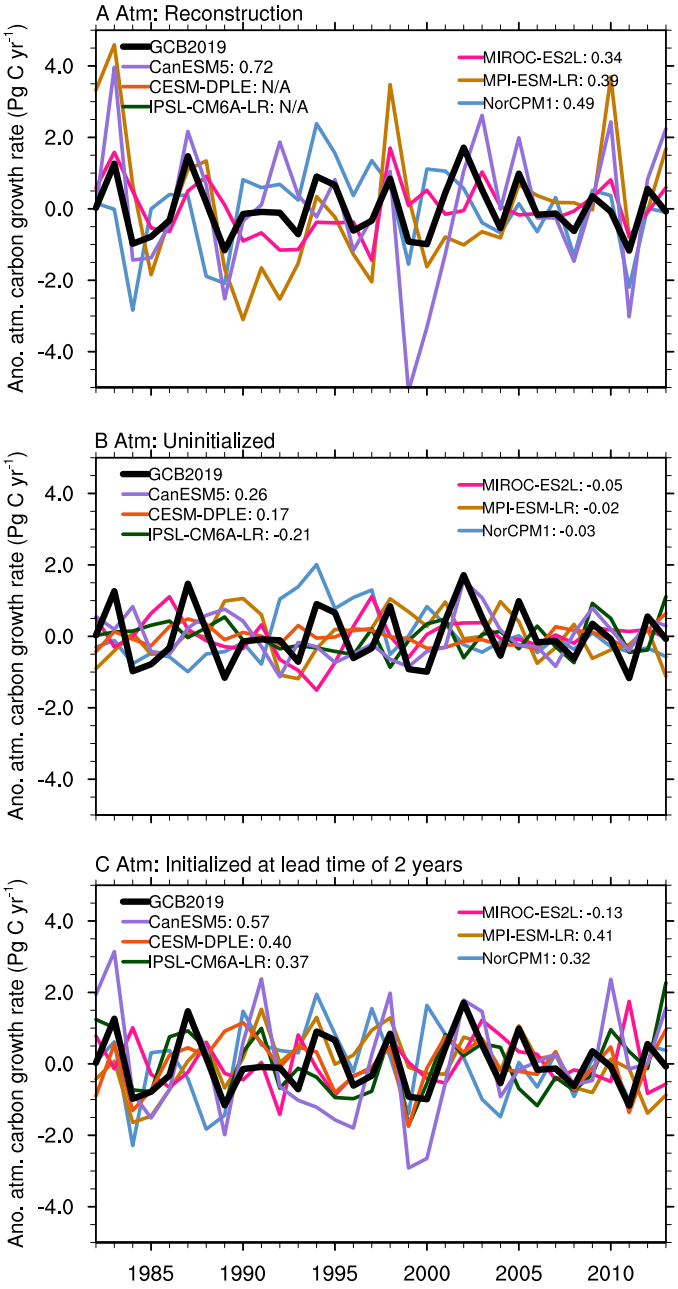

Figure 2. Time series of the anomalous atmospheric carbon growth rate due to natural variations of the ocean and land carbon sinks, represented by the reverse sign of the detrended land and ocean carbon sinks from reconstruction (top), uninitialized simulation, (middle) and initialized retrospective prediction (bottom) simulations at lead time of 2 years. Numbers on the legends show the correlations relative to GCB. Outputs for air-land CO_2 fluxes from the reconstruction simulation were not available from IPSL-CM6A-LR and CESM-DPLE, preventing the computation of the anomalous atmospheric carbon growth rates in these systems.

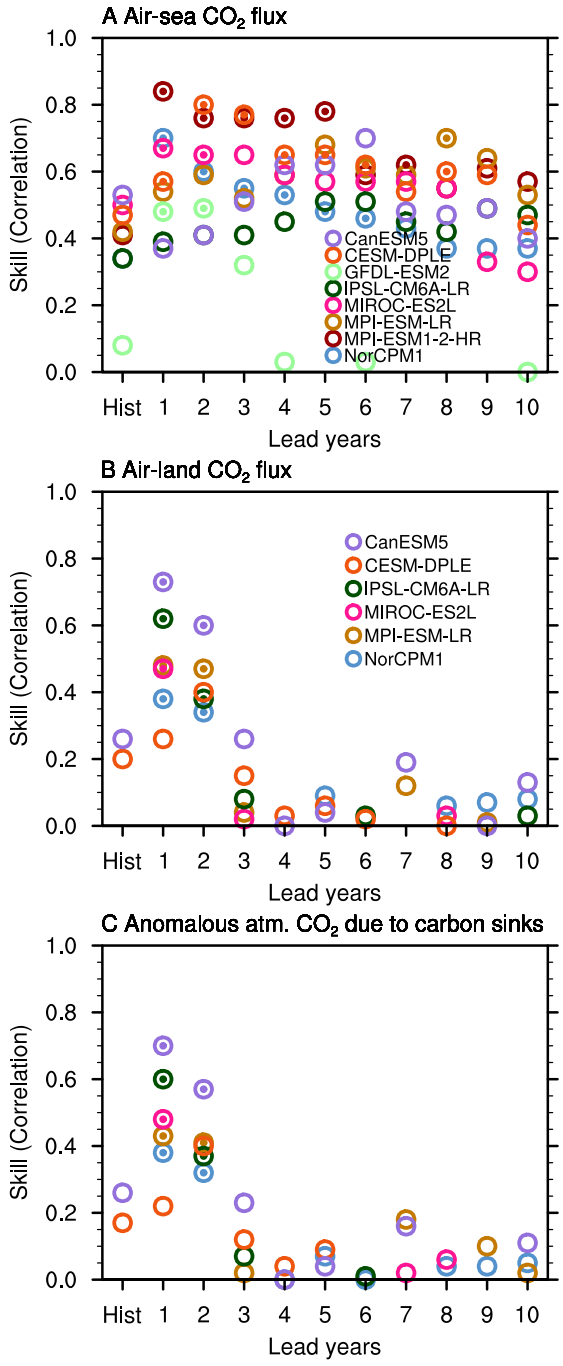

Figure 3. Predictive skill of the detrended CO_2 flux into the ocean (a), CO_2 flux into the land (b), and variations in the inferred atmospheric CO_2 growth (c). Predictive skill is quantified as anomaly correlation coefficients of the model simulations with the SOM-FFN observation-based product for the air-sea CO_2 fluxes (a), and with GCB2019 for the air-land CO_2 flux and anomalous atmospheric CO_2 due to carbon sinks. Significantly improved predictive skill at 95% level for initialized over uninitialized simulations are marked with filled dots, p-values given in Table S2. Note that GFDL-ESM2 and MIROC-ES2L hindcasts start earliest from year 1980, so from lead year 4 the time period is shorter than 1982-2013.

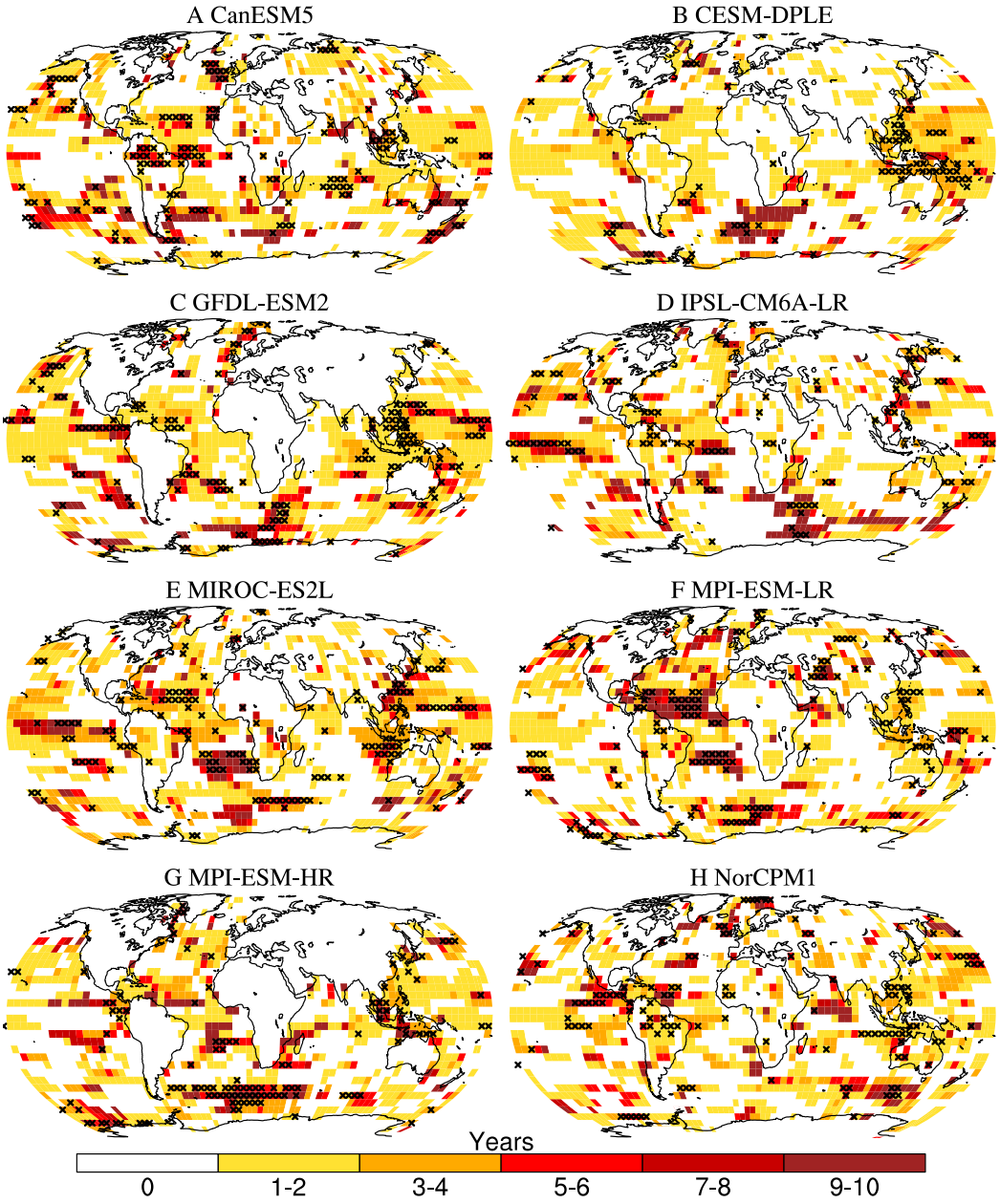

Figure 4. Predictability horizon of the detrended CO_2 flux into the ocean and the land, represented by the lead years with improved predictive skill due to initialization, i.e., when correlations in the initialized simulations are larger than 0 and also larger than those in the uninitialized simulations. Skill is quantified with anomaly correlation coefficient for the period 1982-2013. Predictive skill of the air-sea CO_2 flux gained due to initialization is assessed against SOM-FFN, whereas for the air-land CO_2 flux it is assessed against GCB. Crosses show significance at 95% level for the first 2 years. Note that GFDL-ESM2 and MIROC-ES2L hindcasts start earliest from year 1980, so from lead year 4 the time period is shorter than 1982-2013.