

Supporting Information for
“Predictability Horizons in the Global Carbon Cycle Inferred from a
Perfect-Model Framework”

Aaron Spring^{1,2}, Tatiana Ilyina¹

¹Max Planck Institute for Meteorology, Hamburg, Germany

²International Max Planck Research School of Earth System Modelling, IMPRS, Hamburg, Germany

Corresponding author: Aaron Spring, aaron.spring@mpimet.mpg.de

S0 Computation

The results in this paper were obtained using a number of different software packages. The command line tool known as Climate Data Operators (CDO) [Schulzweida, 2019] was used to aggregate output and perform routine calculations on those files (e.g., the calculation of temporal and spatial means). For more complex analysis and visualization, a Python distribution called Anaconda was used. A Python library called xarray was used for reading/writing netCDF files and data analysis. The xarray-wraper `climpred` was co-developed by Aaron Spring and Riley X. Brady and is publicly available at <https://climpred.readthedocs.io/>. In addition to Matplotlib (the default Python plotting library [Hunter, 2007]), Cartopy [Met Office, 2010] was used to generate the figures.

To facilitate the reproducibility of the results presented here, please find scripts, raw input and intermediate results files archived at <http://hdl.handle.net/21.11116/0000-0004-8276-4>.

This computation section is inspired by Irving [2015] to foster reproducibility in geosciences.

S1 Predictability horizon at atmospheric CO₂ measurement stations

	Lon	Lat	PH RMSE	PH ACC
Alert	82	-62	0	3
Point Barrow	71	-156	0	3
La Jolla	32	-117	4	3
Mauna Loa	19	-155	4	3
Christmas Island	2	-157	3	6
American Samoa	-14	-170	3	3
Kermadec Islands	-29	-177	3	3
Baring Head	-41	174	3	3
South Pole	-89	-24	3	3

Table S1. RMSE- and ACC-based predictability horizon atmospheric CO₂ mixing ratio of location of long-standing atmospheric CO₂ measurement stations in years. Station locations are taken from <https://cdiac.ess-dive.lbl.gov/trends/co2/sio-keel.html>.

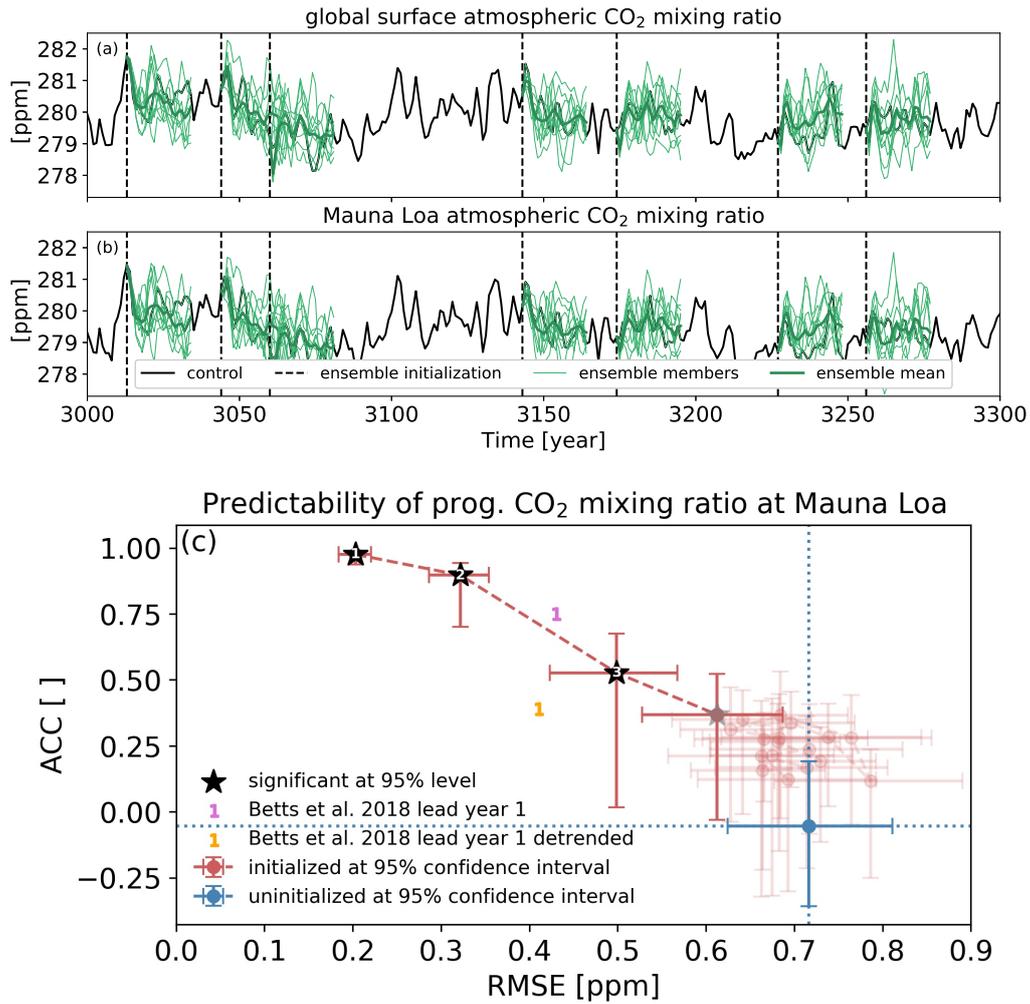


Figure S1. (a, b) Evolution of the annual mixing ratio of atmospheric CO₂: (a) globally-averaged and at Mauna Loa. The ensemble mean (dark green) is taken from individual ensemble members (green), which are branched off a pre-industrial control run (black) at different ensemble initialization years (7 out of 12 shown in dotted gray). (c) Comparison of the mean potential prediction skill of the initialized ensemble (red) versus random uninitialized ensembles (blue) of prognostic atmospheric CO₂ at Mauna Loa, Hawaii with Anomaly correlation coefficient (ACC) on the y-axis and root-mean-square-error (RMSE) on the x-axis for lead years represented as dots. Errorbars show 95% confidence intervals based on bootstrapping with replacement (N=5000). The last lead year with a bootstrapped p-value (which represents that uninitialized ensembles beat initialized ensembles) lower than 5% marks the predictability horizon. Black stars with white integer denote significant lead years in ACC and RMSE, gray stars if only one metric is significant and lead years non-significant in both metrics are blurred. For comparison with *Betts et al.* [2016, 2018], ones show predictability skill of the statistical model for lead-year one for the transient forecast (blue) and the detrended forecast (orange). Non-significant lead years are blurred.

S2 RMSE-based predictability skill surface CO₂ flux

For completeness and comparison, we calculate predictability skill maps of RMSE (for comparison to manuscript Figs. 2 and 4 in ACC) for atmospheric CO₂ mixing ratio and surface CO₂ flux. Predictability skill has similar patterns as ACC, but initialized forecasts perform better than uninitialized forecasts for fewer lead years. Furthermore, there is no emergence in ACC-based predictability skill of surface atmospheric CO₂.

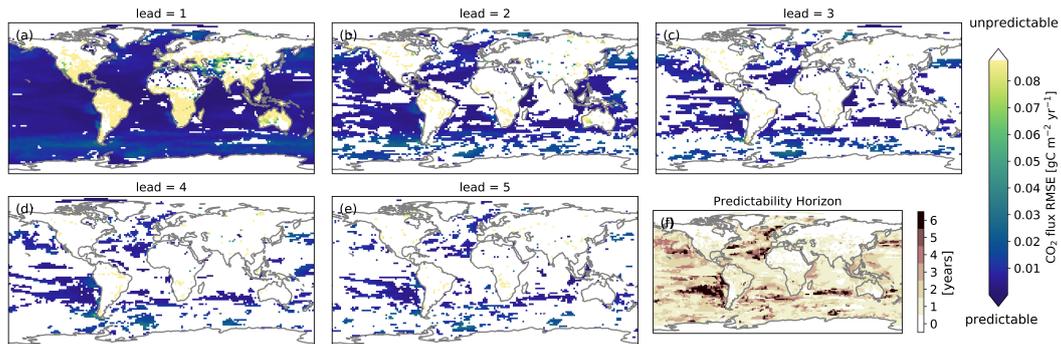


Figure S2. Spatial distribution of RMSE-based surface CO₂ flux predictability: (a-e) predictability skill over the first five lead years. White areas indicate unpredictable areas where the uninitialized predictability skill exceeds the initialized skill at 5% probability based on bootstrapping with replacement (N=1000). (f) The predictability horizon marks the last significant lead year.

S3 Statistics

All skill metric results are based on means over all initialisations and every member is used in turns as verification. Furthermore, we exclude the member being used as verification when calculating the ensemble mean forecast. This approach relies on a super-vector comprised of all available initializations and members and is also used in [Bushuk *et al.*, 2018] to calculate ACC over non-continuous initialization years. Calculating first a distance metric over members or initializations first and then average over the remaining makes only little difference in for perfect-models.

MSE-based predictability has been mostly used in the past to assess potential predictability [Griffies and Bryan, 1997; Pohlmann *et al.*, 2004; Séférian *et al.*, 2018]. However, here for atmospheric CO₂ mixing ratio, ACC (also used in [Bushuk *et al.*, 2018]) predictability comes closer to the raised expectation when predicting something and is therefore primarily used in this study when assessing atmospheric CO₂ predictability.

S3.1 ACC

The anomaly correlation coefficient skill score (ACC) is defined as:

$$ACC(x) = \frac{cov(x, \hat{x})}{\sqrt{var(x), var(\hat{x})}} = \frac{\frac{1}{NM} \sum_{i,j=1}^{N,M} (x_{i,j} - \overline{x_{i,j}})(\hat{x}_j - \overline{\hat{x}_j})}{\sqrt{\frac{\sum_{i,j=1}^{N,M} (x_{i,j} - \overline{x_{i,j}})^2}{MN}} \cdot \sqrt{\frac{\sum_{j=1}^M (\hat{x}_j - \overline{\hat{x}_j})^2}{M}}}$$

where x_i and represent the forecast and reference for each of the N lead years i and \hat{x} denotes the member mean, assess the synchronous evolution of the forecast and the reference [Jolliffe and Stephenson, 2011].

S3.2 RMSE

The root-mean-square-error (RMSE), defined as:

$$RMSE(x) = \sqrt{\frac{\sum_{i,j=1}^{N,M} (x_{i,j} - \hat{x}_j)^2}{NM}},$$

measures the second-order distance between forecast and reference [Jolliffe and Stephenson, 2011].

S3.3 Comparison of predictability horizon definitions

The differences in predictability horizon between *Séférian et al.* [2018] and our study arise from different interpretations of what defines the predictability horizon. While *Séférian et al.* [2018] define the limit at the saturation level of later lead years close to the magnitude of the standard deviation, we define the predictability horizon above a threshold value at which initialized forecasts cease to perform better than random, uninitialized forecasts as defined by [Buizza and Leutbecher, 2015]. The break-point fit of *Séférian et al.* [2018] resembles a 50% bootstrapping and results therefore in by design longer predictability horizons compared to this study [Fig. S3].

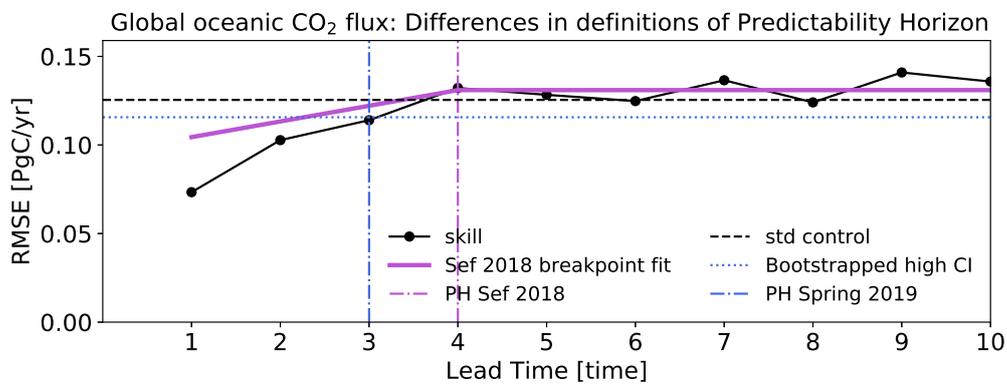


Figure S3. Evolution of global CO₂ flux RMSE over lead time. Definitions of predictability horizon used in *Séférian et al.* [2018] (orchid) leads to systematically higher predictability horizon than when using the methodology from our study (blue).

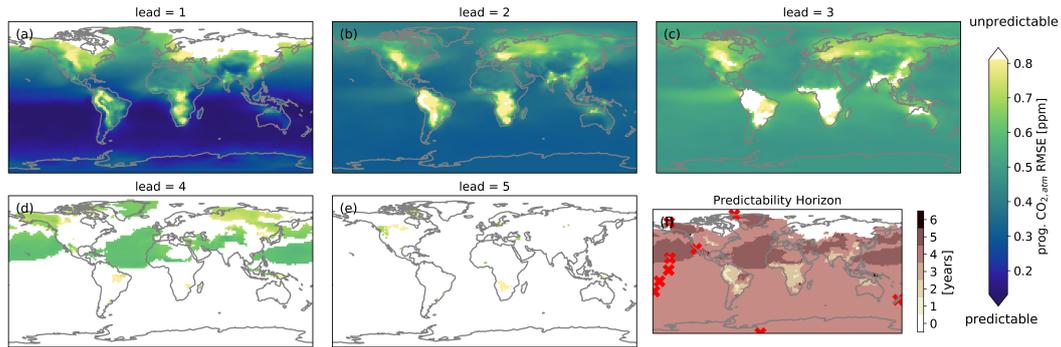
S4 RMSE-based predictability skill surface atmospheric CO₂

Figure S4. Spatial distribution of RMSE-based atmospheric surface CO₂ predictability: (a-e) predictability skill over the first five lead years. White areas indicate unpredictable areas where the uninitialized predictability skill exceeds the initialized skill at 5% probability based on bootstrapping with replacement (N=1000). (f) The predictability horizon marks the last significant lead year. Red crosses show location of long-standing atmospheric CO₂ mixing ratio measurement stations [Keeling *et al.*, 2005].

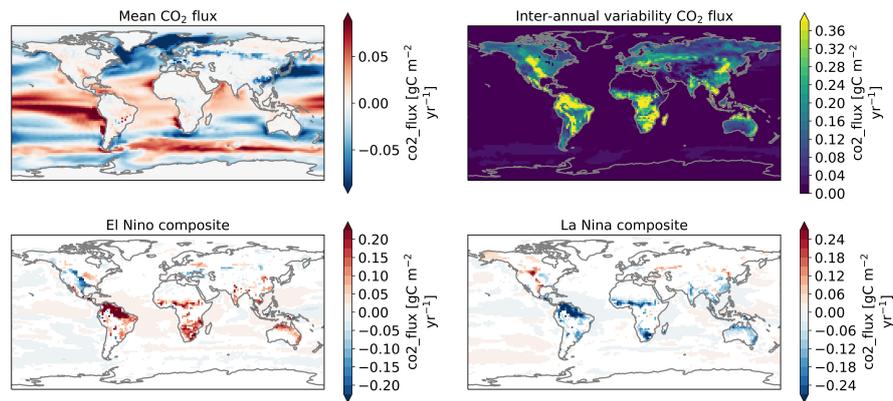
S5 Influence of ENSO on variability

Figure S5. Spatial distribution of surface CO₂ flux: (a) annual mean, (b) inter-annual variability determined as standard deviation (c, d) the composite of the positive/negative ENSO 3.4 index states. Colored areas indicate that the composite is different from the neutral ENSO 3.4 state assessed with a t-test at 95% significance.

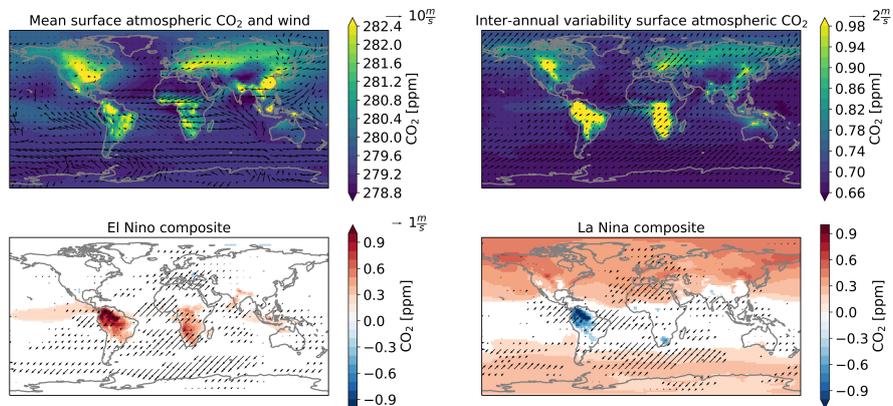


Figure S6. Spatial distribution of atmospheric CO₂ mixing ratio overlain with 10m wind indicated as arrows: (a) annual mean, (b) inter-annual variability determined as standard deviation (c, d) the composite of the positive/negative ENSO 3.4 index states. Colored areas indicate that the composite is different from the neutral ENSO 3.4 state assessed with a t-test at 95% significance.

S6 Comparison of modelled inter-annual variability with observations

Transferring the time-scales of perfect-model predictability to the real Earth system assumes that the model can reproduce observed variability. Here, we compare internal variability of MPI-ESM from the 300-year pre-industrial control simulation with detrended observation-based products under the transient climate. Due to the shorter time period of 34 to 62 years, the observation-based products are likely to show less variations compared to the modelled inter-annual variability.

Furthermore, observations are subject to the climate change trend whereas our control simulation is stable. To compare the variability of the model with observations-based products, we need to remove the trend from the data products. Here, the choice of the detrending polynomial introduces an additional uncertainty. Also changing from a linear to a 4th order polynomial trend removal is not a priori more correct and lead to varying results.

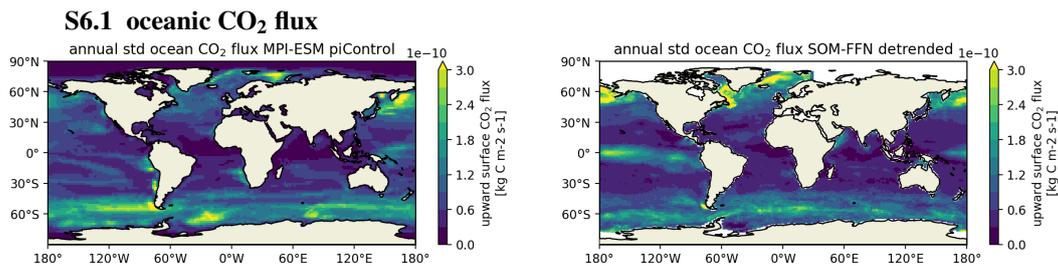


Figure S7. Spatial distribution of inter-annual variability computed as standard deviation of annual mean oceanic CO₂ flux: (a) MPI-ESM pi-esmControl, (b) linear detrended SOM-FFN (1982-2015) [Landschützer *et al.*, 2016].

Hotspots of oceanic CO₂ flux variability in SOM-FFN [Landschützer *et al.*, 2016] in the Southern Ocean and North Pacific are captured by MPI-ESM. MPI-ESM underestimates oceanic CO₂ flux variability in the equatorial Pacific with respect to the detrended SOM-FFN. This feature is less pronounced after 4th-order detrending. Furthermore, SOM-FFN is just one of several SOCOM data products [Rödenbeck *et al.*, 2015] which fill the various gaps of the gridded measurement data. Given the existing uncertainty in data filling methods, a precise estimation of variability is not conclusive. Furthermore, Landschützer *et al.* show that the length of the observational records is insufficient to fully capture natural variability signals.

S6.2 terrestrial CO₂ flux

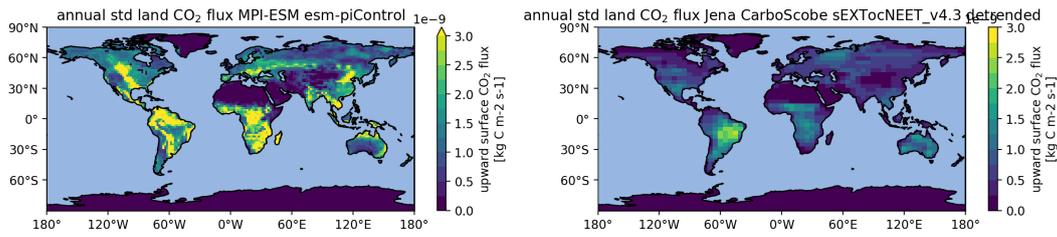


Figure S8. Spatial distribution of inter-annual variability computed as standard deviation of annual mean atmospheric CO₂ mixing ratio: (a) MPI-ESM pi-esmControl, (b) second-order detrended atmospheric CO₂ inversion Jena CarboScope sEXTocNEET_v4.3 (1957-2018) [Rödenbeck *et al.*, 2018].

Terrestrial CO₂ flux variability is highly unconstrained, as there is currently no direct observation-based terrestrial CO₂ flux product available. However, for comparison here, we use the observations-based atm. CO₂ inversion [Rödenbeck *et al.*, 2018] as an estimate for spatio-temporal gridded terrestrial CO₂ flux. While MPI-ESM generally overestimates the magnitudes in terrestrial CO₂ flux variability, the origins of high variability in the tropics and mid-latitudes are well captured.

S6.3 surface atm. CO₂

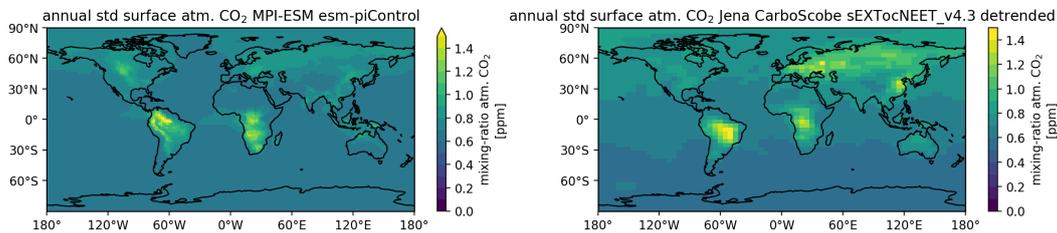


Figure S9. Spatial distribution of inter-annual variability computed as standard deviation of annual mean atmospheric CO₂ mixing ratio: (a) MPI-ESM pi-esmControl, (b) second-order detrended atmospheric CO₂ inversion Jena CarboScope sEXTocNEET_v4.3 (1957-2017) [Rödenbeck *et al.*, 2018].

The effect of the internal variability of both oceanic and terrestrial CO₂ flux on atmospheric CO₂ is well captured by MPI-ESM. The higher variability over the northern Hemisphere based on observations may be explained by the anthropogenic emissions not present in MPI-ESM esm-piControl. The other hotspots of variability from the CO₂ inversion also appear in the model.

S7 Drivers of terrestrial CO₂ flux predictability

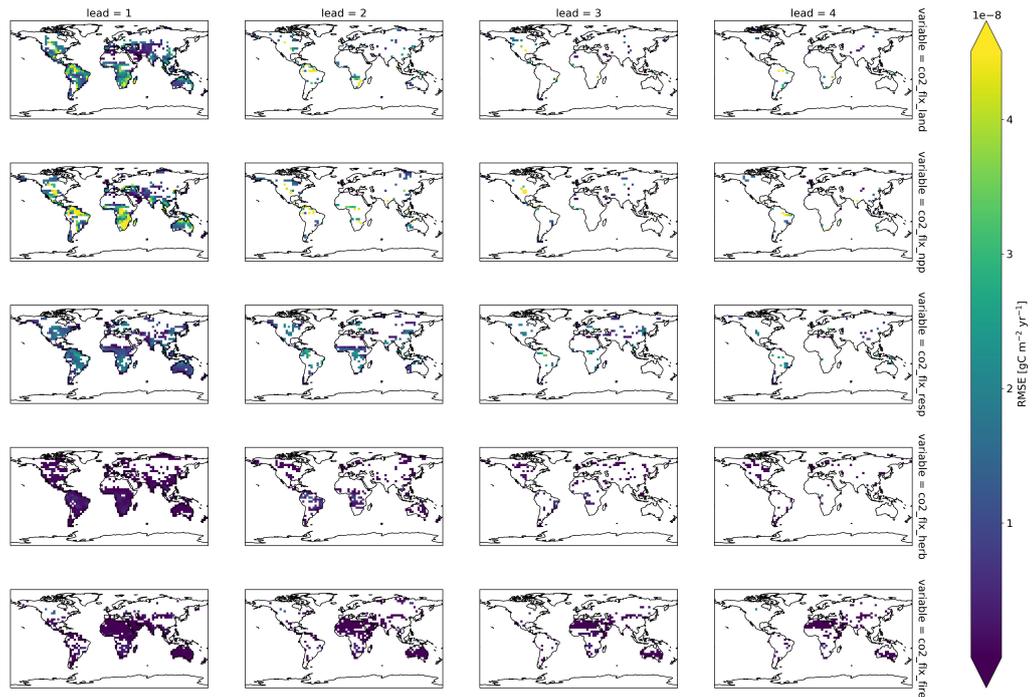


Figure S10. Spatial distribution of the contributors to RMSE-based terrestrial surface CO₂ predictability: the vertical row shows skill at different lead years, the horizontal rows show different variables: total CO₂ flux, CO₂ flux due to net primary production, CO₂ flux due to heterotrophic respiration, CO₂ flux due to herbivory and CO₂ flux due to fires. White areas indicate unpredictable areas where the uninitialized predictability skill exceeds the initialized skill at 5% probability based on bootstrapping with replacement (N=1000).

Figs. S10 and S11 show the dominance of net primary production for terrestrial CO₂ flux in RMSE-based predictability in the first lead years. The metric RMSE is used here to show that large magnitude in RMSE, which shows the large spread between ensemble members. However, the predictability horizon of heterotrophic respiration, the second strongest contributor to terrestrial CO₂ flux predictability, can extend until lead year 3 in the tropics.

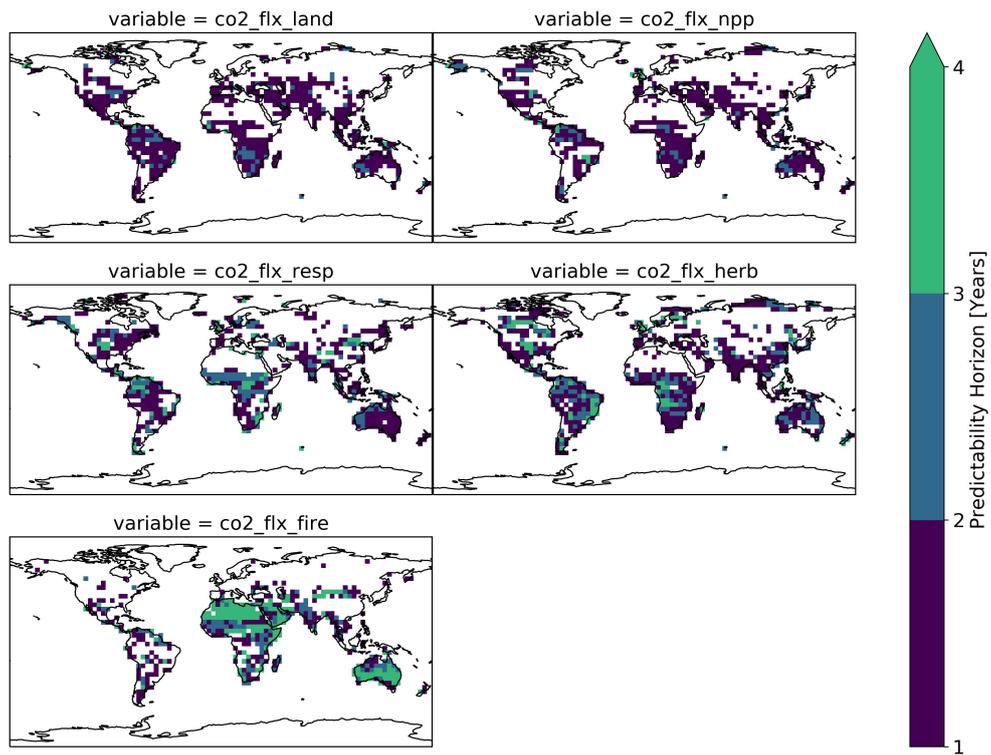


Figure S11. Spatial distribution of the predictability horizon of contributors to RMSE-based terrestrial surface CO₂ predictability: total CO₂ flux, CO₂ flux due to net primary production, CO₂ flux due to heterotrophic respiration, CO₂ flux due to herbivory and CO₂ flux due to fires.

S8 Re-emergence

The ACC-based predictability of all ENSO indices seems to have a small rebound from lead year 2 to lead year 3. We attribute these slightly constrained ENSO states to the weak atm. CO₂ re-emergence pattern in lead year 4. Note that until lead year 3 tropical CO₂ flux is still quite high (above .5) [Figs. 2, S12], whereas in the other areas over land predictability dropped earlier. Also we see that the larger share of positive ENSO initial conditions dominates the predictability skill. However, due to the meridional and zonal transport of atm. CO₂ by wind and the time lag between ENSO and the terrestrial CO₂ flux response a clean attribution of ENSO to atm. CO₂ is quite challenging.

Figure S13 shows the temporal evolution of the Nino 3.4 index and equatorial (35S-35N) terrestrial CO₂ flux of the ensemble members after a few initializations. Here, we can see a weak re-emergence two to four years after initialization. The distance between members in each initialization is the RMSE metric used. Fig. S14 also shows the re-emergence in RMSE skill in Nino 3.4. This re-emergence in Nino 3.4 is not totally surprising as the autocorrelation based on the control simulation shows a oscillatory behaviour with negative autocorrelations for lead two and three in Nino 3.4 [Fig. S15]. The equatorial (35S-35N) terrestrial CO₂ flux responds to El Nino [*Jones et al.*, 2001; *Zeng et al.*, 2005, 2008].

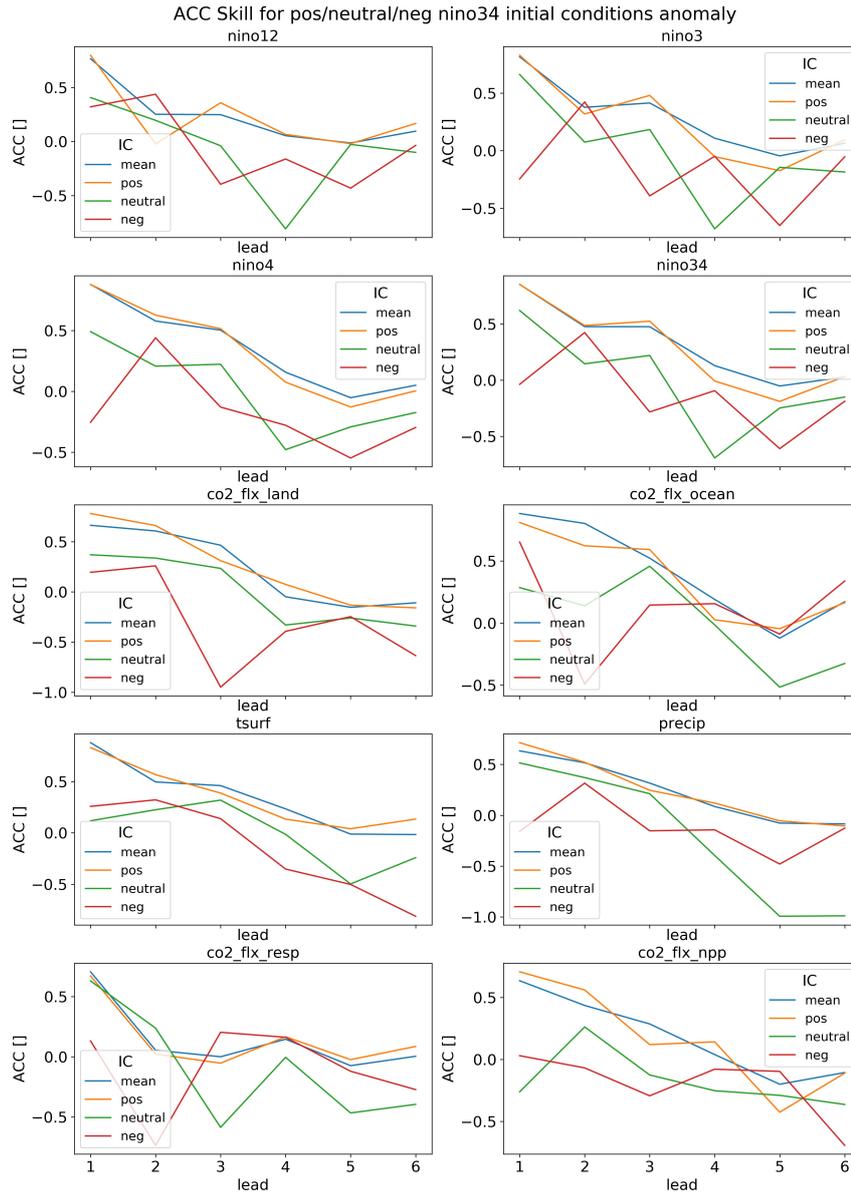


Figure S12. Area-average ACC-based predictability of positive/neutral/negative ENSO 3.4 initial states for different ENSO indices and impacted variables in the tropics 30°S-30°N: (from top left to bottom right) Nino 1.2, Nino 3, Nino4, Nino 3.4, Nino, oceanic CO₂ flux, terrestrial CO₂ flux, surface air temperature, precipitation, terrestrial CO₂ flux due to heterotrophic respiration, terrestrial CO₂ flux due to net primary production.

Figure S13. Evolution of the annual Nino 3.4 index as in Fig. S1(a).

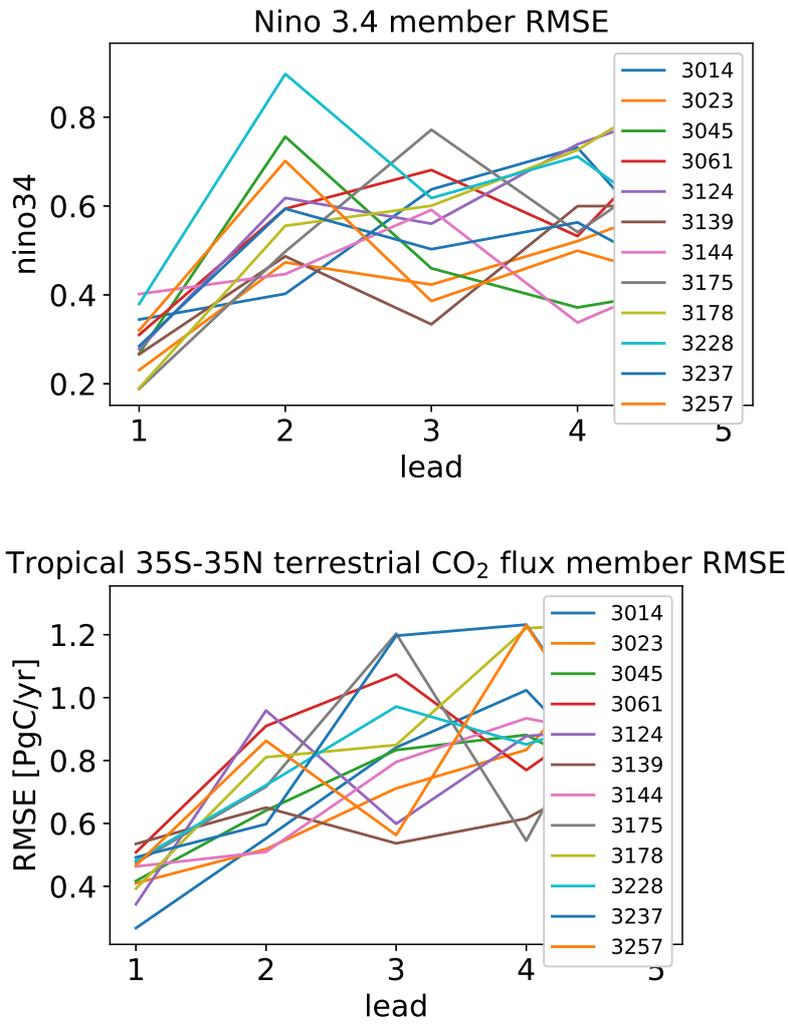


Figure S14. Inter-member distance for all initializations of the annual (a) Nino 3.4 index and (b) 35S-35N terrestrial CO₂ flux.

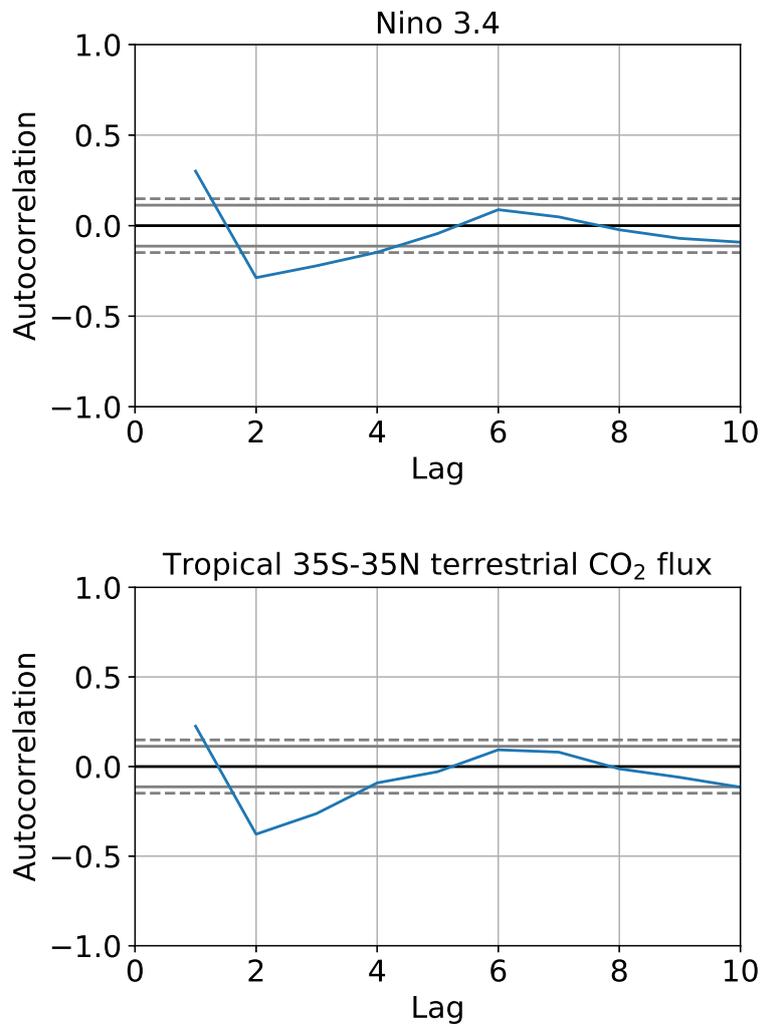


Figure S15. Autocorrelation function of the annual (a) Nino 3.4 index and (b) 35S-35N terrestrial CO₂ flux.

References

- Betts, R. A., C. D. Jones, J. R. Knight, R. F. Keeling, and J. J. Kennedy (2016), El Niño and a record CO₂ rise, *Nature Climate Change*, *6*(9), 806–810, doi:10/gddfcr, 00000.
- Betts, R. A., Jones Chris D., Knight Jeff. R., Keeling Ralph. F., Kennedy John. J., Wiltshire Andrew J., Andrew Robbie M., and Aragão Luiz E. O. C. (2018), A successful prediction of the record CO₂ rise associated with the 2015/2016 El Niño, *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1760), 20170,301, doi: 10/gfthxx.
- Buizza, R., and M. Leutbecher (2015), The forecast skill horizon: The Forecast Skill Horizon, *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 3366–3382, doi: 10/gfx52f.
- Bushuk, M., R. Msadek, M. Winton, G. Vecchi, X. Yang, A. Rosati, and R. Gudgel (2018), Regional Arctic sea–ice prediction: Potential versus operational seasonal forecast skill, *Climate Dynamics*, doi:10/gd7hfq.
- Griffies, S. M., and K. Bryan (1997), A predictability study of simulated North Atlantic multidecadal variability, *Climate Dynamics*, *13*(7-8), 459–487, doi:10/ch4kc4.
- Hunter, J. D. (2007), Matplotlib: A 2D Graphics Environment, *Computing in Science Engineering*, *9*(3), 90–95, doi:10/drbjhg.
- Irving, D. (2015), A Minimum Standard for Publishing Computational Results in the Weather and Climate Sciences, *Bulletin of the American Meteorological Society*, *97*(7), 1149–1158, doi:10/gf4wzh.
- Jolliffe, I. T., and D. B. Stephenson (2011), *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, John Wiley & Sons, Ltd, Chichester, UK, doi: 10.1002/9781119960003.
- Jones, C. D., M. Collins, P. M. Cox, and S. A. Spall (2001), The Carbon Cycle Response to ENSO: A Coupled Climate–Carbon Cycle Model Study, *Journal of Climate*, *14*(21), 4113–4129, doi:10/bwrw6c, 00142.
- Keeling, C. D., S. C. Piper, R. B. Bacastow, M. Wahlen, T. P. Whorf, M. Heimann, and H. A. Meijer (2005), Atmospheric CO₂ and ¹³CO₂ Exchange with the Terrestrial Biosphere and Oceans from 1978 to 2000: Observations and Carbon Cycle Implications, in *A History of Atmospheric CO₂ and Its Effects on Plants, Animals, and Ecosystems*, edited by I. Baldwin, M. Caldwell, G. Heldmaier, R. B. Jackson, O. Lange, H. Mooney, E.-D. Schulze, U. Sommer, J. R. Ehleringer, M. Denise Dearing, and T. E. Cerling, Ecological

- Studies, pp. 83–113, Springer New York, New York, NY, doi:10.1007/0-387-27048-55.
- Landschützer, P., Gruber Nicolas, and Bakker Dorothee C. E. (2016), Decadal variations and trends of the global ocean carbon sink, *Global Biogeochemical Cycles*, 30(10), 1396–1417, doi:10/f9csrg.
- Landschützer, P., T. Ilyina, and N. S. Lovenduski (2019), Detecting Regional Modes of Variability in Observation-Based Surface Ocean pCO₂, *Geophysical Research Letters*, 46(5), 2670–2679, doi:10/gfxkj6.
- Met Office (2010), Cartopy: A cartographic python library with a Matplotlib interface.
- Pohlmann, H., M. Botzet, M. Latif, A. Roesch, M. Wild, and P. Tschuck (2004), Estimating the Decadal Predictability of a Coupled AOGCM, *Journal of Climate*, 17(22), 4463–4472, doi:10/d2qf62.
- Rödenbeck, C., D. C. E. Bakker, N. Gruber, Y. Iida, A. R. Jacobson, S. Jones, P. Landschützer, N. Metzl, S. Nakaoka, A. Olsen, G.-H. Park, P. Peylin, K. B. Rodgers, T. P. Sasse, U. Schuster, J. D. Shutler, V. Valsala, R. Wanninkhof, and J. Zeng (2015), Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean pCO₂ Mapping intercomparison (SOCOM), *12(23)*, 7251–7278, doi:10/gb9bnx, 00000.
- Rödenbeck, C., Z. S., K. R., and H. M. (2018), History of El Niño impacts on the global carbon cycle 1957–2017: A quantification from atmospheric CO₂ data, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1760), 20170,303, doi:10/gf4j8j.
- Schulzweida, U. (2019), CDO: Climate Data Operators.
- Séférián, R., S. Berthet, and M. Chevallier (2018), Assessing the Decadal Predictability of Land and Ocean Carbon Uptake, *Geophysical Research Letters*, doi:10/gdb424.
- Zeng, N., A. Mariotti, and P. Wetzel (2005), Terrestrial mechanisms of interannual CO₂ variability, *Global Biogeochemical Cycles*, 19(1), doi:10/bjr69v.
- Zeng, N., J.-H. Yoon, A. Vintzileos, G. J. Collatz, E. Kalnay, A. Mariotti, A. Kumar, A. Busalacchi, and S. Lord (2008), Dynamical prediction of terrestrial ecosystems and the global carbon cycle: A 25-year hindcast experiment, *Global Biogeochemical Cycles*, 22(4), doi:10/cf445j.