Response to Referees

We are extremely grateful to the two anonymous reviewers for providing such thorough reviews of our manuscript and suggesting changes that have substantially improved the science. Below, we include the original reviewer comments in black, with our response below in red.

We have made the following substantial changes to the manuscript:

1) We now more thoroughly assess the statistically significance of predictability from the initialized forecasts relative to that of the uninitialized and persistence forecasts using a z-test statistic for the 95% confidence interval. All figures and tables now report the results of this significance test.

2) We include a supplemental figure that illustrates air-sea CO_2 flux predictability for lead years 2 through 10, as requested by both reviewers.

3) We have clarified the difference between the potential predictability and the skill relative to observational metrics.

4) We now adhere to the DCPP protocols for calculating the predictability from the uninitialized correlation, as suggested by both reviewers.

Anonymous Referee #1

Received and published: 30 October 2018

In the manuscript 'Predicting near-term changes in ocean carbon uptake' Lovenduski and coauthors assess the predictability of the ocean carbon sink over the last decades using CESM-DPLE, a new large ensemble decadal prediction platform developed at NCAR. By realizing 40 decade-long ensemble members each year from 1954 to 2015, the authors estimate that the global ocean carbon sink is predictable up to 7 years in advance which is in the line of recent published estimates. The authors also investigate the drivers of this predictability and explain that it arises from the predictability of carbon-related fields (DIC and Alkalinity, setting _pCO2 and hence carbon fluxes).

The paper is well-written and the analyses are sounds. I much appreciate this work which explores the predictive capability of the current generation of ESM. Nevertheless, I think this paper needs some clarification that have to be addressed first, and which prevent me of accepting this paper in its present form.

General Major Comments:

1- My first comment concerns the assessment of the initialization procedure which is the core of decadal prediction. Here the other briefly describe this step but do not provide a complete evaluation. ESD paper are not limited by the length. Thus I recommend to include a new section to discuss the initialization procedure because: (1) full fields restoring implies a model drift; this is an interesting to document how it impacts the biogeochemical fields, especially the carbon related fields and nutrients fields that could generates non-linearity in the drift (2) your initialization strategy fails at capturing the recent variability in the ocean carbon sink as suggested by SOM-FFN dataproduct. It could be interesting to show other variables such as SST or the AMOC to support the fact that your initialization procedure is doing a good job.

We have added a more complete description/evaluation of the initialization procedure to Section 2 (see also our response to your Page 3, line 23 comment):

CESM-DPLE initializes an ensemble of 40 simulations each year using round-off level (order 10^{-14}) perturbations in the initial air temperature field (Figure 1). Previous work indicates that this small perturbation in the initial conditions generates a wide divergence in global mean surface temperatures across the ensemble members within about 30 days (V. Yettella, pers. comm., 2018), and the average divergence in globally-integrated, annual-mean forecast CO_2 flux across the ensemble members (0.53 Pg C yr^{-1}) is an order of magnitude greater than that generated by the preindustrial control simulation of CESM (0.09 Pg C yr^{-1}; Lovenduski et al.,2015b).

Indeed, the full-field initialization procedure generates a drift in the forecasts. As stated in the original manuscript, we correct this drift by transforming to anomalies from a drifting climatology. The drifting climatology is statistically robust due to the large number (n=40) of ensemble members in each forecast and the large number (n=62) of start years. In the revised manuscript, we now note that the drifting climatology need not be linear, as its temporal behavior is determined by the variable of interest:

Note that this method does not assume that the drift is linear, and disregards potential dependence of the drift on the external forcing.

We include for the reviewer the companion to Figure 1 that shows the non-drift corrected forecast (see Figure R1 below). It is nearly identical to the original Figure 1, indicating that the drift in CO_2 flux is not particularly large, and thus further description of the drift is unnecessary.

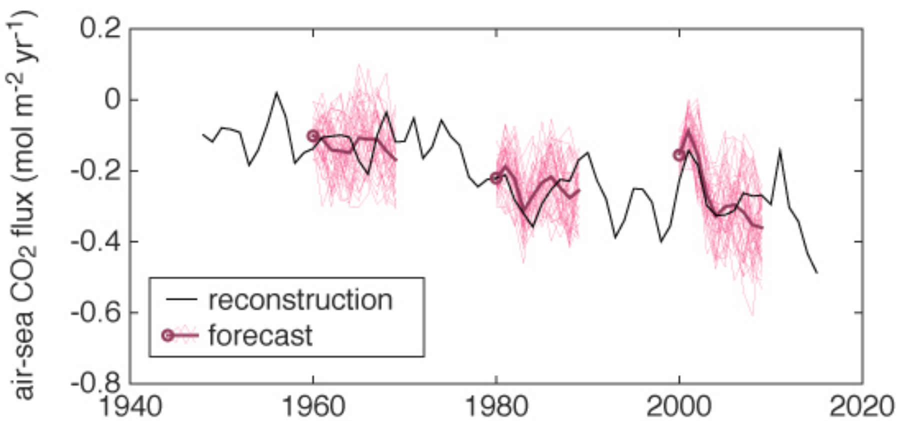

Figure R1. Annual-mean air-sea CO_2 flux (mol m^{-2} yr^{-1}) in the South Pacific subtropical permanently stratified biome for the (black) model reconstruction and (pink) CESM-DPLE decadal forecasts initiated in 1960, 1980, and 2000 (other forecasts omitted for visual clarity). Thick magenta line represents the ensemble-mean forecast; open circles show the ensemble mean in forecast year 1. Positive fluxes denote ocean outgassing. Forecasts have <u>not</u> been drift corrected.

We note that the reconstruction fails to capture the recent variability in air-sea CO_2 flux (compare black and blue lines in Figures 3a and 3b), so the initialization procedure for the forecast is not to blame for the poor skill.

We include for the reviewer below Figure R2 from Yeager et al. (2018) which illustrates the forecast skill of the CESM-DPLE SST over a range of lead years. This figure demonstrates that the initialization improves the skill of the SST forecast.

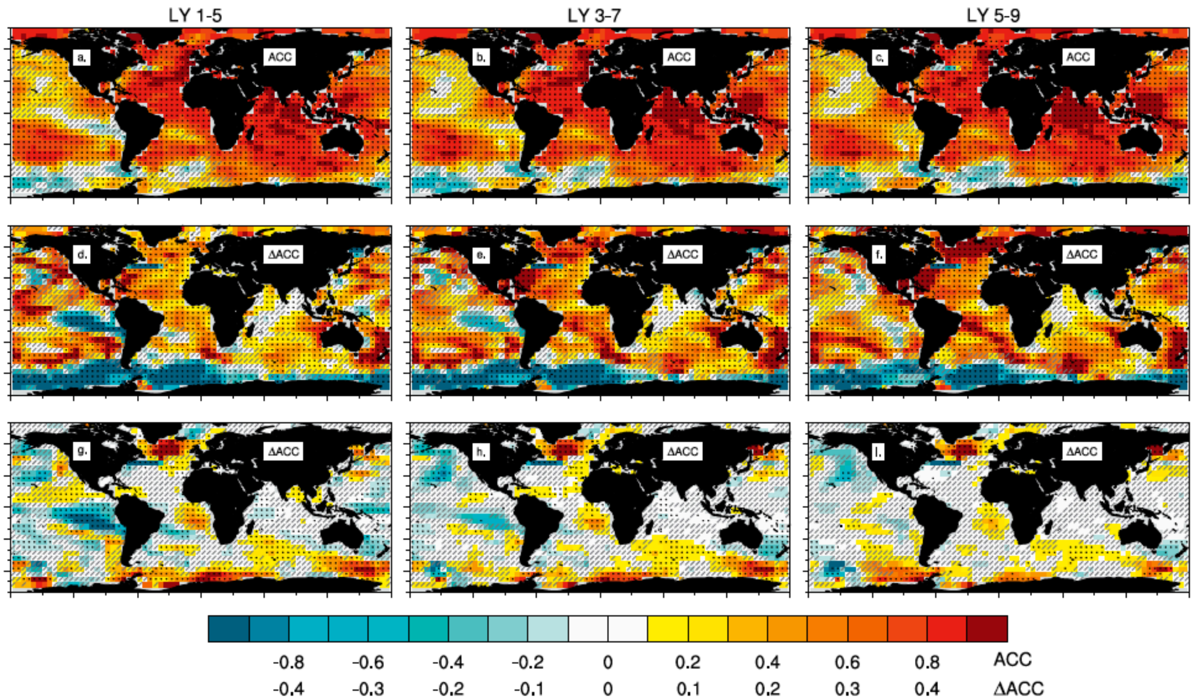

Figure R2. (a) – (c) Anomaly correlation coefficient of annual SST from CESM-DPLE relative to ERSSTv5 observations for lead times of 1-5, 3-7, and 5-9 years, respectively. Anomaly correlation coefficient skill score differences (d) – (f) between CESM-DPLE and persistence and (g) – (i) between CESM-DPLE and CESM-LE. From Yeager et al. (2018).

2- Further discussions is needed when discussing the drivers. It could be interesting to document if your model gives a longer predictability horizon for SST, DIC, Alk : : : than that of ocean carbon sink and compares this result to the persistence. At least for DIC and Alk which have a long-lasting memory it would be helpful to demonstrate that your model's predictability beats the persistence for those fields otherwise it might suggest that the predictability in ocean carbon sink is supported by the persistence of DIC and Alkalinity anomalies.

Indeed this would be interesting -- one could imagine performing a separate analysis to determine the predictability horizons of all relevant physical and biogeochemical variables, but we feel this is beyond the scope of our study. Here, we have demonstrated that the air-sea CO_2 flux predictability originates from pCO_2 and is linked to predictability in DIC and Alk – thus, if one can accurately predict DIC and Alk, one can then predict CO_2 flux variability. Perhaps a future publication could explore this topic in greater detail (see response to point 3 below). Manuscript unchanged in response to comment.

3- Finally, further discussions are needed to discuss how this work compared with previous works based on different prediction system such as Li et al. (2016), Séférian et al., (2018) [using ESMs] but also all the recent studies focusing on ocean physical variables (e.g., Kim et al., 2012)

We agree that a discussion of the various prediction systems and their findings for CO_2 flux is important, and we report the conclusions of these first two studies in our Conclusions section. Two of us (Lovenduski and Yeager) are co-authors on a review paper being drafted on this very topic (lead author is Tatiana Ilyina, journal is Current Climate Change Reports) that will cover the differences between the decadal prediction systems at the various modeling centers and the main findings for ocean carbon uptake. Perhaps this review paper could also include an analysis of the predictability horizons (as suggested in the previous comment). Alas, we await the findings of this review. Manuscript unchanged in response to comment.

Specific comments:
Page 1
Title: I suggest to modify the text because "near-term changes" could also implies anthropogenic carbon sink. This latter is rather well captured by ESM without initialization. I suggest to use "multi-annual variations" instead.

Thank you for the suggestion. We have modified the title to read "Predicting near-term variability in ocean carbon uptake".

L2 : please define somewhere what your mean by "near-term"

Good catch. We have modified the first sentence of the abstract to read:

Interannual variations in air-sea fluxes of carbon dioxide (CO_2) impact the global carbon cycle and climate system, and previous studies suggest that these variations may be predictable in the near-term (from a year to a decade in advance).

L4: of an Earth system model

Manuscript changed as suggested.

L4: initialized forecast=please explain the initialization somewhere in the abstract and avoid the terminology initialized forecast because of forecast is generally initialized

We have excised the word "initialized" from the abstract. The abstract now reads:

Interannual variations in air-sea fluxes of carbon dioxide (CO_2) impact the global carbon cycle and climate system, and previous studies suggest that these variations may be predictable in the near-term (from a year to a decade in advance). Here, we quantify and understand the sources of near-term predictability and predictive skill in air -sea CO_2 flux on global and regional scales by analyzing output from a novel set of retrospective decadal forecasts of an Earth system model. These forecasts exhibit the potential to predict year-to-year variations in the globally-integrated air-sea CO_2 flux several years in advance, as indicated by the high correlation of the forecasts with a model reconstruction of past CO_2 flux evolution. This potential predictability exceeds that obtained solely from foreknowledge of variations in *external forcing or a simple persistence forecast, with the longest-lasting forecast enhancement in the subantarctic Southern Ocean and the northern North Atlantic. Potential predictability in CO_2 flux variations are largely driven by predictability in the surface ocean partial pressure of CO_2, which itself is a function of predictability in surface ocean dissolved inorganic carbon and alkalinity. The potential predictability, however, is not realized as predictive skill, as indicated by the moderate to low correlation of the forecasts with an observationally-based CO_2 flux product. Nevertheless, our results suggest that year-to-year variations in ocean carbon uptake have the potential to be predicted well in advance, and establish a precedent for forecasting air-sea CO_2 flux in the near future.*

L9: moderate predictive skill= please explain the predictive skill measure

We now indicate the measure (correlation) in the abstract -- see above.

L11: initialized predictability= predictability also implies initialization. I suggest to remove 'initialized'

We have excised the word "initialized" from the abstract – see above.

L21: I suggest to include observational references instead (e;g., Landshutzer et al., 2016)

We have added a reference to Landschutzer et al. (2016) here.

Page 2
L7-15 Please expand the discussion by including the key limits of the decadal predictability that were highlight in the literature. For example, the first attempt from Keenlyside et al and Smith et al in 2008 which were challenged 10 years after by the recent observations. Besides, you could include a better rationale of the first attempt in the Earth system community such as Li et al. 2016, Séférian et al. 2018 for the ocean carbon sink and Séférian et al. 2014 for the net marine productivity. And the use of statistical model such as Betts et al. (2016, 2018) for atmospheric CO2.

We now make reference to these first attempts at decadal prediction (Smith et al., 2007; Keenlyside et al., 2008) in the introduction. In addition to referencing the Li and Seferian studies, we now also include a discussion of the limitations of decadal predictability in the Conclusions section:

While the ever-expanding field of decadal climate prediction has the potential to inform policy and management decisions moving forward, decadal forecasts come with several caveats. Initialization shock and drift of the coupled model system, inability of Earth system models to realistically simulate internal variability, uncertain future levels of radiative forcing, and imperfect observations are frequently cited as limitations to making accurate forecasts of the future (Meehl et al., 2014). In the case of ocean carbon, it is important to note that potential predictability in regional CO_2 flux may be driven by initialization of the physical (e.g., SST) or *biogeochemical (e.g., DIC) ocean state (Li et al., 2016), and that the spatiotemporal coverage of CO_2 flux observations is insufficient to fully address predictive skill in our forecast systems.*

L15 Please add Resplandy et al. (2014) which describes how far the decadal variability in ocean carbon fluxes differs between models

We have added a reference to Resplandy et al. (2015) here.

Page 3
L7-13: Please expand this paragraph, see my major comments

We have expanded the discussion of the initialization procedure in Section 2 (see also our response to your Page 3, line 23 comment):

CESM-DPLE initializes an ensemble of 40 simulations each year using round-off level (order 10^{-14}) perturbations in the initial air temperature field (Figure 1). Previous work indicates that this small perturbation in the initial conditions generates a wide divergence in global mean surface temperatures across the ensemble members within about 30 days (V. Yettella, pers. comm., 2018), and the average divergence in globally-integrated, annual-mean forecast CO_2 flux across the ensemble members (0.53 Pg C yr^{-1}) is an order of magnitude greater than that generated by the preindustrial control simulation of CESM (0.09 Pg C yr^{-1}; Lovenduski et al.,2015b).

L20: a reasonable job is not enough to determine if a model is fitted for purpose. Could you please provide further details such as the spatial correlation, the RMSE : : :

Thank you for catching this! We now include the spatial correlation coefficient for the reconstruction and the observational product (r = 0.79) in the manuscript text:

In Figure 2, we illustrate the comparison between observationally-based estimates of CO_2 flux (from the Landschutzer et al. (2016) pCO_2 product) and estimates produced by the reconstruction and coupled CESM-LE over 1982-2015. The model reconstruction does a reasonable job (r = 0.79) of representing observed spatial patterns (in both magnitude and direction) of the flux across most oceanic regions. The globally-integrated air-sea CO_2 flux over 1982-2015 from the observational product and model reconstruction are 1.41 and 1.80 Pg C yr^{-1}, respectively (directed into the ocean).

L23: 10-14 Kelvin ? This is a really small perturbation. Have assessed if this initial perturbation lead to populate the full range of model variability as diagnosed from the piControl ? It is important to tell the reader if you chose to populate this uncertainty or instead to stay close from the initial conditions

Indeed, this is a very small perturbation, but work by others suggests that this tiny perturbation generates a wide divergence in global mean surface temperature within approximately 30 days of the initialization (see Figure R3 below). Further, the globally-integrated air-sea CO_2 flux from the preindustrial control run of CESM has a standard deviation of 0.09 PgC yr-^{1}, while the average standard deviation (spread across 40 ensemble members) of the forecast air-sea CO_2 flux is 0.53 PgC yr^{-1}. We have added a description of these findings to Section 2 of the manuscript:

CESM-DPLE initializes an ensemble of 40 simulations each year using round-off level (order 10^{-14}) perturbations in the initial air temperature field (Figure 1). Previous work indicates that this small perturbation in the initial conditions generates a wide divergence in global mean surface temperatures across the ensemble members within about 30 days (V. Yettella, pers. comm., 2018), and the average divergence in globally-integrated, annual-mean forecast CO_2 flux across the ensemble members (0.53 Pg C yr^{-1}) is an order of magnitude greater than that generated by the preindustrial control simulation of CESM (0.09 Pg C yr^{-1}; Lovenduski et al.,2015b).

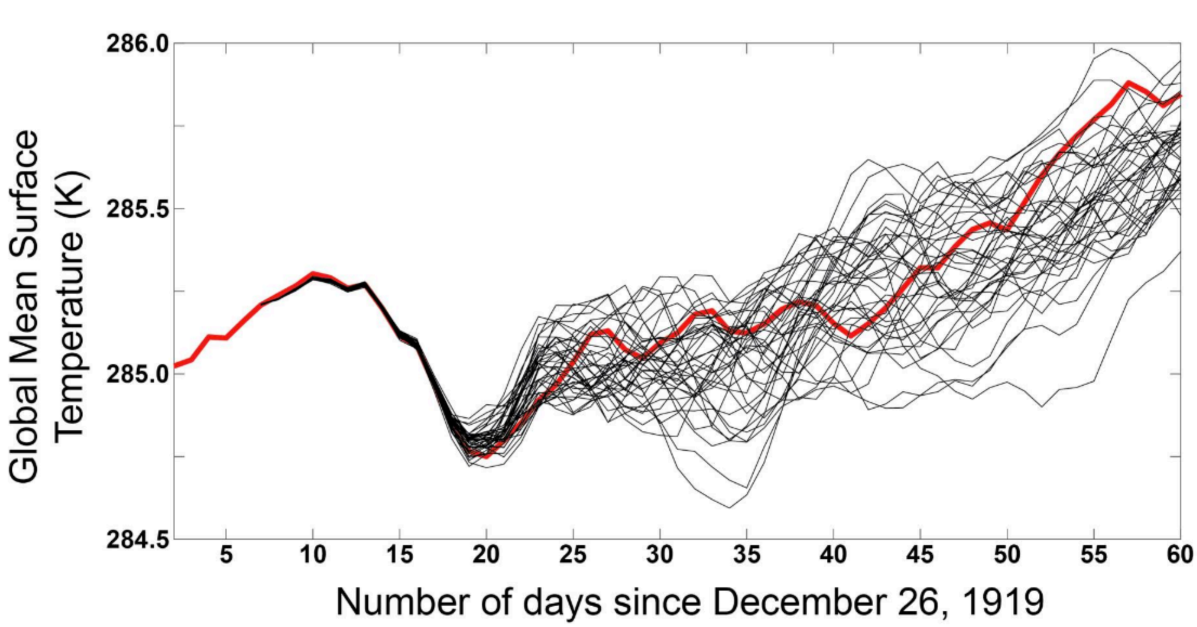

Figure R3. Rapid divergence in global mean surface temperature in the CESM-Large Ensemble, following initialization on January 1, 1920 using order 10^{-14} Kelvin perturbations in the initial air temperature field. Black lines correspond to individual ensemble members, and red line shows ensemble member 1. Figure from V. Yettella (personal communication, 2018).

L33: Please provide a figure of the ensemble with the drift as in Kim et al 2012 in addition to the de-drifted ensemble. You could add on panel on figure 1 to show that.

We include for the reviewer the companion to Figure 1 that shows the non-drift corrected forecast (see Figure R1 above). As it is nearly identical to the drift-corrected version, we have elected not to include it in the revised manuscript. Manuscript unchanged in response to comment.

Page 4

L9-11: I'm a bit puzzled here. Unless I misunderstood McKinley et al 2016 and Lovenduski et al. 2016 used a CMIP5-style CESM and hence CMIP5 forcings to performed their analyses. Here, CESM-DPLE is setup for CMIP6 and hence use CMIP6 forcings, right ? If it does, several external forcing have been revised between CMIP5 and CMIP6. This is the case for the volcanoes (which influence the predictability). Could you please comment this point ?

The CESM-DPLE was generated using the same code base, component model configurations, and historical and projected radiative forcings as in the CESM Large Ensemble (i.e., CMIP5). To avoid reader confusion, we have removed the reference to the DCPP CMIP6 protocols in Section 2:

CESM-DPLE consists of a set of initialized, fully-coupled integrations of CESM that adhere to the protocols for Component A of the Decadal Climate Prediction Project (Boer et al., 2016).

Page 5

L10-11 What happens if you consider the full observational time-series ? Besides could you please explain why the correlation of the uninitialized simulation slightly increases with lead time ?

We are not considering the observational time-series here, but rather the modeled time-series. In response to Reviewer 2, we have modified the uninitialized simulation correlations to align with the DCPP protocols – it is now the same correlation coefficient for all forecast lead times.

On Figure 3 8 and 9 please indicate the correlation limits (R*) on the graphs and indicates the level of confidence (and the number of degree of freedom used for the t-test) employed; this information is missing.

On Figures 3 and 9 (Figure 8 does not show correlation) we have added asterisks and circles to indicate the maximum lead time in the initialized forecast that is statistically separable from the uninitialized and persistence forecasts. We perform a Fisher's r to z transformation on the correlation coefficients and compare the resulting z test statistic to the value for the 95% confidence interval (1.96). We have also modified the text describing Figure 3 to read:

Figure 3 indicates that the initialized forecast exhibits higher predictability than the uninitialized forecast and the persistence forecast for a lead time of 10 years, though this initialized predictability is only statistically separable from the uninitialized and persistence forecasts for lead years 1-2 and 2, respectively (statistical separation determined via a Fisher's r to z transformation and a comparison of the resulting z test statistic to the value (1.96) for the 95% confidence interval).

L15: why are you talking about emissions and terrestrial CO2 uptake. Your model set-up employs CO2 concentration as prescribed in the forcing, correct ?

Thanks for catching this – you are correct. We have excised this text.

L25: linearly detrended forecast: have you done this detrended at grill-cell scale or have you applied the same detrending globally ?

The detrending is done on a grid cell by grid cell basis. We have modified the manuscript to better describe this procedure:

To capture the predictability on interannual timescales, we perform analysis on linearly detrended forecasts in each model grid cell.

L26: as you suggested that your modelling plateform is able to predict ocean carbon sink up to 7 years in advance it could be usefull to show what happens at lead time greater than 1 year. Could you please replace the figures showing lead-time (LT) 1 by LT 7 and/or moving LT1 Figures as supplemental data

We now include a supplemental figure (Figure S1) showing the predictability for CO_2 flux on forecast lead times 2-10 years.

Page 7
L2: please explain what are your forecast skill and what is the limit for a skillfull predictability at a given confidence

Forecast skill is defined in the previous paragraph as a measure of the ability of the forecast to reproduce the observational record. Since "some skill" is not quantitative, we provide the correlation coefficients in line with the text. Manuscript unchanged in response to comment.

L11: predictability or persistence, please see my major comments

The decision to aggregate the results on the biome scale is not dependent on the drivers of DIC and Alk predictability. Manuscript unchanged in response to comment.

L23: You could state that this biomes does not see an impacts of your initialization procedure. Maybe the sea-ice influence regions are not restored to the observations? This is why I suggest to further develop this point with a new section in the ms, see my major comments

Sea ice is initialized in the same way as the other state variables and thus is not treated differently. As indicated in the following paragraph of the original manuscript, the external forcing dominates the predictability here. Manuscript unchanged in response to comment.

Page 8
L1: forecast and the uninitialized= remove 'forecast and'

Text excised as suggested.

L15-21: Please further discuss the limit of your approach= for example your initialization procedure fails at capturing the observed variability. You estimate a predictability of 7 years with only 20 years of data (which is not enough). I suggest the authors to discuss this point and hence to highlight the most of the results presented in this work relates to the potential predictability rather than an effective predictability.

In response to this and other comments, we have substantially modified the Conclusions section:

We analyze output from the CESM-DPLE system to quantify and understand the sources of predictability and predictive skill in global and regional air-sea CO_2 flux on annual to decadal timescales. We find high potential predictability in globally-integrated CO_2 flux several years in advance that is engendered by initialization. This potential predictability is evident across much of the global ocean, driven by predictability in ΔpCO_2 which itself is primarily driven by predictability in surface ocean DIC and Alk. While the CESM-DPLE system exhibits strong potential predictability, model skill as compared to an observationally-based product remains a challenge to developing useful forecasts.

…

While the ever-expanding field of decadal climate prediction has the potential to inform policy and management decisions moving forward, decadal forecasts come with several caveats. Initialization shock and drift of the coupled model system, inability of Earth system models to realistically simulate internal variability, uncertain future levels of radiative forcing, and imperfect observations are frequently cited as limitations to making accurate forecasts of the future (Meehl et al., 2014). In the case of ocean carbon, it is important to note that potential predictability in regional CO_2 flux may be driven by initialization of the physical (e.g., SST) or biogeochemical (e.g., DIC) ocean state (Li et al., 2016), and that the spatiotemporal coverage of CO_2 flux observations is insufficient to fully address predictive skill in our forecast systems.

Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., Kennedy, J. J., Wiltshire, A. J., : : : Aragão, L. E. O. C. (2018). A successful prediction of the record CO 2 rise associated with the 2015/2016 El Niño. Philosophical Transactions of the Royal Society B: Biological Sciences, 373(1760), 20170301. https://doi.org/10.1098/rstb.2017.0301
Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., & Kennedy, J. J. (2016). El Niño and a record CO2 rise. Nature Climate Change, 6(9), 806–810. https://doi.org/10.1038/nclimate3063 Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L., & Roeckner, E. (2008). Advancing decadal-scale climate prediction in the North Atlantic sector. Nature, 453(7191), 84–88. https://doi.org/10.1038/nature06921
Kim, H. M., Webster, P. J., & Curry, J. A. (2012). Evaluation of shortterm climate change prediction in multi-model CMIP5 decadal hindcasts. Geophysical Research Letters, 39(10), L10701. https://doi.org/10.1029/2012GL051644

Li, H., Ilyina, T., Müller, W. A., & Sienz, F. (2016). Decadal predictions of the North Atlantic CO2 uptake. Nature Communications, 7(May 2015), 11076. https://doi.org/10.1038/ncomms11076 Resplandy, L., Séférian, R., & Bopp, L. (2015). Natural variability of CO<inf>2</inf>and O<inf>2</inf>fluxes: What can we learn from centuries-long climate models simulations? Journal of Geophysical Research: Oceans, 120(1). https://doi.org/10.1002/2014JC010463 Séférian, R., Berthet, S., & Chevallier, M. (2018). Assessing the Decadal Predictability of Land and Ocean Carbon Uptake. Geophysical Research Letters, 45(5), 2455–2466. https://doi.org/10.1002/2017GL076092 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 of the United States of America, 111(32). https://doi.org/10.1073/pnas.1315855111 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. https://doi.org/10.1126/science.1139540

Anonymous Referee #2

Received and published: 5 November 2018

The authors have investigated the predictability and predictive skill of the ocean carbon uptake by using a large ensemble of 40-member decadal prediction and historical simulations based on NCAR CESM. They found a prominent improved predictability of the ocean carbon uptake in the initialized simulations in comparing with the uninitialized historical simulations and the persistence forecast. Furthermore, they attribute the predictability of ocean carbon uptake to the dissolved inorganic carbon and alkalinity. The outcome of this study is an important contribution for understanding and predicting variations of the ocean carbon uptake and the global carbon cycle, which are crucial for estimating climate change. Moreover, reconstruction and near-term predictions of global carbon cycle show large potential for supporting the future carbon stocktaking. Therefore the study on this topic merits publication on the Earth System Dynamics.

The manuscript is well written and the results are clearly stated, however, the conclusions are not quantitatively precise and not statistically robust from the results. This together with some other issues listed below prevents me accepting this manuscript at its present format.

1. The authors claimed a potential predictive skill of up to 7 years in the abstract and conclusions. Is this conclusion from Fig. 3d by comparing the initialized forecast to the uninitialized forecast? The authors did not do a statistical test if the difference between initialized and the uninitialized forecast is significant. The red circles only show if the correlation of the initialized forecast itself is significant. It seems to me that the initialized forecast (red dots) at lead time of 6 and 7 years are very close to the uninitialized forecast, these are probably not significantly distinguishable. As the improved skill due to initialization is a main quantitative conclusion in this manuscript, it requires a sound significant test, such as the commonly used bootstrap method (Goddard et al., 2013), which is also suggested by the Decadal Climate Prediction Project (DCPP) (Boer et al., 2016). In addition, the authors only show time-series and maps of predictability at lead time of 1 year. Given the high predictability of ocean carbon uptake as stated in this study, time-series and maps of predictability at longer lead time at least of 2 years are more representative.

Thank you for bringing this to our attention. We have revisited the statistical framework for the correlations and now include a robust assessment of the statistical separation of the anomaly correlation coefficients. To do this, we perform a Fisher's r to z transformation on the correlation coefficients and compare the resulting z test statistic to the value for the 95% confidence interval (1.96). We have modified Figures 3, 9, and 10 and Table 1 to reflect this change, and have modified the text in the abstract, conclusions, and throughout the main manuscript to indicate the findings of this statistical test.

2. The authors estimated both potential predictability against reconstruction and predictive skill against observation-based data product. The two results were separately discussed in the main text, however, the conclusions are mixed especially in the abstract. It's quite difficult for

the readers to distinguish the origin of the conclusions, they are from potential skill or skill against observation. For this reason, the abstract needs to be reorganized and make it clearer. Furthermore, the connections between potential predictabiltiy and predictive skill are weak in the manuscript. How consistent/inconsistent are the predictability and the predictive skill? What would be the implication of potential predictability to the predictive skill versus observation?

Thank you for this suggestion. We have reorganized the abstract to more clearly state the difference between the potential to predict and the actual skill:

Interannual variations in air-sea fluxes of carbon dioxide (CO_2) impact the global carbon cycle and climate system, and previous studies suggest that these variations may be predictable in the near-term (from a year to a decade in advance). Here, we quantify and understand the sources of near-term predictability and predictive skill in air -sea CO_2 flux on global and regional scales by analyzing output from a novel set of retrospective decadal forecasts of an Earth system model. These forecasts exhibit the potential to predict year-to-year variations in the globally-integrated air-sea CO_2 flux several years in advance, as indicated by the high correlation of the forecasts with a model reconstruction of past CO_2 flux evolution. This potential predictability exceeds that obtained solely from foreknowledge of variations in external forcing or a simple persistence forecast, with the longest-lasting forecast enhancement in the subantarctic Southern Ocean and the northern North Atlantic. Potential predictability in CO_2 flux variations are largely driven by predictability in the surface ocean partial pressure of CO_2, which itself is a function of predictability in surface ocean dissolved inorganic carbon and alkalinity. The potential predictability, however, is not realized as predictive skill, as indicated by the moderate to low correlation of the forecasts with an observationally-based CO_2 flux product. Nevertheless, our results suggest that year-to-year variations in ocean carbon uptake have the potential to be predicted well in advance, and establish a precedent for forecasting air-sea CO_2 flux in the near future.

While we maintain the separation of potential predictability and the predictive skill in the manuscript (for reader clarity), we have rewritten the first paragraph of the conclusions to better integrate the two:

We analyze output from the CESM-DPLE system to quantify and understand the sources of predictability and predictive skill in global and regional air-sea CO_2 flux on annual to decadal timescales. We find high potential predictability in globally-integrated CO_2 flux several years in advance that is engendered by initialization. This potential predictability is evident across much of the global ocean, driven by predictability in ΔpCO_2, which itself is primarily driven by predictability in surface ocean DIC and Alk. While the CESM-DPLE system exhibits strong potential predictability, model skill as compared to an observationally-based product remains a challenge to developing useful forecasts.

3. The initialized simulations were started from a forced ocean-sea ice simulation for the ocean-sea ice component, but were started from the CESM Large Ensemble for the atmosphere and

the land components (details were described in page 3 lines 8-13). This means that the ocean and the atmosphere and land are most probably in different climate state/phase, they need to adjust to each other and approach a new equalibrium. The mismatch of initial conditions in the ocean and in the atmosphere and land would affect the variations and predictions of the system, especially for the carbon flux across the boundaries. Discussions of the effects of mismatch in the ocean and the atmophere and land are necessary. Can the model drift due to the mismatch be largely eliminated by the drift correction?

Indeed, the full-field initialization procedure generates a drift in the forecasts. As stated in the original manuscript, we correct this drift by transforming to anomalies from a drifting climatology. The drifting climatology is statistically robust due to the large number (n=40) of ensemble members in each forecast and the large number (n=62) of start years. We include for the reviewer the companion to Figure 1 that shows the non-drift corrected forecast (see Figure R1 above). It is nearly identical to the original Figure 1, indicating that the drift in CO_2 flux is not particularly large, and thus further description of the drift is unnecessary.

4. As stated in McKinley et al. (2016), some ensemble members of the CESM-LE have problem in the ocean biogeochemical outputs. McKinley et al. (2016) used only 32 ensemble members of the CESM-LE, because some ensemble members were discarded due to a setup error which leads to corrupts of ocean biogeochemical output. In this study, the authors use 40 ensemble members as written in Page 4 lines 5-9. How do the authors treat the ensemble members with setup error in this study?

Thank you for catching this! We have excluded the LE ensemble members that have corrupted biogeochemistry. We have modified the text to reflect this.

5. The numbers in Fig. 10 are not significant and deducible from Fig. 9. For instance, the maximum forecast lead time in biome 3 (NP STSS) is 8 years in Fig. 10, but if we look at Fig. 9a, the correlations at lead time beyond 4 years are not significant and end up with less than 0.2 at lead time of 8 years. As for biome 4 (NP STPS), the maximum forecast lead time is 7 years in Fig. 10, but the initialized forecast skill is not significantly higher than the uninitialized forecast skill at lead time of 5 years in Fig. 9b. Therefore, I think the numbers in Fig. 10 need to be carefully checked by taking into account the significant test and the relative magnitude of the correlations.

We have added hatching to Figure 10 to indicate the maximum lead time in the initialized forecast that is statistically separable from the uninitialized and persistence forecasts. To do this, we perform a Fisher's r to z transformation on the correlation coefficients and compare the resulting z test statistic to the value for the 95% confidence interval (1.96). As suggested by Reviewer 1, we now include a supplemental figure (Figure S1) that shows maps of predictability for forecast lead years 2-10.

6. Table 1: the table caption and the title of the columns are unclear. I guess the "Initialized

forecast" and the "uninitialized forecast" refer to forecast skill versus reconstruction, and the "Forecast skill" refer to forecast skill versus observation-based products. The time period used to calculate the correlations needs to be specified, especially for the "Forecast skill" which use much shorter period. In addition, statistical significant test information by highlighting of the numbers will be also helpful. Moreover, a table of predictability for the maximum forecast lead time will be necessary as supplementary information to Fig. 10.

We have clarified the table column headings with footnotes, and added a column that reports the maximum forecast lead time and the maximum statistically separable lead time. As suggested by Reviewer 1, we now include a supplemental figure (Figure S1) that shows maps of predictability for forecast lead years 2-10.

7. Fig. 2: how different is the reconstruction comparing to the uninitialized simulations? Is the reconstruction closer to observations than the uninitialized simulaitons? It would be more informative to also include the climatology of the uninitialized simulations.

We have added a panel to Figure 2 that shows the climatology from the uninitialized CESM-LE simulations. We note very minor differences between the two.

8. It is not introduced but I guess the authors use different time period for the drift correction and correlation calculation along different lead time. As shown in Fig. 3d, the red dashed line has a slightly positive trend, which indicates that the authors use different time period for the correlation calculation for different lead time. To make a consistent estimate of predictive skill along all the forecast range, it is better to use the same time period for all the lead years as suggested by DCPP (Boer et al., 2016, Appendix E) and previous studies focusing on the physical predictions (Hawkins et al., 2014; Smith et al., 2013).

Thank you for bringing this to our attention. We now use a consistent time period for all of the correlation calculations from the uninitialized simulation, and have updated Figures 3, 9, and 10 and Table 1 accordingly.

9. Page 3 line 7: are the historical external forcings from CMIP5 or CMIP6 (i.e., the 5th of 6th Coupled Model Intercomparison Project)?

The CESM-DPLE was generated using the same code base, component model configurations, and historical and projected radiative forcings as in the CESM Large Ensemble (i.e., CMIP5). To avoid reader confusion, we have removed the reference to the DCPP CMIP6 protocols in Section 2:

CESM-DPLE consists of a set of initialized, fully-coupled integrations of CESM that adhere to the protocols for Component A of the Decadal Climate Prediction Project (Boer et al., 2016).

10. Page 5 line 16: ": : :for those forecasting year-to-year changes: : :" should be ": : :for those reproducing year-to-year changes: : :"

The Global Carbon Project forecasts ocean carbon uptake for the current year before it is measured/quantified, and we intended to draw attention to this in the text. Manuscript unchanged in response to comment.

11. Page 6 line 5: ": : :the anomaly correlation coefficients are scaled to CO2 flux units: : :" The correlation coefficient itself is uniformed and has no unit, there is no need to further scale it. What are the results based on the correlation coefficients without the scaling? I think the results without scaling are similar to those based on the scaled correlations. It worths to check. One more question on the scaling formular: how do the authors calculate the @_/@x, how long is the time step?

We include for the reviewer a figure (Figure R4 below) that shows the predictability/correlation coefficients before applying the scaling. This figure suggests a more important role for solubility than the scaled version, and gives the reader a false impression of the role of solubility in the predictability of CO_2 flux. We have opted to maintain the scaled version in the manuscript.

With regard to the scaling formula: the timestep is annual-mean. We now indicate this in the manuscript text.

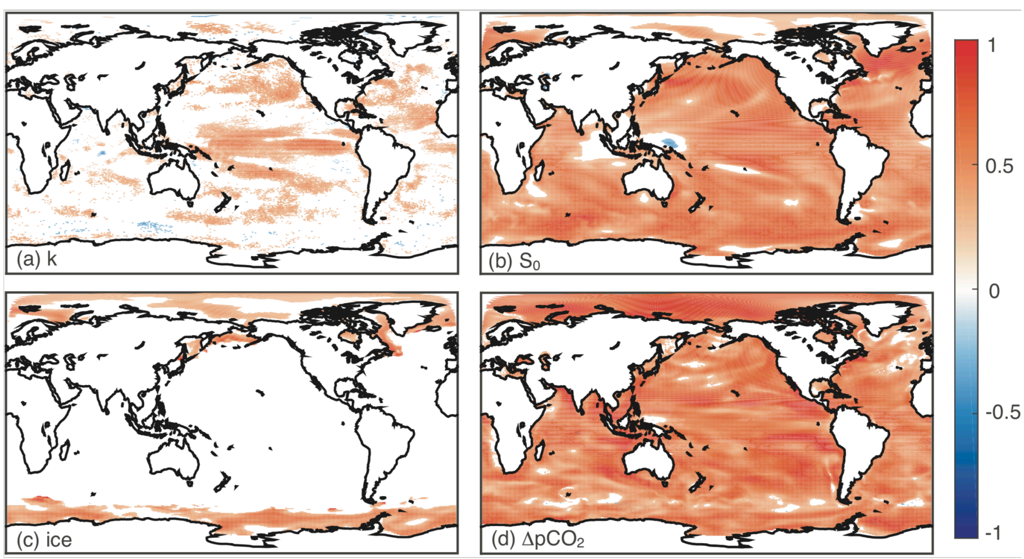

Figure R4. Predictability in the drivers of air-sea CO_2 flux during forecast year 1, as indicated by the correlation of forecast and reconstruction anomalies in the (a) gas-exchange coefficient, (b) solubility, (c) sea ice fraction, and (d) ΔpCO_2. Correlation coefficients that are not statistically significant at the 95% level using a t test are assigned a value of zeros.

12. Page 6 line 22-23: "The similar predictability of DIC and Alk across many regions hints at an important role for ocean circulation, rather than biological productivity: : :, in CO2 flux predictability." From this I understand that the biological productivity is a secondary regulation

of CO2 flux, therefore the biome division is probably not a proper way to divide the global ocean for CO2 flux predictions. The last sentence is the same as line 8-9 on Page 7.

We have overlain a map of the biome boundaries in Figure 2a to illustrate that these boundaries appropriately capture differences in mean air-sea CO_2 flux.

13. Page 8 line 26: "Li and Ilyina (2018)" should be "Li et al. (2016)", right?

Good catch! We now make reference to Li et al. (2016).

14. Figure 4 caption: "CESM-DPLE initialized forecast lead year 1" needs to be revised and includes information of the counterpart of the correlation, e.g., "CESM-DPLE initialized forecast for lead year 1 with the reconstruction".

We have modified the caption as suggested.

15. Figure 9: are the correlations based on detrended time series?

Yes. We have modified the caption to reflect this.

References: Boer, G. J., et al.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geosci. Model Dev., 9, 3751-3777, https://doi.org/10.5194/gmd-9-3751-2016, 2016. Goddard, L., et al.: A verification framework for interannual to decadal predictions experiments, Clim. Dynam., 40, 245–272, doi:10.1007/s00382-012-1481-2, 2013. Hawkins, E., Dong, B., Robson, J., Sutton, R., and Smith, D.: The interpretation and use of biases in decadal climate predictions, J. Climate, 27, 2931–2947, doi:10.1175/JCLI-D-13-00473.1, 2014. McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., and Lovenduski, N. S.: Timescales for detection of trends in the ocean carbon sink, Nature, 530, 469–472, http://dx.doi.org/10.1038/nature16958, 2016. Smith, D. M., Eade, R., and Pohlmann, H.: A comparison of fullfield and anomaly initialization for seasonal to decadal climate prediction, Clim. Dynam., 41, 3325–3338, doi:10.1007/s00382-013-1683-2, 2013.

Predicting near-term ~~changes~~ variability in ocean carbon uptake

Nicole S. Lovenduski[1], Stephen G. Yeager[2], Keith Lindsay[2], and Matthew C. Long[2]

[1]Department of Atmospheric and Oceanic Sciences and Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado, USA

[1]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado, USA

Correspondence to: Nicole S. Lovenduski (nicole.lovenduski@colorado.edu)

Abstract. ~~Annual to decadal~~ Interannual variations in air-sea fluxes of carbon dioxide (CO_2) impact the global carbon cycle and climate system, and previous studies suggest that these variations may be predictable in the near-term (from a year to a decade in advance). Here, we quantify and understand the sources of near-term ~~(annual to decadal)~~ predictability and predictive skill in air -sea CO_2 flux on global and regional scales by analyzing output from a novel set of retrospective decadal forecasts of ~~the Earth system . These initialized~~ an Earth system model. These forecasts exhibit the potential to predict year-to-year variations in the globally-integrated air-sea CO_2 flux ~~up to ∼7~~ several years in advance~~. This initialized predictability exceeds the predictability~~ , as indicated by the high correlation of the forecasts with a model reconstruction of past CO_2 flux evolution. This potential predictability exceeds that obtained solely from foreknowledge of variations in external forcing or a simple persistence forecast~~. The near-term~~ , with the longest-lasting forecast enhancement in the subantarctic Southern Ocean and the northern North Atlantic. Potential predictability in CO_2 flux ~~predictability is~~ variations are largely driven by predictability in the surface ocean partial pressure of CO_2, which itself is a function of predictability in surface ocean dissolved inorganic carbon and alkalinity. ~~Comparison with an~~ The potential predictability, however, is not realized as predictive skill, as indicated by the moderate to low correlation of the forecasts with an observationally-based ~~product suggests that the initialized forecasts exhibit moderate predictive skillin the tropics and subtropics, but low skill elsewhere. In the subantarctic Southern Ocean and northern North Atlantic, we find long-lasting initialized predictability that beats that derived from uninitialized and persistence forecasts . Our~~ CO_2 flux product. Nevertheless, our results suggest that year-to-year variations in ocean carbon uptake ~~may be predictable~~ have the potential to be predicted well in advance, and establish a precedent for forecasting air-sea CO_2 flux in the near future.

1 Introduction

Observations collected over the past few decades indicate that the ocean has absorbed 160 Pg of excess carbon from the atmosphere since the beginning of the industrial revolution (Le Quéré et al., 2018); projections from climate models suggest that ∼540 Pg of excess carbon will reside in the ocean by the end of the century (under the RCP8.5 emission scenario; Ciais and Sabine, 2013). Accurate projections of past and future air-sea CO_2 flux are important for quantifying and understanding the changing global carbon cycle and for estimating future global climate change (Le Quéré et al., 2018). Superimposed on the background of long-term changes in ocean carbon uptake is substantial variability on global and regional scales ~~(McKinley et al., 2017)~~(McKinley et al., 2017; Landschützer et al., 2016). The recent literature highlights ocean carbon uptake variability that manifests on timescales of years to decades. Interannual variability in globally-integrated air-sea CO_2 flux has been estimated to have a standard deviation of 0.31 Pg C yr^{-1} and 0.2 Pg C yr^{-1} from observationally-based products (Rödenbeck et al., 2015) and ocean biogeochemical models (Wanninkhof et al., 2013), respectively, which is on the order of 10% of the global-mean CO_2 flux (2.3 Pg C yr^{-1}). A global extrapolation of sparse pCO$_2$ observations suggests that there is large variability on decadal timescales (Landschützer et al., 2016). On regional scales, Southern Ocean studies have highlighted recent air-sea CO_2 flux variability on interannual (Wetzel et al., 2005; Lenton and Matear, 2007; Lovenduski et al., 2007, 2013, 2015a; Verdy et al., 2007; Wang and Moore, 2012; Hauck et al., 2013; Lenton et al., 2013) and decadal (Fay et al., 2014; Landschützer et al., 2015; Munro et al., 2015) timescales. In the North Atlantic, high air-sea CO_2 flux variability has been linked to the North Atlantic Oscillation (Thomas et al., 2008; Ullman et al., 2009) and the Atlantic Multidecadal Oscillation (Metzl et al., 2010; Breeden and McKinley, 2016), whose spectra peak at interannual and multi-decadal timescales.

Near-term predictions of the climate system (so-called "decadal predictions") are forecasts of climate variability and change on annual, multi-annual, and decadal timescales from global climate models (Meehl et al., 2014). These forecasts are sensitive to both initial conditions (e.g., the atmospheric temperature used to initialize the forecasts) and external forcing (e.g., the long-term increase in atmospheric temperature associated with increasing greenhouse gas concentrations; Kirtman et al., 2013). Recent publications highlight near-term predictability and predictive skill in regional surface air temperature, precipitation, Arctic sea ice concentration, oceanic heat content, and the large-scale Atlantic Ocean circulation ~~(Meehl et al., 2009; Robson et al., 2012; Yeager e~~ (Smith et al., 2007; Keenlyside et al., 2008; Meehl et al., 2009; Robson et al., 2012; Yeager et al., 2012; Meehl et al., 2014; Yeager et al., . As prior literature has established a strong link between air-sea CO_2 flux and variability in the physical climate system on these timescales ~~(e.g., McKinley et al., 2017)~~(e.g., Resplandy et al., 2015; McKinley et al., 2017), it follows that air-sea CO_2 flux may be predictable in the near-term.

Here, we analyze a novel set of decadal prediction simulations from an Earth System Model (ESM) to investigate near-term predictions of global and regional ocean carbon uptake. On annual to decadal timescales, ESM predictions of the past (so-called "retrospective forecasts") are used to assess both predictability and predictive skill in air-sea CO_2 flux. Predictability is the potential to predict the system, based on forecast verification against a model reconstruction. Predictive skill is based on forecast verification against observations. We further assess the role of external forcing in the predictability of CO_2 flux by analyzing a set of uninitialized forecasts run under identical external forcing. By analyzing forecasts of the past, our study establishes a precedent for making skillful predictions of ocean carbon uptake in the near future.

2 Community Earth System Model Decadal Prediction System

Our primary numerical tool is the Community Earth System Model Decadal Prediction Large Ensemble (CESM-DPLE; Yeager et al., 2018). In this section, we describe the model and provide details about forecast initialization, ensemble generation, and drift correction. Importantly, we note that this is the first CESM decadal prediction system to include a representation of ocean biogeochemistry. CESM-DPLE uses the same code base as the CESM Large Ensemble (CESM-LE; Kay et al., 2015).

The CESM is a state-of-the-art coupled climate model consisting of atmosphere, ocean, land, and sea ice component models (Hurrell et al., 2013; Danabasoglu et al., 2012; Lawrence et al., 2012; Hunke and Lipscomb, 2008). The ocean physical model (version 2 of the Parallel Ocean Program; Danabasoglu et al., 2012) has nominal 1° horizontal resolution and 60 vertical levels. The biogeochemical ocean model represents the lower trophic levels of the marine ecosystem (Moore et al., 2004, 2013), full carbonate system thermodynamics (Long et al., 2013), air-sea CO_2 fluxes, and a dynamic iron cycle (Doney et al., 2006; Moore and Braucher, 2008).

CESM-DPLE consists of a set of initialized, fully-coupled integrations of CESM that adhere to the protocols for Component A of the Decadal Climate Prediction Project ~~(DCPP), a contribution to the 6^{th} Coupled Model Intercomparison Project~~ (Boer et al., 2016). We use the CESM-DPLE system (Yeager et al., 2018) that builds on previous CESM decadal prediction efforts (Yeager et al., 2012, 2015) with some modifications (including the addition of ocean biogeochemistry, as noted above). CESM-DPLE initiates 40 decade long "forecasts" of the Earth system each year from 1954-2015; the start date for each forecast is November 1, in accordance with the DCPP protocols. Each of the model integrations are subject to a common set of historical external forcings (e.g., greenhouse gas concentrations).

The ocean physical and biogeochemical initial conditions for the DP experiments are generated from a forced ocean - sea ice simulation of the CESM. That is, a simulation of the ocean and ice components of the CESM that has been forced with fluxes computed from the observed atmospheric state over 1948-2015. This simulation is therefore meant to reconstruct the historical evolution of the ocean physical and biogeochemical state over the 1948-2015 period (Figure 1). Hereafter, we refer to this simulation as the "reconstruction". Initial conditions from the atmosphere and land components of the DP experiments are obtained from a 20^{th} century simulation of the CESM Large Ensemble (Kay et al., 2015).

Ocean biogeochemistry in the version of the CESM used for CESM-DPLE has been extensively validated in the literature (Long et al., 2016; Lovenduski et al., 2016; McKinley et al., 2016; Krumhardt et al., 2017; Freeman et al., 2018). In particular, the simulated mean, variability, and trends in surface ocean pCO$_2$ and air-sea CO_2 flux from CESM over 1982-2011 compare favorably to estimates from observations for the global average and over most ocean biogeochemical biomes (McKinley et al., 2016; Lovenduski et al., 2016). In Figure 2, we illustrate the comparison between observationally-based estimates of CO_2 flux (from the Landschützer et al. (2016) pCO$_2$ product) and estimates produced by the reconstruction and coupled CESM-LE over 1982-2015~~, which indicates that the~~. The model reconstruction does a reasonable job (r = 0.79) of representing observed spatial patterns (in both magnitude and direction) of the flux across most oceanic regions. The globally-integrated air-sea CO_2 flux over 1982-2015 from the observational product and model reconstruction are 1.41 and 1.80 Pg C yr^{-1}, respectively (directed into the ocean).

CESM-DPLE initializes an ensemble of 40 simulations each year using round-off level (order 10^{-14}) perturbations in the initial air temperature field (Figure 1). Previous work indicates that this small perturbation in the initial conditions generates a wide divergence in global mean surface temperatures across the ensemble members within about 30 days (V. Yettella, pers. comm., 2018), and the average divergence in globally-integrated, annual-mean forecast CO_2 flux across the ensemble members (0.53 Pg C yr^{-1}) is an order of magnitude greater than that generated by the preindustrial control simulation of CESM (0.09 Pg C yr^{-1}; Lovenduski et al., 2015b). Each ensemble member is subject to identical external forcing. The number of ensemble members in each forecast ensures statistically robust drift estimates (see below; Boer et al., 2013; Kirtman et al., 2013; Yeager et al., 2018).

Following initialization, the coupled model drifts toward its preferred state over the decadal forecast. This is a common problem for full-field initialization decadal prediction experiments (Meehl et al., 2014) and requires a drift correction to be applied to the model forecasts before predictability and predictive skill may be analyzed. We correct the drift by transforming to anomalies from a drifting climatology, as in Yeager et al. (2012) and Yeager et al. (2018). For a given forecast, $X(L,M,S)$, where L is the forecast length, M is the ensemble member, and S is the start year of the forecast, the drift-corrected forecast anomaly, $X'(L,M,S)$ is defined as

$$X'(L,M,S) = X(L,M,S) - \overline{X(L,M,S)}^{M,S}, \tag{1}$$

where $\overline{X(L,M,S)}^{M,S}$ is the average rate of drift over all forecasts. Note that this method does not assume that the drift is linear, and disregards potential dependence of the drift on the external forcing.

Predictive skill in CESM-DPLE may be enabled by external forcing (e.g. the time evolution of atmospheric greenhouse gases) as well as by initialization. To assess the role of initialization in predictability, we compare CESM-DPLE air-sea CO_2 flux (generated with the initialization procedure described above) with air-sea CO_2 flux from the CESM-LE (McKinley et al., 2016; Lovenduski et al., 2016) over the same historical period. The CESM-LE is a ~~40-member~~ 32-member ensemble of the CESM with fully resolved ocean biogeochemistry that evolves the Earth system from 1920 to 2100 under historical and RCP8.5 forcing (Kay et al., 2015). As such, CESM-LE represents the uninitialized counterpart to the CESM-DPLE system; output from CESM-LE can tell us how the modeled air-sea CO_2 flux would evolve over a given decade in the absence of initialization, but under the same external forcing.

3 Results

3.1 Predictability

Predictability is a property of a system that characterizes the ~~ability~~ potential for its future evolution to be predicted; this concept is distinct from that of model skill. We quantify predictability by evaluating the ability of the CESM-DPLE initialized forecasts to predict variations in air-sea CO_2 flux from the reconstruction. For a given forecast anomaly, $X'(L,M,S)$, predictability is defined as the correlation coefficient of $X'(L,M,S)$ with the corresponding anomaly in the reconstruction; the reconstruction anomaly is obtained by subtracting the climatological mean value over 1955-2015.

The globally-integrated, air-sea CO_2 flux anomaly from the initialized CESM-DPLE in forecast year 1 exhibits high correlation with the CO_2 flux anomaly from the reconstruction (Figure 3a; r = 0.98). This correlation remains high and statistically significant (at the 95% level, using a two-sided student t test while accounting for autocorrelation in the sample size) for 10 forecast lead years (Figure 3c), suggesting high, long-lasting predictability in the globally-integrated air-sea CO_2 flux.

We further investigate whether the predictability in the globally-integrated air-sea CO_2 flux is a function of initialization by (1) correlating integrated CO_2 flux anomalies from the ensemble mean of the uninitialized CESM-LE simulation with anomalies from the reconstruction, and (2) generating a persistence forecast (autocorrelation as a function of lead time) for the CO_2 flux anomalies from the reconstruction. Figures 3a and 3c reveal that the initialization of the forecast does not much improve the prediction from the uninitialized forecast. This is because the strong externally-forced component of the forecast (e.g., the rising CO_2 concentration in the atmosphere) provides an important source of predictability in both the initialized and uninitialized forecasts. While the persistence forecast also yields high correlation coefficients, both the initialized and uninitialized forecasts beat persistence for all prediction lead times (Figure 3c).

Figure 3a also reveals interannual variability in the globally-integrated air-sea CO_2 flux. While this variability is swamped by the externally forced signal (i.e., the increasing CO_2 uptake due to rising atmospheric CO_2), we are nevertheless interested in the ability of CESM-DPLE to forecast this year-to-year variability. To accomplish this, we remove the linear trend from the forecasts and the reconstruction before computing predictability; this method produces estimates of correlation that are not dominated by the trend induced by external forcing. The globally-integrated, detrended, air-sea CO_2 flux anomaly from the initialized CESM-DPLE in lead year 1 exhibits high correlation with CO_2 flux from the reconstruction (Figure 3b; r = 0.70), suggesting high predictability of ocean carbon uptake variability on interannual timescales, as well. While this predictability drops off with forecast lead time, we nevertheless find high correlations ($r >$ ~~0.4~~0.5) between the annual-mean CO_2 flux forecast anomalies and detrended reconstruction anomalies that extend for ~~7 years , and statistically significant correlations that extend for 10 years~~ 4 years (Figure 3d).

Interannual variability in global air-sea CO_2 flux may also be affected by interannual variability in external forcing (e.g., volcanoes). As above, we evaluate the role of initialization by calculating uninitialized predictability and estimating persistence. Figure 3 indicates that the initialized forecast exhibits higher predictability than the uninitialized forecast and the persistence forecast for ~~lead times of 7 and~~ a lead time of 10 years, ~~respectively~~though this initialized predictability is only statistically separable from the uninitialized and persistence forecasts for lead years 1-2 and 2, respectively (statistical separation determined via a Fisher's r to z transformation and a comparison of the resulting z test statistic to the value for the 95% confidence interval (1.96)). Thus, the CESM-DPLE initialized forecasts have the potential to predict year-to-year variations of globally-integrated air-sea CO_2 flux ~~up to 7~~ several years in advance.

The results from our analysis of the globally-integrated air-sea CO_2 flux suggest that interannual variations in global ocean carbon uptake may be predictable ~~well~~ in advance. They further indicate that initialization of the forecasts enhances the predictability of future interannual variations over and above the predictability from variations in the external forcing, such as those imposed by volcanic eruptions~~, fluctuations in the rate of emissions, or variability in terrestrial CO_2 fluxes~~. This is a particularly meaningful result for those forecasting year-to-year changes in the global carbon budget (e.g., Le Quéré et al., 2018), especially as these forecasting efforts are blind to the externally forced variability in advance (i.e., the external forcing of the future is unknown). In this way, near-term predictions of air-sea CO_2 flux variations can help to inform future predictions of land-air CO_2 flux and atmospheric CO_2.

Given the high predictability and the important role of initialization in forecasts of interannual air-sea CO_2 fluxes on a global scale, we next investigate the spatial patterns of air-sea CO_2 flux predictability across the global ocean. Here, we use the same statistical techniques as for the global flux, but instead perform analysis in each model grid cell. On a global scale, the evolution of air-sea CO_2 flux is dominated by the long-term increase in ocean uptake (see, e.g., Figure 3a), whereas on local and regional scales, the evolution is dominated by interannual variability (Figure 1; see also, e.g., Lovenduski et al., 2016). To capture the predictability on interannual timescales, we perform analysis on linearly detrended forecasts in each model grid cell. Figure 4a illustrates large predictability of initialized CO_2 flux across much of the global ocean for forecast lead year ~~1.~~ 1 (additional forecast lead years shown in Figure S1). The uninitialized forecast (Figure 4b) and the persistence forecast (Figure 4c) indicate lower predictability.

If not external forcing or persistence, what drives the high predictability in air-sea CO_2 flux interannual variability? We decompose the predictability of air-sea CO_2 flux (Φ) over forecast lead year 1 by considering the predictability of its drivers:

$$\Phi = k \cdot S_0 \cdot (1 - ice) \cdot \Delta pCO_2, \qquad (2)$$

where k is the piston velocity (also known as the gas transfer coefficient), S_0 is the solubility of CO_2 in seawater, ice is the fraction of the ocean covered by sea ice, and ΔpCO_2 is the difference between the oceanic $p\mathrm{CO}_2$ and the atmospheric $p\mathrm{CO}_2$. As for CO_2 flux, predictability is defined as the anomaly correlation coefficient of each driver variable in forecast year 1 with the corresponding anomaly of that driver variable in the reconstruction, e.g., the correlation of anomalous piston velocities from the forecast with those from the reconstruction. Figure 5 shows the predictability of each of the CO_2 flux driver variables, where the anomaly correlation coefficients are scaled to CO_2 flux units (mol m^{-2} yr^{-1}) and can be easily compared. The predictability scaling is achieved by multiplying the anomaly correlation coefficient (r) by the sensitivity of CO_2 flux to each driver variabile (x) and the standard deviation of the driver variable timeseries:

$$r \cdot \frac{\partial \Phi}{\partial x} \cdot \sigma_x, \qquad (3)$$

where the sensitivities and standard deviations are established from model-estimated, annual-mean quantities in each grid cell (as in, e.g., Lovenduski et al., 2007, 2013, 2015a), using annual averages from the reconstruction. The CO_2 flux predictability is largely driven by predictability in $\Delta p\mathrm{CO}_2$ across the global ocean (Figure 5). Our results suggest secondary roles for the piston velocity in the equatorial Pacific, solubility in the North Atlantic subpolar gyre, and sea ice fraction in the Arctic/North Atlantic and high latitude Southern Ocean. Elsewhere, these other driver variables play only minor roles in CO_2 flux predictability.

As the large predictability in $\Delta p\mathrm{CO}_2$ is caused by predictability of surface ocean $p\mathrm{CO}_2$ in our model framework (i.e., atmospheric CO_2 concentration is prescribed, rather than predicted), we next investigate the drivers of interannual predictability in surface ocean $p\mathrm{CO}_2$: dissolved inorganic carbon (DIC), alkalinity (Alk), temperature (T), and salinity (S). We use a similar approach as for CO_2 flux, but here the sensitivities are derived from carbonate chemistry approximations (Lovenduski et al., 2007; Doney et al., 2009; Long et al., 2013), and all drivers are scaled to pCO$_2$ units (μatm) for ease of comparison:

$$r \cdot \frac{\partial p\mathrm{CO}_2}{\partial x} \cdot \sigma_x. \tag{4}$$

The surface ocean pCO$_2$, and thus the air-sea CO_2 flux predictability for forecast lead year 1 is largely driven by predictability in surface ocean DIC and Alk, with temperature playing a secondary role, and salinity a minor role (Figure 6). The similar predictability of DIC and Alk across many regions hints at an important role for ocean circulation, rather than biological productivity (which has a much larger impact on DIC than Alk), in CO_2 flux predictability.

3.2 Predictive skill

We next evaluate the predictive skill of the CESM-DPLE forecasts; the skill is a measure of the ability of the forecast to reproduce the observational record. For air-sea CO_2 flux, direct observations are rare, and we are constrained to estimates of flux from observations of sparsely sampled surface ocean pCO$_2$. Here, we use as our observational metric the CO_2 flux estimated from the Landschützer et al. (2016) surface ocean pCO$_2$ product. This product is a gap-filled estimate of surface ocean pCO$_2$, which, when combined with measurements of atmospheric pCO$_2$, sea surface temperature, salinity, and wind, yields a monthly estimate of air-sea CO_2 flux at 1° x 1° horizontal resolution from 1982-2015 (see also Figure 2a). As the pCO$_2$ observations are rather sparse prior to 1995 (see Figure 2 of Bakker et al., 2016), we calculate skill for the period between 1995 and 2015 only, but show for the interested reader the full observational product timeseries.

The CESM-DPLE initialized predictions exhibit some skill at representing the globally-integrated air-sea CO_2 flux in forecast lead year 1 (Figure 3a,b; initialized forecast skill = 0.88; detrended, initialized forecast skill = 0.66). Our comparison indicates that CESM-DPLE (and the reconstruction, for that matter) struggles to produce the pronounced trends toward anomalous CO_2 outgassing in the 1990s and anomalous CO_2 uptake in the 2000s. The ability (or lack thereof) of ESMs to reproduce the observationally-derived multi-decadal air-sea CO_2 flux variability has been the subject of recent publications (e.g., Li and Ilyina, 2018; Gruber et al., 2017), though no robust mechanisms seem to explain the (mis)match. The CESM-DPLE initialized forecast in forecast lead year 1 exhibits moderate predictive skill in the tropics and subtropics (Figure 7), and low skill elsewhere.

3.3 Predictability and predictive skill on the biome scale

Because the predictability of air-sea CO_2 flux is primarily driven by predictability of the biogeochemical state variables DIC and Alk, it makes sense to aggregate predictability across biogeographical biomes. We probe the limits of predictability and predictive skill in regional air-sea CO_2 flux by averaging the local flux across 17 biogeographical biomes. This is achieved by re-gridding the Fay and McKinley (2014) mean biome mask to the CESM model grid and computing the area-weighted, average CO_2 flux from the reconstruction, CESM-DPLE initialized forecasts, and observationally-derived pCO$_2$ product. The detrended CO_2 flux anomalies for three of the biomes are shown for forecast lead year 1 in Figure 8, and the predictability and predictive skill across all biomes is detailed in Table 1. These three biomes were chosen to contrast their predictability and/or predictive skill.

The biome-averaged CO_2 flux anomalies from the CESM-DPLE initialized forecast in forecast lead year 1 exhibit high correlations with the reconstruction anomalies in the North Pacific Subtropical biomes, and in the Southern Ocean Ice biome (Figure 8; Table 1), indicating high potential for prediction of CO_2 flux anomalies. This predictability decreases with increasing forecast lead time in the North Pacific Subtropical biomes, but persists for the Southern Ocean ice biome through forecast years 7-9 (Figure 8). Indeed, the Southern Ocean Ice biome is an anomaly in this regard; in the other 16 biomes, predictability drops off with prediction lead time (not shown).

Initialization engenders predictability of air-sea CO_2 flux variability the North Pacific Subtropical biomes, as we find low correlation between the uninitialized CESM-LE forecast CO_2 flux anomalies and the reconstruction anomalies here (Figure 8a,b; Table 1). The initialized forecast for these biomes has higher predictability than the uninitialized forecast and the persistence forecast for 7-8 years (Figure 9). These conclusions hold for most of the other ocean biomes (Table 1), with a few exceptions where the uninitialized forecast and/or persistence forecast are similar to the initialized forecast (e.g., the East Pacific Equatorial biome). In the Southern Ocean Ice biome, the CO_2 flux predictability is almost entirely driven by external forcing, and the persistence forecast indicates high predictability, as well (Figure 8; Figure 9, Table 1). Thus, the high and long-lasting predictability in this biome must be interpreted with caution, given the importance of external forcing in predicting CO_2 flux anomalies here.

The predictive skill of CESM-DPLE in forecast lead year 1 is illustrated for three biomes in Figure 8 and Table 1. Again, we note the moderate skill in the tropics and subtropics, and lower skill elsewhere.

The difference in the predictability between the initialized~~forecasts and the uninitialized~~, uninitialized, and persistence forecasts reveals the impact of initialization on predictions of air-sea CO_2 flux variability on the biome scale (Figure 9). We probe the limits of initialized predictability in each biome by calculating the maximum forecast lead time for which the initialized CESM-DPLE CO_2 flux forecast has both higher predictability than the uninitialized CESM-LE and persistence forecasts and present the results in Figure 10. Our results indicate that initialization improves the forecast for the longest lead times in the subantarctic Southern Ocean and the northern North Atlantic, where the initialized forecast beats the other two forecasts out to forecast lead times of 10 and 9 years, respectively~~(Figure 10)~~. We note, however, that the improvement in the North Atlantic is only statistically significant for 1 lead year, and in the Southern Ocean for 2-3 lead years. Given the important role of these two regions for the global ocean uptake of anthropogenic carbon, and the numerous studies linking climate variability to air-sea CO_2 flux variability in these regions, this long-lasting predictability is ~~akin to that of the global CO_2 flux integral and very~~ encouraging. In other regions, however, such as the Southern Ocean Ice or East Equatorial Pacific biomes, the initialized forecast only beats the uninitialized or persistence forecast for a single year, indicating little benefit of forecast initialization on CO_2 flux forecasts here.

4 Conclusions

We analyze output from the CESM-DPLE system to quantify and understand the sources of predictability and predictive skill in global and regional air-sea CO_2 flux on annual to decadal timescales. We find high potential predictability in globally-integrated CO_2 flux several years in advance that is engendered by initialization~~and extends to forecast lead times of ~7 years. This~~. This potential predictability is evident across much of the global ocean, driven by predictability in $\Delta p CO_2$, which itself is primarily driven by predictability in surface ocean DIC and Alk. While the CESM-DPLE system exhibits strong potential predictability, model skill as compared to an observationally-based product remains a challenge to developing useful forecasts.

~~On the biome scale, we find particularly long-lasting predictability in the northern North Atlantic and subantarctic Southern Ocean that is engendered by initialization.~~

Our study complements two recent studies of ocean carbon decadal predictions conducted at different modeling centers. Li et al. (2016) use decadal predictions from MPI-ESM to investigate near-term changes in North Atlantic CO_2 flux, while Séférian et al. (2018) use CNRM-ESM1 to assess the predictability horizon of globally-integrated ocean and land carbon fluxes. While these studies use different prediction systems, we nevertheless come to some of the same conclusions. For example, Séférian et al. (2018) find that global ocean carbon uptake is potentially predictable for up to 6 years, and ~~Li and Ilyina (2018)~~ Li et al. (2016) find high potential predictability in the North Atlantic that is engendered by initialization. These studies collectively suggest predictability for near-term ocean carbon uptake on global and regional scales, which is beneficial for forecasting the future global carbon budget and climate system.

While the ever-expanding field of decadal climate prediction has the potential to inform policy and management decisions moving forward, decadal forecasts come with several caveats. Initialization shock and drift of the coupled model system, inability of Earth system models to realistically simulate internal variability, uncertain future levels of radiative forcing, and imperfect observations are frequently cited as limitations to making accurate forecasts of the future (Meehl et al., 2014). In the case of ocean carbon, it is important to note that potential predictability in regional CO_2 flux may be driven by initialization of the physical (e.g., SST) or biogeochemical (e.g., DIC) ocean state (Li et al., 2016), and that the spatiotemporal coverage of CO_2 flux observations is insufficient to fully address predictive skill in our forecast systems.

Acknowledgements. 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 U.S. Department of Energy under Contract No. DE-AC02-05CH11231, as well as by an Accelerated Scientific Discovery grant for Cheyenne (doi:10.5065/ D6RX99HX) that was awarded by NCAR's Computational and Information Systems Laboratory. The NCAR contribution to this study was supported by the National Oceanic and Atmospheric Administration Climate Program Office under Climate Variability and Predictability Program grant NA09OAR4310163, the National Science Foundation (NSF) Collaborative Research EaSM2 grant OCE-1243015, and the NSF through its sponsorship of NCAR. NSL is grateful for funding from NSF (OCE-1752724, OCE-1558225).

References

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S.-I., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibánhez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., van Heuven, S. M. A. C., Vandemark, D., Ward, B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality $f\mathrm{CO_2}$ data in version 3 of the Surface Ocean $\mathrm{CO_2}$ Atlas (SOCAT), Earth Syst. Sci. Data, 8, 383–413, doi:10.5194/essd-8-383-2016, http://www.earth-syst-sci-data.net/8/383/2016/, 2016.

Boer, G. J., Kharin, V. V., and Merryfield, W. J.: Decadal predictability and forecast skill, Clim. Dynam., 41, 1817–1833, doi:10.1007/s00382-013-1705-0, http://dx.doi.org/10.1007/s00382-013-1705-0, 2013.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geosci. Model Dev., 9, 3751–3777, doi:10.5194/gmd-9-3751-2016, http://www.geosci-model-dev.net/9/3751/2016/, 2016.

Breeden, M. L. and McKinley, G. A.: Climate impacts on multidecadal $p\mathrm{CO_2}$ variability in the North Atlantic: 1948–2009, Biogeosciences, 13, 3387–3396, doi:10.5194/bg-13-3387-2016, https://www.biogeosciences.net/13/3387/2016/, 2016.

Ciais, P. and Sabine, C.: Chapter 6: Carbon and Other Biogeochemical Cycles, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M. B., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., p. 1535 pp, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., Peacock, S., and Yeager, S. G.: The CCSM4 Ocean Component, J. Climate, 25, 1361–1389, 2012.

Doney, S. C., Lindsay, K., Fung, I., and John, J.: Natural Variability in a Stable, 1000-Yr Global Coupled Climate–Carbon Cycle Simulation, J. Climate, 19, 3033–3054, http://dx.doi.org/10.1175/JCLI3783.1, 2006.

Doney, S. C., Lima, I., Feely, R. A., Glover, D. M., Lindsay, K., Mahowald, N., Moore, J. K., and Wanninkhof, R.: Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air-sea $\mathrm{CO_2}$ fluxes: Physical climate and atmospheric dust, Deep-Sea Res. II, 56, 640–655, http://www.sciencedirect.com/science/article/pii/S096706450800427X, 2009.

Fay, A. R. and McKinley, G. A.: Global open-ocean biomes: mean and temporal variability, Earth Syst. Sci. Data, 6, 273–284, doi:10.5194/essd-6-273-2014, http://www.earth-syst-sci-data.net/6/273/2014/, 2014.

Fay, A. R., McKinley, G. A., and Lovenduski, N. S.: Southern Ocean carbon trends: Sensitivity to methods, Geophys. Res. Lett., 41, 6833–6840, doi:10.1002/2014GL061324, http://dx.doi.org/10.1002/2014GL061324, 2014.

Freeman, N. M., Lovenduski, N. S., Munro, D. R., Krumhardt, K. M., Lindsay, K., Long, M. C., and Maclennan, M.: The Variable and Changing Southern Ocean Silicate Front: Insights From the CESM Large Ensemble, Global Biogeochem. Cycles, 32, 752–768, doi:10.1029/2017GB005816, https://doi.org/10.1029/2017GB005816, 2018.

Gruber, N., Landschützer, P., and Lovenduski, N. S.: The Variable Southern Ocean Carbon Sink, Annu. Rev. Mar. Sci., doi:10.1146/annurev-marine-121916-063407, https://doi.org/10.1146/annurev-marine-121916-063407, 2017.

Hauck, J., Völker, C., Wang, T., Hoppema, M., Losch, M., and Wolf-Gladrow, D. A.: Seasonally different carbon flux changes in the Southern Ocean in response to the Southern Annular Mode, Global Biogeochem. Cycles, 27, 1236–1245, doi:10.1002/2013GB004600, 2013.

Hunke, E. C. and Lipscomb, W. H.: CICE: the Los Alamos sea ice model user's manual, version 4, Los Alamos Natl. Lab. Tech. Report, LA-CC-06-012, 2008.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J. F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research, B. Am. Meteorol. Soc., 94, 1339–1360, doi:10.1175/BAMS-D-12-00121.1, 2013.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J. F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability, B. Am. Meteorol. Soc., 96, 1333–1349, doi:10.1175/BAMS-D-13-00255.1, 2015.

Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L., and Roeckner, E.: Advancing decadal-scale climate prediction in the North Atlantic sector, Nature, 453, 84 EP –, https://doi.org/10.1038/nature06921, 2008.

Kirtman, B., Power, S. B., Adedoyin, J. A., Boer, G. J., Bojariu, R., Camilloni, I., Doblas-Reyes, F. J., Fiore, A. M., Kimoto, M., Meehl, G. A., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G. J., Vecchi, G., and Wang, H. J.: Near-term Climate Change: Projections and Predictability, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, 2013.

Krumhardt, K. M., Lovenduski, N. S., Long, M. C., and Lindsay, K.: Avoidable impacts of ocean warming on marine primary production: Insights from the CESM ensembles, Global Biogeochem. Cycles, 31, 114–133, doi:10.1002/2016GB005528, http://dx.doi.org/10.1002/2016GB005528, 2017.

Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., van Heuven, S., Hoppema, M., Metzl, N., Sweeney, C., Takahashi, T., Tilbrook, B., and Wanninkhof, R.: The reinvigoration of the Southern Ocean carbon sink, Science, 349, 1221–1224, http://www.sciencemag.org/content/349/6253/1221.abstractN2-SeveralstudieshavesuggestedthatthecarbonsinkintheSouthernOcean\T1\textemdashtheo 2015.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochem. Cycles, 30, 1396–1417, doi:10.1002/2015GB005359, http://dx.doi.org/10.1002/2015GB005359, 2015GB005359, 2016.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Fletcher, C. G., Lawrence, P. J., Levis, S., Swenson, S. C., and Bonan, G. B.: The CCSM4 Land Simulation, 1850–2005: Assessment of Surface Climate and New Capabilities, J. Climate, 25, 2240–2260, doi:10.1175/JCLI-D-11-00103.1, http://dx.doi.org/10.1175/JCLI-D-11-00103.1, 2012.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Jackson, R. B., Boden, T. A., Tans, P. P., Andrews, O. D., Arora, V. K., Bakker, D. C. E., Barbero, L., Becker, M., Betts, R. A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Cosca, C. E., Cross, J., Currie, K., Gasser, T., Harris, I., Hauck, J., Haverd, V., Houghton, R. A., Hunt, C. W., Hurtt, G., Ilyina, T., Jain, A. K., Kato, E., Kautz, M., Keeling, R. F., Klein Goldewijk, K., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lima, I., Lombardozzi, D., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Nojiri, Y., Padin, X. A., Peregon, A., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Reimer, J., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Viovy, N., Vuichard, N., Walker, A. P., Watson, A. J., Wiltshire, A. J., Zaehle, S., and Zhu, D.: Global Carbon Budget 2017, Earth Syst. Sci. Data, 10, 405–448, doi:10.5194/essd-10-405-2018, 2018.

Lenton, A. and Matear, R. J.: Role of the Southern Annular Mode (SAM) in Southern Ocean CO_2 uptake, Global Biogeochem. Cycles, 21, GB2016, doi:10.1029/2006GB002714, http://dx.doi.org/10.1029/2006GB002714, 2007.

Lenton, A., Tilbrook, B., Law, R. M., Bakker, D., Doney, S. C., Gruber, N., Ishii, M., Hoppema, M., Lovenduski, N. S., Matear, R. J., McNeil, B. I., Metzl, N., Mikaloff Fletcher, S. E., Monteiro, P. M. S., Rödenbeck, C., Sweeney, C., and Takahashi, T.: Sea–air CO_2 fluxes in the Southern Ocean for the period 1990-2009, Biogeosciences, 10, 4037–4054, doi:10.5194/bg-10-4037-2013, http://www.biogeosciences.net/10/4037/2013/, 2013.

Li, H. and Ilyina, T.: Current and Future Decadal Trends in the Oceanic Carbon Uptake Are Dominated by Internal Variability, Geophys. Res. Lett., 45, 916–925, doi:10.1002/2017GL075370, https://doi.org/10.1002/2017GL075370, 2018.

Li, H., Ilyina, T., Müller, W. A., and Sienz, F.: Decadal predictions of the North Atlantic CO_2 uptake, Nature Comm., 7, 11 076 EP –, 2016.

Long, M. C., Lindsay, K., Peacock, S., Moore, J. K., and Doney, S. C.: Twentieth-Century Oceanic Carbon Uptake and Storage in CESM1(BGC), J. Climate, 26, 6775–6800, doi:10.1175/JCLI-D-12-00184.1, http://dx.doi.org/10.1175/JCLI-D-12-00184.1, 2013.

Long, M. C., Deutsch, C., and Ito, T.: Finding forced trends in oceanic oxygen, Global Biogeochem. Cycles, 30, 381–397, doi:10.1002/2015GB005310, http://dx.doi.org/10.1002/2015GB005310, 2015GB005310, 2016.

Lovenduski, N. S., Gruber, N., Doney, S. C., and Lima, I. D.: Enhanced CO_2 outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, Global Biogeochem. Cycles, 21, GB2026, doi:10.1029/2006GB002900, 2007.

Lovenduski, N. S., Long, M. C., Gent, P. R., and Lindsay, K.: Multi-decadal trends in the advection and mixing of natural carbon in the Southern Ocean, Geophys. Res. Lett., 40, 139–142, doi:10.1029/2012GL054483, http://dx.doi.org/10.1029/2012GL054483, 2013.

Lovenduski, N. S., Fay, A. R., and McKinley, G. A.: Observing multidecadal trends in Southern Ocean CO_2 uptake: What can we learn from an ocean model?, Global Biogeochem. Cycles, 29, 416–426, doi:10.1002/2014GB004933, http://dx.doi.org/10.1002/2014GB004933, 2015a.

Lovenduski, N. S., Long, M. C., and Lindsay, K.: Natural variability in the surface ocean carbonate ion concentration, Biogeosciences, 12, 6321–6335, doi:10.5194/bg-12-6321-2015, http://www.biogeosciences.net/12/6321/2015/, 2015b.

Lovenduski, N. S., McKinley, G. A., Fay, A. R., Lindsay, K., and Long, M. C.: Partitioning uncertainty in ocean carbon uptake projections: Internal variability, emission scenario, and model structure, Global Biogeochem. Cycles, 30, 1276–1287, doi:10.1002/2016GB005426, 2016GB005426, 2016.

McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., and Lovenduski, N. S.: Timescales for detection of trends in the ocean carbon sink, Nature, 530, 469–472, 2016.

McKinley, G. A., Fay, A. R., Lovenduski, N. S., and Pilcher, D. J.: Natural Variability and Anthropogenic Trends in the Ocean Carbon Sink, Annu. Rev. Mar. Sci., 9, 125–150, doi:10.1146/annurev-marine-010816-060529, 2017.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., and Stockdale, T.: Decadal prediction: Can it be skillful?, B. Am. Meteorol. Soc., 90, 1467–1485, doi:10.1175/2009BAMS2778.1, https://doi.org/10.1175/2009BAMS2778.1, 2009.

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., van Oldenborgh, G. J., Vecchi, G., and Yeager, S.: Decadal Climate Prediction: An Update from the Trenches, B. Am. Meteorol. Soc., 95, 243–267, doi:10.1175/BAMS-D-12-00241.1, 2014.

Metzl, N., Corbière, A., Reverdin, G., Lenton, A., Takahashi, T., Olsen, A., Johannessen, T., Pierrot, D., Wanninkhof, R., Ólafsdóttir, S. R., Olafsson, J., and Ramonet, M.: Recent acceleration of the sea surface fCO_2 growth rate in the North Atlantic subpolar gyre (1993–2008) revealed by winter observations, Global Biogeochem. Cycles, 24, doi:10.1029/2009GB003658, http://dx.doi.org/10.1029/2009GB003658, 2010.

Moore, J. K. and Braucher, O.: Sedimentary and mineral dust sources of dissolved iron to the world ocean, Biogeosciences, 5, 631–656, http://www.biogeosciences.net/5/631/2008/, 2008.

Moore, J. K., Doney, S. C., and Lindsay, K.: Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, Global Biogeochem. Cycles, 18, GB4028, doi:10.1029/2004GB002220, http://dx.doi.org/10.1029/2004GB002220, 2004.

Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., and Misumi, K.: Marine Ecosystem Dynamics and Biogeochemical Cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 Scenarios, J. Climate, 26, 9291–9312, doi:10.1175/JCLI-D-12-00566.1, http://dx.doi.org/10.1175/JCLI-D-12-00566.1, 2013.

Munro, D. R., Lovenduski, N. S., Takahashi, T., Stephens, B. B., Newberger, T., and Sweeney, C.: Recent evidence for a strengthening CO_2 sink in the Southern Ocean from carbonate system measurements in the Drake Passage (2002–2015), Geophys. Res. Lett., 42, 7623–7630, doi:10.1002/2015GL065194, http://dx.doi.org/10.1002/2015GL065194, 2015GL065194, 2015.

Resplandy, L., Séférian, R., and Bopp, L.: Natural variability of CO_2 and O_2 fluxes: What can we learn from centuries-long climate models simulations?, J. Geophys. Res. Oceans, 120, 384–404, doi:10.1002/2014JC010463, http://dx.doi.org/10.1002/2014JC010463, 2015.

Robson, J. I., Sutton, R. T., and Smith, D. M.: Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s, Geophys. Res. Lett., 39, doi:10.1029/2012GL053370, http://dx.doi.org/10.1029/2012GL053370, l19713, 2012.

Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., Landschützer, P., Metzl, N., Nakaoka, S., Olsen, A., Park, G.-H., Peylin, P., Rodgers, K. B., Sasse, T. P., Schuster, U., Shutler, J. D., Valsala, V., Wanninkhof, R., and Zeng, J.: Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean pCO_2 Mapping intercomparison (SOCOM), Biogeosciences, 12, 7251–7278, doi:10.5194/bg-12-7251-2015, https://www.biogeosciences.net/12/7251/2015/, 2015.

Séférian, R., Berthet, S., and Chevallier, M.: Assessing the Decadal Predictability of Land and Ocean Carbon Uptake, Geophys. Res. Lett., 45, 2455–2466, doi:10.1002/2017GL076092, 2018.

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., and Murphy, J. M.: Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model, Science, 317, 796–799, doi:10.1126/science.1139540, http://science.sciencemag.org/content/317/5839/796, 2007.

Thomas, H., Friederike Prowe, A. E., Lima, I. D., Doney, S. C., Wanninkhof, R., Greatbatch, R. J., Schuster, U., and Corbiére, A.: Changes in the North Atlantic Oscillation influence CO_2 uptake in the North Atlantic over the past 2 decades, Global Biogeochem. Cycles, 22, GB4027, doi:10.1029/2007GB003167, http://dx.doi.org/10.1029/2007GB003167, 2008.

Ullman, D. J., McKinley, G. A., Bennington, V., and Dutkiewicz, S.: Trends in the North Atlantic carbon sink: 1992–2006, Global Biogeochem. Cycles, 23, doi:10.1029/2008GB003383, http://dx.doi.org/10.1029/2008GB003383, 2009.

Verdy, A., Dutkiewicz, S., Follows, M. J., Marshall, J., and Czaja, A.: Carbon dioxide and oxygen fluxes in the Southern Ocean: Mechanisms of interannual variability, Global Biogeochem. Cycles, 21, GB2020, doi:10.1029/2006GB002916, http://dx.doi.org/10.1029/2006GB002916, 2007.

Wang, S. and Moore, J. K.: Variability of primary production and air-sea CO_2 flux in the Southern Ocean, Global Biogeochem. Cycles, 26, GB1008, doi:10.1029/2010GB003981, 2012.

Wanninkhof, R., Park, G.-H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., Gruber, N., Doney, S. C., McKinley, G. A., Lenton, A., Le Quéré, C., Heinze, C., Schwinger, J., Graven, H., and Khatiwala, S.: Global ocean carbon uptake: magnitude, variability and trends, Biogeosciences, 10, 1983–2000, doi:10.5194/bg-10-1983-2013, https://www.biogeosciences.net/10/1983/2013/, 2013.

Wetzel, P., Winguth, A., and Maier-Reimer, E.: Sea-to-air CO_2 flux from 1948 to 2003: A model study, Global Biogeochem. Cycles, 19, doi:10.1029/2004GB002339, http://dx.doi.org/10.1029/2004GB002339, 2005.

Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., and Teng, H.: A Decadal Prediction Case Study: Late Twentieth-Century North Atlantic Ocean Heat Content, J. Climate, 25, 5173–5189, doi:10.1175/JCLI-D-11-00595.1, 2012.

Yeager, S. G. and Robson, J. I.: Recent Progress in Understanding and Predicting Atlantic Decadal Climate Variability, Curr. Clim. Chang. Rep., 3, 112–127, doi:10.1007/s40641-017-0064-z, http://dx.doi.org/10.1007/s40641-017-0064-z, 2017.

Yeager, S. G., Karspeck, A. R., and Danabasoglu, G.: Predicted slowdown in the rate of Atlantic sea ice loss, Geophys. Res. Lett., 42, 10,704–10,713, doi:10.1002/2015GL065364, 2015GL065364, 2015.

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., and Lovenduski, N. S.: Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model, B. Am. Meteorol. Soc., 99, 1867–1886, doi:10.1175/BAMS-D-17-0098.1, https://doi.org/10.1175/BAMS-D-17-0098.1, 2018.

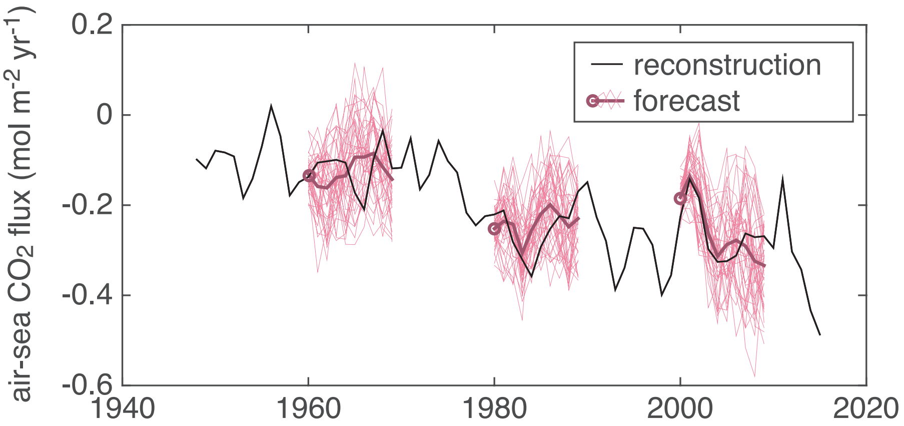

Figure 1. Annual mean air-sea CO_2 flux (mol m^{-2} yr^{-1}) in the South Pacific subtropical permanently stratified biome for the (black) model reconstruction, and (pink) CESM-DPLE decadal forecasts initiated in 1960, 1980, and 2000 (other forecasts omitted for visual clarity). Thick magenta line represents the ensemble-mean forecast; open circles show the ensemble mean in forecast year 1. Positive fluxes denote ocean outgassing. Forecasts have been drift-corrected and adjusted to match the reconstruction climatological mean for ease of visual comparison.

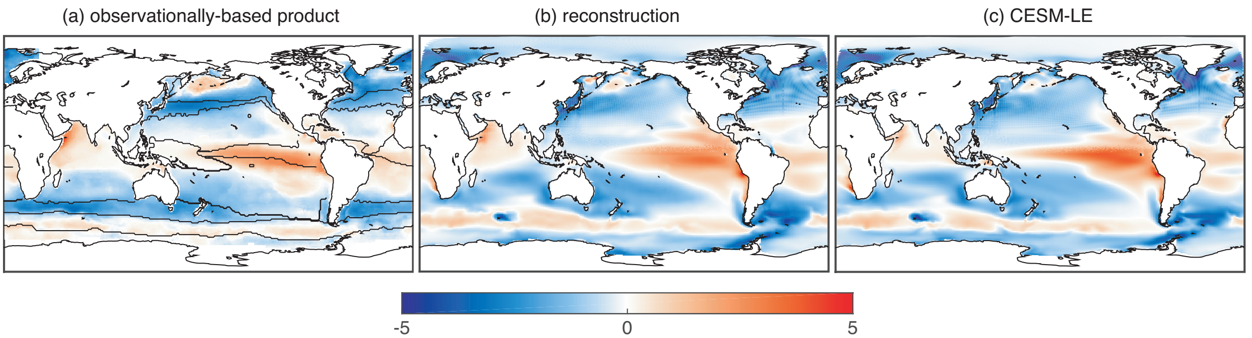

Figure 2. Annual-mean air-sea CO_2 flux (mol m^{-2} yr^{-1}) over the period 1982-2015 as estimated by (a) the Landschützer et al. (2016) observationally-based product, ~~and~~ (b) the model reconstruction, and (c) the CESM-LE. Positive fluxes denote ocean outgassing, and black contours in (a) show biome boundaries.

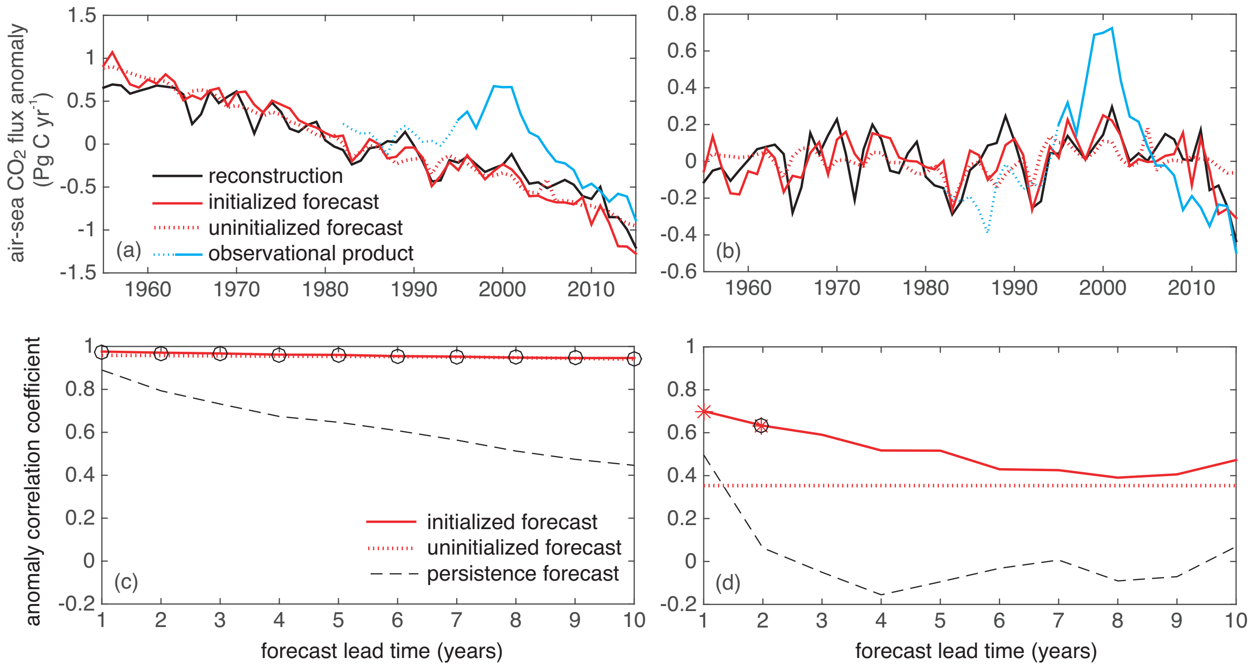

Figure 3. (a) Temporal evolution of the globally-integrated air-sea CO_2 flux anomaly, as estimated by the (black) reconstruction, (red) CESM-DPLE initialized forecast, (red dotted) CESM-LE uninitialized forecast, and (blue) Landschützer et al. (2016) observationally-based product. The CESM-DPLE time series is the drift-corrected, ensemble mean forecast anomalies over lead year 1, and the reconstruction, uninitialized forecast, and observational product have been transformed to anomalies by subtracting their respective climatological means. Observations prior to 1995 are dotted, due to lower observation density. Positive anomalies indicate anomalous ocean outgassing. (b) Same as (a), but with long-term linear trends removed from each time series. (c) Predictability of globally integrated CO_2 flux as a function of lead time, as indicated by the correlation coefficient of CO_2 flux anomalies from the (red) CESM-DPLE initialized forecast, and (red dotted) CESM-LE uninitialized forecast with the reconstruction. Black dashed line shows indicates the correlation coefficient of the persistence forecast as a function of lead time. Red asterisks (black circles) on the initialized forecast indicate ~~statistically significant~~ predictability that is statistically different from the initialized (persistence) forecast at the 95% level using a z test. (d) Same as (c), but with linear trends removed from each time series.

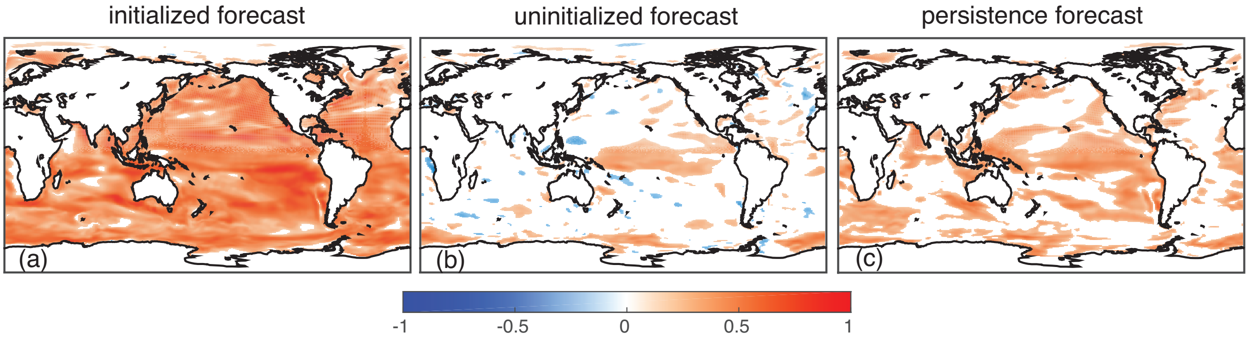

Figure 4. Predictability of air-sea CO_2 flux, as indicated by the correlation coefficient of detrended, air-sea CO_2 flux anomalies from the (a) CESM-DPLE initialized forecast lead year 1 with the reconstruction, and (b) CESM-LE uninitialized forecast with the reconstruction. (c) Correlation coefficient of the persistence forecast for lead year 1. Correlation coefficients that are not statistically significant at the 95% level using a *t* test are assigned a value of zero.

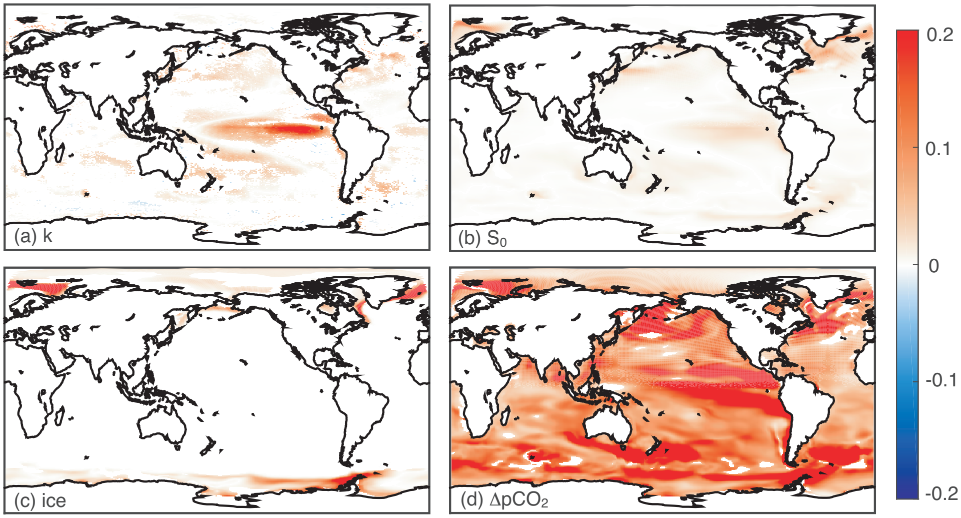

Figure 5. Drivers of predictability in air-sea CO_2 flux during forecast year 1, as indicated by the predictability of the (a) gas-exchange coefficient, (b) solubility, (c) sea ice fraction, and (d) ΔpCO_2, scaled to CO_2 flux units (mol m^{-2} yr^{-1}). Correlation coefficients that are not statistically significant at the 95% level using a t test are assigned a value of zero.

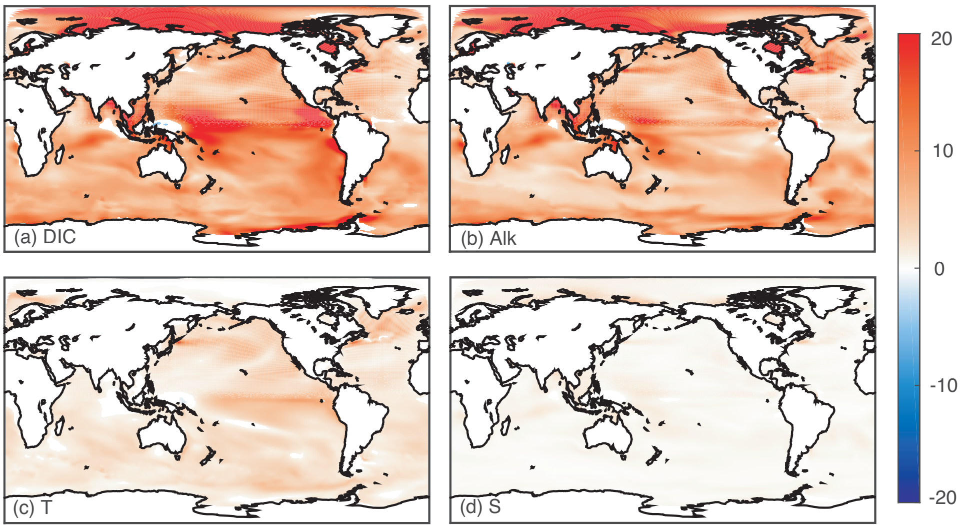

Figure 6. Drivers of predictability in surface ocean pCO_2 during forecast year 1, as indicated by the predictability of surface ocean (a) DIC, (b) Alk, (c) temperature, and (d) salinity, scaled to pCO_2 units (μatm). Correlation coefficients that are not statistically significant at the 95% level using a t test are assigned a value of zero.

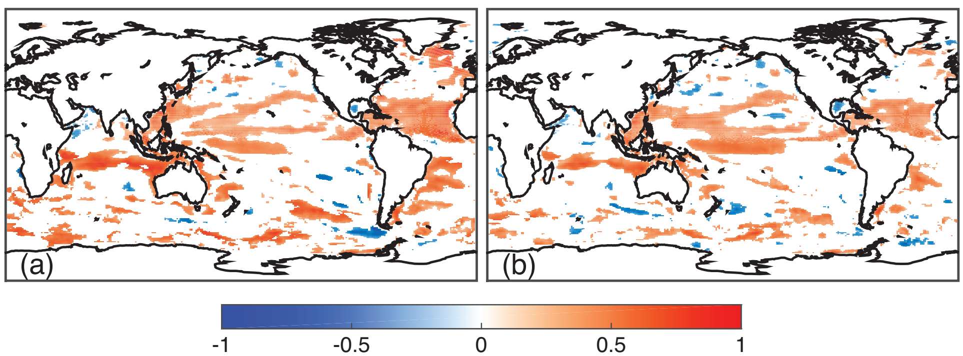

Figure 7. Air-sea CO_2 flux predictive skill, as indicated by the correlation coefficient of air-sea CO_2 flux (a) anomalies, and (b) linearly detrended anomalies from the CESM-DPLE initialized forecast in year 1 with the Landschützer et al. (2016) observational product over 1995-2015. Correlation coefficients that are not statistically significant at the 95% level using a t test are assigned a value of zero.

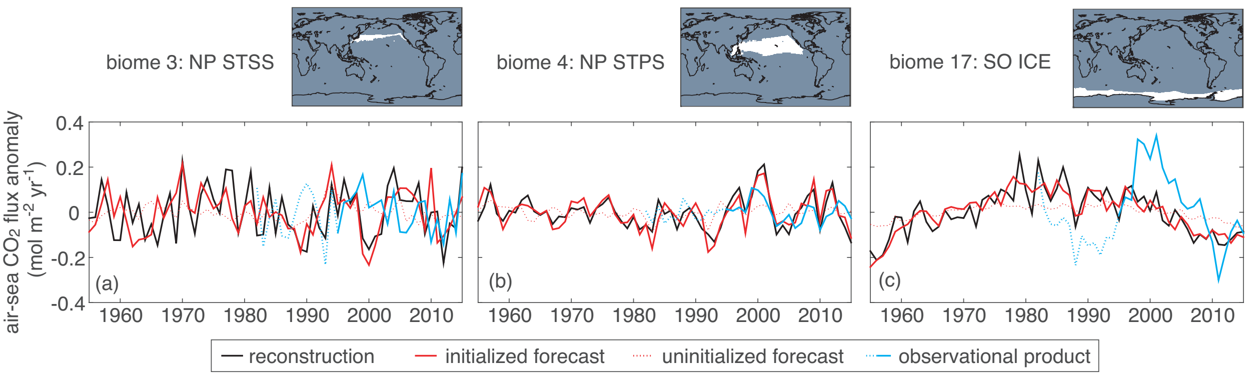

Figure 8. Temporal evolution of the biome-averaged air-sea CO_2 flux anomalies in the (a) NP STSS, (b) NP STPS, and (c) SO ICE biomes (mol m^{-2} yr^{-1}). The following time series are plotted: (black) reconstruction, (red) CESM-DPLE initialized forecast, (red dotted) CESM-LE uninitialized forecast, and (blue) Landschützer et al. (2016) observationally-based product. The CESM-DPLE time series is the linearly detrended, drift-corrected, ensemble mean forecast anomalies in year 1; the reconstruction, CESM-LE ensemble mean, and observed time-series have been transformed to anomalies by removing the linear trend. Observations prior to 1995 are dotted, due to lower observation density. Positive anomalies indicate anomalous ocean outgassing.

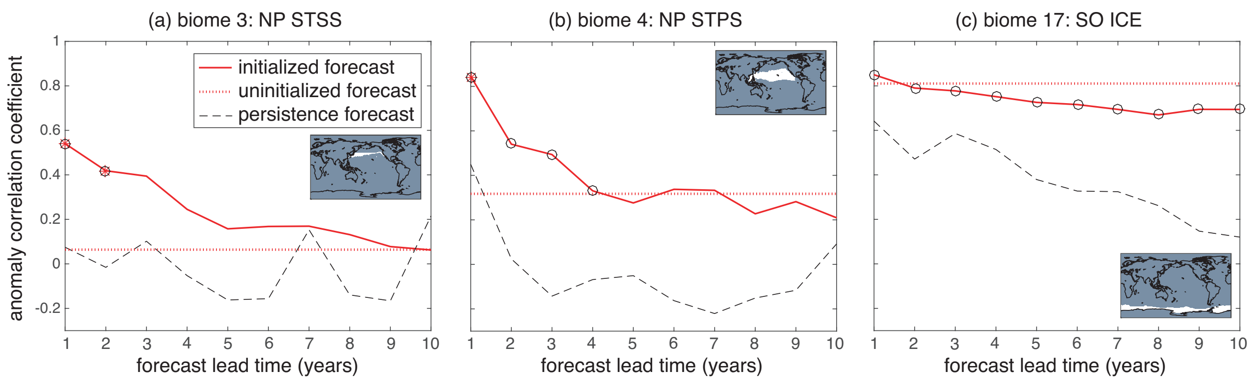

Figure 9. Predictability of biome-average CO_2 flux as a function of lead time in the (a) NP STSS, (b) NP STPS, and (c) SO ICE biomes, as indicated by the correlation coefficient of detrended CO_2 flux anomalies from the (red) CESM-DPLE initialized forecast, and (red dotted) CESM-LE uninitialized forecast with the reconstruction. Black dashed line shows the correlation coefficient of the persistence forecast as a function of lead time. Red asterisks (black circles) on the initialized forecast indicate ~~statistically significant~~ predictability that is statistically different from the uninitialized (persistence) forecast at the 95% level using a z test.

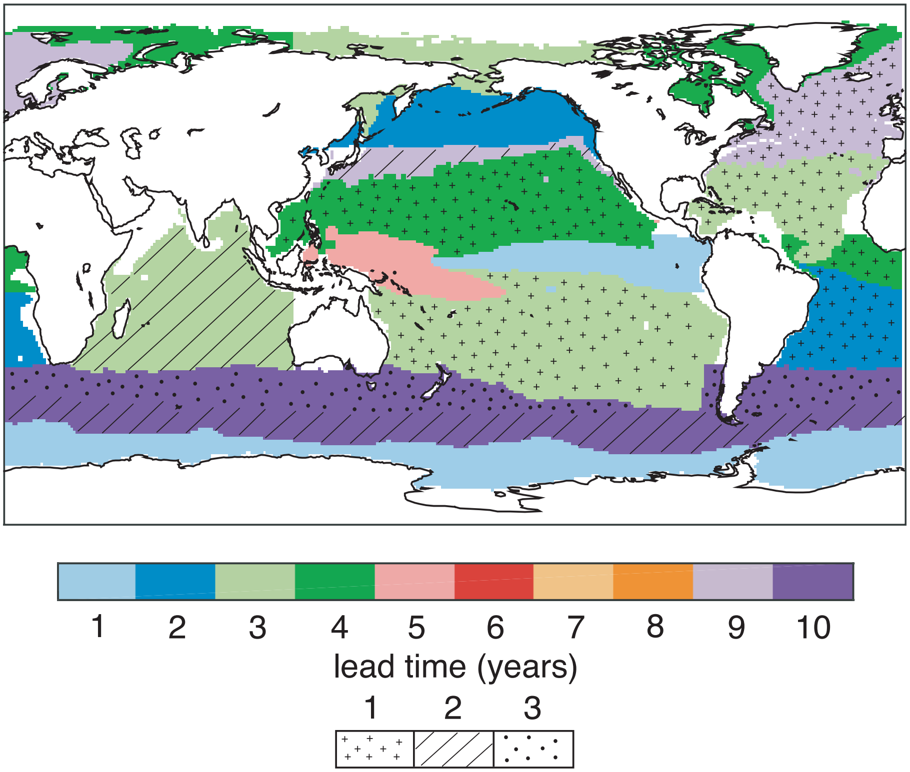

Figure 10. For each biome, the maximum forecast lead time (years) in which the initialized CESM-DPLE CO_2 flux forecast has both higher predictability than the uninitialized CESM-LE forecast and a higher correlation coefficient than the persistence forecast. Hatching shows the maximum forecast lead time while accounting for statistical separation of correlation coefficients at the 95% level using a z test.

Table 1. ~~Predictability and predictive skill of biome-averaged~~ Biome-averaged air-sea CO_2 flux ~~, as indicated by correlation coefficients of flux anomalies for forecast year 1.~~ statistics.

Biome Name	Biome Acronym	Biome Number	Initialized Forecast[a]	Uninitialized Forecast[b]	Persistence Forecast[c]	Forecast Skill[d]	Maximum Lead Time[e]
North Pacific Ice	NP ICE	1	0.29	-0.22	0.25	0.43	3 (0)
North Pacific Subpolar Seasonally Stratified	NP SPSS	2	0.54	-0.12	0.47	-0.45	2 (0)
North Pacific Subtropical Seasonally Stratified	NP STSS	3	**0.54**	0.06	0.07	-0.28	9 (2)
North Pacific Subtropical Permanently Stratified	NP STPS	4	**0.85**	0.32	0.45	0.60	4 (1)
West Pacific Equatorial	PEQU-W	5	0.73	0.31	0.52	0.66	5 (0)
East Pacific Equatorial	PEQU-E	6	0.64	0.35	0.50	0.53	1 (0)
South Pacific Subtropical Permanently Stratified	SP STPS	7	**0.81**	0.33	0.50	0.19	3 (1)
North Atlantic Ice	NA ICE	8	0.49	0.07	0.24	0.36	4 (0)
North Atlantic Subpolar Seasonally Stratified	NA SPSS	9	**0.55**	0.10	0.17	-0.28	9 (1)
North Atlantic Subtropical Seasonally Stratified	NA STSS	10	**0.53**	-0.08	0.01	-0.10	9 (1)
North Atlantic Subtropical Permanently Stratified	NA STPS	11	**0.72**	0.35	0.18	0.56	3 (1)
Atlantic Equatorial	AEQU	12	**0.55**	0.17	0.27	-0.04	4 (1)
South Atlantic Subtropical Permanently Stratified	SA STPS	13	**0.60**	0.09	0.16	0.49	2 (1)
Indian Ocean Subtropical Permanently Stratified	IND STPS	14	**0.16**	-0.11	0.05	0.31	3 (2)
Southern Ocean Subtropical Seasonally Stratified	SO STSS	15	**0.70**	-0.02	0.20	0.26	10 (3)
Southern Ocean Subpolar Seasonally Stratified	SO SPSS	16	0.47	0.08	0.32	0.47	10 (2)
Southern Ocean Ice	SO ICE	17	0.85	0.81	0.64	0.60	1 (0)

[a]Correlation of CO_2 flux anomalies from the CESM-DPLE initialized forecast in lead year 1 with the reconstruction. Boldface indicates correlation coefficient is statistically different from both the uninitialized and persistence forecasts at the 95% level using a z test.

[b]Correlation of CO_2 flux anomalies from the CESM-LE uninitialized forecast with the reconstruction.

[c]Autocorrelation of the persistence forecast at lead year 1.

[d]Correlation of CO_2 flux anomalies from the CESM-DPLE initialized forecast in lead year 1 with the observational product over 1995-2015.

[e]The maximum forecast lead time (years) in which the CESM-DPLE initialized forecast has both higher predictability than the uninitialized CESM-LE forecast and a higher correlation coefficient than the persistence forecast. Lead times in parenthesis account for statistical separation in correlation coefficients at the 95% level using a z test.

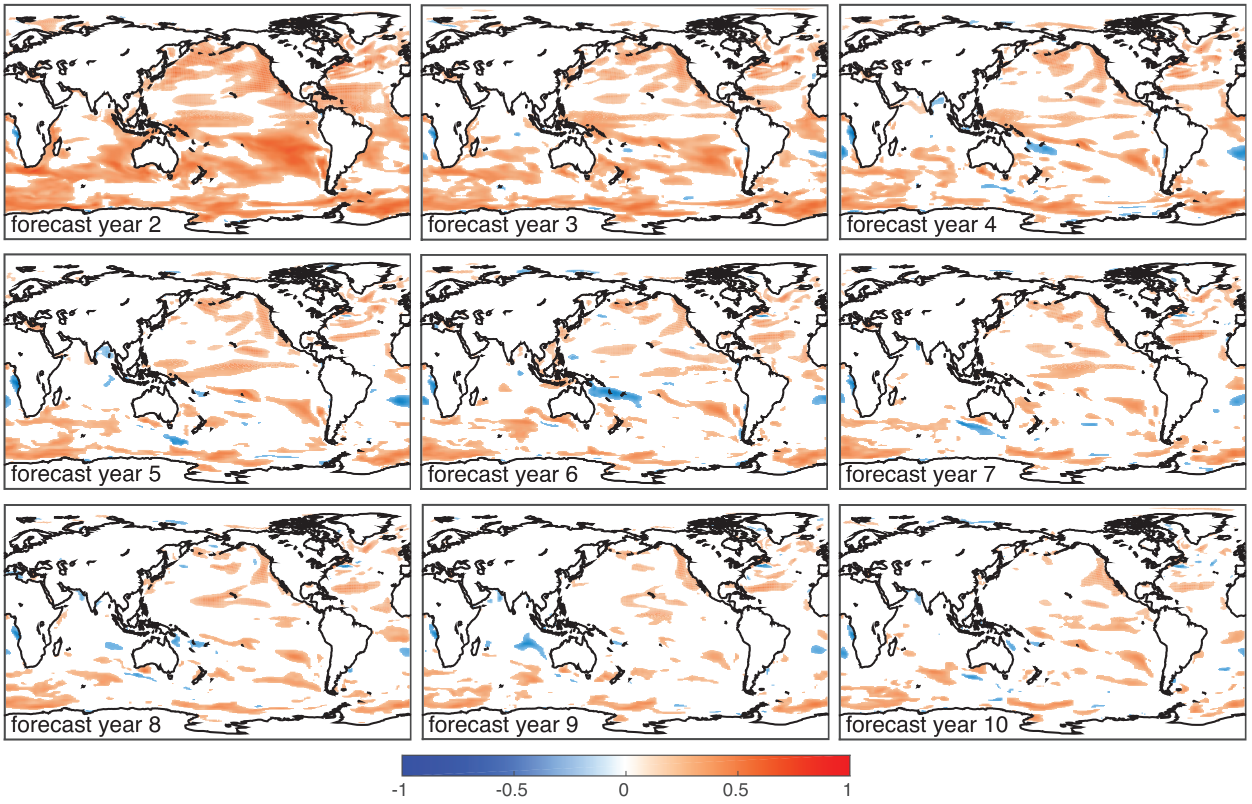

Figure S1. Predictability of air-sea CO_2 flux, as indicated by the correlation coefficient of detrended, air-sea CO_2 flux anomalies from the CESM-DPLE initialized forecast for a range of lead years with the reconstruction. Correlation coefficients that are not statistically significant at the 95% level using a t test are assigned a value of zero.