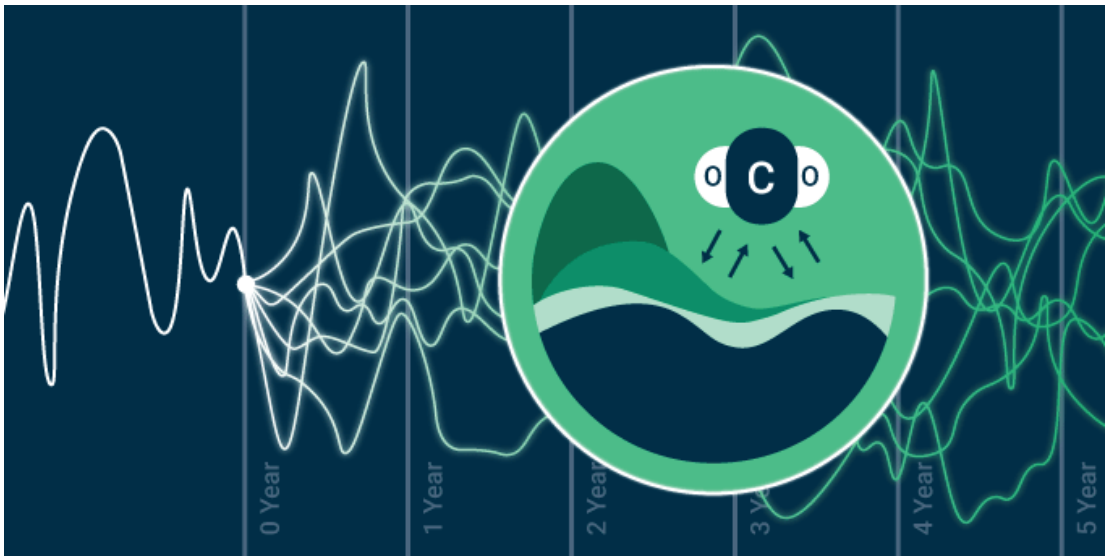
Max-Planck-Institut für Meteorologie

MAX PLANCK GESELLSCHAFT

International Max Planck Research School on Earth System Modelling

Internal Variability and Potential Predictability of the Global Carbon Cycle in a Perfect-Model Framework



Aaron Spring

Hamburg 2021

Berichte zur Erdsystemforschung
Reports on Earth System Science

243
2021

Hinweis

Die Berichte zur Erdsystemforschung werden vom Max-Planck-Institut für Meteorologie in Hamburg in unregelmäßiger Abfolge herausgegeben.

Sie enthalten wissenschaftliche und technische Beiträge, inklusive Dissertationen.

Die Beiträge geben nicht notwendigerweise die Auffassung des Instituts wieder.

Die "Berichte zur Erdsystemforschung" führen die vorherigen Reihen "Reports" und "Examensarbeiten" weiter.

Notice

The Reports on Earth System Science are published by the Max Planck Institute for Meteorology in Hamburg. They appear in irregular intervals.

They contain scientific and technical contributions, including Ph. D. theses.

The Reports do not necessarily reflect the opinion of the Institute.

The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.

Anschrift / Address

Max-Planck-Institut für Meteorologie
Bundesstrasse 53
20146 Hamburg
Deutschland

Tel./Phone: +49 (0)40 4 11 73 - 0
Fax: +49 (0)40 4 11 73 - 298

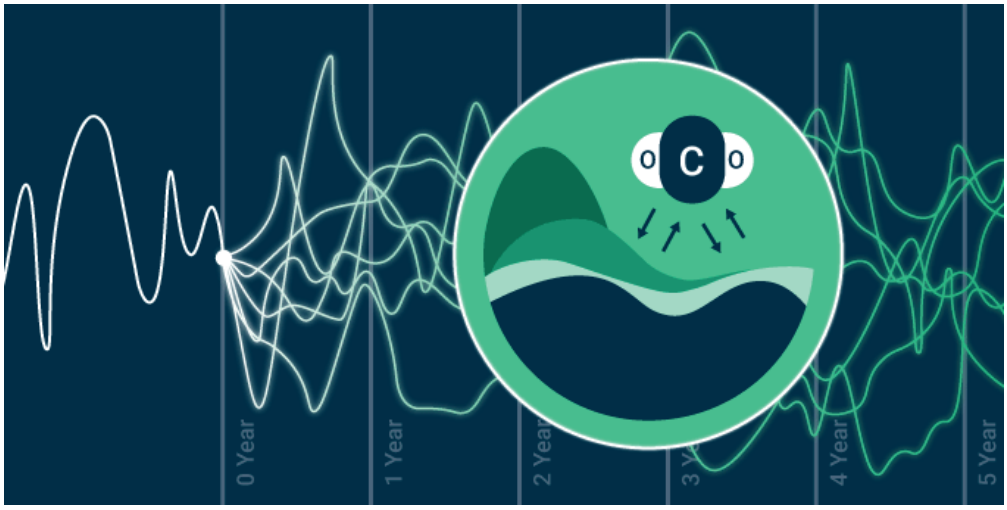
name.surname@mpimet.mpg.de
www.mpimet.mpg.de

Layout

Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

Internal Variability and Potential Predictability of the Global Carbon Cycle in a Perfect-Model Framework



Aaron Spring

Hamburg 2021

Aaron Spring
aus Frankfurt / Main, Deutschland

Max-Planck-Institut für Meteorologie
The International Max Planck Research School on Earth System Modelling
(IMPRS-ESM)
Bundesstrasse 53
20146 Hamburg

Tag der Disputation: 27. April 2021

Folgende Gutachter empfehlen die Annahme der Dissertation:
Dr. Tatiana Ilyina
Prof. Dr. Jochem Marotzke

Vorsitzender des Promotionsausschusses:
Prof. Dr. Dirk Gajewski

Dekan der MIN-Fakultät:
Prof. Dr. Heinrich Graener

Titelgrafik von Yvonne Schrader / Max-Planck-Institut für Meteorologie

Berichte zur Erdsystemforschung / Max-Planck-Institut für Meteorologie 243
Reports on Earth System Science / Max Planck Institute for Meteorology 2021

ISSN 1614-1199

Aaron Spring

Internal Variability and Potential Predictability of the Global Carbon Cycle in a Perfect-Model Framework

ABSTRACT

Reductions of anthropogenic CO_2 emissions are required to stabilize and reduce atmospheric CO_2 concentrations. But variations in the land and ocean carbon sinks, which are triggered by inherent climate variability, disguise CO_2 emission reductions in global atmospheric CO_2 in the near-term. Therefore, an independent verification of CO_2 emission reductions in atmospheric CO_2 needs to take internal variability into account. Initialized predictions of the evolution of atmospheric CO_2 have the potential to constrain this internal variability and enable an ahead-of-time estimate under different proposed CO_2 emission reductions pathways. To that end, this dissertation focuses on near-term variability in the global carbon cycle, and how initialized CO_2 predictions can guide policy-makers onto a pathway of limiting global warming below 2°C.

First, I motivate initialized CO_2 predictions by showcasing the large envelope of internal variability of the global carbon cycle, which initialized prediction can partly constrain. I deduce time-scales in the near-term on which CO_2 emission reductions cause atmospheric CO_2 growth to decelerate. I find that CO_2 emission reductions compatible with the Paris Agreement only cause a deceleration of atmospheric CO_2 after a decade. Particularly on the five-year scale, on which the global stocktake assesses the efficacy of CO_2 emission reductions, internal variability disguises certainty in causation. This study shows how expectations about the near-term efficacy of CO_2 emission reductions need to consider internal variability.

Then, I evaluate initialized predictive skill of prognostic atmospheric CO_2 and its drivers, which were previously unknown, by performing initialized ensemble Earth System Model (ESM) simulations in a perfect-model predictability framework. I show that internal variability of prognostic global atmospheric CO_2 is predictable up to three years in advance. ESM-based predictions are feasible and surpass a statistical regression forecast in scope and accuracy.

Finally, I test the realism of state-of-the-art carbon cycle predictions, which initialize the carbon cycle only indirectly via the reconstruction of the physical climate dynamics. In a perfect-model reconstruction framework, I test the commonly stated assumption that direct carbon cycle reconstruction improves its predictability. I find that indirect reconstruction tracks the target reasonably well. While direct reconstruction improves global carbon cycle initial conditions and predictability slightly, a mean bias reduction achieves similar improvements. This adds confidence to the current practice of indirect carbon cycle reconstruction and refutes the need for direct carbon cycle reconstruction.

My results demonstrate that internal variability of global carbon cycle can disguise CO_2 emission reductions and thereby mislead the evaluation on mitigation efficacy. As a partial solution, initialized predictions can constrain this internal variability for up to three years into the future, which can guide policy-makers navigating on the path towards well below 2°C global warming.

ZUSAMMENFASSUNG

Um die Konzentration von CO_2 in der Atmosphäre zu stabilisieren und zu reduzieren, müssen die menschengemachten CO_2-Emissionen gesenkt werden. Schwankung der Land- und Ozeankohlenstoffsenken, die durch die inhärente Klimavariabilität ausgelöst werden, verdecken allerdings kurzfristig die CO_2-Emissionsreduktionen im globalen atmosphärischen CO_2. Daher muss für eine unabhängige Verifizierung der CO_2-Emissionsminderungen im atmosphärischen CO_2 die interne Variabilität berücksichtigt werden. Initialisierte Vorhersagen von atmosphärischem CO_2 haben das Potenzial, diese interne Variabilität einzugrenzen und damit eine vorausschauende Abschätzung der Wirkung von verschiedenen Szenarien der CO_2-Emissionsreduktion zu ermöglichen. Daher konzentriert sich diese Dissertation auf die kurzfristige Variabilität im globalen Kohlenstoffkreislauf und darauf, wie initialisierte CO_2-Vorhersagen politischen Entscheidungsträgern helfen können, einen Kurs zur Begrenzung der globalen Erwärmung auf unter 2°C zu finden und zu halten.

Zunächst rechtfertige ich initialisierte CO_2-Vorhersagen, indem ich die große Bandbreite an interner Variabilität des globalen Kohlenstoffkreislaufs aufzeige, die durch initialisierte Vorhersagen zum Teil eingegrenzt werden kann. Ich leite Zeitskalen in der nahen Zukunft ab, auf denen CO_2-Emissionsreduktionen eine Verlangsamung des atmosphärischen CO_2 verursachen. Eine Erkenntnis ist, dass CO_2-Emissionsminderungen, die mit dem Pariser Klimaabkommen kompatibel sind, erst nach einem Jahrzehnt das Abbremsen des atmosphärisches CO_2 Wachstums mit Sicherheit verursachen. Insbesondere im Zeitraum von fünf Jahren, in dem die Wirksamkeit von CO_2-Emissionsreduktionen in Folge des Pariser Klimaabkommens bewertet wird, kann die Kausalität wegen der internen Variabilität nicht klar nachgewiesen werden. Diese Studie zeigt, dass Erwartungen an die kurzfristige Wirksamkeit von CO_2-Emissionsminderungen die interne Variabilität berücksichtigen müssen.

Anschließend untersuche ich initialisierte Vorhersagbarkeit von atmosphärischem CO_2 und dessen Treibern, indem ich Simulationen in einem idealisierten, initialisierten Erdsystemmodell (ESM) durchführe. Darin zeige ich, dass die interne Variabilität des prognostizierten globalen atmosphärischen CO_2 bis zu drei Jahre im Voraus vorhersagbar ist. ESM-basierte Vorhersagen sind praktikabel und übertreffen eine statistische Regressionsvorhersage in Reichweite und Genauigkeit.

Schließlich teste ich die Realitätsnähe von neuartigen Kohlenstoffkreislaufvorhersagen, die den Kohlenstoffkreislauf nur indirekt über die Rekonstruktion der physikalischen Klimadynamik initialisieren.

In idealisierten Rekonstruktionen teste ich die häufig geäußerte Annahme, dass die direkte Rekonstruktion des Kohlenstoffkreislaufs dessen Vorhersagbarkeit verbessert. Ich kann zeigen, dass die indirekte Rekonstruktion die eigentlichen Startbedingungen recht gut wiederherstellt. Während die direkte Rekonstruktion die Startbedingungen und die Vorhersagbarkeit des globalen Kohlenstoffkreislaufs leicht verbessert, erzielt eine Angleichung der mittleren Abweichung ähnliche Verbesserungen. Dies stärkt die Zuversicht in die derzeitige Praxis der indirekten Rekonstruktion des Kohlenstoffkreislaufes und widerlegt die Notwendigkeit von dessen direkter Rekonstruktion.

Meine Ergebnisse zeigen, dass die interne Variabilität des globalen Kohlenstoffkreislaufs die globalen CO_2-Emissionsreduktionen verschleiern und damit die Bewertung von dessen Wirksamkeit in die Irre führen kann. Als Teillösung können initialisierte Vorhersagen diese interne Variabilität für bis zu drei Jahre in die Zukunft begrenzen. Diese Vorhersagen können politische Entscheidungsträger leiten, die globale Erwärmung auf deutlich unter 2°C zu begrenzen.

PUBLICATIONS RELATED TO THIS DISSERTATION

As corresponding author:

- Spring, A., & Ilyina, T. [2020]. Predictability Horizons in the Global Carbon Cycle Inferred From a Perfect-Model Framework. *Geophysical Research Letters*, *47*[9], e2019GL085311. doi:10/ggtbv2
- Spring, A., Ilyina, T., & Marotzke, J. [2020]. Inherent uncertainty disguises attribution of reduced atmospheric CO_2 growth to CO_2 emission reductions for up to a decade. *Environmental Research Letters*, *15*[11], 114058. doi:10/ghmjp8
- Spring, A., Dunkl, I., Li, H., Brovkin, V., & Ilyina, T. [2021]. Trivial improvements of predictive skill due to direct reconstruction of the global carbon cycle. *Earth System Dynamics Discussions*, 1–36. doi:10/gh3tn3
- Brady, R. X., & Spring, A. [2021]. Climpred: Verification of weather and climate forecasts. *Journal of Open Source Software*, *6*[59], 2781. doi:10/gh9646

As contributing author:

- Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., … Yeager, S. [2020]. Current and Emerging Developments in Subseasonal to Decadal Prediction. *Bulletin of the American Meteorological Society*, *101*[6], 869–896. doi:10/ggvcqv
- Ilyina, T., Li, H., Spring, A., Müller, W. A., Bopp, L., Chikamoto, M. O., … Yeager, S. [2021]. Predictable Variations of the Carbon Sinks and Atmospheric CO2 Growth in a Multi-Model Framework. *Geophysical Research Letters*, *48*[6], e2020GL090695. doi:10/ghsn7h

CONTENTS

ESSAY

Forecasts possess no intrinsic value. They acquire value through their ability to influence the decisions made by users of the forecasts.

— Murphy [1993]

E.1 INTRODUCTION

This dissertation addresses near-term variability of two-to-fifteen years in the global carbon cycle, and how initialized predictions can constrain the near-term evolution of atmospheric CO_2 concentrations. I explore the variability and predictability of changes in atmospheric CO_2 concentrations over short time scales, which had previously not been investigated in an Earth System Model setting. I pose timely questions in support of the verification of CO_2 emission reductions in atmospheric CO_2 and demonstrate how carbon cycle predictions could inform decision-makers about the efficacy of CO_2 emission mitigation strategies.

This unifying essay characterizes how variability, predictability and reconstruction of the global carbon cycle matter for carbon-related climate policy. I furthermore introduce the statistical framework of forecast verification and the "perfect-model framework" as the modeling approach used. Then, I summarize the three[1] research papers attached as chapters I, II and III. The first paper Spring et al. [2020] quantifies the substantial variations due to climate variability in atmospheric CO_2 and hence the use-case of initialized CO_2 predictions. The second paper Spring and Ilyina [2020] investigates the limits and drivers of atmospheric CO_2 predictability. In a further step towards providing carbon cycle predictions, the third paper Spring et al. [2021] explores the initial conditions of carbon cycle predictions and whether reconstructing the carbon cycle directly improves its predictive skill. Finally, I discuss the importance of my results for future studies and applications of carbon cycle predictability.

I guide the reader through this essay about carbon cycle predictability using two recurring themes. The first addresses "where the carbon goes" [Marotzke et al., 2017] by characterizing how the global carbon

1 The fourth short paper Brady and Spring [2021] introduces a python software that streamlines and standardizes forecast verification in chapter IV. Because it is only short and descriptive paper, it is not summarized in this essay, but motivated in subsection E.2.2.

cycle shapes near-term atmospheric CO_2. The second, "the signal and the noise" [Silver, 2012], characterizes variability and predictability.

E.1.1 *Global Carbon Cycle*

the CO_2 problem

Anthropogenic CO_2 emissions, which mainly stem from fossil-fuel combustion and land-use changes, increase the concentration of CO_2 in the atmosphere, which has been proven to enhance the Earth System's radiative forcing, thus warming the Earth's climate [Stocker et al., 2013]. However, not all anthropogenic CO_2 emissions remain in the atmosphere, as atmospheric CO_2 is modulated by the uptake of carbon by land and ocean. The exchanges of carbon within and between these components, which additionally includes geological storage, is known as the global carbon cycle [Archer, 2010]. The combustion of fossil fuels perturbs the natural state of the carbon cycle by adding carbon from the geological reservoir, which has been secluded for millennia, to the atmosphere. As a result, the perturbed system re-equilibrates the additional source of carbon through air-land and air-sea CO_2 exchange. Thus, the land and ocean reservoirs take up parts of the CO_2 emissions, reducing the increasing CO_2 burden of the atmosphere.

carbon sinks

Between 1850 and 2019, the Global Carbon Budget estimates that 40% of anthropogenic fossil-fuel CO_2 emissions have remained in the atmosphere (the so-called airborne fraction of CO_2 emissions), while the terrestrial and oceanic carbon sinks absorbed 32% and 25%, respectively [Friedlingstein et al., 2020][2].

ocean carbon sink processes

The air-sea CO_2 flux is driven by the difference of the atmospheric partial pressure[3] of CO_2 ($pCO_{2,atm}$) and the oceanic partial pressure[4] of CO_2 ($pCO_{2,ocean}$) at the ocean's surface [Heinze et al., 2015]. Carbon in the ocean is redistributed by the carbon pumps: The organic carbon pump describes the uptake of carbon in surface waters by phytoplankton photosynthesis, the sinking of particulate organic carbon to deeper ocean layers, its remineralization to nutrients and dissolved or-

2 There remains a cumulative budget imbalance of 3% representing imperfect data and imperfect understanding of the contemporary global carbon cycle. Cumulative land use-change CO_2 emissions are broadly equal to the cumulative land carbon sink [Friedlingstein et al., 2020].

3 There are different measures of atmospheric composition in atmospheric chemistry describing concentration [Jacob, 1999]: The partial pressure pCO_2 in a mixture of gases of total pressure P is defined as the pressure that would be exerted by the CO_2 molecules if all the other gases were removed. Dalton's law relates $pCO_2 = XCO_2 \cdot P$. The mixing ratio XCO_2 describes the volumetric ratio of CO_2 molecules per one million molecules [ppm(v)] in air. In this thesis, I use concentration and mixing ratio synonymous for atmospheric CO_2.

4 The partial pressure of oceanic CO_2 ($pCO_{2,ocean}$), i. e. CO_2 in sea-water, above a liquid in thermodynamic equilibrium is proportional to the concentration in the liquid by its solubility S_{CO_2}: $pCO^{eq}_{2,ocean} = S_{CO_2} \cdot pCO_{2,atm}$. Air-sea CO_2 flux aims to equilibrate the difference between $pCO^{eq}_{2,ocean}$ and $pCO_{2,ocean}$ [Sarmiento and Gruber, 2006].

ganic carbon. Its subsequent wind- and bathymetry-driven upwelling sustains another phytoplankton bloom at the surface. Similarly, the carbonate pump transports calcium carbonate ($CaCO_3$) shells, produced by skeleton-building marine organisms while releasing CO_2, from the surface to the deep ocean and therefore counteracts the carbon transfer of the biological pump. The solubility pump is driven by the solubility of CO_2 in oceanic water, which is inversely proportional to ocean temperature. The thermohaline circulation, which re-distributes heat and carbon among the global oceans, is driven by deep-water formation in the cold and therefore high pCO_2 high-latitudes. High pCO_2 waters are upwelled from the deeper ocean by atmospheric wind forcing and bathymetry; e. g. in the tropical warm surface oceans, where these waters release CO_2 to the atmosphere.

land carbon sink processes

Carbon is taken up by the the land reservoir via photosynthesis, which is sensitive to temperature, water and nutrient availability [Schimel, 1995]. Terrestrial carbon is static in the horizontal and passes vertically between the vegetation, litter and soil carbon pools. Anthropogenic emissions enhance plant growth by CO_2 fertilisation [Winkler, 2020] and reactive nitrogen deposition [Reay et al., 2008]. Carbon is released by ecosystem respiration, fire, erosion by rivers and land-use changes [Schimel, 1995; Hong et al., 2021].

long-term CO_2 residence in atmosphere

These processes of the land and ocean carbon uptake maintain a stable airborne fraction of cumulative atmospheric CO_2 emissions for decades and centuries [Canadell et al., 2007; Archer et al., 2009; Friedlingstein et al., 2020]. Only on time-scales of thousands to tens of thousands of years, do geological weathering of carbonates on land or re-dissolution of carbonates in the ocean sediment provide a negative feedback and draw down atmospheric CO_2 [Archer et al., 2009; Heinze et al., 2015]. As an inert gas, CO_2 is only transported by the global atmospheric dynamics and only changes via air-land and air-sea CO_2 fluxes [Jacob, 1999]. However, when investigating the land and ocean carbon uptake and the atmospheric growth rate, i. e. increment of additional carbon in the atmosphere each year, the partitioning of the Global Carbon Budget varies substantially for individual years [Friedlingstein et al., 2020].

E.1.2 *Internal Variability*

Internal variability describes the variations of the climate system, which are not driven by external forcings, such as volcanoes, human activities or the sun[5], but instead are inherent to coupled Earth System dynamics.

ENSO variability

The most dominant mode of inter-annual climate variability is the El Niño–Southern Oscillation (ENSO) [Rasmusson and Wallace, 1983], an irregular variation of winds and sea surface temperatures in the Pacific Ocean. The positive phase El Niño is characterized by high sea surface temperatures and westerly trade winds weakening upwelling in the Pacific. At the same time, the Pacific jet stream shifts northward changing precipitation and temperature in the Americas. ENSO is just the most prominent of such climate modes, others including the North Atlantic Oscillation [Hurrell, 1995], Atlantic Multidecadal Oscillation [Schlesinger and Ramankutty, 1994], Pacific Decadal Oscillation [Mantua et al., 1997] or Southern Annular Mode [Karoly, 1990]. The land and ocean carbon cycles, and in particular the carbon cycle processes described above, are sensitive to such internal climate variability.

land sink variability

The terrestrial carbon sink is characterized by stronger magnitudes and higher frequencies of near-term internal variability compared to the ocean carbon sink [Friedlingstein et al., 2020]. Photosynthesis, ecosystem respiration and fire are sensitive to physical climate variability [Schimel, 1995]. Therefore, internal variability hotspots of the terrestrial carbon cycle are the tropical forests and mid-latitudes, which are driven by ENSO variability [Betts et al., 2018]. In response to El Niño, the terrestrial carbon cycle becomes a net carbon source due to increased respiration and reduced productivity [Betts et al., 2018]. Therefore, one standardized Niño 3 index unit results in a 1.8 PgC/yr terrestrial carbon release, resulting in an global atmospheric CO_2 increase of 0.8 ppm/yr [Jones et al., 2001].

ocean sink variability

The oceanic carbon sink is characterized by smaller magnitudes and lower-frequency variations compared to the land carbon sink [McKinley et al., 2017; Friedlingstein et al., 2020]. Air-sea CO_2 flux is sensitive to internal variability, as the variable ocean transport and sea-surface temperature modulate pCO_2. While the surface ocean carbon cycle inter-annual variability is most pronounced in the tropics driven by ENSO, decadal variations are strongest in the extra-tropical oceans [Landschützer et al., 2016, 2019]. The imprint of the oceanic carbon cycle variations on the atmosphere are not as clearly recognizable as for land, but still detectable: Gruber et al. [2019] attribute a 4 PgC anomalously lower carbon uptake to the Southern Ocean carbon

5 The largest (externally-driven) variability in the carbon cycle is the seasonal cycle driven by the sun that I do not focus on in this thesis. Instead I focus on the multi-year decadal policy-relevant time-scale of two to fifteen years, see subsection E.1.3.

sink from 1995 to 2010 - a magnitude resulting in a 2 ppm global atmospheric CO_2 increase over these 15 years.

Beyond the monotonic increase of atmospheric CO_2, its growth rate varies substantially [Bacastow, 1976; Canadell et al., 2007]. These variations stem from the global land and ocean carbon cycles responding to internal climate variability.

E.1.3 *Global Climate Policy*

Paris Agreement

The Paris Agreement aims to reduce global warming to well below 2°C [UNFCCC, 2015]. In the (multi-decadal and centennial) long-term, anthropogenic CO_2 emissions and carbon-climate feedbacks determine the future of Earth's climate. The reduction of CO_2 emissions is the major measure for mitigating climate change. Therefore, the Paris Agreement [UNFCCC, 2015] requires countries to submit intended nationally determined contributions (NDC) of greenhouse gas reductions, marking a shift in narrative and approach of climate change mitigation [Aykut et al., 2020]. As current pledges are inconsistent with the 2°C target [Rogelj et al., 2016b], a ratcheting mechanism was implemented that requires states to continuously intensify their efforts. Therefore, starting from 2023, the so-called "global stocktake" will take place every five years to track the mitigation progress "in light of the best available science" [UNFCCC, 2015, Art.14] measuring the collective effort and efficacy of past mitigation.

global stocktake

Hence, science is requested to inform policy-makers on the prospects of climate change mitigation, which involves two time-scales:

remaining carbon budget

On the multi-decadal long-term perspective, the remaining carbon budget[6] (RCB) shows the range of cumulative CO_2 emissions compatible with policy goals. This near-linear relationship of cumulative emissions and temperature increase is independent of emission pathway and depends on a probabilistic range of geophysical feedbacks [Allen et al., 2009; Matthews et al., 2009, 2020][7]. Converting the range of the remaining 1.5°C carbon budget to current CO_2 emissions levels yields either an already exceeded carbon budget eight years ago (5^{th} percentile) or 33 years RCB (95^{th} percentile), while the median allows for ten more years at 2019 emission levels [Matthews et al., 2021]. However, such a conversion is misleading, as the RCB is required to lead to

6 The carbon climate community established two similar concepts of carbon budgets for different periods: The Global Carbon Project annually releases the historical or total Global Carbon Budget explaining where the carbon went in the past ten years and since 1850 [Friedlingstein et al., 2020]. In the context of future emissions, the remaining carbon budgets describe how much remaining carbon emissions are compatible with a temperature target [Matthews et al., 2020].

7 On this multi-decadal time-scale relevant for the concept of the RCB, climate model uncertainty, estimating these feedbacks, dwarfs internal variability [Tokarska et al., 2020].

net zero CO_2 emissions and hence the CO_2 emission path ways must be continuously declining. Also this message "years remaining" communicates a cliff-edge deadline, whereas climate change will rather unfold intermittent, slow and gradual [Asayama et al., 2019]. Furthermore, the remaining carbon budget[8] is not fixed anymore [Peters, 2016, 2018]: The RCB is sensitive to the temperature threshold definition [Rogelj et al., 2016a; Millar et al., 2017]. Also, negative emissions [Peters and Geden, 2017] and temperature overshoot scenarios [Geden and Löschel, 2017] allow more flexible success definitions hampering immediate action [McLaren et al., 2019]. Even the communicated likelihood of staying below a temperature target shrank to 66% or 50% [Peters, 2018; Matthews et al., 2021]. Understanding the incentives of politics, carbon budgets help to postpone action to the future rather than inspiring action [Geden, 2018; Peters, 2018]. Due to the range of definitions, choices and uncertainties, the RCB should not be used to reduce climate change to a single magic number, but rather as an elevator pitch explaining the importance of net zero emissions [Peters, 2018] - a more actionable policy target [Geden, 2016]. What reconciles RCBs and net zero emissions is the call for immediate CO_2 emission reductions, i. e. CO_2 emission reductions in the near-term or best now.

verification of CO_2 emissions

Science also needs to inform policy-makers about the dynamics of the Earth System in the multi-year short-term perspective [Peters et al., 2017; Peters, 2018]. How can policy and science judge whether past CO_2 emission reduction efforts have been effective? The global temperature signal shall be limited in the long-term, but global temperature has large inertia and internal variability, making it impossible to track near-term progress in a timely manner [Tebaldi and Friedlingstein, 2013; Marotzke, 2019]. Prior to that, CO_2 emission reductions will also be reflected in decelerating atmospheric CO_2 growth. Therefore, a more direct causal link towards the efficacy of CO_2 emission reductions is the observable atmospheric CO_2 signal [Peters et al., 2017]. As CO_2 is well-mixed in the atmosphere after one to two years [Jacob, 1999; Ciais et al., 2019], atmospheric CO_2 can be regarded as a Global Common [Edenhofer et al., 2014] because only the collective CO_2 emission reductions, i. e. the sum of all NDCs, determine the evolution of atmospheric CO_2. Furthermore, global atmospheric CO_2 provides an *independent* top-down verification of NDC-based bottom-up CO_2 emission reductions, which are reported by countries themselves [Janssens-Maenhout et al., 2020].

internal variability noise

However, there is one substantial challenge when using atmospheric CO_2 as a verification of CO_2 emission reductions: The noise of internal variability in the global land and ocean carbon cycles can disguise the CO_2 emission reduction signal in the near-term, which covers the

8 The term budget stems from a financial *fixed* budget where over-spending in one period needs to be equalized by savings in the following period.

policy-relevant pentadal global stocktakes [Peters et al., 2017; Schwartzman and Keeling, 2020]. Consequentially, I focus this dissertation on the near-term variability and predictability of the carbon cycle in the time-frame of two to fifteen years[9]. From the global policy perspective taken for this dissertation, I disregard the micro perspective, i. e. the details of carbon cycling within the reservoirs, and focus on the macro perspective of the global carbon cycle - the exchange of carbon at the Earth's surface as air-land and air-sea CO_2 flux - which reflects the sum of all reservoir-internal processes at the boundary. Therefore, when discerning the global CO_2 emission reduction signal from the carbon cycle internal variability noise, only the *global* land and ocean carbon cycles, aggregating all spatial CO_2 fluxes, matter and are therefore emphasized in this dissertation.

E.1.4 *Effect of CO_2 Internal Variability on the Signal of CO_2 Emission Reductions*

As a consequence of CO_2 emission reductions, policy-makers and the public will expect that CO_2 emission reductions cause atmospheric CO_2 growth to decelerate[10]. However, internal variability complicates the picture in the near-term: Carbon released to the atmosphere by the global carbon cycle may counteract CO_2 emission reductions, or carbon uptake may slow down atmospheric CO_2 growth despite the lack of CO_2 emission reductions. This raises the question whether CO_2 emission reductions cause decelerating atmospheric CO_2 in the near-term.

pitching paper I

While observations can only describe the present and past climate, climate modeling allows us to simulate the future. Earth-System-Models are numerical models representing the Earth System by coupling physical and biogeochemical processes under external forcings [Flato et al., 2013]. In particular, ESMs allow us to model climate trajectories with or without CO_2 emission reductions. They can therefore be used to tackle the question of CO_2 deceleration causation, which is a matter of detection and attribution [Bindoff et al., 2013]. By inference, chapter I [Spring et al., 2020] showcases the need for initialized carbon cycle predictions by exposing the large range of atmospheric CO_2 internal variability and is therefore an appropriate introductory paper to this thesis.

9 The IPCC's predictability chapter Kirtman et al. [2013] defines the near-term as the next two decades.

10 Over the past decades, atmospheric CO_2 concentrations are not just increasing linearly, but have been accelerating due to CO_2 emissions increasing at 1% per year over decades [Canadell et al., 2007; Peters et al., 2019].

When policy-makers want to evaluate the efficacy of CO_2 emission reductions, they either need to wait until internal variability balances within the following few years or they look into initialized carbon cycle predictions, which provide an outlook on multi-year future outlooks based on alternative CO_2 emissions pathways. The ongoing COVID-19 pandemic provides an unprecedented example for the challenges attributing CO_2 emission reductions in global atmospheric CO_2.

E.1.5 *Case-study: Atmospheric CO_2 Signal in the COVID-19 Pandemic*

little variability in CO_2 emissions

I mentioned previously only internal variability of the global carbon cycle as noise in the atmospheric CO_2 signal. The annual fossil-fuel CO_2 emissions, the bulk of CO_2 emissions, have evolved rather smoothly so far [Fig. E1, gray line]. Even disruptive economic crises, like the global financial crisis in 2008 and its aftermath, have only marginally decreased fossil-fuel CO_2 emissions [Friedlingstein et al., 2020] [Fig. E1].

The current COVID-19 crisis marks an unexpected instance of CO_2 emission reductions, as CO_2 emissions dropped unprecedentedly [Le Quéré et al., 2020; Liu et al., 2020]. Can this signal in CO_2 emissions be identified in annual atmospheric CO_2 observations?

back-of-the-envelope calculation

In this framing, the annual COVID-19 CO_2 emission reductions in 2020 are the signal, whereas the internal variability of land and ocean CO_2 uptake is the noise. With a back-of-the-envelope calculation based on the Global Carbon Budget data [Friedlingstein et al., 2020], I show that internal variability most likely disguises CO_2 emission reductions in atmospheric CO_2 measurements in the year 2020 [Fig. E1].

COVID-19 CO_2 emission reductions signal

The annual COVID-19 related CO_2 emissions are estimated by different groups and methods: $-6.9^{-2.7}_{-10.8}$% ($\pm 1\sigma$) [Le Quéré et al., 2020, updated in Friedlingstein et al. 2020], -7% [IEA, 2020], -13% [Forster et al., 2020], -5.8% [Friedlingstein et al., 2020] and -6.9% [Liu et al., 2020]. Averaging these forecasts results in CO_2 emission reductions of 7.9±3.6% ($\pm 1\sigma$), which amounts to a 0.8±0.3 PgC emission decrease. Due to the land and ocean carbon sinks, the expected CO_2 emission reductions as seen in global atmospheric CO_2 need to be discounted by the airborne fraction of fossil fuel CO_2 emissions[11], which averages to 55% (1958-2019) [Friedlingstein et al., 2020]. Therefore the expected reduction of global atmospheric CO_2 growth rate due to COVID-19 is only 0.4±0.2 PgC, which translates to 0.2±0.1 ppm converted to global atmospheric CO_2 units[12].

internal variability noise

I infer the annual carbon cycle internal variability from the detrended 1959-2019 atmospheric CO_2 growth rate and estimate the 2σ

11 I exclude land-use changes in the airborne fraction, although they are usually included in the airborne fraction (of all anthropogenic CO_2 emissions), because here I do not assume any changes in land-use change specifically due to COVID-19.

12 1 PgC = 1 GtC = 2.12 ppm = 0.27 $GtCO_2$ [Keeling et al., 1982; Ballantyne et al., 2012; Friedlingstein et al., 2020], see also equation 1 in subsection E.2.1.

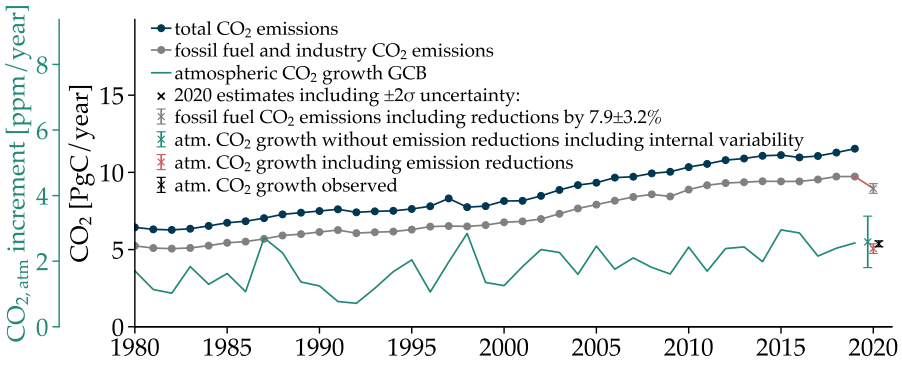


Figure E1: **COVID-19 CO_2 emission reductions not seen in the annual 2020 global atmospheric CO_2 growth rate.** Historical annual CO_2 emissions from fossil fuels and industry (gray), total CO_2 emissions by additionally including land-use changes (dark blue) and annual increase of CO_2 in the atmosphere (green) based on Global Carbon Budget data [Friedlingstein et al., 2020]. Note that the atmospheric CO_2 growth can be interpreted as increase of carbon in the atmosphere (black scale) and also as an increase of global atmospheric CO_2 mixing ratio (green scale). CO_2 emissions drive long-term atmospheric CO_2 growth, but internal variability of the terrestrial and oceanic carbon sink drive variations in the annual atmospheric CO_2 increment. Crosses show back-of-the-envelope estimates for the year 2020 with 2σ uncertainties: how COVID-19 related CO_2 emission reductions (red line) result in an atmospheric CO_2 growth reduction (red), where emission reductions have to be discounted by the mean airborne fraction of fossil-fuel emissions (55% over 1958-2019) [Friedlingstein et al., 2020]. The COVID-19 CO_2 emission reductions are averaged over estimates by Liu et al. [2020] (-6.9%), Le Quéré et al. [2020, updated in Friedlingstein et al. 2020] ($-6.9^{-2.7}_{-10.8}$%) ($\pm1\sigma$), -13% [Forster et al., 2020], -5.9% [Friedlingstein et al., 2020] and IEA [2020] (-7%) to a combined mean reduction estimate of 7.9±3.6% ($\pm1\sigma$). This COVID-19 signal in atmospheric CO_2 is masked by internal variability of the global carbon cycle (green, 2σ uncertainty of detrended past variability in atmospheric CO_2 growth). The observed global atmospheric CO_2 growth rate did not change from 2019 to 2020 despite unprecedented CO_2 emission reductions (black) [Dlugokencky, 2021].

interval as ±1.7 PgC = ±0.8 ppm. Given this large range of internal variability, the COVID-19 CO_2 signal of 0.2 ppm is dwarfed by global carbon cycle noise. Indeed, recent global atmospheric CO_2 observations reveal that the growth rate did not change from 2019 to 2020 [Fig. E1, black cross]. The reported growth rate reduction of 0.01 ppm is inside the measurement uncertainty of ±0.11 ppm [Dlugokencky, 2021]. Hence, internal variability balanced COVID-19 CO_2 emission reductions.

In conclusion, this example clearly illustrates how crucial internal carbon cycle variability is for near-term atmospheric CO_2 predictions. The annual 2020 CO_2 emissions reduction signal cannot be identified in global annual 2020 observations due to the noise in the carbon cycle[13]. In order to be successful with this prediction and to distill the signal, one would need to constrain the noise. This is where initialized carbon cycle predictions come into play.

E.1.6 *Predictability of the Global Carbon Cycle*

The decadal prediction community strictly differentiates between initialized predictions and uninitialized projections. Lorenz [1975] characterizes them as predictability of the first and second kind.

initial value predictability

Like in numerical weather prediction, an initialized climate prediction integrates the initial conditions forward and thereby attempts to predict the near-term [Meehl et al., 2009]. Predictive skill arises in an "initial value problem" where the memory of the system stems from its initial conditions, which constrain internal variability at the beginning of the simulation.

boundary value predictability

Uninitialized projections attempt to foresee the long-term climate state independent of initial conditions. In a "boundary value problem" predictability arises from the boundary conditions, e. g. the pathway-dependent CO_2 concentration at the end of the century, which mainly determines the external forcing for the new climate.

Decadal climate prediction has an ambivalent relation to these two sources of predictability. On the one hand, climate prediction sits at the sweet spot, benefiting from both sources of predictability. On the other hand, both sources of predictability are weaker on the decadal two-to-fifteen-year time-scale compared to the predictive skill in their respective limits.

Decadal predictions have proven to constrain near-term uncertainty from internal climate variability in multi-year retrospective forecasts, so-called hindcasts, in which forecasts aim to predict the past evolution of climate as a means to verify prediction systems [Smith et al., 2007; Keenlyside et al., 2008; Pohlmann et al., 2009; Marotzke et al., 2016].

The ocean carbon cycle community is leveraging this established prediction skill from the physical oceanography community, because predictive skill in the ocean physics is carried into the physics-dependent ocean carbon cycle. The global carbon sink is predictable for two to three years [Li et al., 2019; Lovenduski et al., 2019a; Ilyina et al., 2021]. Also the global terrestrial carbon sink is predictable against reconstruction simulations or the global carbon budget estimates for two years

13 However, regional CO_2 emissions reductions were observed locally [Tohjima et al., 2020; Chevallier et al., 2020].

[Lovenduski et al., 2019b; Ilyina et al., 2021].

CO_2 predictions for policy-makers

How can carbon predictions inform policy-makers? Modeling in general explores alternative scenarios by changing boundary conditions. Initialized ESMs can showcase alternative scenarios[14] of CO_2 emission reductions reflecting more or less progressive aggregated NDCs. Furthermore, initialized ESM predictions constrain the near-term carbon cycle variability to anticipate the short-term CO_2 reductions signal, because the noise of internal variability is dampened. Initialized CO_2 predictions here provide policy-makers with an evidence-based forecasting tool to explore policy alternatives. By knowing about internal variability in advance, atmospheric CO_2 can be anticipated ahead-of-time, i. e. the predictions show the most likely evolution of atmospheric CO_2 based on the current climate state and proposed CO_2 emission reductions.

pitching paper II

As a first step towards ESM-based initialized CO_2 predictions, I perform idealized perfect-model initialized potential predictability experiments to estimate the limits and drivers of atmospheric CO_2 predictability in chapter II [Spring and Ilyina, 2020].

E.1.7 *Reconstruction of the Global Carbon Cycle*

Climate predictions require two ingredients: an Earth System Model and initial conditions representing the observed climate state to start predictions from. In assimilation simulations, so-called reconstructions, climate predictions are confined to track the reanalysis, i. e. observation-constrained model simulations, closely [Meehl et al., 2009]. Reconstructions describe where anthropogenic carbon went and therefore represent a near real-time verification of CO_2 emissions [Peters et al., 2017]. Such reconstructions serve as the initial conditions for initialized predictions representing ahead-of-time verification of CO_2 emissions. Physical oceanographers rely on annual initializations since the late 1950s in their hindcasts to establish multi-year predictive skill[15], which is enabled by global physical climate reanalysis projects [like Balmaseda et al., 2013, starting in 1958]. However, globally-reaching observations of the carbon cycle are limited.

limited ocean carbon observations

The spatio-temporal resolution of the observed state of the oceanic CO_2 system is insufficient to initialize hindcasts for predictions, despite efforts of the observations community to develop gridded data products [Bakker et al., 2016]. The largest problem in creating reliable observations-based products is the lack of bygone observations. Continuous monitoring of the oceanic carbon started in the 1980s at the

14 like the Representative Concentration Pathways (RCP) in CMIP5 [Meinshausen et al., 2011] in CMIP5 [Taylor et al., 2011] and Shared Socioeconomic Pathways (SSP) [Meinshausen et al., 2019] in CMIP6 [Eyring et al., 2016]

15 Seasonal and numerical weather predictive skill is backed by much larger initialization and ensemble member statistics and thus more robust.

islands of Hawaii and Bermuda [Karl and Lukas, 1996; Phillips and Joyce, 2007]. The earliest gridded oceanic CO_2 dataset, dating back to 1982, is powered by machine-learning gap-filling [Landschützer et al., 2015], and is taken as verification for reconstructions and predictions only [Li et al., 2019; Ilyina et al., 2021].

limited land carbon observations

Historically, the land carbon cycle is observed at individual locations. Air-land CO_2 flux is measured at eddy-covariance flux towers. Above ground terrestrial biomass is estimated via remote sensing [Issa et al., 2020]. Inversion models can constrain the geo-spatial patterns of CO_2 flux based on atmospheric CO_2 concentrations top-down [Rayner et al., 1999], but do not describe the sources of CO_2 flux bottom-up. Due to the sparse record of air-land CO_2 flux observations and the complex land surface structure, there currently exists no community-wide adopted gridded air-land CO_2 flux observation-based product. Commonly the multi-model global carbon budget estimate is taken as a pseudo-observational verification of the global land carbon sink by modelers [Ilyina et al., 2021].

The atmospheric carbon cycle is continuously observed only in individual measurement stations, e. g. in Mauna Loa, Hawaii since 1958. Atmospheric CO_2 concentrations can be measured via remote sensing with global atmospheric CO_2 coverage, which was introduced in 1996 [Kobayashi et al., 1999]. However, the satellite coverage is too short for hindcast initialization. On the other hand, carbon cycle forecasters would not assimilate atmospheric CO_2 directly, as the predictability of atmospheric CO_2 does not arise from processes in the atmosphere, but from the land and ocean carbon sinks below. Atmospheric CO_2 is transported around the globe as an inert gas and mixes globally within 1-2 years [Keeling et al., 1989b]. As atmospheric CO_2 integrates the land and ocean CO_2 fluxes, atmospheric CO_2 observations should rather be used to verify the quality of the land and ocean carbon cycle reconstructions.

atm. CO_2 as verification rather than initial value reconstruction

carbon cycle observations too sparse to initialize

These observational carbon cycle records are insufficient to initialize long-lasting hindcasts. A first marine biogeochemical assimilation by Carroll et al. [2020] was just released and only runs back to 1995. Furthermore, assimilating biogeochemical properties into an ESM easily triggers drifts [Toggweiler et al., 1989; Park et al., 2018]. Therefore, current state-of-the-art prediction models do not reconstruct the carbon cycle initial states from observations, presuming that the reconstructed climate initializes the carbon cycle indirectly [Ilyina et al., 2021]. The often mentioned hypothesis that carbon cycle predictions would benefit from direct carbon cycle initialization [Séférian et al., 2018; Ilyina et al., 2021] is challenged in chapter III [Spring et al., 2021].

pitching paper III

E.2 OVERVIEW OF METHODOLOGIES

I now introduce the main concepts used in this dissertation and present a brief overview on their main characteristics.

E.2.1 *Disentangling the Origin of Variations in Atmospheric CO_2*

Disentangling the sources of observed inter-annual variability in atmospheric CO_2 remains a challenge. Both the land and the ocean carbon sinks respond to climate variability and the resulting local surface CO_2 flux anomalies dissipate quickly [Rödenbeck et al., 2003]. Bacastow [1976] and Bacastow et al. [1980] hypothesized that the oceans caused the inter-annual variations in atmospheric CO_2, which correlate with ENSO. Here, general circulation models advanced the understanding of the carbon cycle. Only the mechanistic ocean carbon cycle modeling study by Winguth et al. [1994][16] could confirm the terrestrial biosphere origin of atmospheric CO_2 increase due to El Niño, which was previously suggested by a transport model [Keeling et al., 1989a] and later confirmed by observations [Keeling et al., 1995; Francey et al., 1995].

past challenges of attributing near-term CO_2 variations

The dynamics of the Earth System including the carbon cycle are nowadays described by Earth System Models (ESMs), numerical models coupling physical and biogeochemical processes [Flato et al., 2013]. Only coupling of land and ocean carbon cycles by resolving atmospheric CO_2 shows the full interplay of climate and carbon [Cox et al., 2000; Jones et al., 2001]. To further understand atmospheric CO_2 variability, also in the context of predictability, the contributions from ocean and land variability need to be isolated.

I disentangle the contributions of land and ocean carbon sinks on atmospheric CO_2 by diagnosing atmospheric CO_2 (XCO_2^{diag}) variations compatible with the carbon sinks. Utilizing the conservation of mass, I aggregate the deviations from the ensemble member i, scenario s and temporal t mean of the global land and/or ocean carbon sink $G_{i,s,t}^{\text{land and/or ocean}}$ over time. The conversion factor 2.124 converts CO_2 flux magnitudes in units of petagram (1 PgC = 1 GtC = 2.124 ppm) carbon to a globally dispersed atmospheric CO_2 in units ppm [Keeling et al., 1982; Ballantyne et al., 2012]. The resulting atmospheric CO_2 variations are then added to a forced emissions signal of atmospheric CO_2 f_s:

separate land and ocean in atm. CO_2 variations

$$XCO_{2,i,s}^{diag}(t) = \sum_{t'}^{t} G_{i,s,t}^{\text{land and/or ocean}}(t') \cdot \frac{\text{ppm}}{2.124\text{PgC}} + f_s. \tag{1}$$

In ESMs, the carbon cycle coupling can be configured in two different setups concentration-driven and emission-driven simulations.

carbon cycle model configuration

16 The ocean biogeochemistry model was the Hamburg Ocean Carbon Cycle model version 3, which is a predecessor of HAMOCC 6 used for this thesis [Ilyina et al., 2013].

In *concentration-driven* simulations, atmospheric CO_2 concentrations, which are calculated by simplified climate models (Integrated Assessment Models (IAMs)) based on anthropogenic emission pathways, are prescribed to the model. Therefore, the land and ocean carbon cycle do not alter atmospheric CO_2 levels in these simulations. Hence, CO_2 fluxes are diagnostic from the atmosphere's point of view. In order to still analyze the expected global atmospheric CO_2 variations of these simulations, such as in Spring et al. [2020], I diagnose a variable global atmospheric CO_2 tracer from both land and ocean CO_2 fluxes combined [Ch. I].

Alternatively, the atmospheric CO_2 mixing-ratio can be calculated based on CO_2 emissions and the response of the land and ocean carbon sinks. These *emission-driven* or *prognostic* atmospheric CO_2 simulations resolve a spatio-temporal variations of atmospheric CO_2, but do not reveal their driver. Therefore, I use equation 1 in prognostic atmospheric CO_2 simulations, such as in [Spring and Ilyina, 2020] under pre-industrial conditions (f_i=0), to disentangle the individual contributions of the land and ocean carbon sink to the atmospheric CO_2 evolution [Ch. II]. This allows me to discern whether atmospheric CO_2 is predictable due to the ocean or the land carbon sink. A thorough verification of this method can be found in section I.SI.1.

E.2.2 *Forecast Verification*

The value of forecasts depends on whether or how forecasts are used [Murphy, 1993]. Why should users like business or governmental resource managers trust a forecast? Anyone could issue forecasts, e. g. for atmospheric CO_2 levels in two years.

verification to trust forecasts

Forecasts must be reliable for users to trust it. Forecasts are falsifiable [Palmer, 2020] and repeated evaluation of forecasts lends the forecaster credibility [Murphy, 1993], thereby influencing decision makers and hence becoming valuable. If a forecast worked in the past, and the climate system did not change considerably, it ought to work in the near future. This process is named forecast verification and achieved by issuing retrospective forecasts, which are then evaluated based on past observations [Boer et al., 2016]. Many past forecasts are compared to simpler reference forecasts. If the initialized forecasts outperform a persistence forecast or the uninitialized historical projection two years after starting the forecast, the initialized model has predictive skill over the persistence or uninitialized forecast at lead year two.

current challenges of forecast verification

Forecast verification is a standardized exercise from numerical weather to decadal climate prediction. Many different attributes of forecasting skill can be evaluated. The increasing volume of prediction data calls for the use of distributed and parallelized software and

hardware. Furthermore, the replication of forecast verification code is unnecessarily error-prone and not inherently advancing the prediction science.

pitching paper IV

To standardize the code base, leverage on new software frameworks in python [Hoyer and Hamman, 2017; Rocklin, 2015] and enable interactive workflow, I co-created[17] CLIMPRED, a forecast verification software, which is developed open-source on GitHub. The reference paper [Brady and Spring, 2021] is attached as chapter IV, but not summarized further in this essay.

E.2.3 *Perfect-Model Framework*

what makes my model perfect

The conceptual modeling framework overarching the three research papers is the perfect-model framework. The term perfect-model confronts the popular quote that "all models are wrong" [Box, 1976]. When climate models are evaluated against the observed Earth System, indeed these models represent processes imperfectly, calculate on too coarse spatio-temporal resolution, are limited by imperfect boundary conditions and start from imperfect initial conditions. However, by defining the model itself as the truth to evaluate against, the model now represents processes perfectly, perfect spatio-temporal resolution, perfect boundary conditions, even perfect initial conditions. When started from identical initial conditions on the same software and hardware architecture, a climate model is even deterministic - being able to reproduce an identical simulation - a truly perfect model in its limited scope.

what a perfect-model fails at

However, this perfect-model framework cannot predict the real world, i. e. the observed Earth System. It is merely a studying tool, from which the forecaster can learn about the tool used for real world forecasts in an idealized setup, which is not convoluted by technical burdens of reconstruction to the reanalysis product. Given that the model reproduces observed variability [see section II.SI.7], the perfect-model framework can estimate the limits of predictability of a given ESM, which are eventually imposed by the chaotic atmosphere.

Historically, the predictability science started with perfect-models [Pohlmann et al., 2004], which were followed by retrospective predictions [Pohlmann et al., 2009; Scaife and Smith, 2018] and finally operational predictions [Marotzke et al., 2016]. Hence, perfect-model studies [Séférian et al., 2018; Spring and Ilyina, 2020; Fransner et al., 2020] lay the foundations for carbon cycle predictions [Li et al., 2016; Li et al., 2019; Lovenduski et al., 2019a; Lovenduski et al., 2019b; Ilyina et al., 2021].

perfect-model used in papers

Chapter I uses a perfect-model, because the factual and counter-

17 This software was co-developed with Riley X. Brady, then PhD candidate from Boulder University. Code: https://github.com/pangeo-data/climpred. Documentation: https://climpred.readthedocs.io/.

factual[18] worlds are both described by the model. Chapter II marks the classic perfect-model predictability study for the global carbon cycle. In chapter III, I reconstruct the initial conditions of the global carbon cycle, which I aim to recreate in perfect-model simulations, i. e. the same model as the perfect pseudo-observations come from and test whether direct carbon cycle reconstruction improves carbon cycle predictability.

With these methods in mind, I will now summarize the three previously announced papers in individual sections.

18 Please find an introduction of the factual and counter-factual world in sections E.3 and I.2.1.

E.3 PAPER I: VARIABILITY OF THE GLOBAL CARBON CYCLE

motivation

Atmospheric CO_2 concentrations driven by CO_2 emissions are the carbon signal in the atmosphere, but how strong are atmospheric CO_2 variations compared to changes in CO_2 emissions? Once CO_2 emission reductions are implemented on a global scale, questions about the immediate effects of these emission reductions will be raised [Tebaldi and Friedlingstein, 2013; Marotzke, 2019; McKenna et al., 2020]: When will global mean surface temperature stop increasing? Even more closely linked and along the lines of Peters et al. [2017] and Schwartzman and Keeling [2020], I add: When will CO_2 emission reductions be detectable in atmospheric CO_2? And how can it be ensured that it is the emission reductions that caused an atmospheric CO_2 slowdown and not notorious internal variability?

This question is similarly raised about potential imprints of a reduced CO_2 emission signal due to COVID-19 lock-downs, but when aiming to detect a signal, the noise also needs to be considered. The growth rate of atmospheric CO_2 on inter-annual time scales is largely controlled by the response of the land and ocean carbon sinks to climate variability. Only on longer time-scales, substantially longer than the ENSO cycle of two to seven years, do CO_2 emissions determine atmospheric CO_2 trends [Keeling et al., 1989a; Keeling et al., 1995]. Therefore, internal variability might disguise CO_2 emission reductions. Here, I ask what the probability is that a slowdown in near-term atmospheric CO_2 growth is attributable to a policy change implementing emission reductions, such as a transition from scenario Representative Concentration Pathway (RCP) RCP4.5 to RCP2.6, in the face of internal climate variability. This causation question is desirable to answer, as policy-makers will want to ensure that CO_2 emission reductions are the cause of decelerating CO_2 growth and not internal variability. In particular, I ask whether CO_2 emission reductions cause a CO_2 deceleration in the near-term. I formulate the following guiding research questions:

research questions

1. **What is the probability that CO_2 emission reductions cause a slowdown in atmospheric CO_2 growth over five years in the face of internal climate variability?**

2. **After how many years do such CO_2 emission reductions cause reduced atmospheric CO_2 trends for certain?**

These questions become policy-relevant once policy-makers assess the efficacy of emission reductions in the global stocktake [UNFCCC, 2015, Art.14] every five years [Peters et al., 2017; Schwartzman and Keeling, 2020].

methods: diagnosed atm. CO_2

In the MPI-ESM Grand Ensemble (MPI-ESM GE) simulations [Maher et al., 2019], I diagnose atmospheric CO_2 variations *compatible* with the natural carbon sinks variations as calculated by equation 1.

detecting an atm. CO_2 deceleration

I define a slowdown of atmospheric CO_2 increase as a first robust response to declining CO_2 emissions, reasoning that smaller CO_2 emissions should lead to a weaker atmospheric CO_2 increase. Please note that this response is not a reversal to an actual reduction in atmospheric CO_2 concentrations, which is mathematically described by a negative first derivative. In contrast, I detect a slowdown in the still increasing atmospheric CO_2 growth, mathematically described by a negative second derivative. Thus, decelerating atmospheric CO_2 grwoth is a first step towards stable atmospheric CO_2, which is a sufficient condition[19] for a stable climate.

causation

I aim to detect decelerating atmospheric CO_2 growth in comprehensive causation framework [Pearl, 2000; Hannart et al., 2016; Marotzke, 2019]. I compare growth trends of atmospheric CO_2 before and after the onset of CO_2 emission reductions in 2020. I use the scenario RCP2.6 implementing CO_2 emission reductions in contrast to scenario RCP4.5, which is closest to the pledged and current policies [Rogelj et al., 2016b] without CO_2 emission reductions until 2035 [Thomson et al., 2011].

From probabilities of CO_2 deceleration in a factual and counterfactual world represented by either RCP2.6 or RCP4.5, I compute the probability that CO_2 emission reductions cause the atmospheric CO_2 deceleration in a sufficient, necessary, or sufficient and necessary causation sense. Translated into the quest of CO_2 deceleration detection, *sufficient* means asking in advance in an RCP4.5 world whether CO_2 emission reductions will cause the slow-down; *necessary* means asking in retrospect in an RCP2.6 world whether CO_2 emission reductions were needed for the slowdown; *necessary and sufficient* combine the two causation attributes describing that only CO_2 emission reductions can decelerate atmospheric CO_2.

I only use atmospheric CO_2 observations to verify whether the diagnosed atmospheric CO_2 growth from the model is inside the observed range [Fig. I.SI.7]. This analysis is carried out in an idealized perfect-model framework because only initial conditions and two different CO_2 emission pathways distinguish the simulations. Therefore, model uncertainty and bias with respect to the real world are not subjects of this study. However, all models, as well as the observable Earth System, are subject to this internal variability, which hinders clear causation in near-term observations.

19 Sufficient and/or necessary causation as in Pearl [2000] and Hannart et al. [2016] is different from sufficient and/or necessary condition in logic meant here [Bloch, 2011, p.8-9].

results: deceleration probability

I find a 70% probability that atmospheric CO_2 decelerates in the CO_2 emission reductions scenario RCP2.6, while trends remain impartial without CO_2 emission reductions. This implies a 30% chance that atmospheric CO_2 accelerates despite CO_2 emission reductions. The global carbon cycle response to climate variability can trigger this counter-intuitive response, e. g. when tropical forests release more CO_2 in response to an El Niño event than is mitigated by CO_2 emission reductions [Jones et al., 2001; Betts et al., 2018] [Fig. I.SI.8].

causation probability

These probabilities translate into CO_2 emission reductions being sufficient to cause a five-year trend reduction beforehand by 42% and in hindsight necessary by 31%. The probability that a policy change from RCP2.6 to RCP4.5 is indeed needed and will suffice to bring the desired outcome considering five-year trends is only 22%. These probabilities are far from certain (implying 100%). Analysing increasing periods centering around the year 2020 of CO_2 emission reduction separation, I find that virtually certain sufficient causation is reached regarding ten year trends and necessary causation regarding 16 year trends.

conclusion: internal variability hinders clean causation

The pentadal global stocktakes make the five-year internal variability highlighted in this study especially relevant for policy-makers. This analysis shows that a five-years assessment cycle is likely dominated by internal variability noise and not by a potential CO_2 emission reductions signal.

key findings

1. **CO_2 emission reductions are sufficient by 42%, necessary by 31% and both necessary and sufficient by 22% to cause reduced atmospheric CO_2 five-year trends.**

2. **Certainty implying sufficient or necessary causation is only reached after, respectively, ten and sixteen years.**

initialized predictions as partial solution

This study demonstrates the inherent uncertainty in atmospheric CO_2 projections [Spring et al., 2020]. Initialized ESM-based prediction systems may reduce this uncertainty by predicting natural variations of the global carbon cycle and thereby provide a partial solution to the challenge of identifying signals in the internal variability noise.

E.4 LINKING GLOBAL CARBON CYCLE VARIABILITY TO GLOBAL CARBON CYCLE PREDICTABILITY

predictability of atmospheric CO_2 in the MPI-ESM GE

As an insertion, linking internal variability of atmospheric CO_2 with the following predictability of atmospheric CO_2 chapter, I want to feature the same set of simulations from the MPI-ESM Grand Ensemble as a perfect-model framework initialized in the year 2006. In particular, I show how initialized ensembles predict atmospheric CO_2 better than uninitialized projections by constraining internal variability.

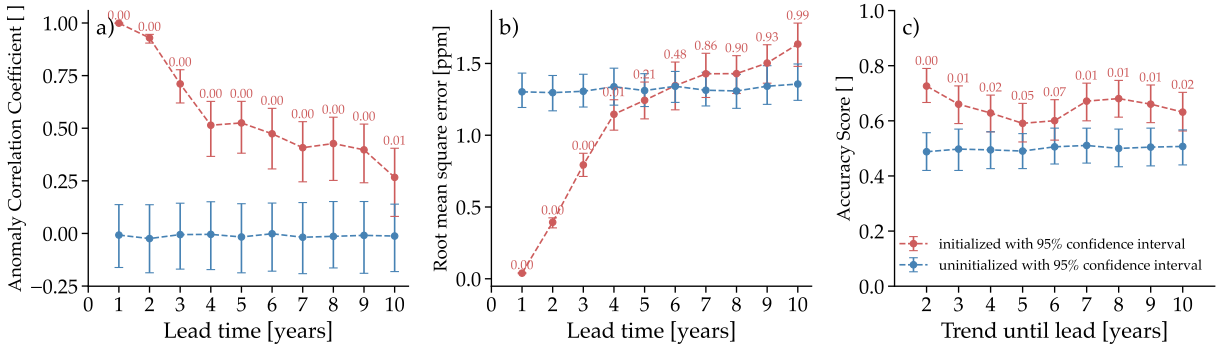


Figure E2: **Predictive skill of detrended global annual diagnostic atmospheric CO_2 in the MPI-ESM Grand Ensemble perfect-model framework initialized in the year 2006**. The initialized skill (red) starts from near perfect and approaches the limits of the resampling uninitialized skill (blue). Red numbers indicate the probability that the uninitialized ensembles are more predictable than the initialized ensembles. For a detailed methodology refer to section II.2.3.

Identifying 100 initializations (formerly seen as ensemble members) with three members (formerly seen as RCP scenarios) each up to 10 lead years, until which the external forcing is nearly indistinguishable, allows me to calculate initialized perfect-model predictability. Different attributes in diagnostic atmospheric CO_2 are predictable for a couple of years [Fig. E2]:

The anomaly correlation coefficient (ACC) measures linear association between a two-member forecast and one-member verification between ensemble members. Initialized atmospheric CO_2 is significantly predictable over all ten lead years. The first three lead years yield ACC values larger than 0.6, commonly considered as useful for predictability [Fig. E2a].

Root-mean-square-error (RMSE) measures the accuracy between forecast and verification. Initialized atmospheric CO_2 is predictable for up to four years. The RMSE most directly shows the potential of how the full range of internal variability of the uninitialized ensemble (blue) can be constrained for four years by initialized ensembles (red) [Fig. E2b].

predictability of the atm. CO_2 trend change

The sign of a trend change measured over an increasing trend length of the initialized ensembles has a higher accuracy score than impartial uninitialized ensembles [Fig. E2c]. These trend changes correspond to the variability assessed in chapter I [Spring and Ilyina, 2020]. Here, the forecaster knows about the past evolution of atmospheric CO_2. The ESM forecasts whether CO_2 will accelerate or decelerate over time due to internal variability alone without any influence of CO_2 emission reductions.

Overall these results indicate a first glimpse of multi-year predictability in atmospheric CO_2. However, these simulations performed in 2013 contain merely three ensemble members, which allow me only to compare a two-member forecast with a single-member verification in turns. In order to gain predictive skill backed by more robust statistics and to investigate spatially-resolved atmospheric CO_2 predictability, new initialized ensemble simulations with prognostic atmospheric CO_2 are required.

E.5 PAPER II: PREDICTABILITY OF THE GLOBAL CARBON CYCLE

motivation

Having demonstrated the large magnitude of diagnostic atmospheric CO_2 variability, I assess the predictability of prognostic atmospheric CO_2 in initialized perfect-model framework simulations. In order to uncover potential anthropogenic CO_2 signals, I specifically aim to predict and constrain the noise from carbon cycle internal variability. Initialized carbon cycle predictions answer the question where the carbon will go in the near-term. Specifically, initialized carbon predictions predict the land and ocean carbon sinks and thereby constrain atmospheric CO_2 variability.

So far, there is only one prediction method for near-term atmospheric CO_2: Betts et al. [2018] regress annual atmospheric CO_2 increments on the El Niño 3.4 index to predict next year's atmospheric CO_2, based on extrapolated historical CO_2 emissions and a six-month Niño 3.4 prediction [Betts et al., 2016]. This prediction is computationally cheap to perform, as they diagnostically predict CO_2 based on existing operational predictions. In contrast, ESM-based predictions consume vast amounts of energy. To justify this effort, they promise better and longer predictions, which the seasonal to decadal prediction community in oceanography delivers [Merryfield et al., 2020]. While operational ESM-based carbon cycle predictions are not yet available, previous ESM-based studies find predictability in the global land [Zeng et al., 2008; Séférian et al., 2018; Lovenduski et al., 2019b] and global oceanic [Séférian et al., 2018; Li et al., 2019; Lovenduski et al., 2019a] carbon sinks. However, multi-year predictability of prognostic atmospheric CO_2, i. e. resolving its variability and spatial distribution, was unexplored.

Furthermore, the main driver of atmospheric CO_2 predictability is unknown. While the air-land CO_2 flux has a larger variability amplitude than the air-sea CO_2 flux [Friedlingstein et al., 2020], the ocean has much longer memory of variability making it much more predictable than processes on land [Meehl et al., 2013]. Which influence will predominate atmospheric CO_2 predictability? Thus, I formulate the following guiding research questions:

research questions

1. **What are the limits of predictability of air-sea and air-land surface CO2 flux as well as the resulting atmospheric CO_2 mixing ratio?**

2. **What are the roles of the ocean or land carbon cycle in limiting and enabling atmospheric CO_2 predictability?**

method: perfect-model predictability framework

I perform initialized simulations in a perfect-model framework [Griffies and Bryan, 1997]. Predictive skill is claimed when initialized simulations perform better than resampled uninitialized simulations, because of their constrained initial conditions. The last lead year with

predictive skill is called the predictability horizon, which marks a theoretical upper limit of predictability for the given variable in a given prediction system (here MPI-ESM1-2-LR [Mauritsen et al., 2019]).

results: CO_2 flux

I estimate the RMSE-based predictability horizon of the global air-land and air-sea CO_2 fluxes as two years [Fig. II.1a,b], similar to predictive skill evaluated against reconstruction and observations [Li et al., 2019; Lovenduski et al., 2019a; Lovenduski et al., 2019b]. Previous estimates of predictability horizons in a perfect-model were higher [Séférian et al., 2018], but end up comparable when computed by my proposed methodology, which is based on the established prediction frameworks [Goddard et al., 2013; Marotzke et al., 2016; Yeager et al., 2018] [see II.SI.4.3 for comparison of verification frameworks]. The global ocean carbon sink predictability is dominated by the extra-tropical oceans in the North Atlantic, North Pacific and Southern Ocean, which are known for their high-variability low-frequency variations [Landschützer et al., 2016] [Figs. II.2, II.SI.2, II.SI.7]. The global terrestrial CO_2 flux predictability is dominated by the tropical forests and mid-latitudes. Prediction error increases most strongly in the regions susceptible to ENSO [Betts et al., 2018] [Figs. II.2, II.SI.2, II.SI.8].

I specifically focus on an accuracy metric like RMSE, because it takes the magnitudes of the noise and predictability signal into account, whereas correlation measures correspondence independent of scale. For instance, longer regional predictability in the ocean is erased in the noise of global oceanic CO_2 flux. I also find correlation-based predictability of air-sea CO_2 flux of four years and air-land CO_2 flux of three years. But predicting the direction of CO_2 flux has only scientific value. What is of interest for the application by policy-makers is the magnitude of internal variability constrained as aggregated in atmospheric CO_2.

results: atmospheric CO_2

The predictability horizon for the global atmospheric CO_2 mixing ratio is three years [Fig. II.1c]. The spatial distribution of atmospheric CO_2 predictability is homogeneous in the extra-tropics with features of longer predictability up to six years across parts of the tropical oceans. However, due to the sampling uncertainties stemming from only twelve initial conditions, this feature is likely not robust for a meaningful application. Atmospheric CO_2 measured at the Mauna Loa observatory in Hawaii is also predictable up to three years in advance, surpassing the statistical prediction of Betts et al. [2018] in maximum predictive skill lead time and beating their predictive skill at lead year one [Spring and Ilyina, 2020] [Fig. II.SI.1c].

Considering temporally accumulated global surface CO_2 flux as a proxy for global atmospheric CO_2 concentration [see equation 1, and compare figs. II.1c,f], I find that the variations of atmospheric CO_2 only due to the ocean carbon cycle are predictable beyond a decade, whereas they are predictable only for five years for the land

carbon cycle [Fig. II.1d,e]. Therefore, the terrestrial carbon sink limits atmospheric CO_2 predictability.

conclusion: atm. CO_2 potentially predictable for three years

This chapter demonstrates that ESM-based initialized predictions of the global carbon cycle with prognostic atmospheric CO_2 have the potential to deliver skillful multi-year outlooks on the evolution of atmospheric CO_2 mixing ratio, which represent an ahead-of-time verification of potential CO_2 emission reductions. They can also provide longer and more accurate predictions than a statistical regression model [Betts et al., 2018]. Furthermore, I outline the importance to assess CO_2 flux predictability in a distance-based metric like RMSE, when explaining atmospheric CO_2 predictability and confirm the global atmospheric CO_2 predictability limits claimed in section E.4. I conclude:

key findings

1. **Global annual air-land and air-sea CO_2 fluxes are predictable for two years. Atmospheric CO_2 is predictable for three years.**

2. **The land carbon sink is dominated by the tropical forests and limits longer atmospheric CO_2 predictability sustained by the ocean carbon sink, which is mainly limited by the Southern Ocean.**

E.6 PAPER III: RECONSTRUCTION OF THE GLOBAL CARBON CYCLE

motivation

After having established potential predictability in the global carbon cycle, incorporation of initial conditions close to observations is often mentioned as the next step towards near-real-time carbon cycle predictions [Séférian et al., 2018; Ilyina et al., 2021]. Here, I ask how important are carbon cycle initial conditions in determining where the carbon goes next. Thereby explicitly asking whether the carbon cycle signal can be better predicted when carbon cycle initial conditions are constrained.

For state-of-the-art prediction systems, only the climate physics are reconstructed. This *indirect* carbon cycle reconstruction method relies on the assumption that biogeochemical cycles indirectly follow the reconstructed climate variables. While plausible, this assumption is currently untested. The *direct* reconstruction of the carbon cycle by carbon cycle observations is often stated as having the potential to improve carbon cycle prediction systems [Li et al., 2019; Lovenduski et al., 2019a; Ilyina et al., 2021].

Fransner et al. [2020] tackle this challenge from a theoretical standpoint using perfect-twin experiments, where either shuffled ocean carbon cycle initial conditions are integrated with identical physics or identical ocean carbon cycle initial conditions are integrated with slightly perturbed physics. To the surprise of the authors, they find that oceanic carbon cycle initial conditions do not matter for one year ahead predictability. However, this approach does not investigate how initial conditions are generated and how well the initial conditions have in fact been reconstructed - both interesting from the perspective of a modeler and carbon cycle forecaster who wants to understand the initial conditions of their predictions. Here, I take into account the integrated carbon cycle including the land and ocean carbon sink.

Therefore I test indirect and direct reconstructions of the global carbon cycle in a perfect-model reconstruction [Servonnat et al., 2015] and predictability framework [Spring and Ilyina, 2020] and I formulate the following guiding research questions:

research questions

1. **How well can initial conditions be reconstructed in the global carbon cycle?**
2. **Can initialization of biogeochemistry improve predictive skill of biogeochemistry?**

methods: reconstructing an ESM with itself

I therefore conduct reconstruction simulations comparing indirect and direct carbon cycle reconstructions with a focus on the global carbon cycle. I nudge atmospheric and oceanographic dynamics from MPI-ESM-LR into MPI-ESM-LR mimicking an MPI-ESM-LR reanalysis based on perfect MPI-ESM-LR observations. This reconstruction is then compared to the ground truth, the perfect MPI-ESM-LR observations.

Additionally for the direct carbon cycle reconstruction, I nudge oceanic variables determining CO_2 flux, namely dissolved inorganic carbon and total alkalinity. For the lack of a nudging routine in the terrestrial carbon cycle model JSBACH, I reset the restart files each January 1st to the target restart files, which resets the land model to the defined ground truth. I compare how well internal variability is reconstructed against a randomly resampling 95th percentile threshold.

Then I use these reconstructed initial conditions to evaluate implications for initialized predictive skill. To account for offset biases in mean state due to reconstructions, I also remove the mean bias before calculating a debiased accuracy metric. Predictions are calibrated before being given to users [Atger, 2003; Gneiting and Raftery, 2007] and the mean bias reduction is the simplest calibration method.

results: initial conditions reconstruction

Reconstruction of only climate physics introduces systemic biases in climate dynamics themselves [Fig. III.1]. Generally, reconstructions improve association and accuracy with the target [Figs. III.3, III.SI.2], while at the same time introducing regional biases, which are translated into the land and ocean carbon cycles [Fig. III.2]. Indirect carbon cycle reconstruction tracks the carbon cycle in the target better than a resampling 95th percentile threshold. Direct carbon cycle reconstruction improves tracking metrics with respect to the carbon cycle target over large parts of the globe, especially correlation and accuracy after bias reduction. However, aggregated for the global carbon cycle, direct carbon cycle reconstruction only marginally improves accuracy over the indirect method after a mean bias reduction [Fig. III.5].

results: predictive skill after reconstruction

Directly initialized ensembles predict the target better than indirectly initialized ensembles, unless the direct reconstruction further increases the bias as for oceanic pCO_2 and atmospheric CO_2. However, when globally aggregated, the same improvement from directly initialized carbon cycle predictive skill can be achieved by a mean bias reduction [Fig. III.6].

conclusions: confidence for indirect carbon cycle reconstruction

While reconstructions of the carbon cycle involve risks of drifts and biases, the improvements in initial conditions and predictive skill on global scales are limited and similarly achieved by a mean bias correction. Therefore I label these improvements trivial in Spring et al. [2021]. Nonetheless, this result adds confidence to the current practice of indirect reconstruction in carbon cycle prediction systems [Ilyina et al., 2021]. Therefore understanding the physical reconstruction biases and their effects on biogeochemistry is more promising than initializing the carbon cycle to improve carbon cycle predictions. I conclude:

key findings

1. **Indirect reconstruction of carbon cycle by nudging only climate physics tracks the target better than a resampling threshold representing internal variability.**
2. **Direct reconstruction of global carbon cycle initial conditions does not improve predictive skill more than a simple bias reduction method does.**

E.7 CONCLUSIONS

Although my thesis is based on an idealized perfect-model framework, many conclusions can be transferred to real world predictions. First, I demonstrate why this transfer is feasible and then I draw conclusions from each of the three papers.

translating to carbon cycle hindcasts

The largest difference between perfect-model and retrospective hindcast predictions is initialisation. Hindcasts are initialized from observation-based reanalysis products, which are often created using a different ESM in reconstruction simulations and hence follow a different climatology. "Pulling" the ESM towards observations yields accurate initial conditions and hence the hindcasts and forecasts can be evaluated against observations. When hindcasts are evaluated against observations, ESMs suffer from insufficient temporal or spatial resolution and lack processes which would fully represent the real world. While the global carbon cycle is subject to these deficiencies, at least the climatology of the globally aggregated carbon cycle is bound by observational constraints in the process of tuning the carbon cycle. By design, the perfect-model framework does not suffer from the above mentioned deficiencies other that it cannot be used to predict the real world.

Both initialized perfect-model and hindcast predictions use the same ESMs, i. e. the same boundary conditions, spatio-temporal resolution and processes. Therefore, both also share the ultimate reason for ensemble members to diverge: the chaotic nature of the atmosphere [Lorenz, 1965]. Here, perfect-model frameworks can more cleanly identify the processes driving predictability and are free from biases with respect to their verification.

paper I: internal variability may disguise the detection of CO_2 emission reductions in the near-term

With the unprecedentedly large number of ensemble simulations and a novel method to visualize the expected variability of atmospheric CO_2 in concentration-driven simulations, Spring et al. [2020] shows that pronounced internal variability has the capacity to disguise CO_2 emission reductions for up to a decade. While the detection of mitigation is found to be delayed much longer for global mean surface temperature [Tebaldi and Friedlingstein, 2013; Marotzke, 2019], such causation of mitigation detection of atmospheric CO_2, the most directly linked quantity, was previously unknown[20]. In particular, causation of reduced atmospheric CO_2 trends is far from certain over the pentadal periodicity of the global stocktake. Hence, the global stocktake should focus on CO_2 emissions aiming for net zero CO_2 emissions. Recent

20 The publication timelines of Spring et al. [2020] and Schwartzman and Keeling [2020] overlapped, where Schwartzman and Keeling [2020] refined the method by Peters et al. [2017]

observations of atmospheric CO_2 should be interpreted with caution.

paper II: global atmospheric CO_2 is predictable for up to three years

Spring and Ilyina [2020] shows that ESM-based predictions can constrain internal variations in atmospheric CO_2 by up to three years. This, to my knowledge, first estimate of predictive skill of atmospheric CO_2 marks the upper achievable limit due to the chaotic Earth System. As the perfect-model framework includes perfect initialization and a perfect model, slightly lower predictive skill is expected to be found for real-world predictions. I furthermore separate the predictive skill into a predictability-sustaining ocean signal and a predictability-limiting land signal. Therefore, more accurate predictions of the land carbon sink will help to constrain near-term predictions of atmospheric CO_2. By reducing the internal variability uncertainty, the expected atmospheric CO_2 signal can be predicted, given specific CO_2 emission reductions. This will give confidence to short-term outlooks towards future observations and will set realistic expectations of near-term atmospheric CO_2 evolution. Eventually, these CO_2 predictions can guide policy-makers towards ahead-of-time constrained projections of the efficacy of mitigation efforts.

paper III: direct carbon cycle reconstruction does not improve initial conditions and predictability

In order to generate real-world predictions, the global carbon cycle needs to be reconstructed in an ESM without triggering (too large) artificial drifts introduced by assimilation of observations. Hence, Spring et al. [2021] makes a first attempt to quantitatively assess reconstruction skill for the carbon cycle assuming perfect carbon cycle observations in an idealized perfect-model framework. I find limited improvements after initializing the carbon cycle directly. Spring et al. [2021] thereby challenge the previous hypothesis that carbon cycle predictions are unconstrained from an initial conditions point of view [Séférian et al., 2018; Ilyina et al., 2021]. I concluded that if the perfect-model reconstruction does not improve initial conditions sufficiently to increase predictive skill, the reconstruction towards observation-based carbon cycle reanalysis would likely trigger even worse drifts. Practically speaking, we have to accept that drifts occur [see Li et al., 2019, SI Fig. 1] and learn how to minimize their consequences.

Until now predictive skill was mostly based on the bias-invariant anomaly correlation coefficient. The mean bias correction I apply only marks the most basic post-processing technique, bringing considerable improvements and allowing bias-sensitive metrics to be verified after bias reduction. Also the currently indirectly reconstructed carbon predictions can benefit from bias reduction. Likely, the effects of climate physics reconstructions from different reanalyses on the carbon cycle yield larger potential to improve predictions of the global carbon cycle. Here Spring et al. [2021] serves as a test-bed for understanding how an ESM responds to reconstructions.

Reconstruction of the Global Carbon Cycle in ESMs is also important for the Global Carbon Budget, in which separate reanalysis-forced simulations for land and ocean aim to explain "where does the carbon go?" [Marotzke et al., 2017]. Ideally, the Global Carbon Budget simulations should model prognostic atmospheric CO_2 based on past emissions. In this way, the partitioning of carbon between atmosphere, land and ocean is internally consistent within the given ESM, which may reduce the carbon budget imbalance [Friedlingstein et al., 2020], which is required for the verification of real-time CO_2 emission reductions [Peters et al., 2017; Schwartzman and Keeling, 2020].

In conclusion, this dissertation based on idealized Earth System Model simulations tells the story how initialized predictions have the potential to constrain internal variability of the global carbon cycle.

E.8 OUTLOOK

I arrange the implications and inferences from this thesis in two parts: the first, relevant to the carbon cycle variability and predictability research community; and the second, relevant to the climate policy of CO_2 emissions.

E.8.1 *Future Research Opportunities*

In the following, I deduce a few subsequent research questions for fellow carbon cycle variability & predictability research studies and furthermore outline related applications:

My first paper Spring et al. [2020] relies only on one model and one possible transition in CO_2 emission reduction pathways, however the trend sign method is very sensitive towards the ratio of internal variability and the magnitude of emission reductions over time. Therefore, a study exploring the multitude of modeled internal variability [also including observations-based internal variability like Schwartzman and Keeling, 2020] versus a range of CO_2 emission reduction pathways would better set the detectability of CO_2 emission reductions in atmospheric CO_2 in context.

Furthermore I focus on annually and globally averaged atmospheric CO_2. The actual detectability might be larger for specific locations [requiring prognostic atmospheric CO_2 simulations as in Spring and Ilyina, 2020] or seasons [analogous to the recovery of the ozone hole, which was first detectable in September as reported by Solomon et al., 2016].

Following Spring and Ilyina [2020], there are still many properties of carbon cycle predictability than remain unknown. All related papers so far focused on annual aggregates of CO_2 flux. Based on the reported predictability limits of two to three years, seasonal or monthly aggregation is more accurate and useful for most end-users. In general, the carbon cycle prediction community can learn a lot from the sea-ice predictability community, which operates on similar time-scales from months to three years but is methodologically more advanced [Tietsche et al., 2013; Hawkins et al., 2016; Bushuk et al., 2018]. Future carbon cycle research question include: Which target months from which initialisation are better predictable than others? How does carbon cycle predictability depend on initial states? How does near-term predictability change in a high CO_2 concentration world? Which processes sustain and limit predictability? How model-dependent [Ilyina et al., 2021] and model generation-dependent [Borchert et al., 2020] is predictive skill?

Furthermore, understanding of the spatio-temporal near-term variability of atmospheric CO_2 is vital to understand its predictability. However, challenges in modeling near-term variability of atmospheric CO_2 remain [Keeling et al., 1989b; Keppel-Aleks et al., 2013], e. g. in response to climate indices other than ENSO [Jones et al., 2001; Wang et al., 2018] and compared to observations [Keeling et al., 1989a; Gier et al., 2020].

Following Spring et al. [2021], the implications of climate reconstructions from different climate reanalysis products or direct observations (via ensemble Kalman filter assimilation [Brune et al., 2017; Brune and Baehr, 2020]) on the global carbon cycle should be reassessed. Especially the switch between a reconstruction simulation and the freely running model, triggering adjustments from the preferred freely running model climatology to the reconstructed climatology and reverse [Spring et al., 2021], should be taken into account when deciding for the best matching climate reanalysis of the given land and ocean carbon cycle models [Saito et al., 2011; Lee and Biasutti, 2014; Hua et al., 2019].

Spring et al. [2021] show that a simple mean bias reduction can improve the quality of initial conditions and that a lead-time dependent bias reduction effectively reduces drifts from initialization. Likely, more advanced methods from the physical prediction community can be adopted to improve predictive skill [Pasternack et al., 2018; Feldmann et al., 2019].

potential for predictions in biogeochemistry

On a general note, when restricting carbon cycle predictions to predictions of atmospheric CO_2 and CO_2 fluxes, I cannot think of any direct application other than carbon management and monitoring derived from climate policy. But methodologically and tool-wise carbon predictions are identical to predictions of biogeochemistry, which is an emerging field with a large range of application opportunities: marine biogeochemical predictions could be used for fishing management [Park et al., 2019; Krumhardt et al., 2020; Koul, 2020] and terrestrial biogeochemical predictions could predict agricultural yields [Moemken et al., 2021]. Here, the biogeochemical ESM-based prediction community still needs to prove that biogeochemical predictions based on ESMs have additional value over offline biogeochemistry or domain specific (e. g. fishery) models, which are just fed with the physical climate output from ESM-based climate prediction systems [Tommasi et al., 2017a,b; Payne et al., 2017; Årthun et al., 2018; Bett et al., 2020]. Only this would prove that the biogeochemistry coupled to climate needs to be predictable instead of just coupling biogeochemistry to physics predictions.

step towards the usability of predicts

If seasonal-to-decadal predictions of biogeochemistry aim to inform

users, forecasters need to establish direct communication with users [Merryfield et al., 2020]. Furthermore, scaled up hindcast statistics (ensemble members and initializations) are required to establish robust skill [Palmer, 2020]; especially for probabilistic metrics, which evaluate the full range of the ensemble rather than just the ensemble mean and show whether a prediction is accurate and reliable and hence valuable. Reliability, discrimination, the contingency table and the continuous ranked probability score (CRPS) are implemented in CLIMPRED [Brady and Spring, 2021]. Such a verification would also be needed for real-world atmospheric CO_2 predictions, which do not yet exist.

E.8.2 *Implications for Climate Policy*

How should policy-makers interpret this dissertation on variability and predictions? I partition the implications for climate policy into the near-term and long-term.

internal variability determines near-term atm. CO_2

This dissertation demonstrates that near-term internal variability in the global carbon cycle is so large that realistic CO_2 emission reductions cannot, with certainty, cause policy-relevant five-year reduced trends in atmospheric CO_2. Analogous to the current COVID-19 CO_2 emission reductions, it is important that our expectations are managed and science provides robust estimates including uncertainties [Marotzke, 2019; Samset et al., 2020]. The first global stocktake will take place in 2023. Policy-makers should not be distracted by the atmospheric CO_2 signal[21] and focus strictly on anthropogenic CO_2 emissions.

putting numbers in perspective

What are the opportunities and limitations of carbon cycle predictions? In the near-term, internal variability is the largest uncertainty of the global carbon cycle. Carbon cycle predictions can constrain the envelope of internal variability for a few years. This is a substantial reduction opportunity because uninitialized, unconstrained internal variability (2σ = 4.0 PgC/yr = 1.8 ppm) [Fig. E1, green errorbar] is twice as large as the infamous historical mean Global Carbon Budget imbalance[22] (2σ = 1.6 PgC/yr = 0.8 ppm).

21 The quantification of ^{14}C in atmospheric CO_2 provides an alternative method to separate natural carbon cycle-based variability from fossil-fuel CO_2 emissions, as fossil-fuel CO_2 emissions are free of ^{14}C [Levin et al., 2003; Turnbull et al., 2006]. Basu et al. [2020] verifies fossil fuel CO_2 emissions of the United States using inversion modeling. Increasing coverage of the sparse ^{14}C observations network is needed to reduce biases. Instead, ESMs could resolve ^{14}C, which is implemented only in ocean models [Galbraith et al., 2011; Jahn et al., 2015; Tjiputra et al., 2020].

22 The Global Carbon Budget imbalance accumulates to 0% of all CO_2 emissions for 1859-2019 in Friedlingstein et al. [2020], but fluctuated between 1959 and 2019 by 1.6 PgC/yr = 0.8 ppm/yr (2σ) and is also reported with an uncertainty of 1.6 PgC/yr = 0.8 ppm/yr (2σ).

However, regarding mitigation, CO_2 is not the only climate forcing. While other greenhouse gases show earlier detection potentials [Samset et al., 2020], CO_2 is the most important one on the long run[23]. Aiming for a net zero emissions goal, the current CO_2 emissions of 10 PgC/year have a long way to go for zero or likely negative emissions. The current level of CO_2 emissions, corresponding to an increase of 2-3 ppm/year atmospheric CO_2, dwarfs the potential of uncertainty reduction from carbon cycle prediction.

Furthermore, climate policy and the public should turn the spotlight on CO_2 emissions [supported by EU projects: CHE, 2017; VERIFY, 2018, CoCO2]. Historically, CO_2 emissions have mostly been reported annually [Andrew, 2020; Friedlingstein et al., 2020]. But without (continuous) data, there can be no management and no verification of progress [Janssens-Maenhout et al., 2020]. The recent Global Carbon Budget also stresses the need for up-to-date data [Friedlingstein et al., 2020]. Only recently have Liu et al. [2020] established a near-real-time monitoring of CO_2 emissions, which is an important step for CO_2 emissions reporting and management.

how near-term predictions can help the long-term solution

On the one hand, the fate of atmospheric CO_2 in the long-term will be determined by anthropogenic CO_2 emissions. On the other hand, climate-carbon feedbacks are uncertain [Friedlingstein et al., 2013] and emissions from permafrost thawing are not even included in most model projections [Schuur et al., 2015; Burke et al., 2020; Turetsky et al., 2020]. These known unknowns are not directly influenced by climate policy but will unfold as the Earth warms. This uncertainty of the remaining carbon budgets stresses the point that CO_2 emissions, the primary driver of atmospheric CO_2 growth and influenced by climate policy, need to approach zero as soon as possible to achieve the Paris Agreement to limit global mean warming to 1.5 to 2°C [Friedlingstein et al., 2014; UNFCCC, 2015; Rogelj et al., 2016b].

The only long-term solution to avoid dangerous climate change is mitigation of greenhouse gas emissions [UNFCCC, 2015]. How do initialized predictions, which operate on the near-term, help policy-makers in that regard? First, policy-makers need to be better informed about the role of near-term variability in the carbon cycle. Only then will they acknowledge that future near-term CO_2 predictions with ESMs have the capacity to constrain CO_2 uncertainty for up to three years. These predictions bring clarity about whether the current and near-future path is on track to limit warming to below 1.5°C or 2°C or not. While Peters et al. [2017] and Schwartzman and Keeling [2020] discuss real-time verification of CO_2 emission reductions, initialized CO_2 predictions can deliver even ahead-of-time anticipation. Instead of waiting to see what will happen, this approach is more action-

23 Peters [2018] argues "that obsession with cumulative carbon emissions and carbon budgets has perhaps diverted attention from gains through non-CO_2 mitigation."

orientated, data-driven and allows for more frequent re-adjustments [Rogelj et al., 2019]. Hence, relying on initialized near-term predictions can prevent losing yet another decade of progressive mitigation, when the efficacy of mitigation has not yet come into force. Due to internal variability and the delayed impacts on temperature [Tebaldi and Friedlingstein, 2013; Marotzke, 2019], there will not be enough time to wait until mitigation measures show their effect, if the Paris Agreement is to be fulfilled.

Connecting to the introductory quote and speaking for the value of predictions [Murphy, 1993], near-term CO_2 predictions have no value, unless they inform and hopefully influence decisions of policy-makers.

INHERENT UNCERTAINTY DISGUISES ATTRIBUTION OF REDUCED ATMOSPHERIC CO_2 GROWTH TO CO_2 EMISSION REDUCTIONS FOR UP TO A DECADE

The attached article has been published in *Environmental Research Letters*:

Spring, A., Ilyina, T., & Marotzke, J. [2020]. Inherent uncertainty disguises attribution of reduced atmospheric CO_2 growth to CO_2 emission reductions for up to a decade. *Environmental Research Letters*, *15*[11], 114058. doi:10/ghmjp8

Author contributions: A.S. performed the analysis, created the figures and drafted the manuscript. A.S. conceived the study derived from Marotzke [2019]. T.I. and J.M. contributed in manuscript editing and revisions. J.M. provided overall scientific guidance over the topic of causation.

Inherent Uncertainty Disguises Attribution of Reduced Atmospheric CO_2 Growth to CO_2 Emission Reductions for up to a Decade

Aaron Spring[1,2], **Tatiana Ilyina**[1] **and Jochem Marotzke**[1]
[1] Max Planck Institute for Meteorology, Hamburg, Germany
[2] International Max Planck Research School of Earth System Modelling, IMPRS, Hamburg, Germany

Submitted: 13 December 2019, Revised: 28 September 2020, Accepted: 23 October 2020, Published: 24 November 2020

ABSTRACT

The growth rate of atmospheric CO_2 on inter-annual time scales is largely controlled by the response of the land and ocean carbon sinks to climate variability. Therefore, the effect of CO_2 emission reductions to achieve the Paris Agreement on atmospheric CO_2 concentrations may be disguised by internal variability, and the attribution of a reduction in atmospheric CO_2 growth rate to CO_2 emission reductions induced by a policy change is unclear for the near term. We use 100 single-model simulations and interpret CO_2 emission reductions starting in 2020 as a policy change from scenario Representative Concentration Pathway (RCP) 4.5 to 2.6 in a comprehensive causal theory framework. Five-year CO_2 concentration trends grow stronger in 2021-2025 after CO_2 emission reductions than over 2016-2020 in 30% of all realizations in RCP2.6 compared to 52% in RCP4.5 without CO_2 emission reductions. This implies that CO_2 emission reductions are sufficient by 42%, necessary by 31% and both necessary and sufficient by 22% to cause reduced atmospheric CO_2 trends. In the near term, these probabilities are far from certain. Certainty implying sufficient or necessary causation is only reached after, respectively, ten and sixteen years. Assessments of the efficacy of CO_2 emission reductions in the near term are incomplete without quantitatively considering internal variability.

I.1 INTRODUCTION

Substantial year-to-year variations in the growth rate of global atmospheric CO_2 concentrations show variations that cannot be explained by land-use changes, fossil fuel emissions or the increase of carbon sink capacities due to increasing atmospheric CO_2 concentrations [Friedlingstein et al., 2019; Peters et al., 2017]. The variations originate instead from the variability of the global carbon cycle in response to climate variability, which is inherent to the physics of the Earth System. For instance, the variations of the tropical land carbon sink is dominated by the El Niño-Southern Oscillation [Jones et al., 2001; Zeng et al., 2005], and the pronounced Southern Ocean carbon sink is susceptible to changes in atmospheric circulation patterns [Landschützer et al., 2015; McKinley et al., 2017]. Therefore, this internal variability of the global carbon cycle in atmospheric CO_2 may disguise the detection of potential CO_2 emission reductions in atmospheric CO_2 observations. But CO_2 emission reductions are required to achieve the targets of the Paris Agreement [UNFCCC, 2015]. Here we ask what the probability is that a slowdown in atmospheric CO_2 growth is attributable to a policy change implementing CO_2 emission reductions as the difference between Representative Concentration Pathway (RCP) 4.5 and RCP2.6, in the face of internal climate variability. This question becomes policy-relevant as policy-makers assess the efficacy of CO_2 emission reductions in the global stocktake every five years [Peters et al., 2017; Schwartzman and Keeling, 2020]. Furthermore, we ask after how many years this policy change will cause atmospheric CO_2 growth rates to slow down for certain.

The challenge of emissions reductions verification in atmospheric CO_2 concentrations was first outlined by Peters et al. [2017]. We address this challenge by using a large ensemble of Earth System Model (ESM) simulations [Maher et al., 2019]. We integrate 100 simulations based on the code of a single ESM with slightly perturbed initial conditions that serve as different realizations of the Earth System. Our analysis compares RCP4.5, which is close to the pledged and current policies until 2035 [Rogelj et al., 2016b; Hausfather and Peters, 2020], with an emission reductions scenario compatible with the Paris targets under RCP2.6 [Fig. I.SI.4]. We attribute a reduction of trend in atmospheric CO_2 concentrations to CO_2 emission reductions in the comprehensive causation framework of Pearl [2000] and Hannart et al. [2016]. In the context of CO_2 emission reductions, necessary causation means that a factual trend reduction would not have occurred without a policy change. By contrast, sufficient causation implies that while a policy change may trigger a trend reduction, this trend reduction is not certain.

We go beyond approaches in previous studies [Tebaldi and Friedlingstein, 2013; Peters et al., 2017; Marotzke, 2019; Samset et al., 2020;

Schwartzman and Keeling, 2020] by comprehensively diagnosing atmospheric CO_2 variability in an ESM, which is compatible with the terrestrial and oceanic carbon sinks variations. The recently formalized emissions reductions verification of Schwartzman and Keeling [2020] uses an autoregressive model based on the observed carbon imbalance and a different statistical framework. While Tebaldi and Friedlingstein [2013] only focus on causation in a necessary causation sense, we here complete the probabilistic setting by asking also about sufficient causation. We compare two RCP scenarios in a single-model framework; formally only internal variability may undermine the detectability of CO_2 emission reductions. Assessing the contribution of our quantitative results against structural model uncertainty and imperfections is left for future study.

From a policy-maker's perspective looking into the near-term future, necessary and sufficient causation of CO_2 emission reductions slowing down atmospheric CO_2 trends deal with two different questions [Pearl, 2000; Hannart et al., 2016]:

1. Will a policy change towards CO_2 emission reductions suffice to slow down atmospheric CO_2 growth? Other factors, such as a weakening uptake by the natural carbon sinks, may induce an increase in atmospheric CO_2 growth despite policy measures. From the viewpoint of a pathway without CO_2 emission reductions, the uncertainty in this question is based on sufficient causation.

2. Would a factual atmospheric CO_2 growth slowdown have occurred even without the policy change? This question asks whether the policy change was necessary to achieve the policy goal. From the viewpoint of a factual pathway of CO_2 emission reductions and a factual slowdown, the uncertainty in this question is based on necessary causation.

Based on this causation framework, we obtain probabilities that a policy change causes atmospheric CO_2 trends to decline. However, this causation may be far from certain depending on the time-scale assessed. Should CO_2 emission reductions not soon lead to reduced atmospheric CO_2 growth trends, we might face a debate analogous to the warming hiatus debate [Lewandowsky et al., 2015; Fyfe et al., 2016] about why CO_2 rises faster despite falling emissions. Therefore, scientists need to communicate the role of internal variability to policymakers and the public [Deser et al., 2012].

Marotzke [2019] shows the uncertain effect of emission reductions on global mean surface temperature (GMST) 15-year trends. As atmospheric CO_2 drives the forced GMST signal, the emissions reduction signal should become detectable earlier in atmospheric CO_2. Analyzing the effect of individual climate forcers, Samset et al. [2020]

confirms that anthropogenic CO_2 has the highest potential for emission reduction detection.

In our study, we also ask after how many years internal variability can still obscure the identification of CO_2 emission reductions in atmospheric CO_2. In other words, how long does it take until certainty arises in causation? This is a distinctly different question than the classical time-of-emergence of anthropogenic signals, which asks on which timescales the climate change signal emerges from natural variability [McKinley et al., 2016]. Here, we ask on which time-scales a forcing change induced by this policy change causes a climate response considering sufficient and necessary causation [Marotzke, 2019]. These time-scales of CO_2 emission reduction detection might be longer than the periodicity of the global stocktake in which policy-makers will assess the efficiency of mitigation measures.

I.2 METHODS

I.2.1 *Causation Attribution Framework*

To identify whether CO_2 emission reductions cause a reduction in atmospheric CO_2 growth, we apply the concept of event causation [Pearl, 2000; Hannart et al., 2016; Marotzke, 2019]. We use the scenario RCP2.6 as implementing CO_2 emission reductions and RCP4.5 for the near-term future without CO_2 emission reductions [for a detailed justification see section I.2.2]. Taking a decline in atmospheric CO_2 growth as an effective consequence of CO_2 emission reductions policy, we define a reduction in the linear trend in global atmospheric CO_2 concentration as the policy goal, comparing the period before emission reductions started with the period afterwards. While this response is expected as the forced response averaged over all realizations, the trend of single ensemble members could potentially increase due to internal variability. For a five-year trend period and a scenario separation in 2020, compatible with the switch from RCP4.5 to RCP2.6, we hence compare the trends 2016-2020 and 2021-2025. The fraction of responses in a given scenario s yields the probability of trend reduction P_{RCPs}. The two scenarios serve as either the real world, labeled as factual, or the alternative world, labeled as counter-factual. The probabilities of trend reduction can be translated into a probability $P_{\{S,N\}}$ that the trend reduction is caused by the policy change [Pearl, 2000; Hannart et al., 2016]:

- In a currently pledged policy pathway (factual RCP4.5 world) without CO_2 emission reductions in near-term, we ask in advance whether CO_2 emission reductions (a policy change towards the counterfactual RCP2.6) would be *sufficient* to cause a reduction in atmospheric CO_2 trends. The probability P_S means

how likely a policy change would be sufficient to cause a reduced trend:

$$P_S = \frac{P_{RCP2.6} - P_{RCP4.5}}{1 - P_{RCP4.5}}. \tag{2}$$

In our case of CO_2 emission reductions, P_S is important when answering questions whether the policy goal of reduced atmospheric CO_2 growth will be achieved from the perspective of a planner.

- Considering CO_2 emission reductions in RCP2.6 as the factual where in retrospect atmospheric CO_2 trends have indeed declined and no CO_2 emission reductions in RCP4.5 as the counterfactual world, the probability P_N that the policy change was necessary to cause the trend reduction is:

$$P_N = 1 - \frac{P_{RCP4.5}}{P_{RCP2.6}}. \tag{3}$$

In our case of CO_2 emission reductions, P_N is important when answering questions whether the policy goal would not have been reached without the policy change.

- Combining the two aforementioned, P_{NS} describes the probability that the policy change is both *necessary and sufficient* to cause the respective trend reduction:

$$P_{NS} = P_{RCP2.6} - P_{RCP4.5}. \tag{4}$$

This strongest causation probability P_{NS} means how likely a reduced trend would occur in case of a policy change and would not occur without.

These probabilities hence describe probabilities that trend reductions over a given trend length are caused by the policy change, but how long do these trends need to be in order to be virtually certain that CO_2 emission reductions caused them? To answer this question, we define *Time to Detection of CO_2 emission reductions in a causation sense* $D_{\{S,N\}}$ as the trend length around CO_2 emission reductions start in 2020 for which $P_{\{S,N\}} > 99\%$, using the probability framing of Mastrandrea [2010]. This time-scale marks the maximum range of influence of internal variability over changes in the forced signal due to a policy change.

1.2.2 *Choice of Scenarios*

We identify RCP2.6 as roughly compatible with the Paris targets [van Vuuren et al., 2011]. Compared to the pre-industrial control, MPI-ESM Grand Ensemble warms by 1.4±0.2°C in RCP2.6 and 2.2±0.2°C

in RCP4.5 until the end of the century [Suarez-Gutierrez et al., 2018; Maher et al., 2019]. Anthropogenic CO_2 emissions in RCP2.6 are increasing until 2020; after 2020, emissions are expected to decrease by 2% per year in RCP2.6 until 2030 [Fig. I.SI.3]. By contrast, RCP4.5 has similar anthropogenic CO_2 emission levels as RCP2.6 until 2020 and continues a moderate emissions increase until 2040 with a 1% per year increase until 2030 [Figs. I.SI.3, I.SI.4]. This scenario was designed to reach a forcing stabilization at the end of this century [Thomson et al., 2011] at about 3°C warming. Although RCP8.5 is closer to the most recently recorded combined land-use and fossil-fuel emissions, we choose RCP4.5 as a reference scenario, because the differences between RCP4.5 and reported emissions until 2018 originate in land-use change whereas fossil-fuel emissions match [Fig. I.SI.3]. More importantly, the levels of projected fossil-fuel emissions based on current and pledged policy until 2035 parallel RCP4.5 [Rogelj et al., 2016b; Hausfather and Peters, 2020].

In the above-described comparison under the causal theory framework, we compare trends before and after this policy change to assess causality of this policy change on changes in trends with respect to the period before the policy change. This policy change is assumed to happen in one scenario (RCP2.6), and not in the other one (RCP4.5). Therefore, we require simulations under mostly identical forcing before this policy change. We assume that this policy comes into effect as implemented by the RCP scenarios [Meinshausen et al., 2011]. We identify the combination of RCP2.6 vs RCP4.5 with a scenario split in 2020 as suitable scenario comparison. This scenario combination and timing also describes the present quest aiming for an at most 2°C warmer world with net emissions reduction of 3% per year over 10 years.

Comparing RCP2.6 with RCP8.5 would be another possible combination. However, RCP8.5 entails higher fossil-fuel CO_2 emissions than recently observed and much higher levels than what current policies pledge for until 2035 [Rogelj et al., 2016b; Hausfather and Peters, 2020]. Furthermore, RCP2.6 and RCP8.5 separate at a time when emissions in RCP2.6 still grow. This would make the definition of climate event as trend reduction awkward, if our goal is to investigate the effect of policy on possible trend reductions. Therefore we compare RCP2.6 and RCP4.5 and include the comparison against RCP8.5 in the supplementary material.

I.2.3 *Large-Ensemble Simulations*

The Max-Planck-Institute Earth System Model (MPI-ESM) Grand Ensemble comprises 100 members started from different initial conditions branched off a pre-industrial control simulation [Maher et al., 2019]. MPI-ESM contains comprehensive terrestrial and oceanic carbon cycle

sub-models, which capture the dynamics of the global carbon cycle [Ilyina et al., 2013; Schneck et al., 2013; Li and Ilyina, 2018]. The Grand Ensemble is based on a model version close to the CMIP5 version [Giorgetta et al., 2013]. To our knowledge this is the largest ensemble of comprehensive climate models available up to date [Branstator and Selten, 2009; Kay et al., 2015; Rodgers et al., 2015; Kirchmeier-Young et al., 2016; Frankignoul et al., 2017; Stolpe et al., 2018]. Its statistics have proven to be useful in investigating internal variability in the Southern Ocean carbon sink [Li and Ilyina, 2018] and enable a 1% resolution for climate event probabilities [see section I.2.1]. From the year 2006 onward, the 100 historical simulations are extended under RCP2.6 and RCP4.5 [Meinshausen et al., 2011; Taylor et al., 2011]. The Grand Ensemble simulations are forced with a scenario-dependent prescribed atmospheric CO_2 concentration, aerosols, non-CO_2 greenhouse gases and land-use change [Meinshausen et al., 2011; Taylor et al., 2011].

1.2.4 *Diagnostic Atmospheric CO_2 Concentration*

CO_2 concentration-driven simulations do not represent a variable atmospheric CO_2 concentration tracer. To quantify the expected variations in global atmospheric CO_2 concentration that are compatible with variations of the global land and ocean carbon sinks, we diagnose a virtual tracer of global atmospheric CO_2 based on the changes due to internal variability of the land and ocean carbon sinks in atmospheric CO_2 [Fig. I.SI.2]. The global residual CO_2 flux $G_{i,s}$ is the difference of CO_2 flux $F_{i,s}$ of the each ensemble member i to the ensemble mean of CO_2 flux F_s:

$$G_{i,s} = \sum_{global} (F_{i,s} - \frac{1}{M} \sum_{m=1}^{M} F_{m,s}) \tag{5}$$

where $M = 100$ is the number of ensemble members and i the number of a single ensemble member and s the scenario. The ensemble mean F_s is subject to all forcings (anthropogenic fossil-fuel CO_2 emissions, non-CO_2 emissions, land-use change, aerosols) on CO_2 flux, but *no* internal variability. The remaining residual shows the variations of CO_2 flux around neutral flux *only* due to internal variability. The forced atmospheric CO_2 signal f_s is scenario s-dependent and generated by a simplified climate model fed with emissions from integrated assessment models [Meinshausen et al., 2011] incorporating the strengthening of the carbon sinks with higher CO_2 concentrations and land-use CO_2 emissions. This internal variability component of time-accumulated global CO_2 flux is then superimposed on the smooth

atmospheric CO_2 forcing f_s and defines internally varying diagnostic global atmospheric CO_2 concentration $XCO_{2,i,s}$:

$$XCO_{2,i,s}(t) = \sum_{t'}^{t} G_{i,s}(t') \cdot \frac{\text{ppm}}{2.124\text{PgC}} + f_s. \tag{6}$$

We assume that the internal variability of the global carbon cycle to be driven by climate variability. This ignores the short-term effects of atmospheric CO_2 variability on CO_2 flux as in all concentration-driven simulations. Explicitly, for diagnostic atmospheric CO_2, we use as forcing f_s the concentration scenarios generated by the simplified climate model and not directly the CO_2 emissions generated by the integrated assessment models. This assumes that the emission scenarios from the integrated assessment model roughly match the resulting concentration scenarios from the simplified climate model [Fig. I.SI.9]. The hereby diagnosed variations of global atmospheric CO_2 capture the observed global atmospheric CO_2 variations [Figs. I.1, I.SI.7; Spring and Ilyina, 2020]. For a detailed method description and verification in emission-driven simulations, see Supplementary Information section I.SI.1.

1.2.5 *Method Limitations*

The generalisability of our results strongly depend on the strength and timing of the CO_2 emission reductions underlying the two compared scenarios, where causation probabilities $P_{\{N,NS\}}$ are even more sensitive than the probabilities of reducing atmospheric CO_2 trend P_{RCPx} in scenario x. Here, we present one special case of CO_2 emission reductions as the difference between RCP4.5 and RCP2.6 representing a net 3% annual emission reductions until 2030. There is an active debate whether RCP4.5 [argued for by Hausfather and Peters, 2020] or RCP8.5 [argued for by Schwalm et al., 2020] tracks the current anthropogenic CO_2 emissions pathway best. Also the attribution probabilities are contingent upon whether the climate model simulates realistic magnitudes of internal variability [Marotzke, 2019]. Furthermore, our attribution method focuses only on one observable variable under internal variability, although atmospheric CO_2 is the most important indicator for CO_2 emission reductions. Lastly, we use the atmospheric CO_2 concentration prescribed to MPI-ESM generated by the simplified climate model based on CO_2 emission scenarios from the integrated assessment model and not the CO_2 emissions from the integrated assessment models themselves to calculate diagnostic atmospheric CO_2 Meinshausen et al. [2011]. While this is consistent with the forcing applied to the the climate system in MPI-ESM, this leads to small differences between in the compatible emissions of concentration-driven RCPs and actual CO_2 emissions as discussed by Jones et al. [2013] and demonstrated for MPI-ESM [Fig. I.SI.9].

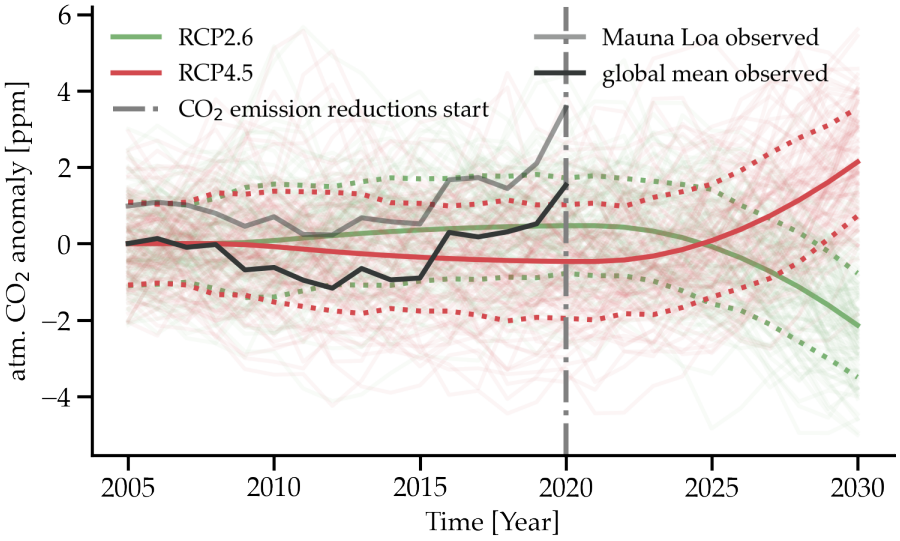


Figure I.1: **The difference in diagnostic atmospheric CO_2 concentration between CO_2 emission reductions scenario RCP2.6 (green) and currently most likely scenario for the near-term RCP4.5 (red) appears after the start of CO_2 emission reductions in 2020 (vertical dashed gray line).** Both scenarios are detrended with the combined scenario mean forcing, see the otherwise identical Fig. I.SI.4 without detrending. Individual ensemble members are shown in faded colors. The forced response (solid green and red) in atmospheric CO_2 follows the prescribed atmospheric CO_2 forcing derived from the a simplified climate model with emissions from integrated assessment models [Meinshausen et al., 2011]. The variation of global mean (black) and Mauna Loa (gray) atmospheric CO_2 observations by NOAA/ESRL [Dlugokencky and Tans, 2019], mostly fall within the ensemble standard deviation (dotted green and red).

I.3 PROBABILITY OF CO_2 EMISSION REDUCTIONS CAUSING CHANGES IN ATMOSPHERIC CO_2 GROWTH TREND

We first assess the frequency distributions of five-year trends in atmospheric CO_2. These distributions over the period 2016-2020 in RCP2.6 and RCP4.5 are nearly indistinguishable [Figs. I.1, I.2a,d]. The most recent 2015-2019 observations-based estimate for global atmospheric CO_2 [Dlugokencky and Tans, 2019] trend is in the upper tercile and thereby captured by our model [Fig. I.2a,d, I.SI.7]. Comparing the distributions before and after CO_2 emission reductions onset in 2020 in RCP2.6, we find overlapping distributions with a tendency towards lower trends after CO_2 emission reductions [Fig. I.2a,b]. The ensemble mean responds to CO_2 emission reductions with a decrease in trend of one ppm over five years. The trend reduces in 70 ensemble members, resulting in $P_{RCP2.6}$=70% [Fig. I.2c]. This implies that with a 30% probability, atmospheric CO_2 growth will strengthen despite emissions reductions. In RCP4.5, the distributions of atmospheric CO_2

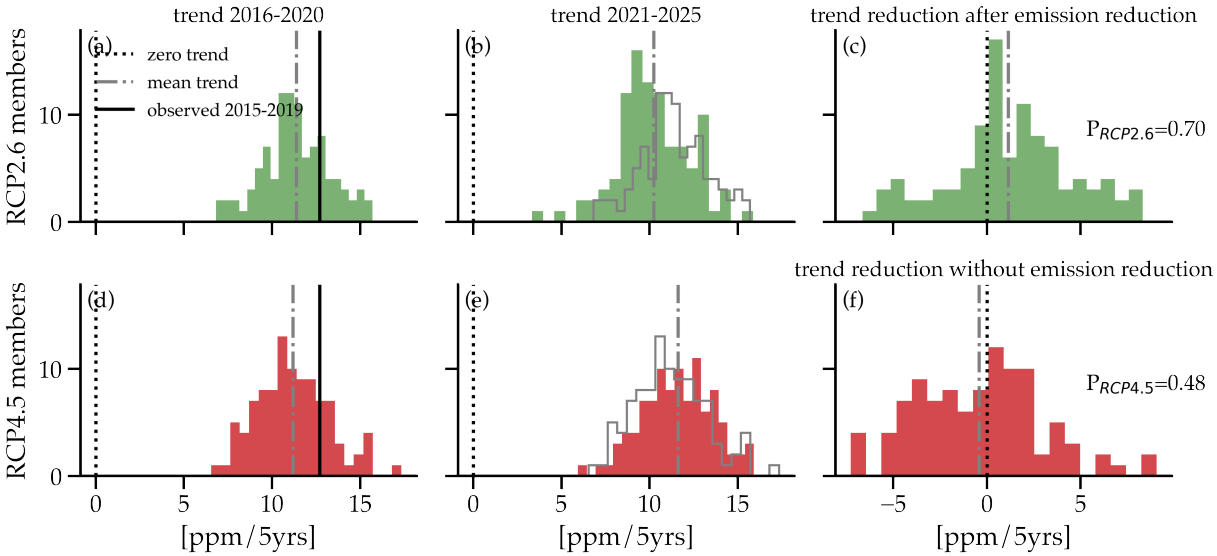


Figure I.2: **Probability distributions of 5-year trends from 2016 to 2020 and 2021 to 2025 in diagnostic atmospheric CO_2 in a 100-member ensemble following scenarios RCP 2.6 and 4.5:** (a) trends before CO_2 emission reductions start in RCP2.6, (b) trend after CO_2 emission reductions started in RCP2.6, distribution of (a) is indicated in gray contours for comparison, (c) trend reduction over time calculated as (a)-(b). (d)-(f) as (a)-(c) but for scenario RCP4.5. $P_{RCP2.6}$ marks the probability of trend reduction over time in the CO_2 emission reductions scenario, $P_{RCP4.5}$ marks the probability of trend reduction over time in RCP4.5. Vertical dashed gray lines show ensemble-mean trends per scenario. Dotted black vertical lines mark zero trends. Recent observational atmospheric CO_2 trends between 2015 and 2019 from NOAA/ESRL [Dlugokencky and Tans, 2019] are shown in solid black.

trends before and after 2020 look similar because the emissions rise steadily. Hence, only roughly half of the ensemble members show a reduced trend, with $P_{RCP4.5}$=48% [Fig. I.2d-f].

The atmospheric CO_2 may increase more strongly despite the onset of CO_2 emission reductions, when the global carbon cycle triggered by internal climate variability releases more CO_2 than CO_2 emission reductions save. For instance, this is possible when the tropical forests react to higher temperature and less precipitation caused by a strong El Niño event [Jones et al., 2001; Zeng et al., 2005]. The released CO_2 from the tropical biosphere persists in the atmosphere and can overwhelm the reduction of anthropogenic emissions [Fig. I.1]. These stronger atmospheric CO_2 growth trends despite CO_2 emission reductions might occur for trend comparisons around the CO_2 emission reductions start of up to ten years [Fig. I.3].

These probabilities of trend reduction of the two scenarios can be converted into probabilities of trend reduction being *caused* by CO_2 emission reductions [see section I.2.1]. If asked in advance in 2015,

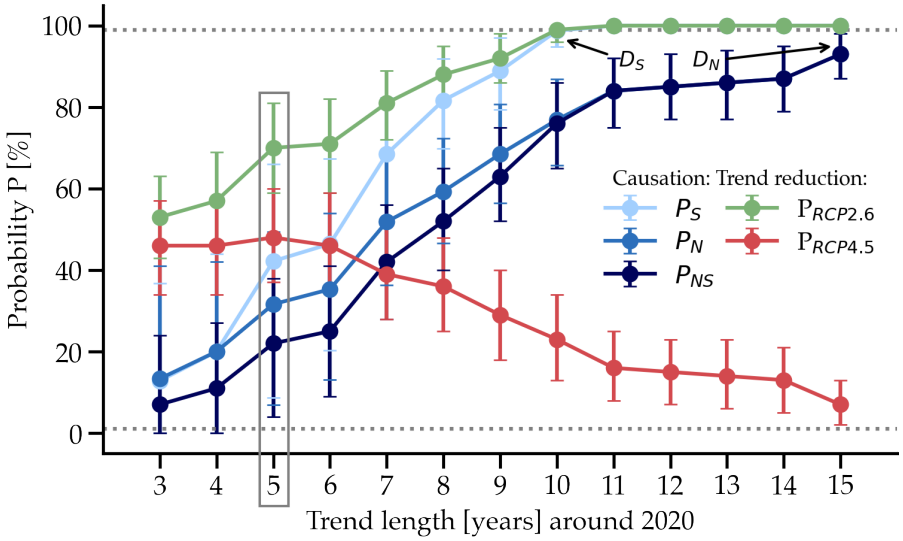


Figure I.3: **Probabilities of trend reduction in diagnostic atmospheric CO_2 between periods of varying trend length before and after CO_2 emission reductions start in 2020.** $P_{RCP2.6}$ (green) shows the probability of trend reduction in CO_2 emission reductions scenario RCP2.6. $P_{RCP4.5}$ (red) shows the probability of trend reduction in the currently most likely scenario for the near-term RCP4.5. P_S (pale blue) show the probability that a change from RCP4.5 to RCP2.6 causes the respective trend reduction in a sufficient causation sense. P_N (blue) show the probability that a change from RCP4.5 to RCP2.6 causes the respective trend reduction in a necessary causation sense. P_{NS} (dark blue) shows the probability that change from RCP4.5 to RCP2.6 causes the respective trend reduction in a sufficient and necessary causation sense. Error bars show the 1% and 99% confidence intervals based on bootstrapping with replacement. Dotted lines show 99% confidence interval for time of virtual certainty in trend reduction or causation ($D_{\{S,N\}}$). Results for policy-relevant five-year trends are highlighted in the gray box.

the answer would be that a policy change from RCP4.5 to RCP2.6 representing CO_2 emission reductions starting in 2020 are *sufficient* to cause a five-year trend reduction in atmospheric CO_2 growth by P_S=42% [Fig. I.3]. Here, this policy change works toward a trend reduction, but the trend reduction might also be prevented by internal variability. Asking from a 2025 perspective looking into the recent past, CO_2 emission reductions in 2020 were *necessary* by P_N=31% to cause trend reductions [Fig. I.3]. This policy change causes the five-year trend reduction in a *necessary and sufficient* sense by P_{NS}=22% [Fig. I.3, dark blue in box]. These results show that CO_2 emission reductions are far from certain to cause trend reductions in global atmospheric CO_2 growth when considering five-year trends.

To estimate the time-scales when CO_2 emission reductions are virtually certain to cause reduced atmospheric CO_2 growth trends, we

consider trends calculated over different time window lengths around the CO_2 emission reductions start. As expected, the shorter the trend-lengths considered, the more dominant internal variability is. Therefore, trend reductions are less likely occurring in the CO_2 emission reductions scenario. The three-year-trend probabilities of trend reduction even overlap with the 50% random forecast [Fig. I.3]. Conversely, when longer trends are considered, the influence of the signal of emissions change becomes stronger. CO_2 emission reductions reduce atmospheric CO_2 trends in RCP2.6 virtually for certain only when considering ten-year-trends [Fig. I.3]. In contrast, trend reductions are still possible due to internal variability despite the absence of CO_2 emission reductions for much longer in RCP4.5 [Fig. I.3]. Note that under RCP4.5 the annual anthropogenic CO_2 emissions increase very little until the 2040s. Therefore, a few members can still have reduced trends over time. Consequently, $P_{RCP4.5}$ does not drop to 1% until 2042.

The low causation probabilities over short time-scales show the inability to clearly attribute reduced atmospheric CO_2 trends to a policy change from RCP4.5 to RCP2.6 due to the large internal variability. The longer the time-scales considered, the stronger the two scenario pathways differ, and the attribution probabilities rise. If $P_{RCP2.6} > P_{RCP4.5}$ as assumed by the response to CO_2 emission reductions, P_S increases more quickly than P_N when $P_{RCP2.6}$ approaches 1 faster than $P_{RCP4.5}$ 0. Therefore, in the context of the scenarios RCP2.6 and RCP4.5, $P_S > P_N$ [Marotzke, 2019]. This means that in our context, sufficient causation is a stronger causation facet than necessary causation. Sufficient causation P_S describes whether the objective of reduced atmospheric CO_2 trends is met, which might be prevented by internal variability. As soon as growth trends decline in all realizations ($P_{RCP2.6} = 1$), also P_S saturates. In contrast, necessary causation P_N describes whether the response of reducing atmospheric CO_2 would only have happened in the presence of CO_2 emission reductions. Therefore, as long as trend reductions are possible even without CO_2 emission reductions, necessary causation will not be certain, that is, if $P_{RCP4.5} > 0$, then $P_N < 1$.

The time to detection of CO_2 emission reductions $D_{\{S,N\}}$ describes after how many years this policy change is virtually certain to cause atmospheric CO_2 growth trends to decline. CO_2 emission reductions sufficiently cause trend reductions after $D_S = 10$ years and necessary cause of reduction after $D_N = 27$ years. We note that once sufficient causation is certain, $P_S = 1$ in 2030 see (1), necessary causation and causation both necessary and sufficient coincide, $P_N = P_{NS}$; compare (2) and (3) with $P_{RCP2.6} = 1$. Virtual certainty in $P_{\{N,NS\}}$ is hindered by $P_{RCP4.5}$ above 1%. Due to the slow increase in emissions in the 2030s, internal variability allows a few members to have increasing trends. Taking a less strict threshold of 95% certainty like in Tebaldi and Friedlingstein [2013], we obtain $D_N = D_{NS} = 16$ years. This time-scale

of CO_2 emission reductions detection in a necessary causation sense D_N is a bit longer than the similarly defined estimate based on IPSL-CM5A-LR [Tebaldi and Friedlingstein, 2013, Table 1]. Our analysis also shows that whether this policy change from RCP4.5 to RCP2.6 can be identified as the cause of reduced atmospheric CO_2 trends after 10 or 16 years depends on the causation attribute. The differently defined emission reduction detection protocol of Schwartzman and Keeling [2020] finds a similar detection delay of 9±4 years for comparable 2% net annual emissions reduction.

I.4 SUMMARY AND CONCLUSIONS

In the context of potential future CO_2 emission reductions, we ask whether atmospheric CO_2 growth trend reductions in the near term can be attributed to a policy change. We focus on one specific pathway of CO_2 emission reductions interpreted as a policy change from scenario RCP4.5 without near-term CO_2 emission reductions to emissions reduction scenario RCP2.6 designed to achieve for the Paris targets representing 3% net annual CO_2 emission reductions until 2030. We apply a causation framework comprising two perspectives of policy elaboration [Hannart et al., 2016; Marotzke, 2019]. We diagnose atmospheric CO_2 variations compatible with the natural carbon sinks variations and compare growth trends of atmospheric CO_2 before and after the onset of CO_2 emission reductions in 2020 in RCP2.6. While five-year trends reduce in 70% of all realizations in the CO_2 emission reductions scenario RCP2.6 (consequentially implying increasing trends despite of CO_2 emission reductions by 30%), there is 48% probability of trend reductions in RCP4.5. This translates into CO_2 emission reductions from RCP4.5 to RCP2.6 being sufficient to cause a five-year trend reduction beforehand by 42% and in hindsight necessary by 31%. The probability that this policy change is both necessary and will suffice to bring the desired outcome considering five-year trends is only 22%. These probabilities are far from certain for up to a decade. It takes ten or 16 years of CO_2 emission reductions from RCP4.5 to RCP2.6 to virtually certainty cause a trend reduction in a sufficient or necessary causation sense, respectively. Communicating these probabilities in a clear manner is challenging but needed to inform policy-makers about the impact of internal variability on CO_2 emission reduction causation in the Earth system [Deser et al., 2012; Hannart et al., 2016; Howe et al., 2019].

The five-year global stocktake following the Paris Agreement [UNFCCC, 2015] makes the five-year internal variability highlighted in this study especially relevant for policy-makers. This study demonstrates the inherent uncertainty in near-term atmospheric CO_2 projections. As a partial solution to this challenge, initialized ESM-based prediction systems can reduce this uncertainty by predicting natural variations

of the global carbon cycle. Global oceanic CO_2 flux is predictable for two to three years [Li et al., 2019; Lovenduski et al., 2019a] and global atmospheric CO_2 variations have the potential to be predicted for up to three years in advance [Spring and Ilyina, 2020]. These multi-year ESM-based predictions of the global carbon cycle thereby bring added value about the expected natural variations of atmospheric CO_2 to policy-makers in the global stocktake process [UNFCCC, 2015].

Our analysis shows that it is crucial to have realistic expectations of the efficacy of climate policy in the near term [Marotzke, 2019; Samset et al., 2020]. Also Schwartzman and Keeling [2020] find a detection delay of up to a decade in a different approach. Even if anthropogenic emissions begin to decline after 2020, there still remains a substantial probability that atmospheric CO_2 trends will not have declined five years afterwards. In this case, the effects of CO_2 emission reductions on other iconic climate variables, such as global mean surface temperature, very likely get delayed even longer [Marotzke, 2019]. The likelihood of this happening is substantial. For instance, there is a three-out-of-ten chance that atmospheric CO_2 rises even stronger in the five years after CO_2 emission reductions started compared to before. Assuming the evolution of the RCPs [Meinshausen et al., 2011] and the magnitude of internal variability in the global CO_2 fluxes in MPI-ESM-LR, such increasing atmospheric CO_2 growth trends despite CO_2 emission reductions from RCP4.5 to RCP2.6 are possible for up to a decade. Although this analysis relies on only a single model, internal variability may disguise CO_2 emission reductions efforts in the Earth System for a couple of years. Should this be the case, climate science should explain the observed atmospheric CO_2 evolution honoring internal variability. policy-makers should rather be informed by initialized predictions about the internal variability in the near-term evolution of atmospheric CO_2 [Betts et al., 2018; Spring and Ilyina, 2020]. Evaluation of CO_2 emission reduction efficacy from an atmospheric CO_2 perspective needs to take internal variability, and therefore longer than five-year trends, into account.

ACKNOWLEDGEMENTS

We acknowledge Luis Kornblueh, Jürgen Kröger and Michael Botzet for producing the historical, RCP2.6 and RCP4.5 MPI-ESM Grand Ensemble simulations, and the Swiss National Computing Centre (CSCS) and the German Climate Computing Center (DKRZ) for providing the necessary computational resources.

We acknowledge funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 821003 "Climate-Carbon Interactions in the Coming Century (4C)" and No 641816 (CRESCENDO). T.I. received funding from the Federal Ministry of Education and Research in Germany (BMBF) through the research programme "MiKlipII" (FKZ: 01LP1517B). J.M. acknowledges long-term support by the Max Planck Society for the Advancement of Science.

We thank Alexander Winkler for internal review and the three anonymous reviewers for improving the manuscript with their comments.

I.SI SUPPLEMENTARY INFORMATION

I.SI.1 *Method Verification of Diagnostic Atmospheric CO_2*

Removing the ensemble mean from each member also removes the increasing strength of the carbon sinks with respect to increasing atmospheric CO_2 concentration. This increase is however already included in the Integration Assessment Models pathways. To minimize the effect of long-term trends from model drift [Maher et al., 2019] in the carbon sinks on atmospheric CO_2, we initialize the method in the year nearest to the analysis start in 2005 when the detrended integrated global carbon sink is close to 0 PgC.

We assume instantaneous global atmospheric mixing [Ballantyne et al., 2012], which is a fair assumption for an annual analysis time-step considering the one to two year time-scales for inter-hemispheric mixing. Furthermore, we assume that variability in the carbon cycle is driven by climate variability, i.e. the internally varying changes in atmospheric and oceanic circulation, temperature and precipitation induce variability in the land and ocean carbon sinks. This also implies that short-term effects of the variability of atmospheric CO_2 variations on terrestrial CO_2 flux (short-term CO_2 fertilization effect) and oceanic CO_2 flux (short-term effect of higher atmospheric CO_2 after strong land or/and ocean outgassing) is ignored as in all historical simulations with prescribed atmospheric CO_2 levels. These assumptions and the method itself are similar to the concept of compatible emissions [Jones et al., 2013]. However, here we diagnose atmospheric CO_2 concentrations (given in ppm) compatible with global carbon sink variability, whereas Jones et al. [2013] diagnose anthropogenic emissions (given in PgC annual CO_2 emissions) compatible with the RCP pathways and internal variability.

We verify our proposed method by comparing our diagnostic atmospheric CO_2 calculated offline from monthly output with prognostic atmospheric CO_2 calculated by the model itself at each time-step in a 9-member simulation with prognostic atmospheric CO_2 of the current CMIP6 version of MPI-ESM [Mauritsen et al., 2019]. Diagnostic atmospheric CO_2 tracks the variations of prognostic atmospheric CO_2 simulated by the model fairly well over the historical period (metric=member mean $\pm$ member standard deviation: anomaly correlation coefficient (ACC) = 0.994 $\pm$ 0.03, mean absolute error (MAE) = 0.130 $\pm$ 0.036 ppm) [Fig. I.SI.1].

This method is applied in a non-transient climate in Spring and Ilyina [2020].

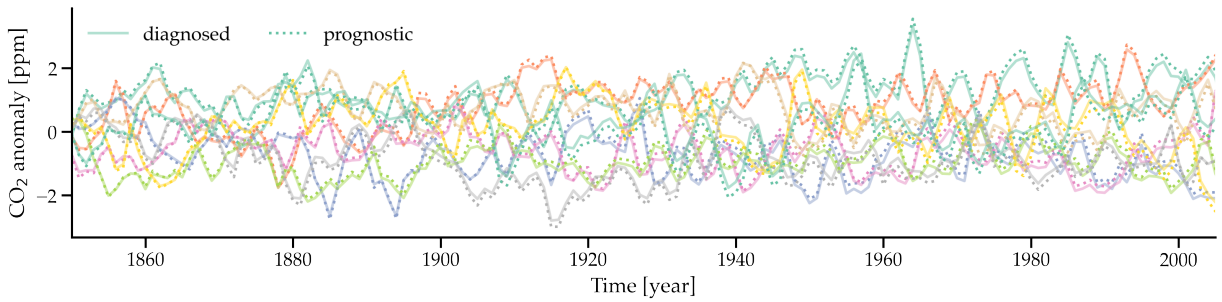


Figure I.SI.1: **Verification of diagnostic atmospheric CO_2:** Comparing diagnostic (solid line) (see section I.2.4) with prognostic atmospheric CO_2 (dotted line) from a 9-member C^4MIP CMIP6 MPI-ESM-LR ensemble, where the members are shown in different colors. Both diagnostic and prognostic atmospheric CO_2 are detrended by subtracting the ensemble mean prognostic atmospheric CO_2 as the forced signal.

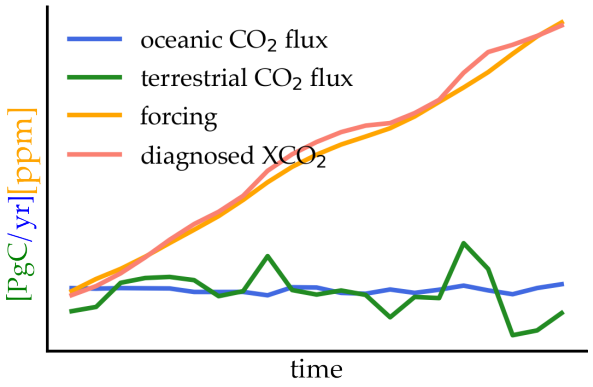


Figure I.SI.2: **Schematic of contributions to diagnostic atmospheric CO_2:** (green) terrestrial surface CO_2 flux in units [PgC/yr], (blue) oceanic surface CO_2 flux in units [PgC/yr], (orange) atmospheric CO_2 forcing from a simplified climate model fed with emissions from integrated assessment models in units [ppm], (salmon) diagnostic atmospheric CO_2 in units [ppm].

I.SI.2 *Differences in Scenarios before 2020*

Note that RCP2.6 has around 1 ppm higher atmospheric CO_2 forcing than RCP4.5 in the period 2015-2022 [Fig. I.1]. This gap is much smaller than the possible ranges due to internal variability across ensemble members. We decided against harmonizing the two scenarios to identical levels up to CO_2 emission reductions in 2020 to keep the methodology straight-forward without modifying the scenarios.

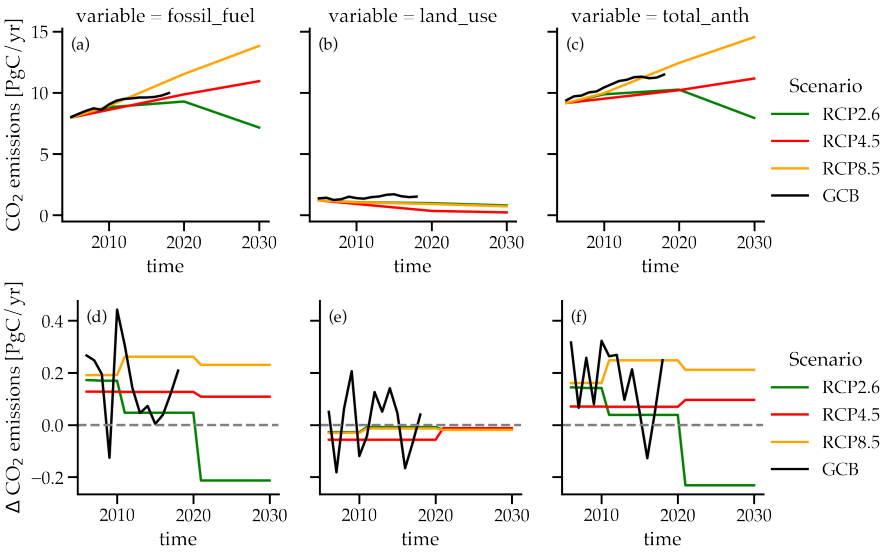


Figure I.SI.3: **Anthropogenic CO_2 emissions:** (a,d) fossil-fuel, (b,e) land-use change and (c,f) combined. Observations (black) taken from the Global Carbon Budget [Friedlingstein et al., 2019] and modelling RCP scenarios (colors) taken from Meinshausen et al. [2011]. The first row (a-c) shows absolute CO_2 emissions per year, the second row (d-f) shows changes in CO_2 emissions per year.

The land-use changes in RCP4.5 include strong reforestation [Fig. I.SI.3] [Thomson et al., 2011]. Therefore, when combining fossil-fuel emissions and land-use emissions into anthropogenic CO_2 emissions, RCP2.6 and RCP4.5 follow very similar combined emission pathways until 2020 [Fig. I.SI.3]. The observed evolution of annual atmospheric CO_2 concentration until 2019 tracks the scenarios RCP2.6 and RCP4.5 more closely than RCP8.5 [Fig. I.SI.4].

I.SI.2.1 *Results for Harmonized Emissions before Emissions Separation Date*

To exclude the effect of differences in the period before the potential CO_2 emission reduction, we create a mean RCP2.6-RCP4.5 forcing until the two forcings are meet in 2025. We therefore also set the date of potential CO_2 emission reductions to 2025, repeat the analysis and find very similar results [Fig. I.SI.5].

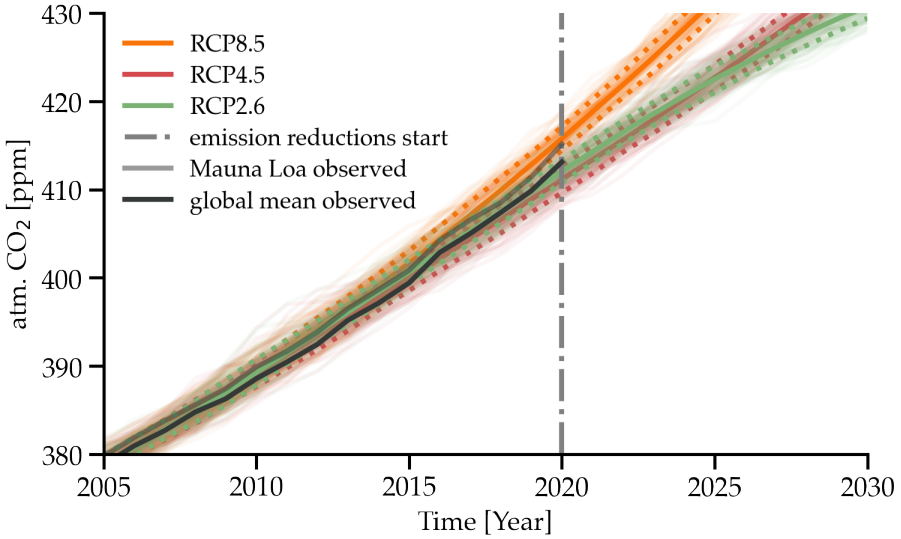


Figure I.SI.4: **Diagnostic atmospheric CO_2 concentration under different scenarios:** CO_2 emission reductions scenario RCP2.6 (green) and currently most likely scenario for the near-term RCP4.5 (red) and non-mitigation scenario RCP8.5 (orange). Individual ensemble members are shown in faded colors. The forced response (solid) in atmospheric CO_2 follows the prescribed atmospheric CO_2 forcing derived from a simplified climate model fed with emissions from integrated assessment models [Meinshausen et al., 2011]. The variations of observed global mean (black) and Mauna Loa (gray) atmospheric CO_2 range within the member standard deviation (dotted green and red) [Dlugokencky and Tans, 2019].

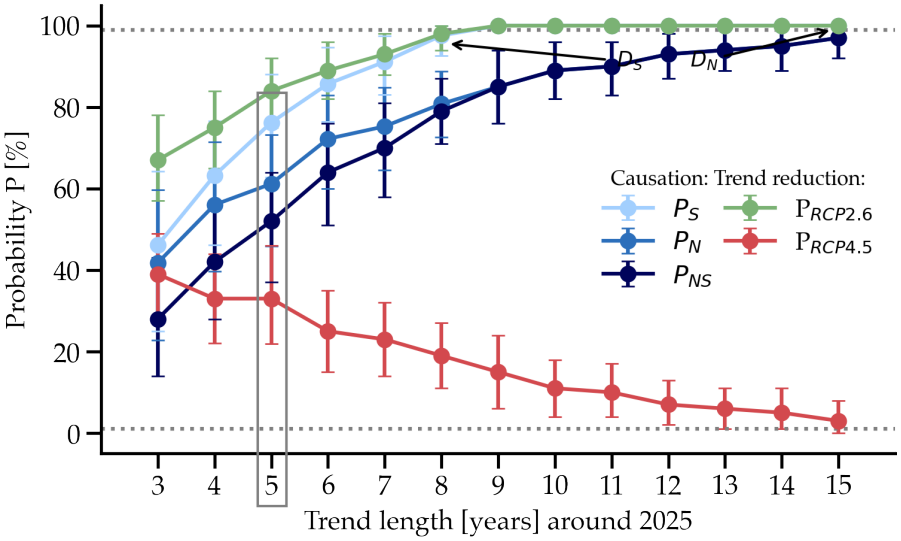


Figure I.SI.5: **Probabilities of trend reduction in harmonized diagnostic atmospheric CO_2 concentrations** for periods of varying trend length before and after CO_2 emission reductions start in 2025, where both scenarios are harmonized before 2025. Labels as in Fig. I.3.

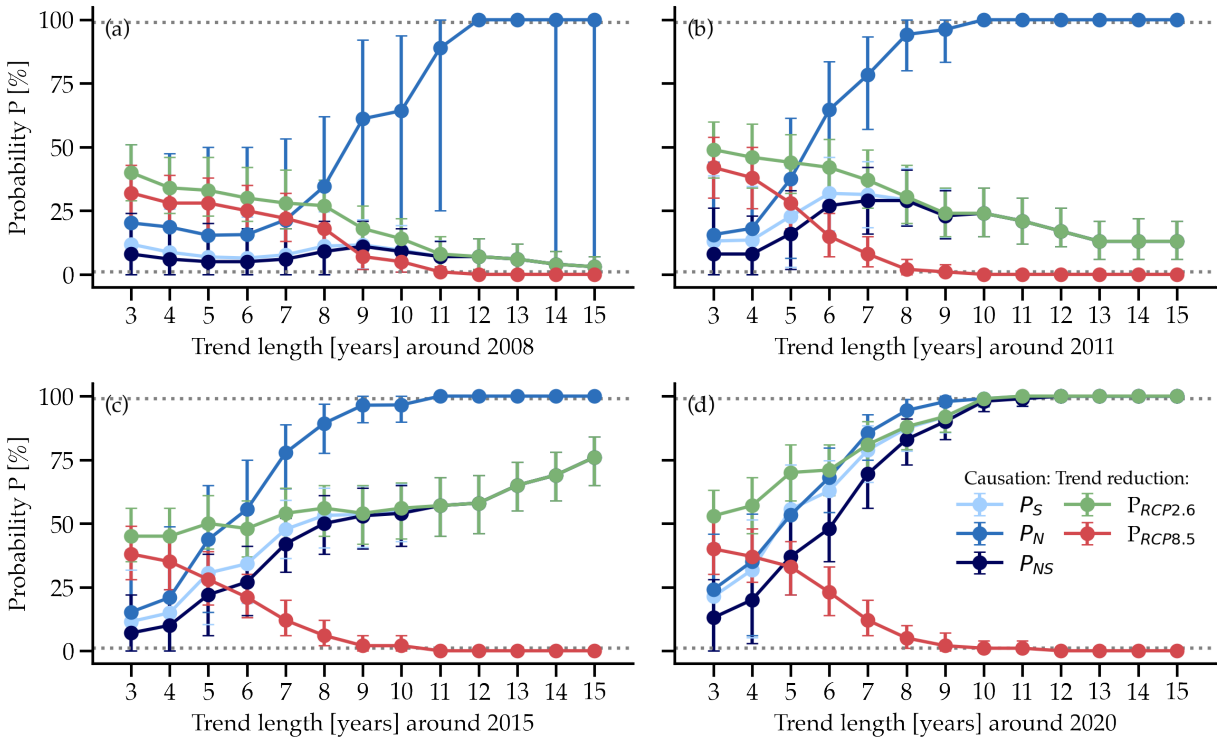


Figure I.SI.6: **Probabilities of trend reduction in diagnostic atmospheric CO_2** for periods of varying trend length before and after CO_2 emission reductions emerge in (a) 2008 (separation of 0.001 ppm), (b) 2011 (separation of 0.1 ppm), (c) 2015 (separation of member standard deviation ~1 ppm) and (d) 2020 (same separation time as for RCP2.6 vs. RCP4.5). Labels as in Fig. 3, but with RCP8.5 instead of RCP4.5.

I.SI.3 *Comparison RCP2.6 vs RCP8.5*

Comparing RCP2.6 to RCP8.5 slightly changes the question to whether reduced emission increase would be visible in atmospheric CO_2, because even in RCP2.6 emissions are rising until 2020 [Fig. I.SI.3].

Proving the limits of feasibility of the method, we here compare RCP2.6 as an CO_2 emission reduction scenario vs RCP8.5 without any mitigation for different years of separation. The differences in Fig. I.SI.6 highlight the sensitivity of the method across the year of separation of trends before and after this policy change. The difference towards Fig. I.3 show the sensitivity towards the chosen scenario combination. Note that the assumption that a change from RCP8.5 to RCP2.6 before 2015 is incompatible with the required climate response of reducing atmospheric CO_2 trends, because RCP2.6 emissions rise until 2020.

Fig. I.SI.6 shows the probabilities of trend reduction $P_{RCP2.6}$ needs to be higher than $P_{RCP8.5}$ to find sufficient or necessary causation. If both probabilities are very close to each other, as when comparing RCP2.6 and RCP8.5 for a difference in scenarios after 2011, trends are not

reducing for trend lengths longer than ten years, because the emission increases in both scenarios are still overwhelming internal variability. For short trend lengths up to seven years however, we identify close to 50% reduced trends after CO_2 emission reductions.

Fig. I.SI.6 shows how appropriate this method is to identify different attributes of causation of CO_2 emission reductions leading to reduced atmospheric CO_2 trends in scenarios at different rates of CO_2 emissions.

I.SI.4 *Comparison of Modeled & Observation-based Atmospheric CO_2 Growth Rate*

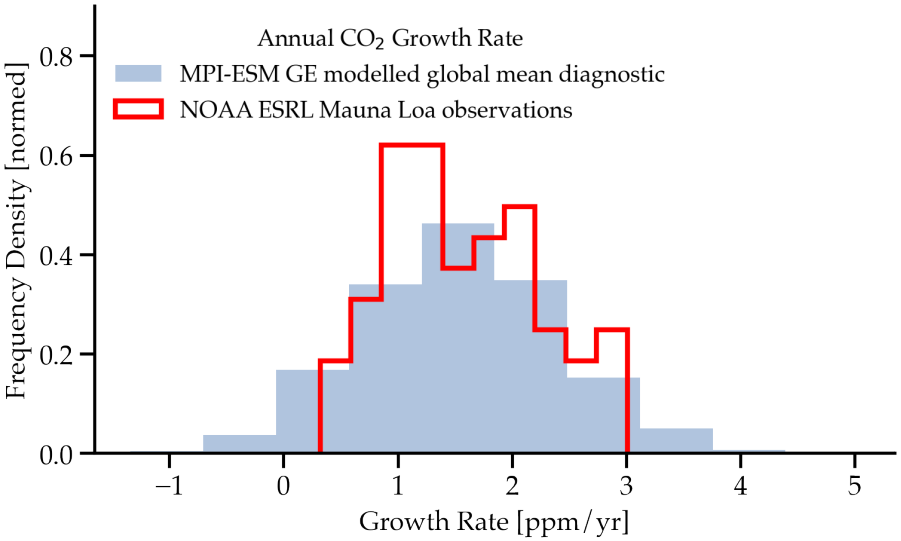


Figure I.SI.7: **Distribution of annual atmospheric CO_2 growth rate from 1958-2018:** (gray) modeled global diagnostic atmospheric CO_2 based on MPI-ESM Grand Ensemble, (red) observed at Mauna Loa (NOAA/ESRL) [Dlugokencky and Tans, 2019]

I.SI.5 *Spatial Origin in CO_2 Flux Trends*

The trends of strong positive or negative CO_2 flux (and its time-accumulating effect on atmospheric CO_2) originate in the tropical forest [Fig. I.SI.8] in a pattern that resembles the ENSO composites for temperature and precipitation, which are known to drive terrestrial CO_2 flux changes.

I.SI.6 *Compatible Emissions*

As the concentration-driven scenarios were created with MAGICC tuned for 3K equilibrium climate sensitivity and medium climate/CO_2-carbon feedback, the compatible emissions [Jones et al., 2013] in an

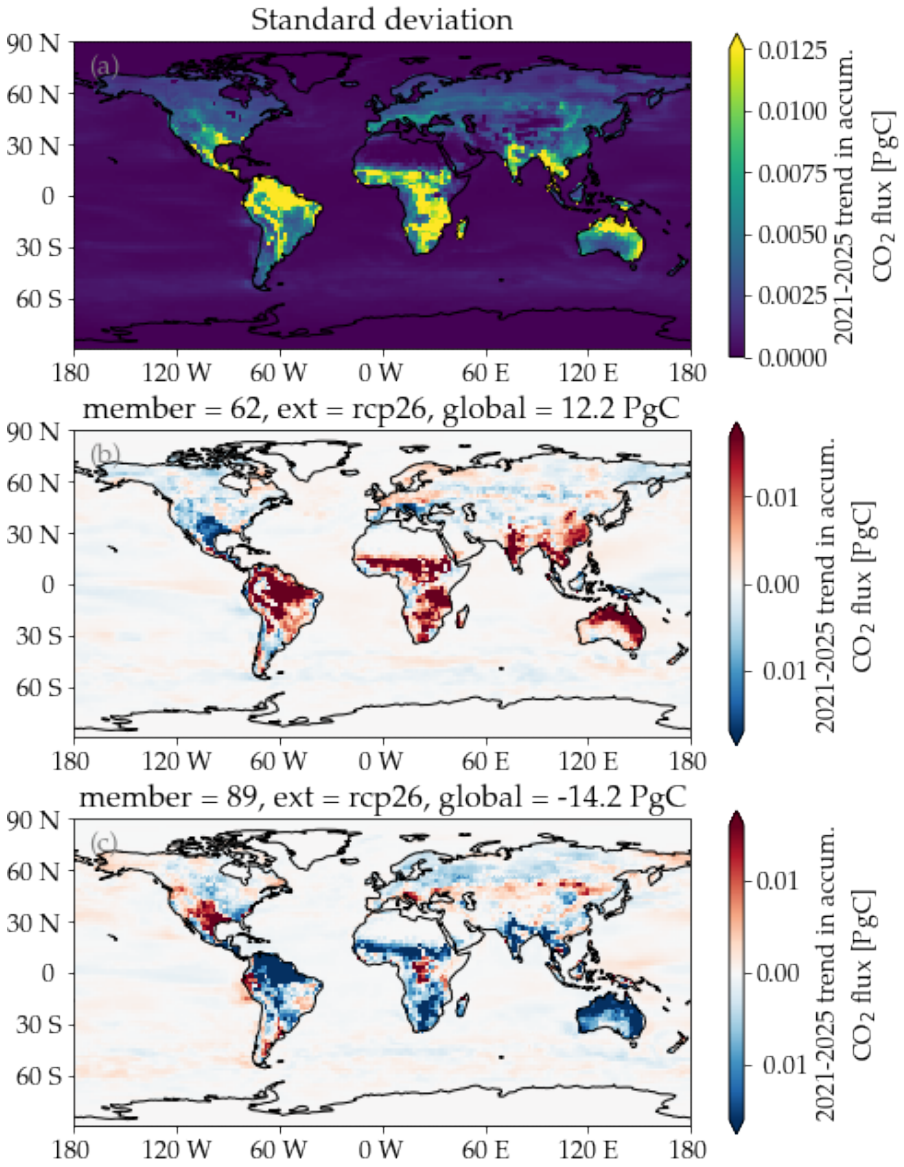


Figure I.SI.8: **Distribution of modelled five-year 2021-2025 accumulated CO_2 flux trends after removing the forced signal in MPI-ESM Grand Ensemble:** (a) ensemble-member standard deviation, (b) most positive trend, (c) most negative trend.

ESM (like in our case here MPI-ESM with its particular ECS and feedbacks) can differ from the emission inputs to MAGICC, which explains the higher compatible emissions in MPI-ESM concentration-driven simulations than emissions used in MAGICC [Fig. I.SI.9]. The three year time difference that compatible emissions decrease later in RCP2.6, we attribute one year of that difference to the differencing of the emissions growth rate and the time of adjustment of the land and ocean carbon sink to changes in atmospheric CO_2 concentration.

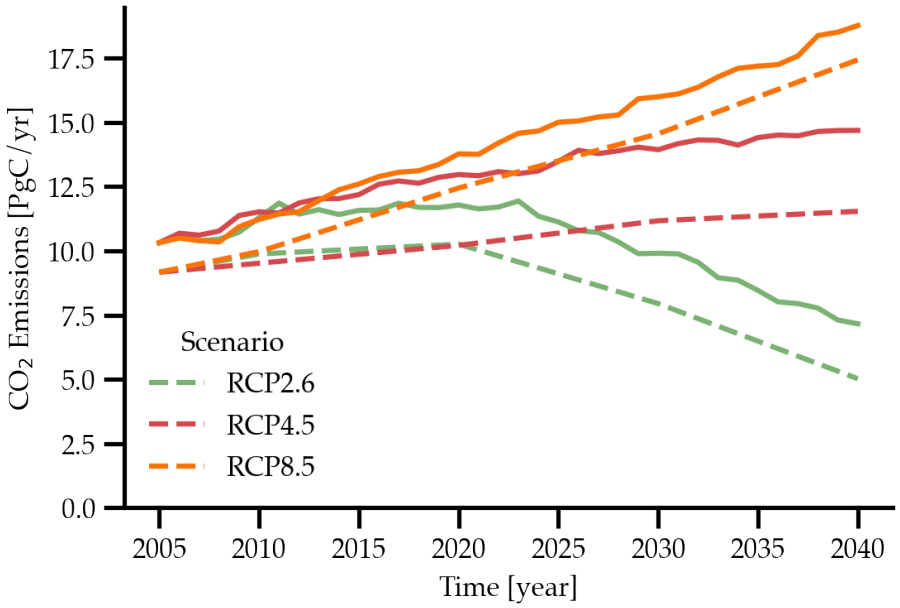


Figure I.SI.9: **Comparison of compatible and integrated assessment model CO_2 emissions:** continuous lines show compatible CO_2 emissions [Jones et al., 2013] from ensemble mean concentration-driven MPI Grand Ensemble simulations and dotted lines show CO_2 emissions from the integrated assessment model per RCP scenario (colors) [Meinshausen et al., 2011].

I.SI.7 *Reproducibility*

Scripts used in the analysis and other supporting information that may be useful in reproducing the authors work are archived by the Max Planck Institute for Meteorology: http://hdl.handle.net/21.11116/0000-0005-467C-2.

PREDICTABILITY HORIZONS IN THE GLOBAL CARBON CYCLE INFERRED FROM A PERFECT-MODEL FRAMEWORK

The attached article has been published in *Geophysical Research Letters*:

Spring, A., & Ilyina, T. [2020]. Predictability Horizons in the Global Carbon Cycle Inferred From a Perfect-Model Framework. *Geophysical Research Letters*, *47*[9], e2019GL085311. doi:10/ggtbv2

Author contributions: A.S. conceived the study, performed the simulations and analysis, created the figures and drafted the manuscript. T.I. contributed in manuscript editing, revisions, and provided overall scientific guidance.

Predictability Horizons in the Global Carbon Cycle Inferred from a Perfect-Model Framework

Aaron Spring[1,2] **and Tatiana Ilyina**[1]
[1] Max Planck Institute for Meteorology, Hamburg, Germany
[2] International Max Planck Research School of Earth System Modelling, IMPRS, Hamburg, Germany

Submitted: 7 September 2019, Accepted: 9 April 2020, Published: 23 April 2020

ABSTRACT

On inter-annual time-scales the growth rate of atmospheric CO_2 is largely controlled by the response of the land and ocean carbon sinks to climate variability. Yet, it is unknown to what extent this variability limits the predictability of atmospheric CO_2 variations. Using perfect-model Earth System Model (ESM) simulations, we show that variations in atmospheric CO_2 are potentially predictable for three years. We find a two-year predictability horizon for global oceanic CO_2 flux with longer regional predictive skill of up to seven years. The two-year predictability horizon of terrestrial CO_2 flux originates in the tropics and mid-latitudes. With the predictive skill of the isolated effects of land and ocean carbon sink on atmospheric CO_2 of five and 12 years respectively, land dampens the overall predictive skill of atmospheric CO_2 variations. Our research shows the potential of ESM-based predictions to forecast multi-year variations in atmospheric CO_2.

PLAIN LANGUAGE SUMMARY

The amount of anthropogenic carbon emissions absorbed by land and ocean from the atmosphere varies annually due to their sensitivity to climate. Therefore, the atmospheric CO_2 growth rate does not strictly follow the emissions signal. Whether decadal prediction systems can also predict variations of atmospheric CO_2 has not been shown yet, but is crucial to inform policy-makers about the efficiency of the implementation of the Paris Agreement. Using numerical Earth System simulations in an idealized prediction framework, we show that global atmospheric CO_2 is predictable up to three years in advance. The global ocean and land CO_2 fluxes are predictable for two years. The isolated effects of the land and ocean carbon sink on atmospheric CO_2 are predictable for five and 12 years, respectively. Therefore, the land carbon cycle limits atmospheric CO_2 predictability. Our study demonstrates that simulation-based multi-year forecasts have the potential to predict natural atmospheric CO_2 variations.

II.1 INTRODUCTION

The atmospheric CO_2 mixing ratio rises in response to increasing anthropogenic carbon emissions. The terrestrial and oceanic carbon sinks modulate the atmospheric CO_2 mixing ratio by currently absorbing about 30% and 25% of the anthropogenic carbon emissions, respectively [Friedlingstein et al., 2019]. However, the inter-annual atmospheric CO_2 growth rate does not strictly follow the changes in anthropogenic CO_2 emissions [Bacastow, 1976; Keeling et al., 1995; Peters et al., 2017]. This discrepancy is due to internal variability of the oceanic and terrestrial carbon sinks driven by the large-scale modes of variability of the climate system [Doney et al., 2006; Heinze et al., 2015; McKinley et al., 2017; Li and Ilyina, 2018] and volcanic eruptions [Jones et al., 2001; Frölicher et al., 2013; Eddebbar et al., 2019]. Indeed, this internal climate variability of the natural carbon sinks might obscure the identification of the anthropogenic emissions signal in atmospheric CO_2 mixing ratio [Peters et al., 2017]. This sensitivity of inter-annual changes in atmospheric CO_2 concentration to natural climate variability is a major uncertainty in our understanding of the near-term evolution of atmospheric CO_2 and the remaining carbon budget. How this sensitivity limits the near-term predictability of the atmospheric CO_2 growth rate has not been addressed before.

The Paris Agreement [UNFCCC, 2015] aims to limit global warming to at most 2°C by mostly mitigation of fossil fuel emissions. Article 14 demands for an assessment of collective progress periodically every five years "in the light (...) of the best available science". The potential of prediction systems with prognostic CO_2 has to be explored in the context of their ability to predict near-term atm. CO_2 variations in response to the changes of the natural carbon sinks. Given that CO_2 predictions are feasible, changes in emissions can be earlier separated from internal variability. Therefore, the near-term efficacy of the planned mitigation efforts can be skillfully forecasted in atm. CO_2 in advance until the prediction horizon, rather than being assessed only in the following global stocktake. As a first step towards a viable atm. CO_2 prediction system, we here show its potential to be successful in a idealized setup.

Over the last decade, multi-year climate predictions have been established. Decadal prediction systems based on comprehensive Earth System Models (ESM) take advantage of the memory of the climate system. Observational products are assimilated into an ESM and used as initial conditions of the forecast [Meehl et al., 2009]. While decadal prediction systems are not perfect, they proved themselves successful for many aspects of the climate system [Meehl et al., 2013; Marotzke et al., 2016; Yeager et al., 2018]. Having established prediction skill for many climate variables, which the carbon cycles are sensitive to,

prediction systems can be used to assess decadal predictability of the global carbon cycle.

In the emerging predictability studies of terrestrial and ocean carbon cycle components, only physical climate variables such as ocean temperature and salinity are nudged towards observations due to the sparse and short observational period of biogeochemical variables. Li et al. [2016] find predictive skill of four to seven years in the winter North Atlantic air-sea CO_2 uptake when evaluating against an ocean reconstruction with uninitialized biogeochemical components. The global oceanic carbon uptake can be skilfully predicted for up to two years evaluated against observations [Li et al., 2019] and ocean reconstruction [Lovenduski et al., 2019a]. Séférian et al. [2018] estimates the potential predictability horizon of the global carbon sink predictive skill of four to six years in idealized perfect-model simulations. Unlike the previously mentioned ESM-based predictions, Betts et al. [2016, 2018] use a statistical model to forecast the growth rate of the atmospheric CO_2 mixing ratio at Mauna Loa. A regression of previous imprints of the El Ninõ-Southern Oscillation (ENSO) on atmospheric CO_2 mixing ratio and anthropogenic emissions predicts the next-year atmospheric CO_2. Although ENSO is one of the main drivers of CO_2 flux variability [Jones and Cox, 2005; Zeng et al., 2005], this method lacks the representation of other dynamic processes and a multi-year outlook. Zeng et al. [2008] demonstrate technical feasibility of ESM-based seasonal air-land CO_2 flux predictions. They find a higher prediction skill for Mauna Loa atmospheric CO_2 than a statistical ENSO regression and thereby indicate the added value of ESM-based carbon cycle predictions. A very recent study finds two-year predictive skill in terrestrial net ecosystem production, which is closely linked to terrestrial CO_2 flux predictive skill [Lovenduski et al., 2019b].

Previous studies on carbon cycle predictability lack a fully-coupled global carbon cycle. Terrestrial carbon cycle based studies [Zeng et al., 2008; Betts et al., 2018] miss the oceanic carbon sink and multi-year lead times. (Oceanic) carbon cycle based predictability studies [Li et al., 2016; Séférian et al., 2018; Li et al., 2019; Lovenduski et al., 2019a; Lovenduski et al., 2019b] rest upon *concentration-driven* simulations, in which atmospheric CO_2 mixing ratio is prescribed. In such simulations, air-sea and air-land CO_2 flux do not alter atmospheric CO_2 mixing ratio. We go beyond previous studies by using fully-coupled carbon cycle ESM simulations. In our simulations, the atmospheric mixing ratio of CO_2 is a prognostic property and can be physically-consistently predicted. We examine the limits of predictive skill of air-sea and air-land surface CO_2 flux as well as the resulting atmospheric CO_2 mixing ratio. Furthermore, we determine the isolated contributions of the terrestrial and oceanic carbon sinks to the limits of predictive skill of

atmospheric CO_2. We examine the relevance of different regions of the overall land and ocean predictive skill based on two metrics.

We use a perfect-model framework, in which we assumes that the model can reproduce observed variability [Griffies and Bryan, 1997]. As it is free of initialization from reanalysis products, the perfect-model framework cannot perform predictions, but only estimate time scales after which the state of initialization is lost in the ensemble due the chaotic nature of the Earth system. The assimilation of reanalysis products in prediction systems inevitably induces drifts from the reanalysis climatology towards the native model climatology due to biases between reanalysis and ESMs [Kröger et al., 2017]. As marine biogeochemical cycles are highly sensitive to changes in ocean circulation, such ocean reconstructions can disrupt biogeochemical tracer distributions [Toggweiler et al., 1989; Park et al., 2018]. This perfect-model framework offers the only self-consistent setup to examine processes leading to predictive skill without methodological artifacts due to initialization from reanalysis. We therefore use it as a first-order approach study to examine the limits of predictive skill of a fully-coupled global carbon cycle.

II.2 METHODS

II.2.1 *Model Description*

Our perfect-model framework study is based on perturbed initial-conditions ensemble simulations from a pre-industrial control simulation with prognostic atmospheric CO_2 mixing ratio, which is known as *esm-piControl* in the Coupled Model Intercomparison Project Phase 6 (CMIP6) framework [Eyring et al., 2016]. We use the close to equilibrium spin-up simulation of the CMIP6 version of the Max Planck Institute ESM (MPI-ESM)[Mauritsen et al., 2019] taken for the Coupled Climate–Carbon Cycle Model Intercomparison Project (C^4MIP) [Jones et al., 2016] in its low-resolution configuration (horizontal resolution of about 1.8°in the atmosphere and on land, and about 1.5°in the ocean).

The ocean general circulation model MPIOM [Jungclaus et al., 2013] provides the circulation field to the marine biogeochemical cycle model HAMOCC [Ilyina et al., 2013]. HAMOCC encompasses carbonate chemistry and an NPZD-type ecosystem including additional nitrogen-fixating cyanobacteria [Paulsen et al., 2017], nutrient-light-temperature co-limitation, calcite and opal export treated explicitly and iron as a macro-nutrient. The terrestrial carbon cycle model JSBACH incorporates wildfires, dynamic vegetation, soil carbon decomposition and storage [Schneck et al., 2013]. The atmospheric general circulation model ECHAM6 relies on a flux-form semi-Lagrangian scheme [Lin and Rood, 1996] to represent the transport of the three-dimensional atmospheric prognostic atmospheric CO_2 tracer [Stevens et al., 2013].

II.2.2 *Perfect-Model Framework*

The simulations in the perfect-model framework are started from 12 randomly chosen initialization states. From each of those states of a stable 300-year control simulation, we branch off 9 ensemble members and integrate them for 20 years. Ensemble members are generated by perturbing the stratospheric horizontal diffusion by 1.0000{member} in the first year after initialization. This initialization approach induces only tiny perturbations to the climate system as the ocean and land initial conditions remain identical. Therefore, this perfect-model initialization presents an upper limit estimate of predictive skill. As a result of such initialization approach, the ensemble members first follow a similar evolution and then diverge until reaching the time scale of the predictability horizon, when temporal variability is statically indistinguishable from variability across members [Fig. II.SI.1a].

II.2.3 *Predictive skill assessment*

Because only slightly perturbed initial conditions distinguish ensemble members, all ensemble members can equally serve as a verification to compute predictive skill according to a certain metric. Therefore, we compute predictive skill as the anomaly correlation coefficient skill score (ACC) and root-mean-square-error (RMSE) between the ensemble member mean, with verification member excluded, as forecast and every single ensemble member as verification [Wilks, 2006; Jolliffe and Stephenson, 2011]. ACC assesses the linear association of the forecast and the verification and therefore measures synchronous evolution. RMSE measures the second-order distance between forecast and verification. RMSE describes how the individual ensemble members spread over lead time and is therefore a more conservative attribute of predictive skill. The two metrics measure related but different attributes of predictive skill [Murphy, 1988]. Find additional details about the metrics in the supplementary information.

We assess the significance of initialized forecast predictive skill by bootstrapping with replacement [Efron and Tibshirani, 1993; Goddard et al., 2013]. To compare against random forecasts, we create hypothetical uninitialized ensembles randomly drawn as 20-year chunks from the control simulation. The p-value denotes the probability that random forecasts are more skillful than the initialized forecast re-sampled with replacement over all initializations based on 5000 iterations (except noted otherwise). This non-parametric approach relies on no further assumptions and is especially free from any normalization choice, which can bias the predictability horizon [Hawkins et al., 2016]. We define the limit of predictive skill, the so-called *predictability horizon* of a metric as the last significant lead year (with p-value <0.05).

II.2.4 *Diagnosing Global Atmospheric CO_2 Mixing Ratio*

We furthermore disentangle the isolated effects of the global terrestrial and oceanic carbon sinks on the global atmospheric CO_2 mixing ratio. For this we convert zero-mean free-of-trend (denoted by ′) temporally accumulated global surface CO_2 flux into diagnosed atmospheric CO_2 mixing ratio by multiplying with $\frac{ppm}{2.12PgC}$ [Ballantyne et al., 2012; Friedlingstein et al., 2019]:

$$XCO_{2,\text{atm,diag}}(t) = XCO_{2,atm}(0) + \sum_{t'}^{t} CO_2\text{flux}'(t') \cdot \frac{ppm}{2.124PgC}$$

where $XCO_{2,atm}(0)$ is a free choice parameter of initial diagnosed atmospheric surface CO_2 mixing ratio. This approach assumes the atmosphere as one instantaneously mixed box and implicitly incorporates direct CO_2 flux from land to ocean. Converting the sum of terrestrial and oceanic CO_2 flux into diagnosed atmospheric CO_2 yields a very similar evolution as prognostic atmospheric CO_2.

II.2.5 *Variance-weighted Mean Period*

We infer potentially predictable regions from a single control simulation following Branstator and Teng [2010]. Accordingly, the variance-weighted mean period highlights regions with low-frequency variations with longer potentially predictable periods:

$$P_x = \sum_k V(f_k, x) / \sum_k f_k V(f_k, x),$$

where $V(f, x)$ is the variance per unit frequency for frequency f at location x and k are the individual frequencies from the power spectrum.

II.3 PREDICTIVE SKILL OF AIR-SEA CO_2 FLUX

We first examine predictive skill in the global ocean carbon sink. Global annual air-sea CO_2 flux is potentially predictable for two and four years, when assessed with ACC and RMSE, respectively [Fig. II.1a]. This means that initialized ensembles predict each other significantly better due to initialization than uninitialized ensembles. When assessing predictive skill with the same metric RMSE and predictability horizon definition as Séférian et al. [2018], global air-sea CO_2 flux predictability horizon in our model MPI-ESM-LR is comparable to CNRM-ESM1 (for details see supplementary information section S3.3).

On a regional scale longer predictability horizons are found. For instance, the North and Subtropical Atlantic as well as the North and Subtropical Pacific have a predictability horizon of more than six years and thereby exceed the global oceanic predictability horizon

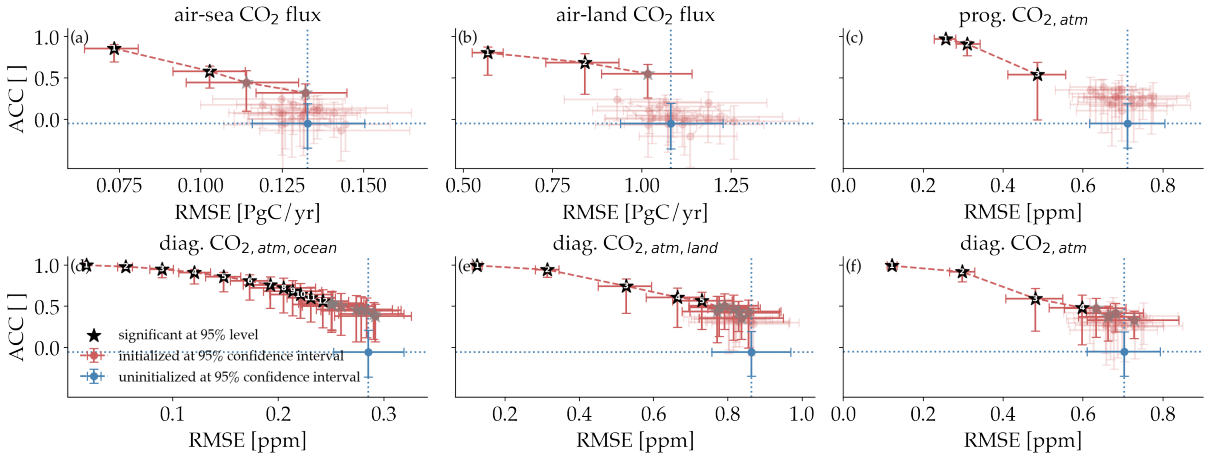


Figure II.1: **Comparison of the mean potential prediction skill of the initialized ensemble (red) versus random uninitialized ensembles (blue) in global annual surface quantities of the carbon cycle**. The anomaly correlation coefficient (ACC) on the y-axis and root-mean-square-error (RMSE) on the x-axis are shown by lead years represented as dots: (a) air-sea CO_2 flux, (b) air-land CO_2 flux, (c) prognostic surface atmospheric CO_2, (d) diagnosed atmospheric CO_2 based on oceanic carbon sink, (e) diagnosed atmospheric CO_2 based on the global terrestrial carbon sink and (f) diagnosed atmospheric CO_2 based on the global oceanic and terrestrial carbon sink [see section II.2.4]. Errorbars show 95% confidence intervals based on bootstrapping with replacement (N=5000). The last lead year with a bootstrapped p-value (which represents that uninitialized ensembles beat initialized ensembles) lower than 5% marks the predictability horizon. Black stars with white integer denote significant lead years in ACC and RMSE, gray stars if only one metric is significant and lead years non-significant in both metrics are blurred.

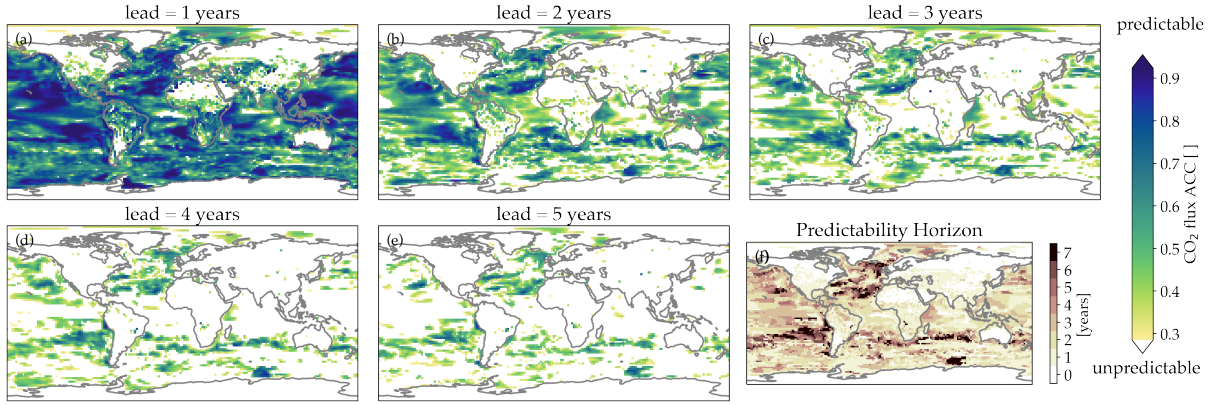


Figure II.2: **Spatial distribution of ACC-based surface CO_2 flux predictive skill:** (a-e) predictive skill over the first five lead years. White areas indicate unpredictable areas where the uninitialized predictive skill exceeds the initialized skill at 5% probability based on bootstrapping with replacement (N=1000). (f) The predictability horizon marks the last significant lead year.

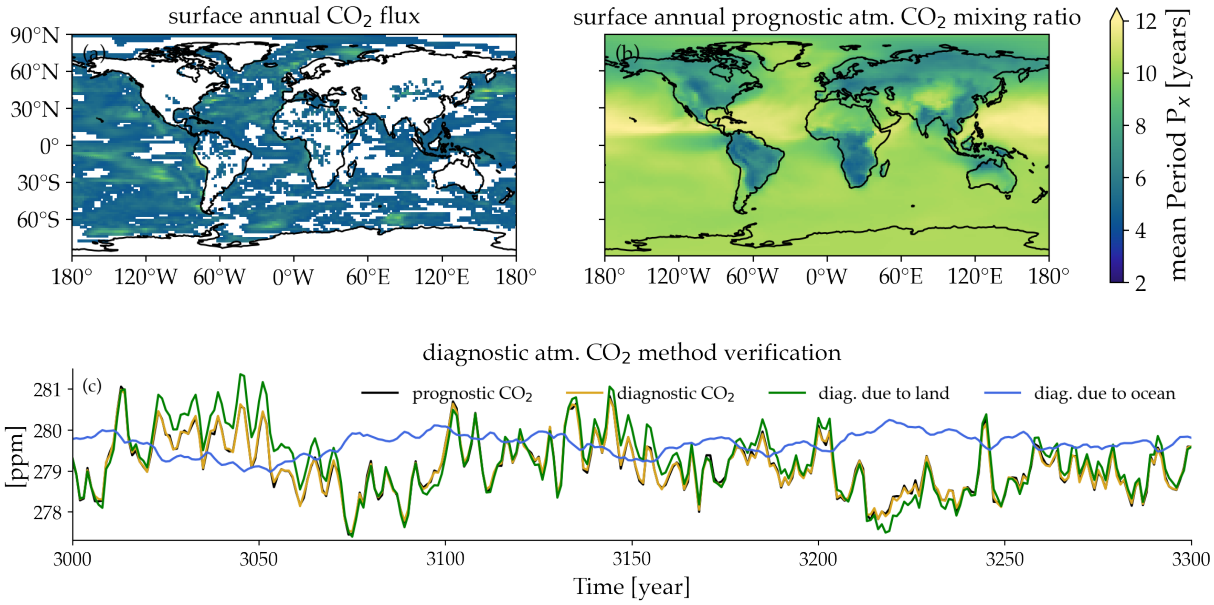


Figure II.3: **(a,b) Spatial distribution of variance-weighted mean period [Branstator and Teng, 2010] from the 300-year control simulation.** In white areas the null hypothesis, that the variable is independent and random, is not rejected at 95% significance level based on N=500 re-samplings with replacement: (a) surface annual CO_2 flux and (b) prognostic annual surface atmospheric CO_2 mixing ratio. (c) Verification of diagnostic global annual atmospheric CO_2 mixing ratio (gold) with prognostic atmospheric CO_2 (black) based on temporally accumulated surface CO_2 flux of ocean (blue) and land (green) in the control simulation [see section II.2.4]

[Figs. II.2, II.SI.2] [Li et al., 2016; Li et al., 2019; Lovenduski et al., 2019a]. These are regions of potentially predictable low-frequency air-sea CO_2 flux variations [Fig. II.3b]. However, which areas support the predictability of air-sea CO_2 flux at the global scale? Assessing the magnitude of CO_2 flux spread between ensemble members measured by RMSE shows very little relevance of the subtropical gyres for the global oceanic carbon sink predictive skill [Fig. II.SI.2]. While ACC shows linear association independent of magnitude, RMSE puts weight on the distance between ensemble members. The subtropical gyres are predictable in ACC up to a decade [Fig. II.2], but this region is irrelevant for the global oceanic carbon sink predictive skill as indicated by the low RMSE from this region in contrast to other regions. The larger areas of the Southern Ocean are predictable only for one year. We attribute this little memory in CO_2 flux to the wind-driven circulation which sets local oceanic pCO_2 in the Southern Ocean, but atmospheric processes with less memory are harder to predict than oceanic processes [Pohlmann et al., 2004; Kirtman et al., 2013; Zhang et al., 2017b,a]. Owing to the large magnitude of Southern Ocean CO_2 flux and short predictability horizon, the Southern Ocean is the main limiting contributor to the predictive skill of the global air-sea CO_2 flux [Fig. II.SI.2, II.SI.5b]. The imprint of longer regional predictive skill in the subtropical gyres is erased in the global predictive skill.

II.4 PREDICTIVE SKILL OF AIR-LAND CO_2 FLUX

Analogous to air-sea CO_2 flux, global air-land CO_2 flux is predictable for three years in ACC and for two years in RMSE, respectively [Fig. II.1b]. These results are likewise comparable to Séférian et al. [2018].

On the regional scale, terrestrial biogeochemical processes have weaker memory towards atmospheric circulation in contrast to variations of marine biogeochemical processes responding to lower-frequency oceanic circulation [Fig. II.3a]. Also, the imprints of initial conditions over land are faster lost than over the ocean as shown by faster and stronger increasing RMSE [Fig. II.1a,b; II.SI.2a-e]. This results in lower regional predictability horizons on land than over the oceans [Fig. II.2f, II.SI.2f]. Terrestrial CO_2 flux RMSE also increases to higher magnitudes due to the larger magnitudes of variability of terrestrial CO_2 flux [Fig. II.SI.5b].

As for the oceanic CO_2 flux predictive skill, we assess the relevance of different region for the global terrestrial CO_2 flux predictive skill. Terrestrial high-latitudes and deserts CO_2 flux is not predictable at lead year one in both metrics indicating no inter-annual memory. The mid-latitudes and tropical forests, which are highly variable [Fig. II.SI.5b], are predictable up to two years and thereby dominate the predictive skill of global terrestrial CO_2 flux [Fig. II.2, II.SI.2]. Zeng et al. [2008] find also highest ACC-based predictive skill in the tropics

and mid-latitudes over three seasons, while skill in other regions drops quickly within three months.

The high-RMSE mid-latitudes and tropics are also strongly variable areas in which net primary production and heterotrophic respiration respond heavily to temperature and precipitation patterns associated with ENSO [Fig. II.SI.5, II.SI.6]. Our results support ENSO as the limiting process of air-land CO_2 flux predictive skill [Keeling et al., 1995; Jones et al., 2001; Zeng et al., 2005; Zeng et al., 2008; Betts et al., 2018; Lovenduski et al., 2019b]. Note that unlike global air-sea CO_2 flux predictive skill, global air-land CO_2 predictive skill is determined by regions which also show the longest predictability horizon.

II.5 PREDICTIVE SKILL OF ATMOSPHERIC CO_2 VARIATIONS

Having examined the predictive skill of surface air-sea and air-land CO_2 fluxes, we can now explain their implications for predictive skill of the mixing ratio of surface atmospheric CO_2. Our results indicate that global atmospheric CO_2 mixing ratio is predictable for three years assessed with both metrics [Fig. II.1c]. Yet, if the combined global terrestrial and oceanic CO_2 flux is predictable for two years only [Fig. II.1a,b], how can atmospheric CO_2 mixing ratio be predictable longer than the global surface CO_2 flux? As surface CO_2 flux directly changes atmospheric CO_2 content, atmospheric CO_2 mixing ratio can be regarded as an integrator of surface CO_2 flux [Keppel-Aleks et al., 2013]. Reliable predictions in CO_2 fluxes up to two years result in longer predictive skill of their imprints on atmospheric CO_2.

To explain the limits of globally averaged atmospheric surface CO_2 predictive skill, we diagnose the individual contributions of the global terrestrial and oceanic carbon sinks to the atmospheric CO_2 mixing ratio. The predictive skill of diagnosed atmospheric CO_2 mixing ratio is very similar to that of prognostic atmospheric surface CO_2 [Fig. II.1c,f]. We find RMSE-based predictability horizons of five and 12 years for the global terrestrial and oceanic carbon sinks, respectively [Fig. II.1d,e]. Consequently, the oceanic imprint on atmospheric CO_2 would be predictable for more than a decade, but this atmospheric CO_2 signal is overlaid by the dominant imprint of the terrestrial carbon sink on atmospheric CO_2 mixing ratio. Note that the predictability horizons of the individual contributions of oceanic and terrestrial carbon sinks to atmospheric CO_2 predictability are longer than when combined [Fig. II.1d-f]. We explain this by a lower combined global carbon sink variability compared to terrestrial CO_2 flux [Fig. II.3c]. This has been also illustrated in a previous study showing that the ocean acts to suppress the atmospheric CO_2 variability. [Doney et al., 2006]. Furthermore, the terrestrial and oceanic carbon sinks contributing to the atmospheric CO_2 reservoir are both driven by climate variability,

but respond on different timescales with different magnitudes of variability.

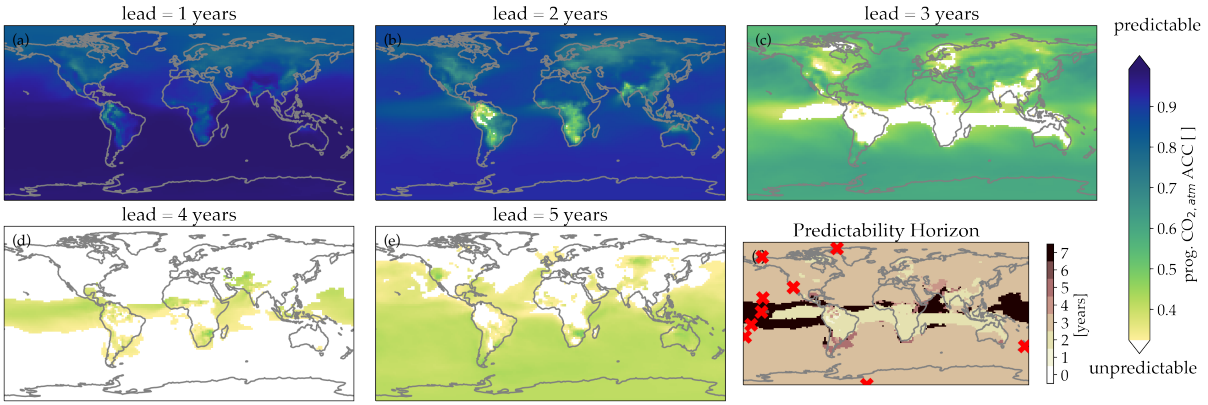


Figure II.4: **Spatial distribution of ACC-based atmospheric surface CO_2 predictive skill:** (a-e) predictive skill over the first five lead years. White areas indicate unpredictable areas where the uninitialized predictive skill exceeds the initialized skill at 5% probability based on bootstrapping with replacement (N=1000). (f) The predictability horizon marks the last significant lead year. Red crosses show location of long-standing atmospheric CO_2 mixing ratio measurement stations [Keeling et al., 2005].

The spatial distribution of atmospheric CO_2 mixing ratio ACC-based predictive skill [Fig. II.4] is strongly influenced by the subjacent surface CO_2 flux predictive skill [Fig. II.2, II.SI.2]. Atmospheric surface CO_2 mixing ratio predictive skill persists over the first two lead years over the whole globe. Predictive skill decreases faster and stronger over areas close to the terrestrial biosphere. We explain this atmospheric CO_2 predictive skill, which is more skillful than for the surface CO_2 fluxes, with the integrator feature of the atmospheric CO_2 leading to lower frequency variations [Fig. II.3a,b]. Strong high-frequency perturbations from underlying air-land CO_2 flux are subsequently transported zonally on the time scale of months [Fig. II.SI.5b, II.SI.6a,b] and dilute as they spread across the globe on annual time scales [Fig. II.2].

Interestingly, the atmospheric surface CO_2 ACC-based predictive skill, which was lost over the tropics in lead year three, re-emerges in the fourth lead year and persists over a few lead years over the tropical oceans and Southern hemisphere [Fig. II.4]. The low probabilities at which initialized forecasts beat uninitialized forecasts at lead years three to seven in ACC indicate this re-emergence also on global scale [Fig. II.1c]. However, this re-emergence, while significant, is only slightly better than uninitialized ACC and therefore unlikely to be exploited as a feature of atmospheric CO_2 predictions. The limited number of twelve initializations, in which positive, neutral and negative ENSO initial states are not normally distributed, might also

have triggered this re-emergence. We explain the weak re-emergence of ACC-based predictive skill with the oscillatory behaviour of ENSO. Strong ENSO events inject a large pulse of CO_2 into the atmosphere in the tropics [Fig. II.SI.5], which is zonally transported within months and dilutes meridionally over years [Fig. II.SI.6]. When the opposite ENSO phase is reached after the typical 3-4 years, ensemble members follow similar evolutions, which is measured with the ACC metric. As RMSE-based predictive skill lacks re-emergence [Fig. II.SI.2], this illustrates the different attributes of predictive skill and also shows the metric-dependence of predictability horizon.

We further examine atmospheric CO_2 predictive skill in the context of the existing monitoring framework [Keeling, 2008]. Due to the dilution and pole-ward transport of atmospheric CO_2 signals, the measurement station's predictability horizon is quite homogeneous over the ocean, but can be higher in the tropics due to re-emergence. The predictability horizon decreases in proximity to terrestrial CO_2 sources explainable by high terrestrial CO_2 flux variability [Fig. II.4f, Table 1]. The highest ACC-based predictability horizon from Christmas Island station of six years is explained by the re-emergence of atmospheric CO_2 predictive skill originating in the tropical forests, which reach this central Pacific island before the initial memory is lost [Fig. II.4f]. The atmospheric CO_2 mixing ratio at the Mauna Loa station, which is an appropriate observational reference for global atmospheric CO_2 variations, is predictable in ACC for three years [Fig. II.4f, II.SI.1b,c].

II.6 SUMMARY AND CONCLUSIONS

This study estimates decadal potential predictive skill of the fully-coupled global carbon cycle. Using a perfect-model approach, we estimate the upper bound of ESM-based predictive skill in the context of two metrics. We find the RMSE-based predictability horizon of the global air-sea and air-land CO_2 flux to be two years [Li et al., 2019; Lovenduski et al., 2019a; Lovenduski et al., 2019b]. While the ACC-based four- and three-year predictability horizons of the air-sea and air-land CO_2 flux signal are longer, RMSE-based predictive skill illustrates the relevance of individual sinks and regions for the predictive skill of atmospheric CO_2. Previous estimates of predictability horizon were higher [Séférian et al., 2018], but end up comparable when computed by our proposed bootstrapping methodology, which is backed by the established prediction frameworks [Goddard et al., 2013; Marotzke et al., 2016; Yeager et al., 2018].

We furthermore show that the global oceanic CO_2 flux predictive skill is mainly controlled by the Southern Ocean. The global terrestrial CO_2 flux predictive skill is dominated by the tropical forests and mid-latitudes affected by ENSO. We find a different regional contribution pattern for land and ocean to global predictive skill of CO_2 flux. On

land, predictive skill is maintained due to regions with the longest predictability horizon. On the contrary in the ocean, regions with longest predictability horizon are less relevant for the global signal.

The predictability horizon for climate modulated variations in the global atmospheric CO_2 mixing ratio is three years. The isolated effect of global terrestrial carbon sink on atmospheric CO_2 of five years dominates over the oceanic contribution of 12 years. Consequently, while the oceanic carbon cycle dampens the imprint of the terrestrial carbon cycle on atmospheric CO_2, the predictive skill of atmospheric CO_2 is limited by the terrestrial carbon sink.

The atmospheric CO_2 predictability horizon is globally quite homogeneous, except over parts of the tropical Pacific, because of a weak but significant re-emergence pattern which might also affect initialized predictions. As the dominant features of terrestrial CO_2 flux predictive skill in the tropics arise from ENSO, multi-year forecasts performed by initialized prediction systems for atmospheric CO_2 variations are expected to be challenging.

In this study, we demonstrate that ESM-based initialized forecasts of the global carbon cycle may deliver multi-year outlooks on the evolution of the atmospheric CO_2 mixing ratio. Hence, ESM-based predictions have the potential to constrain uncertainty of changes in atmospheric CO_2 due to internal climate variability in the near-term future and can thereby inform the pentadal stocktakes [UNFCCC, 2015].

ACKNOWLEDGEMENTS

Forecast verification was computed with the python package climpred https://github.com/bradyrx/climpred/, which was co-developed with Riley X. Brady from University of Colorado, Boulder.

We acknowledge funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 821003 "Climate-Carbon Interactions in the Coming Century (CCiCC)" and No 641816 (CRESCENDO). T.I. received funding from the Federal Ministry of Education and Research in Germany (BMBF) through the research programme "MiKlipII" (FKZ: 01LP1517B). Simulations were performed at the German Climate Computing Center (DKRZ).

We thank Leonard Borchert for internal review.

II.SI SUPPLEMENTARY INFORMATION

II.SI.1 *Reproducibility*

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

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

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

II.SI.2 *Predictability Horizon at Atmospheric CO_2 Measurement Stations*

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

Table 1: **RMSE- and ACC-based predictability horizon for atmospheric CO_2 mixing ratio at locations of long-standing atmospheric CO_2 measurement stations in years.** Station locations are taken from https://cdiac.ess-dive.lbl.gov/trends/co2/sio-keel.html.

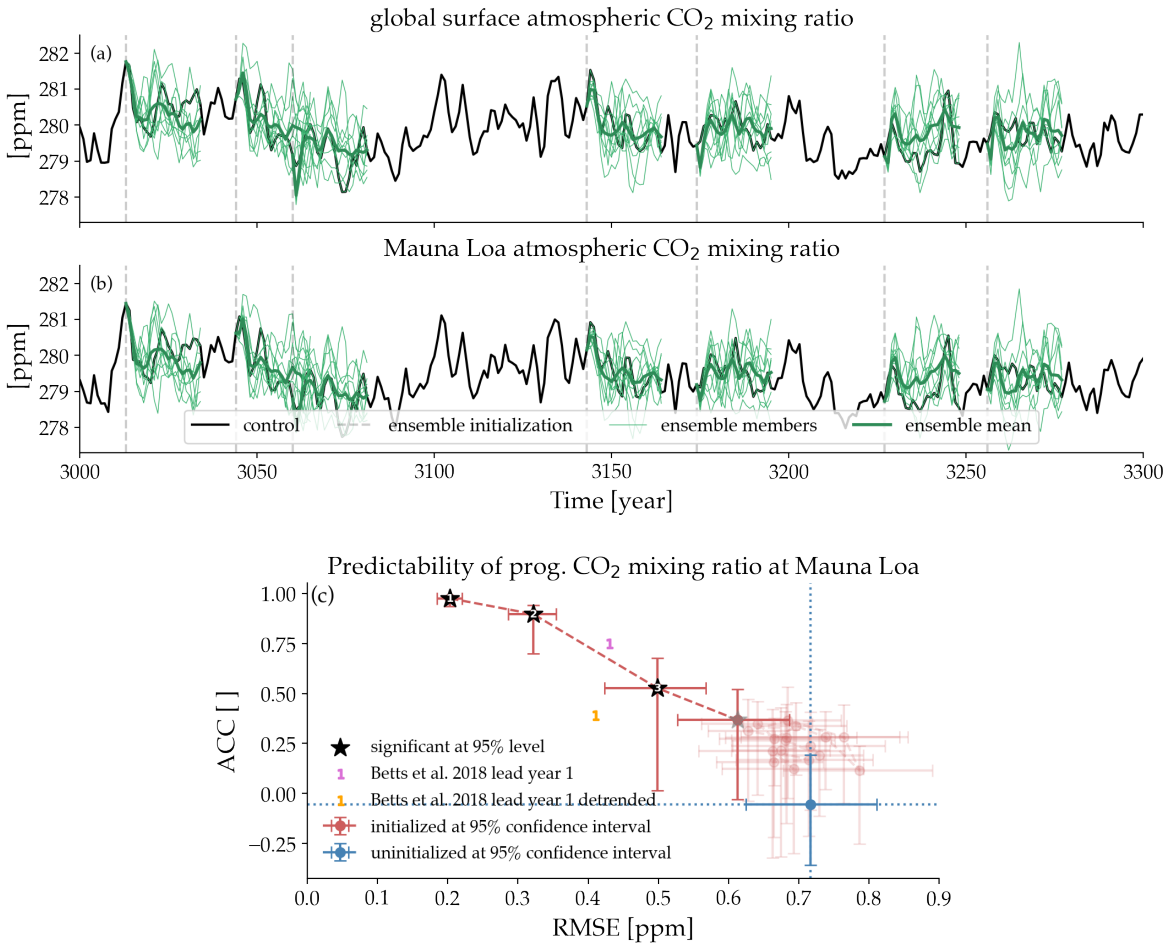


Figure II.SI.1: **Evolution of the annual mixing ratio of atmospheric CO_2:** (a) globally-averaged and at Mauna Loa. The ensemble mean (dark green) is taken from individual ensemble members (green), which are branched off a pre-industrial control run (black) at different ensemble initialization years (7 out of 12 shown in dotted gray). (c) **Comparison of the mean potential prediction skill.** Initialized ensemble (red) versus random uninitialized ensembles (blue) of prognostic atmospheric CO_2 at Mauna Loa, Hawaii with Anomaly correlation coefficient (ACC) on the y-axis and root-mean-square-error (RMSE) on the x-axis for lead years represented as dots. Errorbars show 95% confidence intervals based on bootstrapping with replacement (N=5000). The last lead year with a bootstrapped p-value (which represents that uninitialized ensembles beat initialized ensembles) lower than 5% marks the predictability horizon. Black stars with white integer denote significant lead years in ACC and RMSE, gray stars if only one metric is significant and lead years non-significant in both metrics are blurred. For comparison with Betts et al. [2016, 2018], ones show predictive skill of the statistical model for lead-year one for the transient forecast (blue) and the detrended forecast (orange). Non-significant lead years are blurred.

II.SI.3 *RMSE-based Predictive Skill of Surface CO_2 Flux*

For completeness and comparison, we calculate predictive skill maps of RMSE (for comparison to manuscript Figs. II.2 and II.4 in ACC) for atmospheric CO_2 mixing ratio and surface CO_2 flux. Predictive skill has similar patterns as ACC, but initialized forecasts perform better than uninitialized forecasts for fewer lead years. Furthermore, there is no re-emergence in RMSE-based predictive skill of surface atmospheric CO_2.

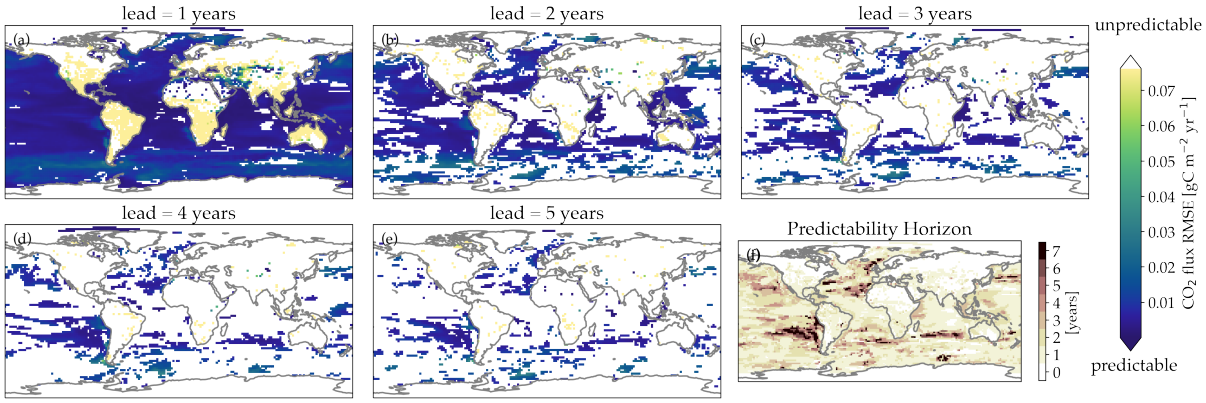


Figure II.SI.2: **Spatial distribution of RMSE-based surface CO_2 flux predictive skill:** (a-e) predictive skill over the first five lead years. White areas indicate unpredictable areas where the uninitialized predictive skill exceeds the initialized skill at 5% probability based on bootstrapping with replacement (N=1000). (f) The predictability horizon marks the last significant lead year.

II.SI.4 *Statistics*

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

(R)MSE-based predictive skill has been mostly used in the past to assess potential predictability [Griffies and Bryan, 1997; Pohlmann et al., 2004; Séférian et al., 2018]. However, here for atmospheric CO_2 mixing ratio, ACC-based (also used in [Bushuk et al., 2018]) predictive skill comes closer to the raised expectation when predicting something and is therefore primarily used in this study when assessing atmospheric CO_2 predictability.

II.SI.4.1 *ACC*

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

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

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

II.SI.4.2 *RMSE*

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

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

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

II.SI.4.3 *Comparison of Predictability Horizon Definitions*

The differences in predictability horizon between Séférian et al. [2018] and our study arise from different interpretations of what defines the predictability horizon. While Séférian et al. [2018] define the limit at the saturation level of later lead years close to the magnitude of the standard deviation, we define the predictability horizon above a

threshold value at which initialized forecasts cease to perform better than random, uninitialized forecasts as defined by Buizza and Leutbecher [2015]. The break-point fit of Séférian et al. [2018] resembles a 50% bootstrapping and results therefore in by design longer predictability horizons compared to this study [Fig. II.SI.3].

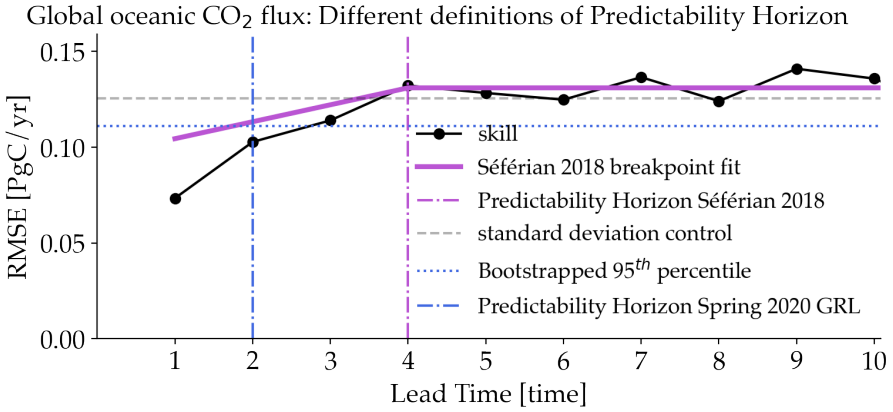


Figure II.SI.3: **Evolution of global CO_2 flux RMSE over lead time with different definitions for predictability horizon.** Definitions of predictability horizon used in Séférian et al. [2018] (orchid) leads to systematically higher predictability horizon than when using the methodology from our study (blue).

II.SI.5 *RMSE-based Predictive Skill Surface Atmospheric CO_2*

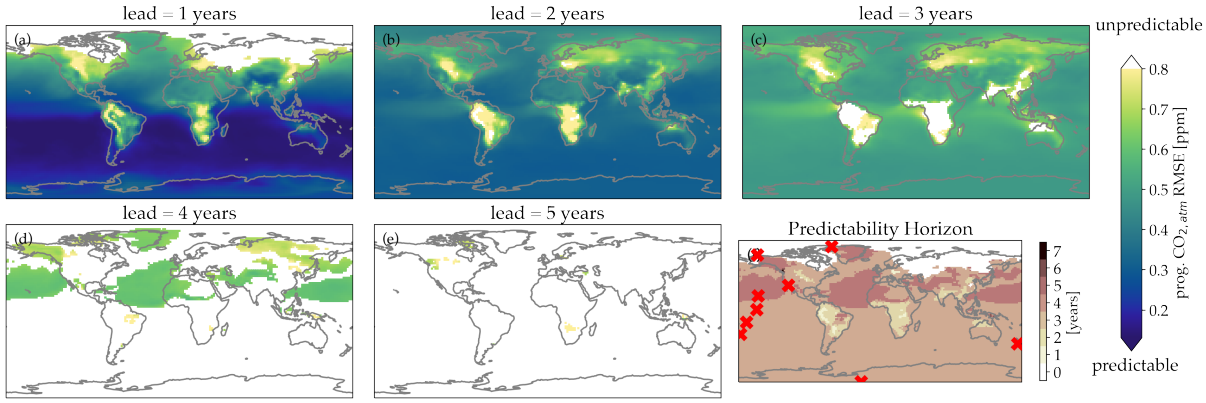


Figure II.SI.4: **Spatial distribution of RMSE-based atmospheric surface CO_2 predictive skill**: (a-e) predictive skill over the first five lead years. White areas indicate unpredictable areas where the uninitialized predictive skill exceeds the initialized skill at 5% probability based on bootstrapping with replacement (N=1000). (f) The predictability horizon marks the last significant lead year. Red crosses show location of long-standing atmospheric CO_2 mixing ratio measurement stations [Keeling et al., 2005].

II.SI.6 *Influence of ENSO on carbon cycle variability*

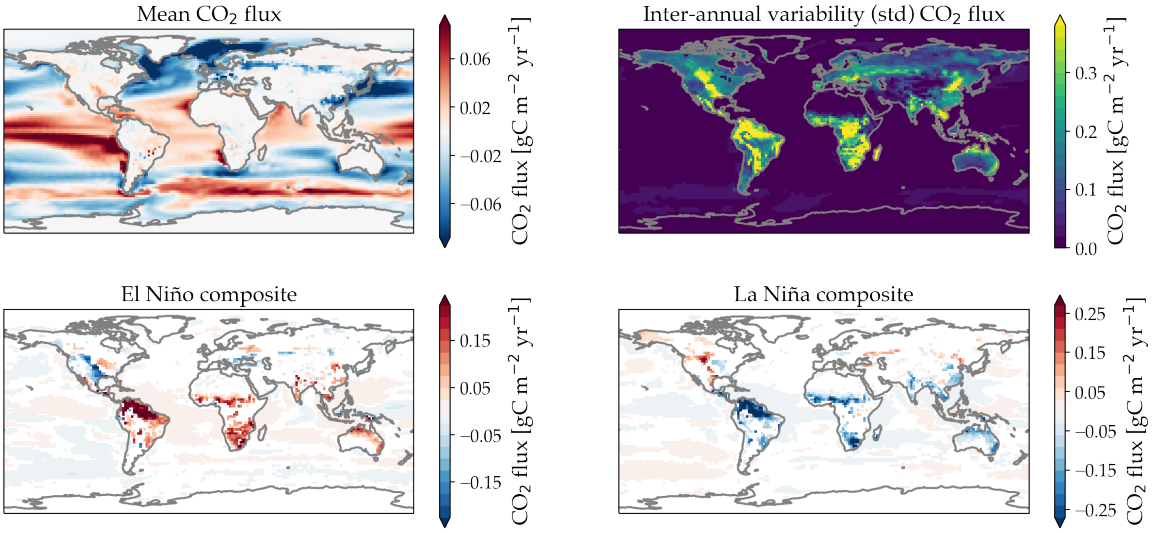


Figure II.SI.5: **Spatial distribution of surface CO_2 flux:** (a) annual mean, (b) inter-annual variability determined as standard deviation (c, d) the composite of the positive/negative ENSO 3.4 index states. Colored areas indicate that the composite is different from the neutral ENSO 3.4 state assessed with a t-test at 95% significance.

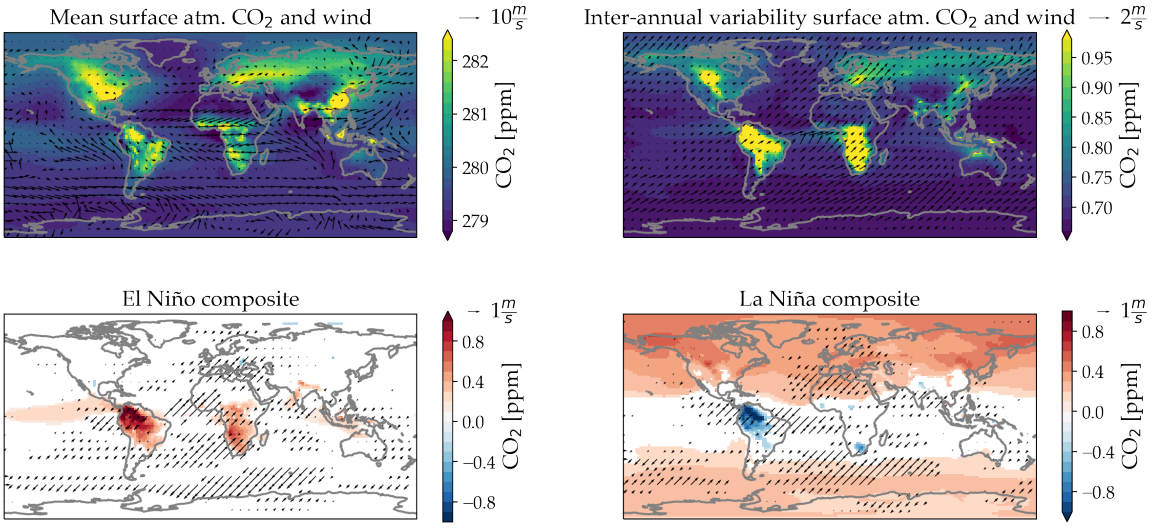


Figure II.SI.6: **Spatial distribution of atmospheric CO_2 mixing ratio overlain with 10m wind indicated as arrows:** (a) annual mean, (b) inter-annual variability determined as standard deviation (c, d) the composite of the positive/negative ENSO 3.4 index states. Colored areas indicate that the composite is different from the neutral ENSO 3.4 state assessed with a t-test at 95% significance.

II.SI.7 *Comparison of Modelled Inter-annual Variability with Observations*

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

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

II.SI.7.1 *Oceanic CO_2 flux*

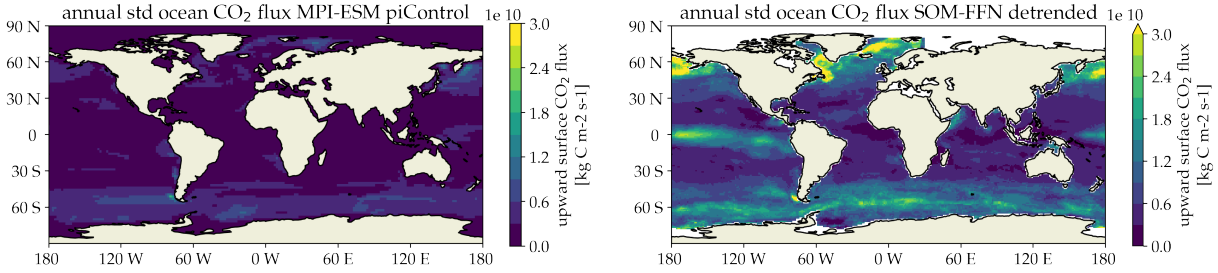


Figure II.SI.7: **Spatial distribution of inter-annual variability computed as standard deviation of annual mean oceanic CO_2 flux:** (a) MPI-ESM pi-esmControl, (b) linear detrended SOM-FFN (1982-2015) [Landschützer et al., 2016].

Hotspots of oceanic CO_2 flux variability in SOM-FFN [Landschützer et al., 2016] in the Southern Ocean and North Pacific are captured by MPI-ESM. MPI-ESM under-estimates oceanic CO_2 flux variability in the equatorial pacific with respect to the detrended SOM-FFN. This feature is less pronounced after 4^{th}-order detrending. Furthermore, SOM-FFN is just one of several SOCOM data products [Rödenbeck et al., 2015] which fill the various gaps of the gridded measurement data. Given the existing uncertainty in data filling methods, a precise estimation of variability is not conclusive. Furthermore, Landschützer et al. [2019] show that the length of the observational records is insufficient to fully capture natural variability signals.

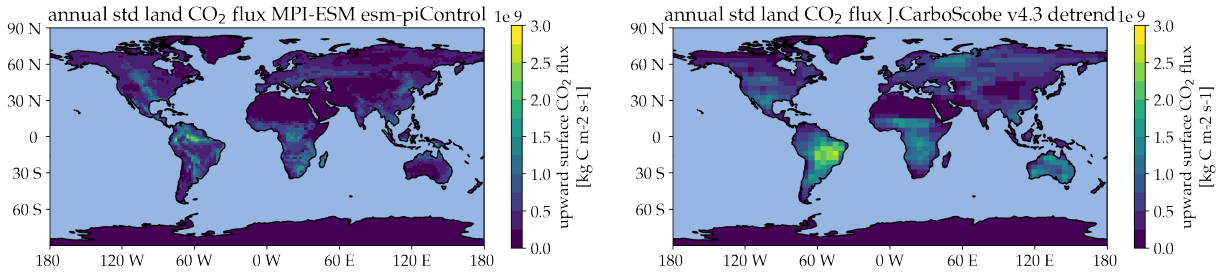


Figure II.SI.8: **Spatial distribution of inter-annual variability computed as standard deviation of annual mean atmospheric CO_2 mixing ratio**: (a) MPI-ESM pi-esmControl, (b) second-order detrended atmospheric CO_2 inversion Jena CarboScope sEXTocNEET_v4.3 (1957-2018) [Rödenbeck et al., 2018].

II.SI.7.2 *Terrestrial CO_2 flux*

Terrestrial CO_2 flux variability is highly unconstrained, as there is currently no direct observation-based terrestrial CO_2 flux product available. However, for comparison here, we use the observations-based atm. CO_2 inversion [Rödenbeck et al., 2018] as an estimate for spatio-temporal gridded terrestrial CO_2 flux. While MPI-ESM generally over-estimates the magnitudes in terrestrial CO_2 flux variability, the origins of high variability in the tropics and mid-latitudes are well captured.

II.SI.7.3 *Surface Atmospheric CO_2*

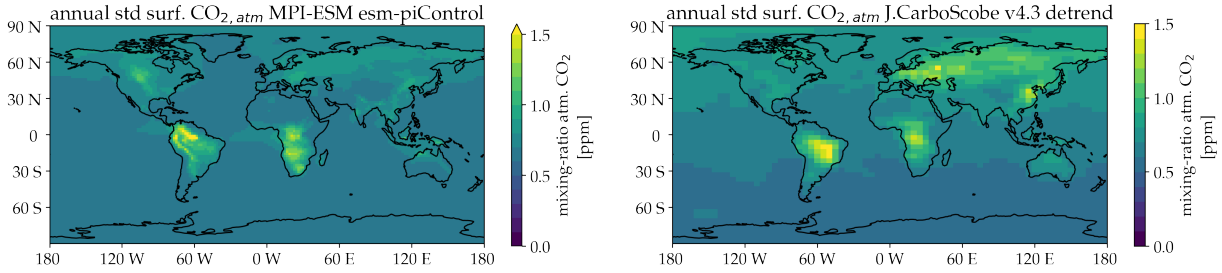


Figure II.SI.9: **Spatial distribution of inter-annual variability computed as standard deviation of annual mean atmospheric CO_2 mixing ratio:** (a) MPI-ESM pi-esmControl, (b) second-order detrended atmospheric CO_2 inversion Jena CarboScope sEXTocNEET_v4.3 (1957-2017) [Rödenbeck et al., 2018].

The effect of the internal variability of both oceanic and terrestrial CO_2 flux on atmospheric CO_2 is well captured by MPI-ESM. The higher variability over the northern Hemisphere based on observations may be explained by the anthropogenic emissions not present in MPI-ESM *esm-piControl*. The other hotspots of variability from the CO_2 inversion also appear in the model.

TRIVIAL IMPROVEMENTS OF PREDICTIVE SKILL DUE TO DIRECT RECONSTRUCTION OF THE GLOBAL CARBON CYCLE

The attached article is under review in *Earth System Dynamics*:

Spring, A., Dunkl, I., Li, H., Brovkin, V., & Ilyina, T. [2021]. Trivial improvements of predictive skill due to direct reconstruction of the global carbon cycle. *Earth System Dynamics Discussions*, 1–36. doi:10/gh3tn3

Author contributions: A.S. and T.I. conceived the study, A.S. performed the simulations and analysis, created the figures and drafted the manuscript. T.I., I.D., H.L. and V.B. contributed in manuscript editing and provided feedback.

Trivial improvements of predictive skill due to direct reconstruction of the global carbon cycle

Aaron Spring[1,2], **István Dunkl**[1,2], **Hongmei Li**[1], **Victor Brovkin**[1,3] **and Tatiana Ilyina**[1]

[1] Max Planck Institute for Meteorology, Hamburg, Germany

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

[3] Center for Earth System Research and Sustainability, University of Hamburg, Germany

Submitted: 8 February 2021

ABSTRACT

State-of-the art climate prediction systems include a carbon component recently. While physical state variables are assimilated in reconstruction simulations, land and ocean biogeochemical state variables adjust to the state acquired through this assimilation indirectly instead of being assimilated themselves. In the absence of comprehensive biogeochemical reanalysis products, such approach is pragmatic. Here we evaluate a potential advantage of having perfect carbon cycle observational products to be used for direct carbon cycle reconstruction.

Within an idealized perfect-model framework, we reconstruct a 50-year target period from a control simulation. We nudge variables from this target onto arbitrary initial conditions, mimicking an assimilation simulation generating initial conditions for hindcast experiments of prediction systems. Interested in the ability to reconstruct global atmospheric CO_2, we focus on the global carbon cycle reconstruction performance and predictive skill.

We find that indirect carbon cycle reconstruction through physical fields reproduces the target variations on a global and regional scale much better than a resampling threshold. While reproducing the large scale variations, nudging introduces systematic regional biases in the physical state variables, on which biogeochemical cycles react very sensitively. Initial conditions in the oceanic carbon cycle are sufficiently good reconstructed indirectly. Direct reconstruction slightly improves initial conditions. Indirect reconstruction of global terrestrial carbon cycle initial conditions are also sufficiently good reconstructed by reconstructing the physics only. Direct reconstruction

improves air-land CO_2 flux negligibly. Atmospheric CO_2 is very well indirectly reconstructed. Direct reconstruction of the marine and terrestrial carbon cycles slightly improve reconstruction while establishing persistent biases at the same time. We find improvements in global carbon cycle predictive skill from direct reconstruction compared to indirect reconstruction. After correcting for mean bias, indirect and direct reconstruction both predict the target similarly well and only moderately worse than perfect initialization after the first lead year.

Our perfect-model study shows that indirect carbon cycle reconstruction yields satisfying initial conditions for global CO_2 flux and atmospheric CO_2. Direct carbon cycle reconstruction adds little improvements in the global carbon cycle, because imperfect reconstruction of the physical climate state impedes better biogeochemical reconstruction. These minor improvements in initial conditions yield little improvement in initialized perfect-model predictive skill. We label these minor improvements due to direct carbon cycle reconstruction *trivial*, as mean bias reduction yields similar improvements. As reconstruction biases in real-world prediction systems are likely stronger, our results add confidence to the current practice of indirect reconstruction in carbon cycle prediction systems.

III.1 INTRODUCTION

Predicting variations in weather and climate yields numerous benefits for economic, social, and environmental decision-making [Merryfield et al., 2020]. Carbon cycle prediction systems have the ability of predicting the near-term evolution of CO_2 fluxes [Li et al., 2019; Lovenduski et al., 2019b; Lovenduski et al., 2019a] and atmospheric CO_2 [Spring and Ilyina, 2020; Ilyina et al., 2021] to constrain the large internal variability of the global carbon cycle [Spring et al., 2020]. Predictions require a forecasting model and initial conditions representing observations. However, due to sparse and temporally incomplete records, there is currently no global biogeochemical reanalysis product to initialize Earth System Models (ESMs). Therefore, direct initialization of the carbon cycle is not possible. State-of-the-art carbon prediction systems initialize the carbon cycle indirectly, by nudging the physical climate, assuming that carbon cycle follows the initialized climate indirectly. Whereas a direct initialization of the carbon cycle is not feasible due to the lack of data. However, this indirect carbon cycle initialization leaves the initial conditions of the carbon cycle unconstrained.

Here, we test how well indirect and direct carbon cycle reconstructions in an ESM initialize the carbon cycle in a perfect-model framework [Table 2 presents an overview which variables are reconstructed in which simulation]. We use the term reconstruction to describe methods of initialization of climate and the carbon cycle. Reconstructions aim to reproduce the evolution of the target, like a reanalysis product, in the ESM. Furthermore, we use the term "carbon cycle" to describe the processes exchanging carbon across the surface boundary between land, atmosphere and ocean, represented here by the air-land and air-sea CO_2 fluxes. We ask the following research questions:

- How well can initial conditions be reconstructed in the global carbon cycle?
- Can initialization of the carbon cycle improve predictive skill of the carbon cycle?

In this perfect-model framework, we have perfect knowledge about the ground truth and a perfect model. Literally speaking, this asks how well could perfect observations be reconstructed in an ESM.

Originally, data assimilation is used to align the model state to an observations-based state, generally a reanalysis product [Schneider and Griffies, 1999; Meehl et al., 2009]. However, here we use the same data assimilation technique to assess how well variables can be reconstructed in an idealized setup.

Thus, reconstruction in a climate model interferes with the freely running climate model yielding gains and drawbacks: The main ad-

vantage of climate reconstruction is that the reconstruction forces the climate model to follow the target [Jeuken et al., 1996; Meehl et al., 2009]. The main handicap associated with reconstruction is that the mass conservation is violated and that the model dynamics and feedbacks are obstructed [Zhu and Kumar, 2018]. Consequently, circulation fields may change, and this has severe consequences for the biogeochemical tracer distributions in the ocean and carbon pools on land, because they are so sensitive and adapted to the previous climate state [Toggweiler et al., 1989]. Therefore, reconstructions often lead to biases. A partial solution can be bias removal by post-processing, which is feasible if the bias does not change the climate or ecosystem regime all together. Another solution is omitting nudging in sensitive biasing regions such as the tropics, as demonstrated by [Park et al., 2018]. Even if biogeochemical reanalysis products were available, it is unclear whether the reconstruction benefits correct these handicaps.

The lack of reanalysis products available for the reconstruction of carbon cycle initial conditions is often assumed as a weakness of the current predictions systems [Li et al., 2016; Séférian et al., 2018; Lovenduski et al., 2019a; Lovenduski et al., 2019b; Li et al., 2019; Ilyina et al., 2021], but to our knowledge an elaborate assessment is missing. The literature presents two alternative approaches to test the quality of reconstructed initial conditions:

In a perfect-model study, Servonnat et al. [2015] nudge only ocean surface temperature, salinity and sea-ice and assess how well this surface reconstruction penetrates into the subsurface ocean physics, without addressing biogeochemistry in their analysis. This target reconstruction approach allows us to directly assess the quality of reconstructed initial conditions, which is useful and practical to know for forecaster issuing a forecast.

In a recent study Fransner et al. [2020] ask whether the initial conditions of ocean biogeochemistry or the initial conditions of ocean physics have a stronger influence on multi-year predictions using perfect-model twin perturbed initial conditions experiments. In the first set of hindcasts, they take identical initial conditions of the ocean physics to ensure identical climate evolution but completely different states from different members for ocean biogeochemistry. In the other set of hindcasts, they slightly perturb the ocean physics to force members on differing climate evolutions while keeping the ocean biogeochemistry initial conditions identical. They find that ocean biogeochemistry initial conditions did not affect predictive skill later than the first lead year. Their approach asks the more theoretical question whether initial conditions of ocean biogeochemistry matter compared to ocean physics initial conditions.

We go beyond previous studies by using the methodology of Servonnat et al. [2015], with the aim to understand the quality of initial

conditions reconstruction. In contrast to Fransner et al. [2020], we aim to answer the questions about quality of initial conditions produced by different reanalysis approaches. We expand the scope by addressing the global carbon cycle, including the land, ocean and atmospheric compartments and the interactive exchange of CO_2 fluxes between them. We then assess the influence of these previously reconstructed carbon cycle initial conditions for initialized predictions of the natural carbon sinks and atmospheric CO_2. We focus on the global carbon cycle, because the land and ocean carbon cycle control the internal variability of atmospheric CO_2 [Friedlingstein et al., 2020].

After explaining the approach of target reconstruction in section III.2, we separate reconstruction and its implication on predictive skill in two parts: We first evaluate physical reconstruction representing indirect biogeochemical reconstruction in sections III.3.2.1, III.3.3.1 and III.3.4.1 for spatial maps and global aggregated values. Furthermore, we test the potential tracking performance increase for direct carbon cycle reconstruction in sections III.3.2.2, III.3.3.2 and III.3.4.2. In part two, we assess the impact of different reconstruction methods on initial conditions predictive skill in section III.4. Finally, the main findings and conclusions of this study are summarized in section III.5.

III.2 METHODS

III.2.1 *Model Description*

We use the Max Planck Institute ESM [Mauritsen et al., 2019, MPI-ESM], which was also used in the Coupled Model Intercomparison Project Phase 6 (CMIP6) framework [Eyring et al., 2016]. We run the model MPI-ESM1-2-LR, the low resolution configuration with 63 spherical harmonics in the atmosphere and with a horizontal resolution of about 1.8° on land, and about 1.5° in the ocean with daily coupling of the compartments. The time steps of the atmosphere/land and the ocean are 600 and 4320 s, respectively. We run the model with prognostic atmospheric CO_2 mixing ratio under pre-industrial conditions (*esm-piControl*).

The marine biogeochemical cycle model HAMOCC [Ilyina et al., 2013] is embedded in the ocean general circulation model MPIOM [Jungclaus et al., 2013]. HAMOCC includes carbonate chemistry and an extended NPZD-type cycle including nutrient-light-temperature co-limitation and nitrogen-fixating cyanobacteria [Paulsen et al., 2017]. The land carbon cycle model JSBACH includes dynamic vegetation, wildfires, soil carbon decomposition and storage [Schneck et al., 2013]. The atmospheric general circulation model ECHAM6 transports the three-dimensional atmospheric prognostic atmospheric CO_2 tracer

with a flux-form semi-Lagrangian scheme [Lin and Rood, 1996; Stevens et al., 2013].

III.2.2 *Perfect-Model Target Reconstruction Framework*

Simulations in a perfect-model target reconstruction framework aim to reproduce the target climate evolution [Griffies and Bryan, 1997; Servonnat et al., 2015], but are started from an independent initial state. Therefore the initial conditions of the reconstruction simulation and the the target to not match. But both target and initial conditions share the same climatology. We choose a 50-year target period from model years 1850 to 1900 and an uncorrelated restart file from model year 2005 from the pre-industrial control simulation (esm-piControl) submitted for the MPI-ESM1-2-LR model for C4MIP [Jones et al., 2016] in CMIP6 [Eyring et al., 2016].

In order to assess how many variables are needed to sufficiently reconstruct climate and biogeochemical cycles, we first perform reconstruction simulations only reconstructing physical state variables in atmosphere and/or ocean [Table 2]. In these simulations, the carbon cycle is only indirectly affected by the reconstruction of physical variables. In further simulations, we test how much carbon cycle states improve with respect to the target when carbon cycle state variables are reconstructed directly.

III.2.3 *Reconstruction Simulations*

Newtonian or Haney [1974] relaxation, which is often called *nudging*, is a simple four-dimensional assimilation technique that dynamically reconstructs variables in an ESM. A non-physical relaxation term with relaxation coefficient R (units $1/s$) is added to the prognostic equation to drag the model variable X, which is subject to model forcing F_m, towards its target X_t:

$$\frac{\delta X}{\delta t} = F_m(X) + R(X_t - X) \tag{7}$$

For reconstruction of the dynamics of the ocean, we reconstruct three-dimensional temperature and salinity as well as sea-ice concentration and thickness [Table 2]. We label this reconstruction indirect [Table 2] from the carbon cycle's perspective, as the carbon cycle is not reconstructed directly, but instead indirectly follows the reconstructed physical climate. Observational ocean data is often not available at each model time step. Therefore, we interpolate (without adjustments preserving the temporal mean) monthly model target output to daily frequency as done in previous studies [Pohlmann et al., 2009]. We choose a 60-day ocean relaxation time (converted to units $1/s$) like Servonnat et al. [2015] in their perfect-model target reconstruction

	Reconstructed variables for each realm (nudging relaxation time-scale)				
Reconstruction simulations	Atmosphere: temperature (24h) surface pressure (24h) vorticity (6h) divergence (48h)	Ocean (60d): temperature salinity	Sea-ice (60d): concentration thickness	Ocean carbon (60d): DIC alkalinity	Land: all JSBACH (reset restart files Jan 1st)
$\text{indirect}_{\text{ATM only}}$	x				
$\text{indirect}_{\text{OCEAN only}}$		x			
indirect	x	x	x		
direct	x	x	x	x	x

Table 2: Overview over different reconstruction simulations. The first column title marks the labels of the experiments as used in the manuscript. The reconstruction strength as relaxation time-scales is noted in brackets, where h denotes hours and d days. The land carbon cycle is not dynamically reconstructed at each time step, but by a hard reset of restart files each January 1st from the target run. These land restart files include carbon and nitrogen pools, soil physics (moisture, temperature, snow cover), vegetation cover (plant functional types distribution), and canopy (leaf area index).

study. Reconstructions towards observations usually choose a stronger nudging strength [Pohlmann et al., 2009; Keenlyside et al., 2008].

We reconstruct the physics of the atmosphere by nudging temperature, vorticity, divergence and the logarithm of surface pressure [Pohlmann et al., 2019]. The high-frequency 6 hourly output serves as the target and is nudged into all 63 spherical harmonics. Temperature and the logarithm of surface pressure are nudged with a relaxation timescale of 24 hours, vorticity is nudged with a relaxation timescale of 6 hours, and divergence is nudged with a relaxation timescale of 48 hours. Relaxation coefficients are converted to units $1/s$ and are taken from previously used setups [Rast et al., 2012; Pohlmann et al., 2019; Li et al., 2019]. Nudging the atmosphere with these quite short relaxation times is similar to the forced simulations, such as the Model Intercomparison Projects for ocean (OMIP) [Griffies et al., 2016; Orr et al., 2017], land (LMIP) [van den Hurk et al., 2016] and Global Carbon Budget [Friedlingstein et al., 2019] simulations, where (atmospheric) external boundary forcing drives the carbon cycle.

For reconstructions of oceanic carbon cycle, we use the same nudging approach and strength as for physical ocean reconstruction but on different variables. To reconstruct the components of the carbonate system, we nudge three-dimensional dissolved inorganic carbon (DIC) and total alkalinity [Table 2].

Unfortunately, there is no nudging module available in the land surface model JSBACH. Here we choose to manually reset the initial conditions every January 1st to the target values instead of the dynamic reconstruction at each time step. We thereby reconstruct land

biogeochemistry and land surface physics such as soil moisture by resetting all restart variables every year. In supplementary information section III.SI.3, we provide several sensitivity analyses by resetting land only every two or five years and resetting the ocean every year in the same way.

III.2.4 *Evaluating Tracking Performance*

We compare the target with reconstructions in the various metrics showing different attributes of tracking performance: bias, anomaly correlation coefficient and root-mean-square-error. The non-physical relaxation terms in the prognostic equations can disturb the dynamics in the ESM and introduce biases defined as the differences in the reconstruction compared to the freely running target over time. The anomaly correlation coefficient skill score (ACC) shows the linear association between the reconstruction and the target over time and therefore measures synchronous evolution while ignoring bias. The root-mean-square-error (RMSE) takes into account bias and measures the second-order euclidian distance between reconstruction and target simulation over time. Under the assumption that persistent biases can be removed by post-processing, we also assess RMSE after having the mean monthly bias removed. For equations please consult the supplementary [sec. III.SI.1]. We calculate tracking performance over running 10-year chunks to capture the variability within tracking performance and reduce the influence of drifts over time.

How do we evaluate that a reconstruction is good enough? While good enough is a subjective judgement, we resample the target simulation along the time dimension with a block length of ten years to check the metric of two randomly compared 10-year chunks. We consider the 95^{th} quantile threshold for ACC and 5^{th} quantile threshold for the remaining distance-based metrics as a baseline of internal variability to be a good enough reconstruction [Efron and Tibshirani, 1993], which we will refer to as "resampling threshold" in the following.

III.2.5 *Perfect-Model Predictive Skill Framework*

In the second part of this study, we perform initialized perfect-model experiments [as in Spring and Ilyina, 2020]. The simulations in the perfect-model framework are started from the indirect and direct reconstructions as well the target representing perfect initial conditions. We take 19 initialization states chosen every second January 1^{st} between 1860 and 1896, after allowing a 10 years adjustment phase after reconstructions were started. From each of those states from different reconstruction simulations, we fork five ensemble members and simulate three lead years. The perfectly initialized ensembles are started from the target initial conditions without any previous reconstruc-

tion simulation. We generate ensemble members by perturbing the stratospheric horizontal diffusion by a factor of 1.0000{member} in the first year. This member generating approach provokes only tiny initial perturbations to the climate system as the ocean and land initial conditions remain identical.

III.2.6 *Predictive Skill Quantification*

We compute predictive skill as the root-mean-square-error (RMSE) between the ensemble mean and the target as verification [Wilks, 2006; Jolliffe and Stephenson, 2011] [sec. III.SI.1]. Please find additional details about the predictive skill metrics and the uninitialized bootstrapping in Spring and Ilyina [2020]. Acknowledging that our reconstruction simulation developed biases and that biases are commonly reduced by post-processing in predictability research, we also apply a simple lead-time dependent mean bias reduction to the initialized ensembles to show whether skill improvements go beyond what a simple post-processing could deliver. For each initialization in turns, we first calculate the mean bias for all but that given initialization and then remove that mean bias from the given initialization. This implies using information about future initializations as in bias-reduced hindcasts [Marotzke et al., 2016]. We also evaluate predictive skill from a perfectly initialized ensemble, which are started from the perfect initial conditions taken from target simulation, whereas the ensembles from reconstructed initial conditions are biased with respect to the target [Fig. III.5]. This initialized predictive skill is also compared with uninitialized ensembles randomly generated from the target simulation representing ensembles without common initialization and hence no memory. This uninitialized reference skill is used in predictability research community to assign whether the skill increase stems from initialization.

III.3 RECONSTRUCTION IN AN EARTH-SYSTEM-MODEL

As the carbon cycle is sensitive to the climate evolution, we first assess how well the physical climate is reconstructed. Therefore, we first evaluate the physical climate state after reconstruction in subsection III.3.1]. Afterwards, we assess how these different reconstructions of physical climate indirectly reconstruct the ocean, land and atmospheric carbon cycle in subsections III.3.2.1, III.3.3.1 and III.3.4.1]. The direct reconstructions are shown in subsections III.3.2.2, III.3.3.2 and III.3.4.2].

III.3.1 *Reconstruction of Physical Climate*

Reconstructing the ocean and/or the atmosphere systematically disturbs the freely evolving model, which leads to annual mean biases

with respect to the original target. We identify atmospheric circulation represented by winds and resulting precipitation and temperature to be descriptive for the impact of circulation on the carbon cycle. The gray stippling in figure III.1 shows where this reconstruction bias is larger than the randomly resampling 5^{th} percentile mean absolute error threshold and therefore labeling the reconstruction not significantly better than internal variability.

All reconstructions yield identical results for winds and precipitation tracking performance. Reconstructing the ocean and/or the atmosphere introduces biases of up to 0.6 m/s in zonal and 0.9 m/s in meridional 10-m wind speed, depicting a southward shift of the Intertropical Convergence Zone (ITCZ). This bias results in a significant weakening of the equator-ward latitudinal winds, whereas extra-tropical latitudinal winds intensify [Fig. III.1a]. The intensification and equator-ward shift of the easterly trade winds and weakening of the southern hemisphere westerlies are both not significant [Fig. III.1b]. Precipitation is heavily impacted by these biases in atmospheric transport across many regions of the globe. Precipitation significantly shifts southward at the equator with changes of more than 1 mm/day and increases in Western Canada, Western Russia and Southern Australia [Fig. III.1c]. Unlike the previously described variables, the 2m-temperature bias depends on whether the ocean is reconstructed or not. Just reconstructing the ocean temperature and salinity ($\text{indirect}_{\text{OCEAN only}}$) leads to small, negative and significant biases in the tropical Atlantic and West Pacific. Also Northern and Southern Africa as well as the Amazon and China are subject to a small cold bias, whereas Saharan Africa and Southeast Asia gets substantially warmer. The polar regions cool significantly [Fig. III.1d]. Only reconstructing the atmosphere ($\text{indirect}_{\text{ATM only}}$) leads to a warm bias nearly across the all oceans, but less cold bias over Northern and Southern Africa as well as China [Fig. III.1e]. Combining atmosphere and ocean reconstruction (indirect) reduces the overall temperature bias, especially over the oceans [Fig. III.1f].

While the above explained biases are liabilities of reconstructions, the linear association measured by the Anomaly Correlation Coefficient (ACC) benefits from reconstruction. Reconstruction recreates climate variability of the target [Fig. III.1g-l]. The running 10-year correlation between the target and the reconstruction in atmospheric variables is in most grid cells above 0.4 and significantly better than the randomly resampling threshold. Reconstruction over the oceans is more successful in the tropics than in the extra-tropics, where the Northern and Southern Hemisphere mid-latitude westerlies have low, but still significant correlation. Generally, the atmosphere above the ocean is better reconstructed than above land, showing the stabilizing effect of an internally consistent ocean reconstruction on the atmo-

sphere [Fig. III.1g-l]. The Southern Hemisphere tropical convergence of winds is well reconstructed, but the meridional winds in central Canada and tropical Africa are not significantly reconstructed [Fig. III.1g]. Also zonal winds across North America, Southern Africa and Siberia have low correlation with the target, but the tropical zonal winds are very well reconstructed [Fig. III.1h]. Precipitation from the central Atlantic over central Africa is worse reconstructed than the resampling threshold, and the extratropical westerlies have low correlation with the target [Fig. III.1i]. Temperature is well reconstructed in the tropical oceans [Fig. III.1j-l]. Reconstructing both atmosphere and ocean (indirect) improves 2m temperature correlation better than only reconstructing a single realm. The indirect carbon cycle reconstruction is significantly better than the resampling threshold except in central Africa, where the ITCZ shift changes the climate regime [Fig. III.1l].

This physical bias due to reconstruction, especially in the tropics, can be explained by the sensitivity of atmosphere-ocean coupling to perturbation induced by nudging [Milinski et al., 2016]. The reconstruction of ocean and atmospheric variables is perfectly aligned with the model climatology into that same model. Hence, the reconstruction error does not arise from inconsistent observations, but from the perturbed interaction of atmospheric and oceanic dynamics. While reconstructing an increasing set of variables shows that nudging can be an efficient way to reconstruct variability [Jeuken et al., 1996], this reconstruction is biasing the climate state in the tropics at the same time [also explained in Zhu and Kumar, 2018].

Nudging atmospheric and ocean dynamics including sea-ice all at once (indirect reconstruction), as is often done in state-of-the-art carbon cycle prediction systems, brings large-scale improvements over random resampling and atmosphere-only ($\text{indirect}_{\text{ATM only}}$) reconstruction, but strong regional biases remain [Fig. III.1].

III.3.2 *Reconstruction of the Oceanic Carbon Cycle*

III.3.2.1 *Indirect Reconstruction of the Oceanic Carbon Cycle*

How do these regional physical biases affect the reconstruction of oceanic carbon cycle? In order to assess the tracking performance in the indirect reconstruction of the oceanic carbon cycle, we focus on air-sea CO_2 flux and surface oceanic pCO_2 as the state variable of the ocean carbon sink, which is the oceanic driver of air-sea CO_2 flux [Lovenduski et al., 2019a].

Reconstructing only the atmospheric dynamics ($\text{indirect}_{\text{ATM only}}$) leads to strong positive biases across large parts of the global ocean, which can be reduced by also reconstructing oceanic temperature and salinity (indirect) [Fig. III.2a,b,d,e]. The weakening of the Southern hemisphere westerly winds decreases the magnitude of air-sea CO_2

flux, but more importantly reduces the Southern hemisphere overturning circulation and upwelling of carbon-rich waters, which leads to increased Southern Ocean carbon uptake [Fig. III.2b,e]. The intensification of easterly trade winds [Fig. III.1b] strengthens upwelling and therefore higher pCO_2 in the tropical Atlantic [Fig. III.2b] [Lefèvre et al., 2013]. The bias pattern of air-sea CO_2 flux is dominated by the bias of pCO_2 [Lovenduski et al., 2019a] [Fig. III.2b,e].

The variations in the oceanic carbon cycle, described by the correlation coefficient, are better reconstructed than the resampling threshold. Indirect reconstruction of oceanic and atmospheric dynamics greatly improves tracking performance over atmosphere-only $\text{indirect}_{\text{ATM only}}$ reconstruction. The additional reconstruction of the physical ocean [Fig. III.1e,f] enables largely a correlation above 0.7 [Fig. III.3b,e]. Only the carbon cycle in the tropical oceans remain difficult to reconstruct due to the strong biases in atmospheric circulation [Fig. III.1a,b,c]. Note that the land and atmospheric carbon bias due to indirect reconstruction are discussed in subfigures III.2g-o in sections III.3.3.1 and III.3.4.1).

III.3.2.2 *Direct Reconstruction of the Oceanic Carbon Cycle*

Next, we compare the previously shown indirect carbon cycle reconstruction with direct carbon cycle reconstruction by nudging dissolved inorganic carbon (DIC) and alkalinity (ALK) towards the target.

While direct oceanic carbon cycle reconstruction reduces the magnitudes of the bias across the ocean, biases are still evident [Fig. III.2c,f]. These biases are caused by the physical biases, which the dynamical oceanic carbon cycle model is sensitive to. Hence, the biased ocean physics inhibits additional improvements in tracking performance from direct ocean carbon reconstruction.

Direct oceanic carbon cycle reconstruction improves the already high correlations across the oceans [Fig. III.3c,f]. The resampling threshold is surpassed nearly everywhere. Only coastal areas, especially in the Eastern tropical Atlantic with strong wind and precipitation biases, have a correlation below 0.7.

Section III.3.2 shows how well indirect and direct reconstruction of the ocean carbon cycle work overall. While the direct reconstruction has slightly larger biases in air-sea CO_2 flux, direct reconstruction also brings higher correlation. Note that the land and atmospheric carbon bias due to direct reconstruction are discussed in subfigures III.2g-o in sections III.3.3.2 and III.3.4.2).

III.3.3 *Reconstruction of the Land Carbon Cycle*

III.3.3.1 *Indirect Reconstruction of Land Carbon Cycle*

How do these regional physical biases affect the reconstruction of the land carbon cycle? In order to assess the tracking performance in the best indirect reconstruction of the land carbon cycle, we focus on the state variable cVeg, which represents carbon storage in vegetation (leaves, stems, roots) and drives air-land CO_2 flux and hence the land carbon sink.

For the land carbon cycle, the reconstruction of the ocean temperature and salinity did not matter, when atmospheric temperature was also reconstructed [Figs. III.2, III.3]. Indirect reconstruction leads to biases compared to the target in carbon storage, and in particular cVeg [Fig. III.2g,h], as the land carbon cycle is very sensitive to changes in atmospheric circulation, which are strongest in the tropics due to the ITCZ shift. In the Amazon and Southern Africa, the air-land CO_2 bias increases, most likely caused by the strong positive precipitation bias in these regions [Fig. III.1c; III.2j,k]. Conversely, the carbon sink in Southeast Asia and central Africa has a carbon release bias due to less precipitation and a warm bias [Fig. III.2j,k].

The reconstruction correlations in the land carbon cycle are much lower than for the oceanic carbon cycle. cVeg is well reconstructed in the extratropics, but the biases in the tropics result in correlations with the target lower than the resampling threshold [Fig. III.3g,h]. Air-land CO_2 shows the same patterns with lower correlations, which are below the resampling threshold in the tropics [Fig. III.3j,k].

III.3.3.2 *Direct Reconstruction of the Land Carbon Cycle*

Direct reconstruction of the land carbon cycle, which is here performed by resetting all restart files of the land carbon sub-model to the target every Jan 1^{st}, greatly enhances tracking performance of cVeg by simulation design. A sensitivity analysis for less frequent resetting can be found in the supplementary information [section III.SI.3].

This direct resetting reconstructs cVeg much better than the resampling threshold in the extra-tropics. However, the physical climate biases during the course of a year even introduce cVeg biases stronger than the resampling threshold in the tropics [Fig. III.2i]. Also, the biases in the air-land CO_2 flux are not improved [Fig. III.2l], which indicates that this hard reset of restart files introduces a shock to the dynamical land model.

On the other hand, correlations in cVeg and air-land CO_2 flux increased to above 0.5 everywhere expect in the tropics, where the ITCZ shift changes the climate regime [Fig. III.3i,l].

Section III.3.3 shows the direct land carbon cycle reconstruction yields stronger correlation improvements than ocean direct carbon cycle reconstruction, because the indirect reconstruction of the ocean was already quite good. Direct reconstruction reduces biases in land carbon cycle state variables, but the resulting air-land CO_2 flux biases becomes worse.

III.3.4 *Reconstruction of the Global Carbon Cycle and Atmospheric CO_2*

Tracking performance for prognostic atmospheric CO_2 integrates the air-sea and air-land CO_2 fluxes over time [Spring and Ilyina, 2020; Spring et al., 2020]. As atmospheric CO_2 mixes fast across the globe, we first examine globally aggregated quantities driving globally averaged atmospheric CO_2 [Fig. III.4].

III.3.4.1 *Indirect Reconstruction of the Oceanic and Land Carbon Cycle*

We first examine the indirect reconstruction represented by the green error bars in figure III.5.

The indirect reconstruction has a negative bias in global pCO_2 in the annual mean [Fig. III.4a], which is negative in boreal winter and positive in boreal summer, indicating that the reconstruction intensifies the seasonal cycle [Fig. III.SI.5]. This bias is mostly larger than the resampling mean absolute error threshold, which resembles the temporal standard deviation [Fig. III.5a]. The global oceanic CO_2 flux is low biased but within the resampling threshold [Figs. III.4b, III.5d].

On the other hand, the variations of the global oceanic carbon cycle measured by ACC are well reconstructed surpassing the resampling threshold [Fig. III.5b,e].

The accuracy or distance measured by root mean squared error (RMSE) has strong seasonal errors especially in boreal winter up to 1.3 ppm. When biases are persistent, they can be reduced by a bias reduction procedure, which is often done when applying climate model output to a real-world application. After applying a simple mean bias reduction, RMSE is well below the resampling threshold at below 0.5 ppm [Fig. III.5c,f].

The indirect reconstruction also leads to biases in the land carbon cycle [Fig. III.4c,d]. Vegetation carbon pools have a positive bias much larger than the resampling threshold [Fig. III.4c]. The bias of global air-land CO_2 flux only surpasses the resampling threshold in August, September and November, reducing the global seasonal cycle by 25% [Figs. III.4d, III.5j].

Global cVeg has a 0.5 correlation with the target, which is lower than the resampling 0.7 correlation. The 10-year running correlation also

with a large range of ±0.3. Global air-land CO_2 is well reconstructed surpassing the resampling threshold [Fig. III.5h,k].

Without bias reduction, accuracy measured by RMSE is worse than the resampling cVeg threshold. After bias reduction, cVeg accuracy is still slightly worse than the threshold, but accuracy improved from 5 PgC to 1 PgC. Global air-land CO_2 flux accuracy is below the threshold until June and up to 0.3 PgC/month. But again when applying the bias mean reduction, tracking accuracy is always below the resampling threshold at 0.1 PgC/month [Fig. III.5i,l].

Global atmospheric CO_2 has larger variations in reconstruction skill, depending on which 10-year chunk is used to calculate the metric. And the skill has a nearly constant level throughout the year. The mean bias is close to zero [Figs. III.4e, III.5m]. Correlation with the target is above 0.7 and slightly above the resampling threshold [Fig. III.5n]. Accuracy is at 0.7 ppm slightly above the threshold, but below the threshold at 0.5 ppm after mean bias reduction [Fig. III.5o].

Understanding the tracking performance of the ocean and land carbon cycle, we can now evaluate the spatial distribution of globally averaged atmospheric CO_2. Reconstructing only the atmosphere warmed the globe and also increased atmospheric CO_2 globally [Figs. III.1k, III.2m]. Reconstructing additionally also the ocean keeps the temperature stable, but introduces a less than 1 ppm low bias across the Southern Hemisphere, reflecting the higher uptake of the Southern Ocean carbon sink and the Southern Hemisphere land carbon sink [Fig. III.2e,k,n]. The variations in atmospheric CO_2 are well reconstructed with correlation coefficients above 0.6 in the Southern Hemisphere, but across the Northern extra-tropics and the land regions with strong physics biases correlation is at 0.5 below the resampling threshold [Fig. III.2m,n].

III.3.4.2 *Direct Reconstruction of the Oceanic and Land Carbon Cycle*

Now, we assess the potential improvements in the global carbon cycle due to direct reconstruction of the global carbon cycle variables shown in orange in figure III.5.

The global ocean carbon cycle improves after direct DIC and alkalinity reconstruction. Monthly biases remain but are now within the resampling threshold [Fig. III.5a]. Correlation improves from 0.8 to above 0.9 in surface pCO_2. Air-sea CO_2 does not improve, but only because of correlations above 0.9 for the indirect reconstruction were already very high. Correlation for boreal winter is above 0.95, indicating that initial conditions in winter are well reconstructable to initialize forecasts with for the oceanic carbon sink [Fig. III.5b]. Direct reconstruction improves pCO_2 accuracy to 0.3 ppm [Fig. III.5c].

All results for the direct reconstruction of the land carbon cycle must be understood in the context of the method chosen for the direct

reconstruction: Because we reset the restart files in Jan 1st to the target, the metrics are near to perfect in January by design. However, then the biogeochemistry is not modified directly for twelve months and only follows the physical climate reconstruction indirectly, so biases triggered by physical biases unaligned with the reset land biogeochemistry pools quickly build up and may approach the metric of the indirect reconstruction. Likewise, there is no bias in global cVeg in January by design. The bias increases with the physical biases, until surpassing the resampling threshold in August increasing until the end of the year [Fig. III.5g]. Global air-land CO_2 flux has even stronger bias than the indirect reconstruction [Fig. III.5j]. Correlation in the global cVeg is near perfect in January by design and slowly decreases to 0.8 in December while still better than the resampling threshold [Fig. III.5h]. Global air-land CO_2 flux maintains a 0.1 higher correlation than indirect reconstruction [Fig. III.5k]. Direct reconstruction improves global cVeg accuracy. Accuracy is better than the resampling threshold after mean bias reduction. Direct reconstruction has worse CO_2 flux accuracy than the indirect, but after mean bias reduction the accuracy is slightly better [Fig. III.5i,l].

The global CO_2 bias in the direct reconstruction increase to +1.8 ppm [Fig. III.5m], but correlation increases from 0.7 to 0.9 [Fig. III.5n]. The direct reconstruction has worse accuracy than the indirect due to the new bias, but after mean bias reduction the accuracy is slightly better at 0.4 ppm [Fig. III.5o].

How does direct carbon cycle reconstruction affect tracking performance in prognostic atmospheric CO_2? Already the time series indicate, that there is a 1-2 ppm atmospheric CO_2 positive bias in the direct reconstruction [Fig. III.4e]. This bias is very homogeneous over the oceans [Fig. III.2o]. However, correlation strongly increased to 0.9 above the oceans and above 0.7 on land except for central Africa with its persistent biases, where the reconstruction is not better than the resampling threshold.

Section III.3.4 shows that atmospheric CO_2 follows the reconstructed land and ocean carbon cycle integrating their respective fluxes over time. The direct carbon cycle reconstruction introduces a large bias in the atmospheric CO_2 distribution that the indirect reconstruction did not suffer from, even after mean bias reduction [Fig. III.SI.3]. Globally averaged atmospheric CO_2 after direct reconstruction had a better accuracy tracking performance after the mean bias reduction, showing how global aggregation can balance regional biases. The direct land and ocean carbon cycle reconstructions track target much better than the indirect reconstruction, when measured by correlation.

Hence, in large, this first part showed how direct carbon cycle reconstruction improves linear association between reconstruction

and target (measured by ACC), but often increases biases degrading accuracy (measured by RMSE). Only after bias reduction, accuracy improves with respect to the indirect carbon cycle reconstruction.

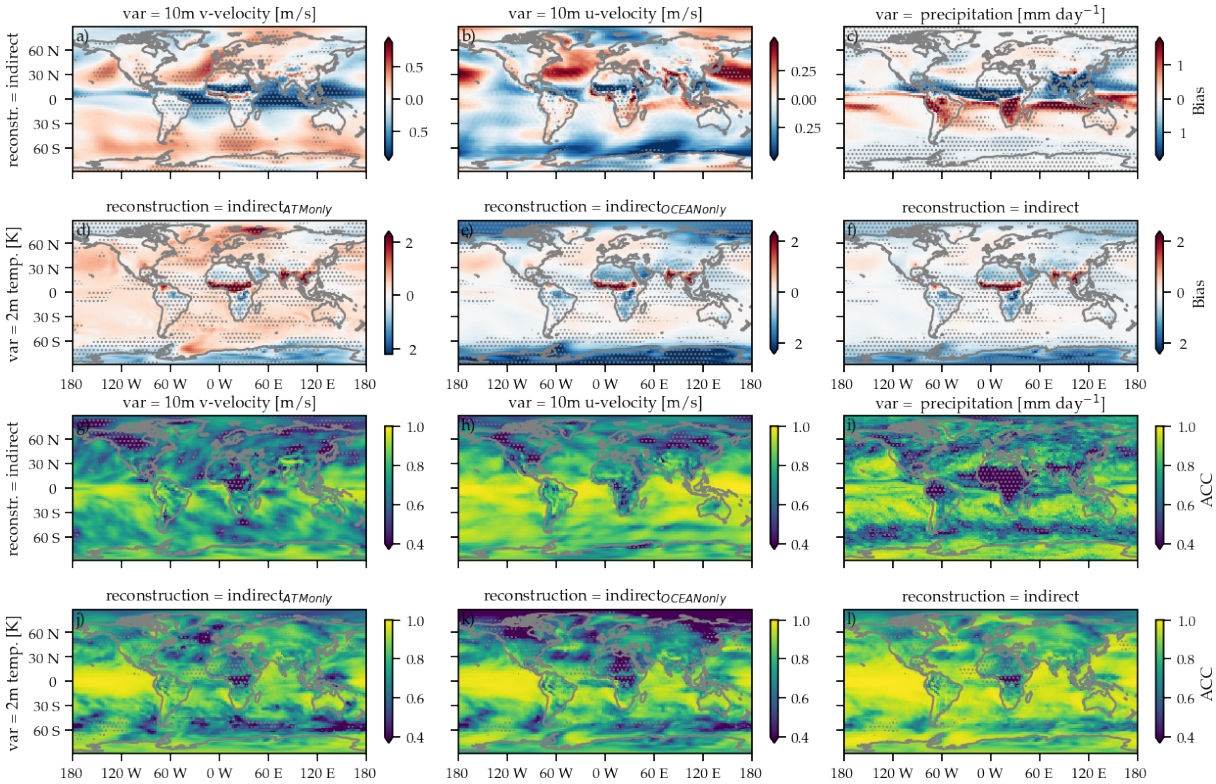


Figure III.1: Spatial distribution of the bias (construction - target) (a-f) and anomaly correlation coefficient (ACC) (g-l) of different indirect carbon cycle reconstructions relative to the target over 10-year running windows of annual means [see III.SI.1]. The reconstruction metrics for 2m temperature are shown for the indirect$_{\text{ATM only}}$ (d,j), indirect$_{\text{OCEAN only}}$ (e,k) and indirect reconstruction (f,l). Because of identical reconstruction skill for all indirect methods, only one indirect reconstruction is shown for other variables, zonal westward 10m wind (a,g), and meridional northward 10m wind (b,h), and precipitation (c,i). Gray stippling shows where the metric exceeds the 5$^{\text{th}}$ (for a-f) or 95$^{\text{th}}$ (for g-l) percentile threshold from random target block resampling, i.e. the reconstruction is not significantly better than internal variability.

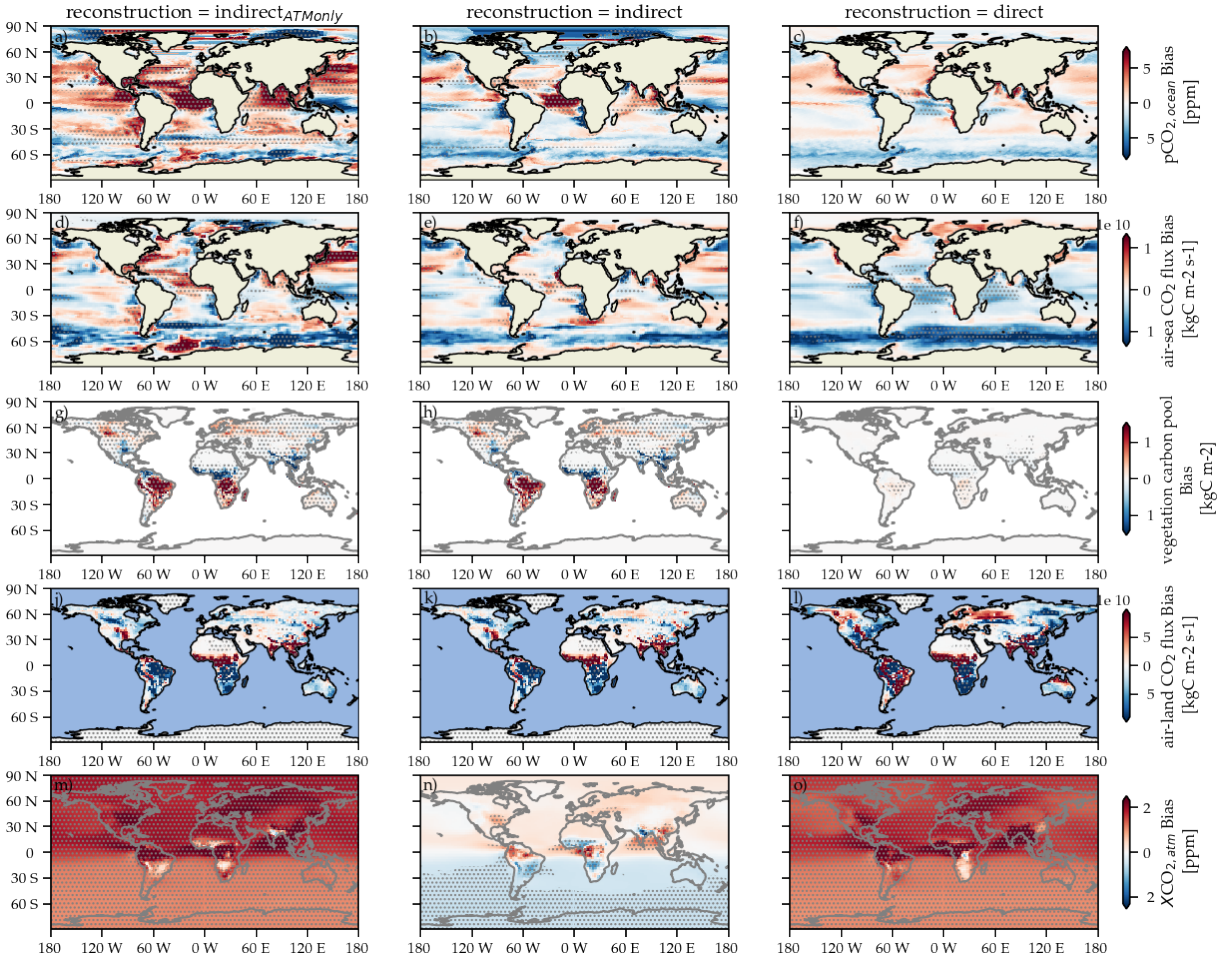


Figure III.2: Spatial distribution of the bias between the target and different indirect carbon cycle reconstruction methods over 10-year running windows of annual means [see III.SI.1]. Columns show the different carbon cycle reconstruction methods [see Table2]. Rows show the different variables: the ocean carbon cycle is represented by (a-c) the partial pressure of surface CO_2 in the ocean (pCO_2) and (d-f) surface air-sea CO_2 flux (negative values indicate carbon uptake by the ocean); the land carbon cycle is represented by (g-i) the vegetation carbon pools and (j-l) air-land surface CO_2 flux (negative values indicate carbon uptake by land); and the atmospheric carbon is represented by (m-o) the atmospheric CO_2 mixing ratio (XCO_2). Gray stippling shows where the bias exceeds the 5^{th} percentile mean absolute error threshold from random target block resampling, i.e. the reconstruction is not significantly better than internal variability.

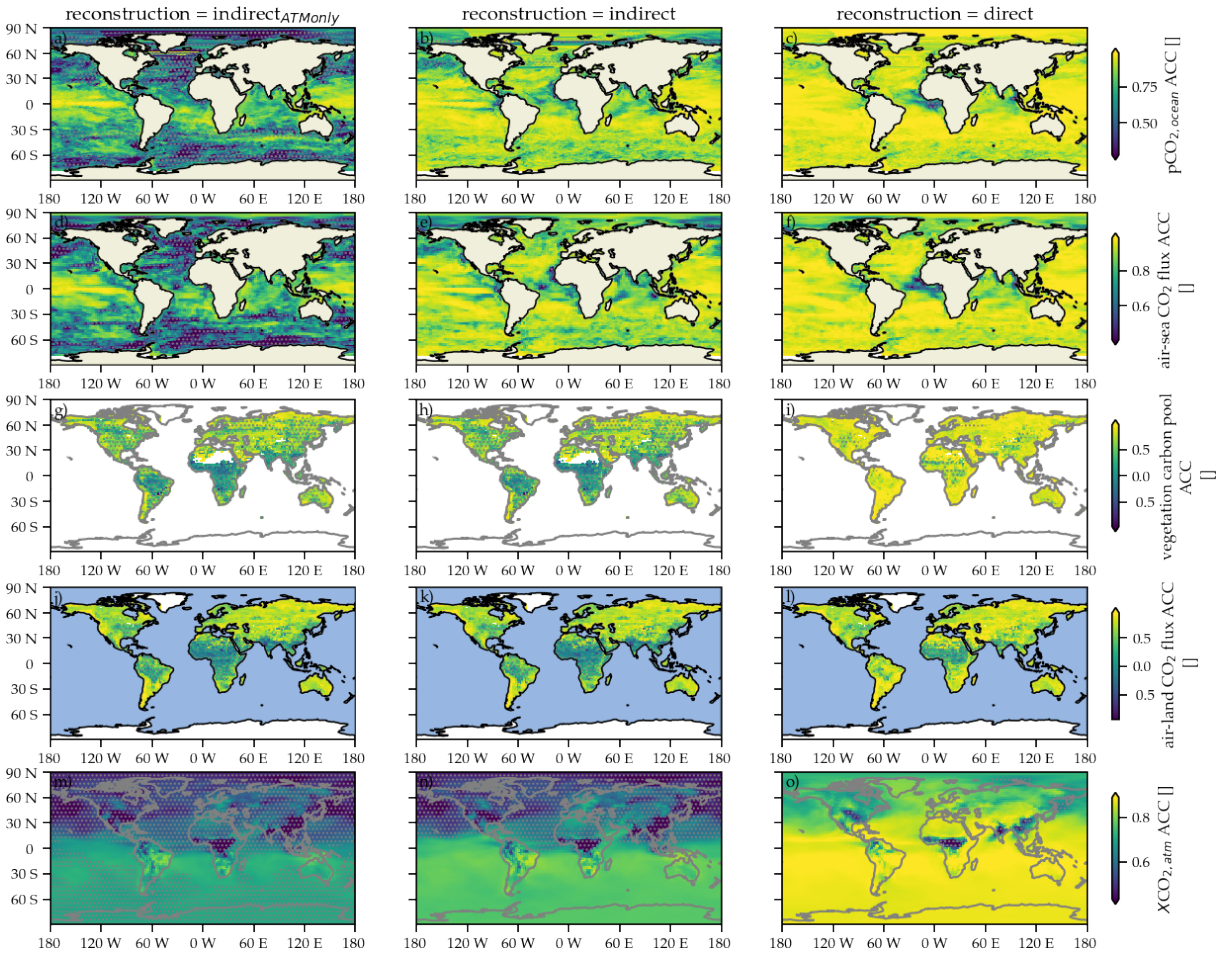


Figure III.3: As Fig. III.2 but for the anomaly correlation coefficient (ACC). Gray stippling shows where the ACC is lower than the 95th percentile ACC threshold from random target block resampling, i.e. the reconstruction is not significantly better than a resampling internal variability threshold.

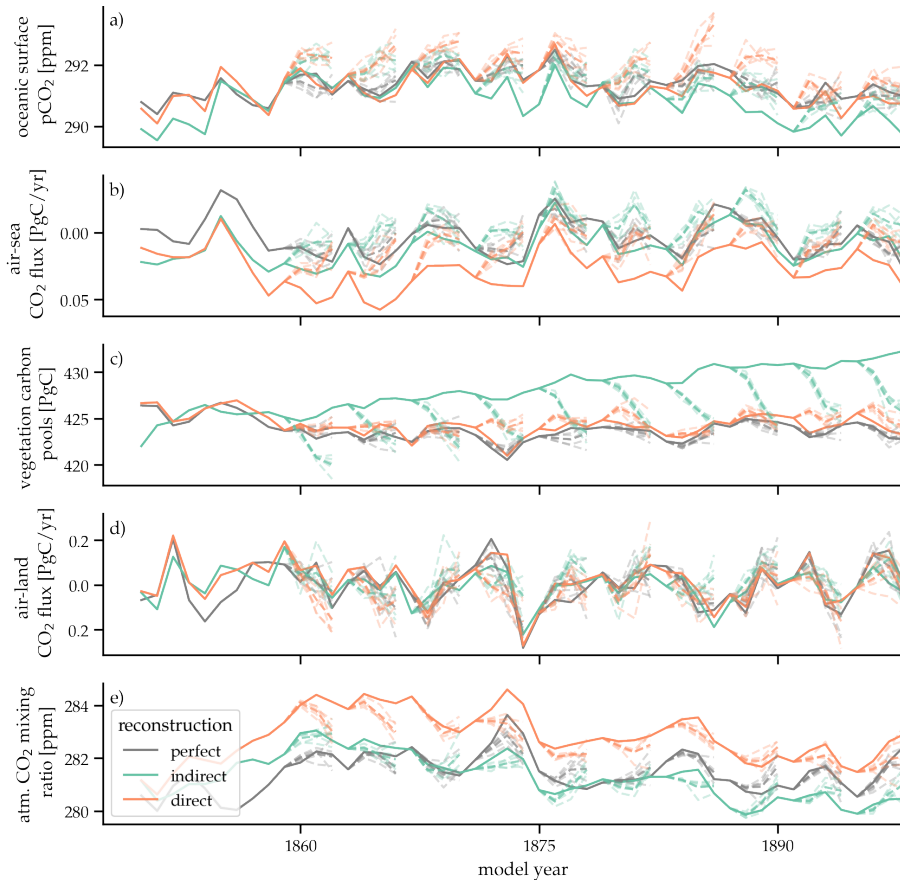


Figure III.4: Evolution in global annual mean of (a) surface ocean pCO_2 (b), air-sea surface CO_2 flux (negative values indicate carbon uptake by the ocean) (c), vegetation carbon pools (g-i), air-land surface CO_2 flux (negative values indicate carbon uptake by land) (d) and atmospheric CO_2 mixing ratio (e). The target (gray) is quite well tracked by the indirect (green) and direct (orange) carbon cycle reconstruction. The solid line shows the different reconstruction simulations, the dashed lines show the initialized ensembles started from the different reconstructions.

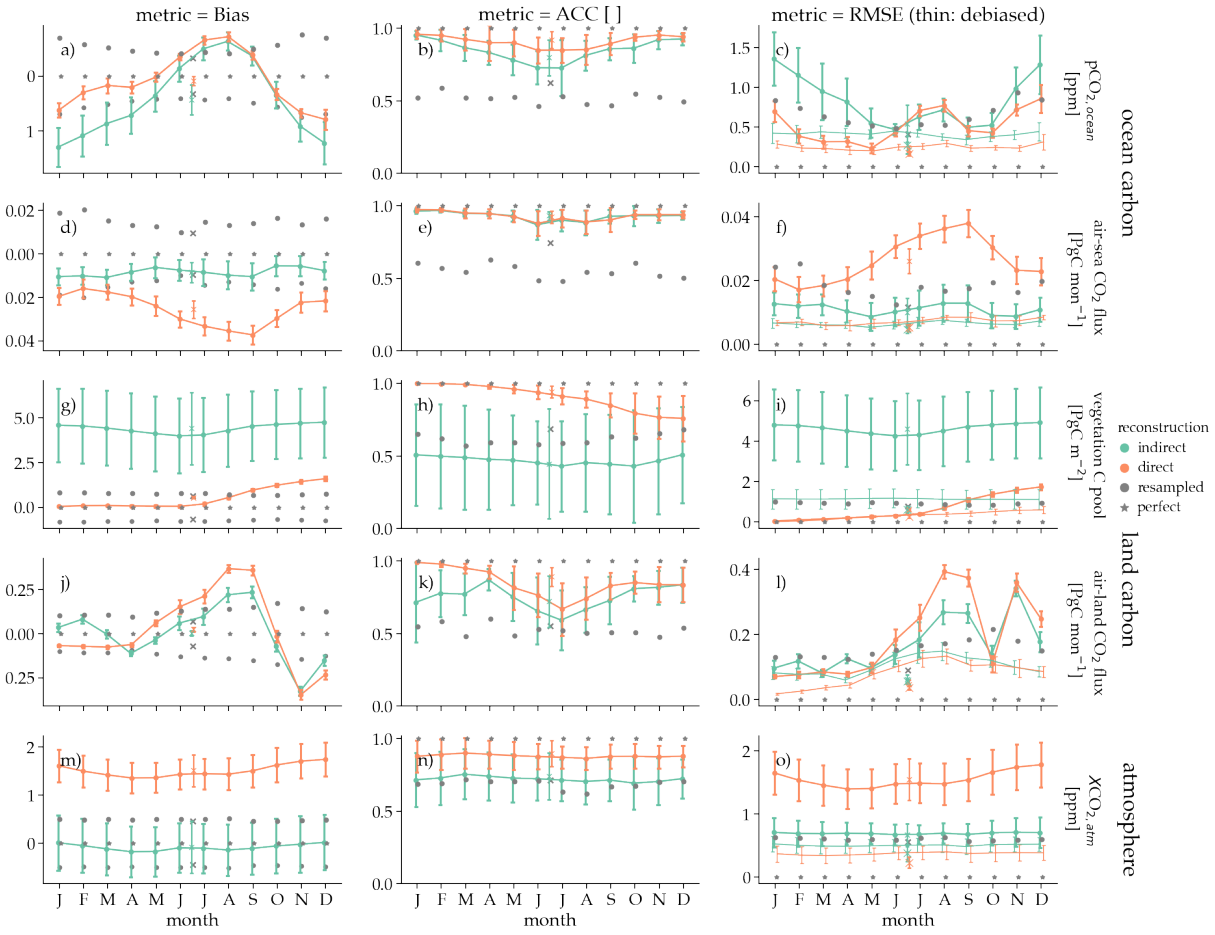


Figure III.5: 10-year running mean reconstruction skill per month in bias (left), anomaly correlation coefficient (ACC, middle) and root-mean-square-error (RMSE, right) for global aggregation of carbon cycle variables: (a-c) surface oceanic partial pressure of CO_2, (d-f) air-sea CO_2 flux (negative values indicate carbon uptake by the ocean), (g-i) vegetation carbon pools, (j-l) air-land CO_2 flux (negative values indicate carbon uptake by land) and (m-o) mixing ratio of atmospheric CO_2. Whiskers show the 5^{th} and 5^{th} percentile of the running skill over time. Colors show different reconstruction methods: indirect (green) and direct (orange). Gray stars indicate perfect skill. Gray dots mark 95^{th} percentile for ACC and 5^{th} percentile for the remaining distance-based metrics of random reconstruction skill block-bootstrapped from the target control simulation as an unskillful reference skill. Crosses show reconstruction skill of annual mean timeseries. Thin lines show monthly RMSE skill after a mean bias reduction.

III.4 IMPACT OF RECONSTRUCTION ON GLOBAL CARBON CYCLE PREDICTIVE SKILL

The second part of the paper assesses how predictive skill improves due to direct initialization of global carbon cycle variables. Specifically, we verify the RMSE between the five ensemble members initialized from the indirect and direct reconstructions across all initializations based on raw and lead-time dependent bias corrected timeseries [Figs. III.4, III.6].

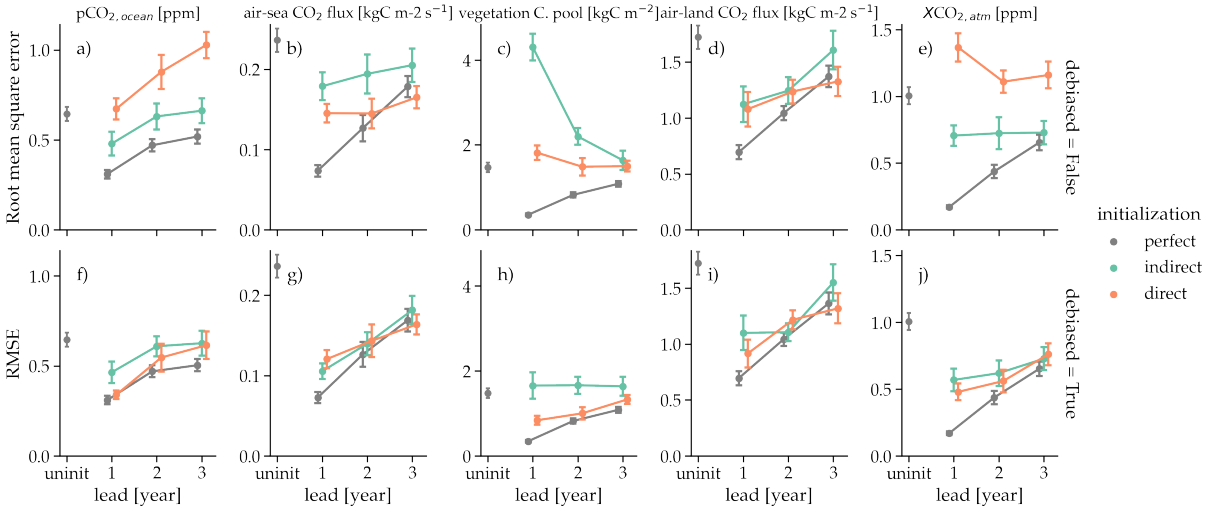


Figure III.6: Predictive skill measured by root-mean-square-error (RMSE) between the initialized ensemble mean and the target as a function of lead year for different initialization setups: perfect indicating no reconstruction and hence perfect initial conditions to predict the target (gray), indirect (green) and direct (orange). Columns show global variables: for the ocean carbon cycle (a) oceanic surface pCO_2, (b) air-sea CO_2 flux; for the land carbon cycle (c) total land carbon pools, (d) air-land CO_2 flux and in the atmosphere (e) atmospheric CO_2 mixing ratio. (f-j) show RMSE-based predictive skill as (a-e) after mean bias reduction. Initialized ensembles are resampled with replacement (N=500) along the initialization dimension to account for initialization sampling uncertainty [see Spring and Ilyina, 2020], where errorbars show the resampled initialization skill uncertainty ($\pm 1\sigma$). Uninitialized ensembles, shown at lead 0, are resampled from the target control simulation and show the reference skill without initialization.

III.4.1 *Oceanic Carbon Cycle*

The RMSE between the initialized ensembles and the target simulations in annual globally-averaged pCO_2 continuously increases from lead year one to lead year three as expected. While perfectly and indirectly initialized ensembles stay below the resampling uninitialized

threshold for the first two lead years indicating that global pCO_2 is predictable due to initialization [Fig. III.6a], the direct initialization has a larger error due to the offsets in global atmospheric CO_2, which pCO_2 tries to equilibrate to [Fig. III.4e]. Therefore this persistent bias causes lead year three to be not predictable. A simple mean bias reduction resolves this issue making all three lead years predictable. Direct initialization only beats indirect initialization for lead year one with RMSE of 0.35±0.05 ppm versus 0.45±0.05 ppm [Fig. III.6f].

Global air-sea CO_2 flux is predictable for three years in all initialization methods, which is one year longer than in Spring and Ilyina, 2020, possibly because here we use more and more equally distributed initialization dates. Direct initialization is advantageous over the indirect initialization, because the initial lead offset is smaller (0.14±0.01 PgC/year vs 0.18±0.02 PgC/year) [Fig. III.6b]. The simple mean bias reduction improves skill of the non-perfect initializations to identical magnitudes [Fig. III.6g].

III.4.2 *Land Carbon Cycle*

Indirect initialization makes cVeg not predictable. The physical reconstruction biases drive larger errors in lead year one than in later lead years, also to a lesser extent for the direct reconstruction where some biases are corrected. But both reconstructed initialized ensembles show decreasing distances towards the target, whereas increasing distances are expected for vanishing predictive skill as in the perfectly initialized ensembles [Fig. III.6c]. Mean bias reduction eliminates the differences between direct and perfect reconstruction making both predictable unlike the indirect reconstruction [Fig. III.6h].

Global air-land CO_2 flux is predictable for three years, again one year longer than found in Spring and Ilyina, 2020. Both reconstructed initializations start with a higher error of 1.1±0.2 PgC/year in lead year one compared to perfect initialized 0.7±0.1 PgC/year [Fig. III.6d]. Mean bias reduction brings non-perfect initializations within the error bars of the perfect initialization after lead year one [Fig. III.6i].

III.4.3 *Atmospheric CO_2*

Perfect and indirect initialization atmospheric CO_2 predict the target for three years as found in Spring and Ilyina, 2020. While the perfect initialization error grows continuously from zero, the indirect initialization error stays nearly constant at 0.7±0.1 ppm, but the error stays below the direct initialization error, which suffers from the bias in the direct reconstruction simulation [Fig. III.6e]. Mean bias reduction improves RMSE, making direct initialization better but still within the margins of the indirect initialization. After lead year one, indirect and direct initializations are similar to perfect-initialization predictive skill

at 0.7 ppm [Fig. III.6j].

These initialized predictive skill results show that indirectly initialized ensembles predict the target quite reasonably. Direct initialization suffers strong shocks in some variables, when reconstruction is started and stopped, but these shocks can be partly reduced by a mean bias reduction. The improvements of direct reconstruction over indirect reconstruction in the global carbon cycle predictive skill after bias reduction are not significant [Fig. III.6f-j].

III.5 SUMMARY AND CONCLUSIONS

In this study, we assess how well the global carbon cycle is reconstructed in an ESM and how well a ground truth target simulation can be predicted by these initializations.

The main limitation of land carbon cycle reconstruction potential is the hard reset of restart files which is fundamentally different to the dynamical nudging applied for ocean and atmospheric physics. Our study represents a first attempt to quantify whether initial conditions reconstruction in land carbon cycle is indeed needed for addressing predictive skill of the global carbon sinks and atmospheric CO_2 concentration. For a real-world application, our direct land carbon reconstruction method should not be used. In practice satellite products of carbon cycle variables could be assimilated into the model periodically or at each time step. However, just strong interference with the model will likely result in strong drifts, especially in dependent variables. For useful real-world applications of land carbon cycle assimilation, sequential [Evensen, 1994; Balmaseda et al., 2007; Zhang et al., 2007] or variational data assimilation techniques [Han et al., 2004]. But still the problem of data availability for the reforecast period remains. Haney reconstruction is the simplest approach to data assimilation allowing little flexibility to the model. Many centers are now transitioning towards the ensemble Kalman filter data assimilation which allows more variability [Park et al., 2019; Brune and Baehr, 2020]. Applying such techniques to the carbon cycle may lead to better reconstructions. A final limitation of the method is that we use a model to reconstruct to itself. Therefore we do not have any structural uncertainty other than the reconstruction method itself and no processes missing in our framework. When reconstructing the real world, our model lacks processes and resolution contributing to structural uncertainty.

We find that reconstruction, which is an interference into the freely evolving model, leads to biases in physical climate. Because of its sensitivity to physical climate, the global carbon cycle is heavily biased itself by these physical biases. In ESMs, first the atmosphere, then the ocean and only then the carbon cycle is equilibrated and tuned for preindustrial control conditions. Once reconstruction slightly modifies the mean state in the physical climate, the sensitive carbon cycle deviates from the near-equilibrium state. A previous study reported biases after reconstruction [Zhu and Kumar, 2018]. Yet, to our knowledge, we present the first attempt at reconstructing in a perfect-model framework, where no biases due to climatology differences are expectable. Zhu and Kumar, 2018 also mention that reconstruction ability likely depends on the model and application area, hence there seems to be no out-of-the-box solution for all ESMs.

We furthermore find that the commonly used indirect reconstruction of carbon cycle, in which only climate physics are reconstructed and the carbon cycle follows indirectly, tracks the target reasonably well. A resampling threshold corresponding to internal variability is surpassed across large parts of the globe. Only the areas with strong physical and consequently carbon cycle biases miss that benchmark occasionally. For the ocean carbon cycle, the reconstruction of the physical ocean fields is critical to reconstruct carbon cycle initial conditions, which explain why current state-of-the-art carbon cycle prediction systems have skill despite not initializing the ocean carbon cycle with ocean carbon cycle observations [Séférian et al., 2014; Park et al., 2018; Li et al., 2019; Lovenduski et al., 2019a].

Direct reconstruction of ocean and land carbon cycle improves bias, association and accuracy on a grid cell level, but aggregated on the global scale, direct reconstruction does not improve over the indirect reconstruction significantly. Also after a mean bias reduction, which is a common post-processing technique applied to model output for real-word use, accuracy measured in RMSE after direct reconstruction is only slightly better, often still overlapping with indirect reconstruction. Because the advantage of direct reconstruction can similarly be achieved by a simple mean bias reduction, we label these direct reconstruction improvements *trivial* with respect to the indirect method on the global scale. More advanced data assimilation methods may yield better reconstruction skill for the carbon cycle [Han et al., 2004; Balmaseda et al., 2007; Zhang et al., 2007].

When the success of atmospheric CO_2 reconstruction is evaluated, caution is needed. Reconstruction of the ocean and land carbon sink can easily introduce offsets from the target, because reconstruction violates conservation of mass by creating or erasing carbon. This can easily lead to offsets in the sinks which quickly accumulate in atmospheric CO_2. If CO_2 reconstruction is the focus, i.e. in reconstructing the transient climate from CO_2 emission, and offsets appear, adjustments of atmospheric CO_2 might be needed to correct for these offsets. However, we find that these offset biases are only of the order of 1-2 ppm in a perfect-model framework, which is small compared to the range of carbon feedbacks seen in atmospheric CO_2 in transient simulations. Hence, these offsets due to the restart files are not in our focus. Rather, equilibrated land and ocean carbon sinks with reconstructed climate determine realistic reconstructed atmospheric CO_2.

In the second part, we find that predictive skill after indirect initialization is similarly good as after direct initialization. This means that oceanic carbon cycle initial conditions are much less important that physical ocean initial conditions for oceanic carbon cycle predictions, which confirms the findings of [Fransner et al., 2020]. Reconstructed initialized predictive skill is close to perfectly initialized predictive skill after mean bias reduction, especially after lead year one.

Because the improved global predictive skill after direct reconstruction can similarly be achieved by a simple mean bias reduction and predictive skill after both reconstructions mostly overlaps, we label these direct reconstruction predictive skill improvements *trivial*, with respect to the indirect method on the global scale. This result is similar to Fransner et al. [2020], who find that ocean carbon cycle initial conditions matter much less than physical ocean initial conditions for annual carbon cycle predictions.

We conclude that the indirect carbon cycle reconstruction serves its purpose of reconstructing variation in the global carbon cycle. However, our study is designed and conducted in an idealized framework. When transferring our results into assimilation of real-world observations and its implications on predictability, structural uncertainties (model resolution in space and time) and missing ecosystem processes need to additionally be dealt with. Future studies, especially those aiming to address regional marine ecosystems, could consider a wider range of assimilation techniques and data breadth. Furthermore, more advanced data assimilation techniques [Evensen, 1994; Han et al., 2004; Balmaseda et al., 2007; Zhang et al., 2007] should be explored. Reducing the physical climate bias with its consequences for the carbon cycle holds more potential for improvements in initial conditions and predictive skill than direct carbon cycle initialization [Saito et al., 2011; Lee and Biasutti, 2014; Hua et al., 2019].

Nevertheless, our results add confidence to the current practice of indirect reconstruction in carbon cycle prediction systems [Ilyina et al., 2021].

ACKNOWLEDGEMENTS

Forecast verification was computed with the python package climpred https://github.com/bradyrx/climpred/ [Brady and Spring, 2021], which was co-developed with Riley X. Brady from University of Colorado, Boulder.

We acknowledge funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 821003 "Climate-Carbon Interactions in the Current Century (4C)", No 820989 "COMFORT" and No 641816 "CRESCENDO". Simulations were performed at the German Climate Computing Center (DKRZ).

We thank Jürgen Bader for internal review.

III.SI SUPPLEMENTARY INFORMATION

III.SI.1 *Metrics*

III.SI.1.1 *ACC*

The anomaly correlation coefficient (ACC) assesses the synchronous evolution over time of the forecast, here reconstruction $x(t)$ and the reference, here target $\hat{x}(t)$, [Jolliffe and Stephenson, 2011] and is defined as:

$$\begin{aligned} ACC(x(t), \hat{x}(t)) &= \frac{cov(x(t), \hat{x}(t))}{\sqrt{var(x(t)) \cdot var(\hat{x}(t))}} \\ &= \frac{\frac{1}{T}\sum_{t=1}^{T}(x(t) - \overline{x(t)})(\hat{x}(t) - \overline{\hat{x}(t)})}{\sqrt{\frac{\sum_{t=1}^{T}(x(t)-\overline{x(t)})^2}{T}} \cdot \sqrt{\frac{\sum_{t=1}^{T}(\hat{x}(t)-\overline{\hat{x}(t)})^2}{T}}}. \end{aligned} \tag{8}$$

III.SI.1.2 *RMSE*

In the initial conditions reconstruction part, the root-mean-square-error (RMSE) measures the second-order distance between forecast $x(t)$, here reconstruction $x(t)$ and the reference, here target $\hat{x}(t)$, [Jolliffe and Stephenson, 2011] and is defined as:

$$RMSE(x(t), \hat{x}(t)) = \sqrt{\frac{\sum_{T=1}^{T}(x(t) - \hat{x}(t))^2}{T}}. \tag{9}$$

As a predictability metric, the root-mean-square-error (RMSE) measures the second-order distance between forecast $x(t)$ and the target $\hat{x}(t)$ over lead time t [Jolliffe and Stephenson, 2011]. RMSE is calculated over all initialisations N and every member M is used as a forecast and verified against the target. RMSE is defined as:

$$RMSE(x(t), \hat{x}(t)) = \sqrt{\frac{\sum_{i,j=1}^{N,M}(x_{i,j}(t) - \hat{x}_j(t))^2}{NM}}. \tag{10}$$

III.SI.1.3 *Bias*

We set the target as the ground truth. Therefore any deviation from the reconstructions $x(t)$ to the target $\hat{x}(t)$ is seen as a bias, analogous to the bias between a model simulation (reconstruction) and observations (ground truth).

$$bias(t) = x(t) - \hat{x}(t) \tag{11}$$

III.SI.1.4 *Removing the Bias*

After removing the mean bias from reconstruction $\overline{x(t)}$ and target $\overline{\hat{x}(t)}$, the RMSE is also calculated as debiased RMSE.

$$RMSE_{\text{debiased}}(t) = RMSE\left((x(t) - \overline{x(t)}, \hat{x}(t) - \overline{\hat{x}(t)}\right) \tag{12}$$

III.SI.1.5 *Running Metric*

We calculate the mean tracking performance (mtp) over time for all metrics as a running mean over $s = 10$ years. This reflects that reconstructions are supposed to reconstruct the given climate states within months to a couple of years and the metric should not be prone to long-term trends that are not captured by the reconstruction. We ignore the first $c = 10$ years (out of $t_{max} = 48$ years) of reconstruction, where the model experiences an initial shock after adjusting to the new reconstructed climate [Kröger et al., 2017].

$$tpm(metric) = \frac{1}{t_{max} - s - c} \sum_{t=c}^{t_{max}-s} metric(x_{(t=t..t+s)}, \hat{x}_{(t=t..t+s)}) \quad (13)$$

III.SI.1.6 *Resampled Threshold*

To get an estimate of random tracking performance due to internal variability, i.e. how well one 10-year chunk tracks just another random 10-year chunk, we randomly resample 10-year chunks from the target simulation and apply the same tracking metrics. As a baseline skill from this random resampling in the figures, we take the 95% threshold for ACC and the 95% for the remaining distance-based metrics to ensure that the tracking performance from a reconstruction simulation is only worse compared to one out of 20 randomly resampled 10-year chunks.

III.SI.2 *Reconstruction RMSE Maps*

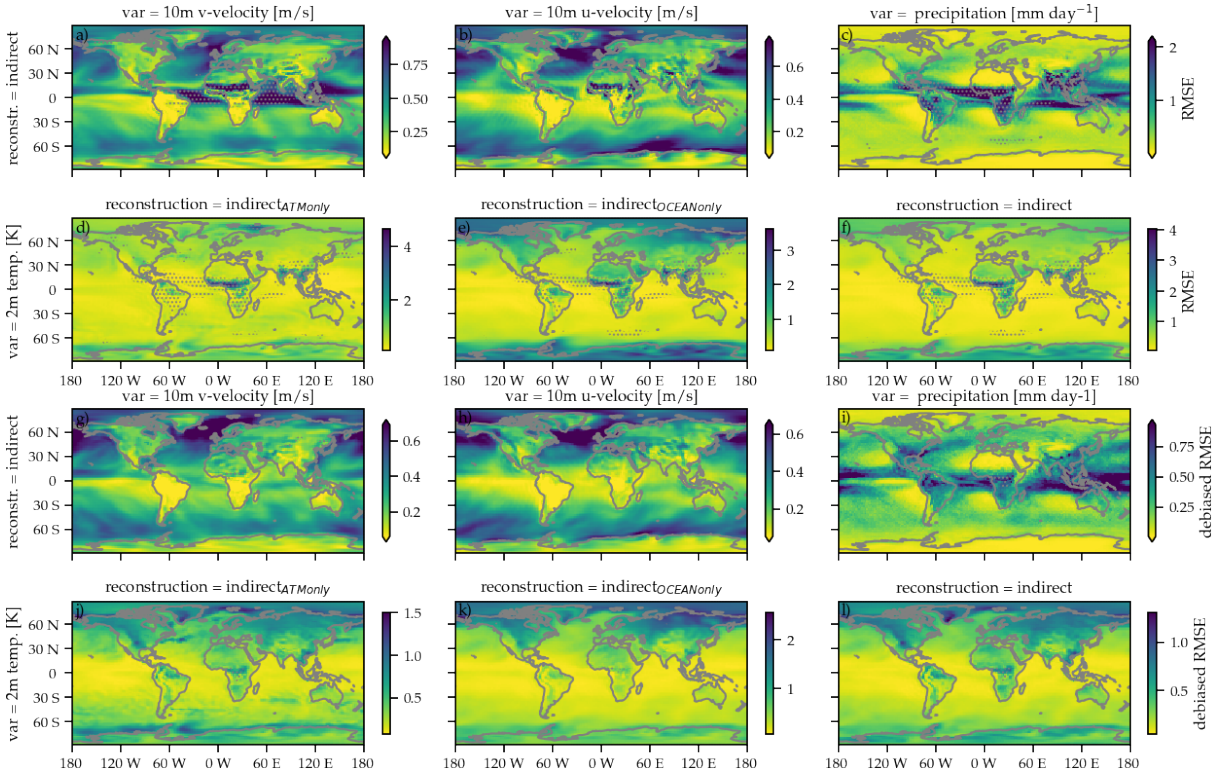


Figure III.SI.1: **As Fig. III.1 but for Root-mean-square-error (RMSE) (a-f) and for RMSE after bias reduction (g-l).**

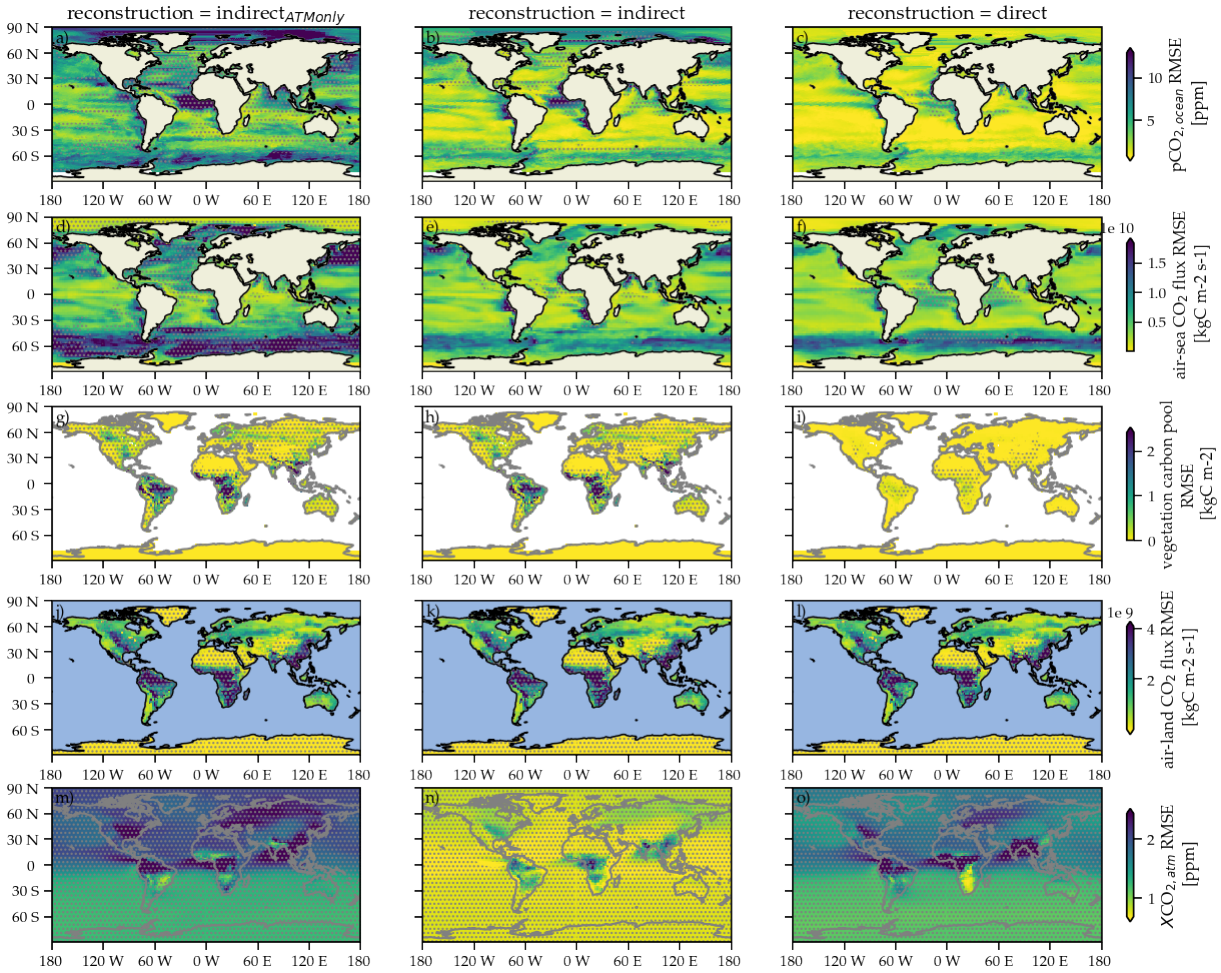


Figure III.SI.2: **As Fig. III.2 but for the Root-mean-square-error (RMSE).** Gray stippling shows where the RMSE is worse than the 5^{th} percentile RMSE threshold from random target block resampling, i. e. the reconstruction is not statistically significantly compared to internal variability.

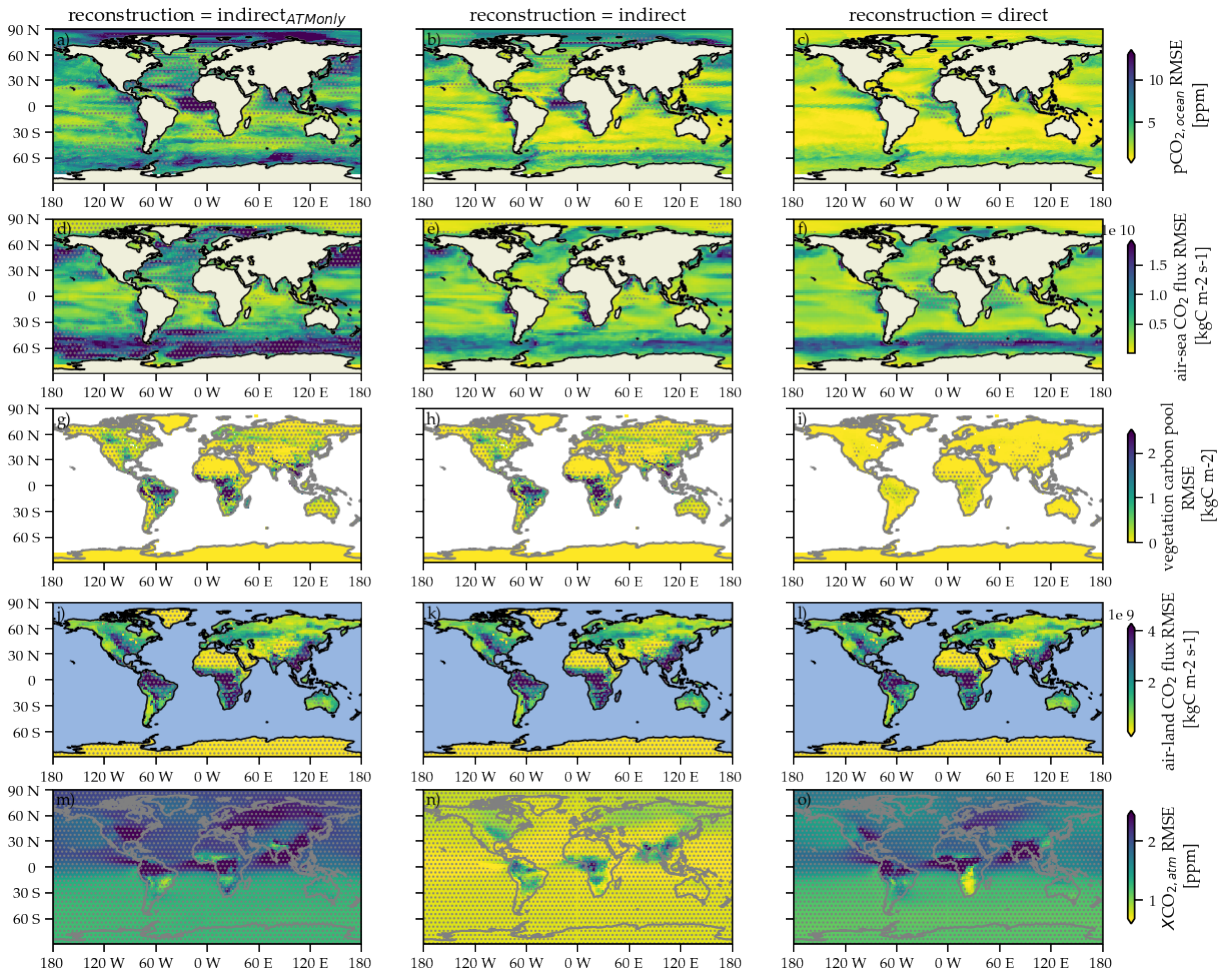


Figure III.SI.3: **As Fig. III.SI.2 but for Root-mean-square-error (RMSE) after bias reduction.**

III.SI.3 *Sensitivity Analysis for Different Reconstruction Timestep in ...*

III.SI.3.1 *... on Land Carbon Cycle*

We perform sensitivity reconstructions of the land restart file resetting to understand how sensitive this reconstruction method to the frequency of resetting. We performed additional simulations resetting the land model on Jan. 1st every second or every fifth year [orange triangles in fig. III.SI.4].

Global cVeg starts by definition with perfect skill in Jan after a reset. When resetting only every second year, the mean January tracking performance is already decreased, and decreases further. The negative correlations for five-year resetting shows the shock to the system if not immediately balanced by further resetting in the every (second) year case.

The global air-land CO_2 flux correlation degrades for less frequent resetting towards the indirect performance, but bias and accuracy improve.

Global atmospheric CO_2 aggregates these results and is also sensitive to biases developing in both sinks. Here, less frequent resetting of the land carbon cycle reduces the bias and therefore accuracy.

The tracking accuracy is of similar magnitude after mean bias reduction.

III.SI.3.2 *... on Ocean Carbon Cycle*

We perform the same kind of restart file resetting reconstruction to the ocean model [blue line in fig. III.SI.4]. The motivation here is to see whether a resetting of the ocean carbon cycle also yields perfect accuracy (RMSE) skill for January. But the ocean carbon cycle is sensitive to the physical climate and hence the direct ocean carbon cycle resetting accuracy degrades compared to the indirect tracking bias and accuracy, only correlation increases [Fig. III.SI.4a-f]. Contrary to resetting restart files in the land model, initial conditions accuracy measured by RMSE does not approach perfect skill of 0, because the physical climate did not experience this hard reset but is nudged dynamically.

In general, this hard reconstruction also seems to work for the ocean carbon cycle, because the tracking performances are not very different from the indirect method [Fig. III.SI.4a-f].

The tracking accuracy is of similar magnitude after mean bias reduction.

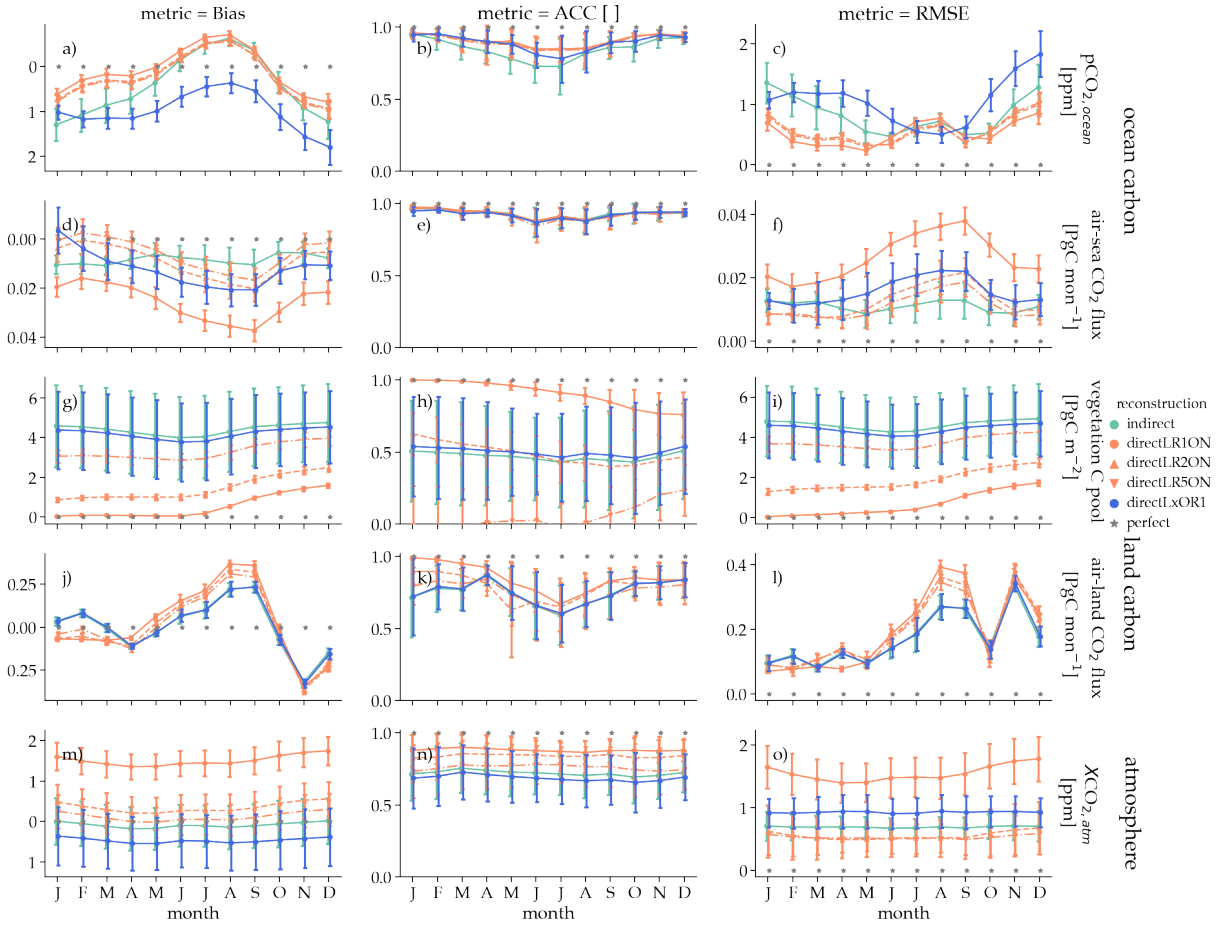


Figure III.SI.4: **As Fig. III.5 but for sensitivity simulations of the restart file resetting reconstruction.** In all simulations the physical climate is nudged as in indirect [Tab. 2]. directLR1ON describes land resetting every year and ocean nudging and is the indirect simulation. directLR2ON describes land resetting every second year and ocean nudging. directLR5ON describes land resetting every fifth year and ocean nudging. directLxOR1 describes no land reconstruction and ocean setting every year.

III.SI.4 *Seasonality*

In reference for figure III.5 to better understand reconstruction skill in context of target seasonality:

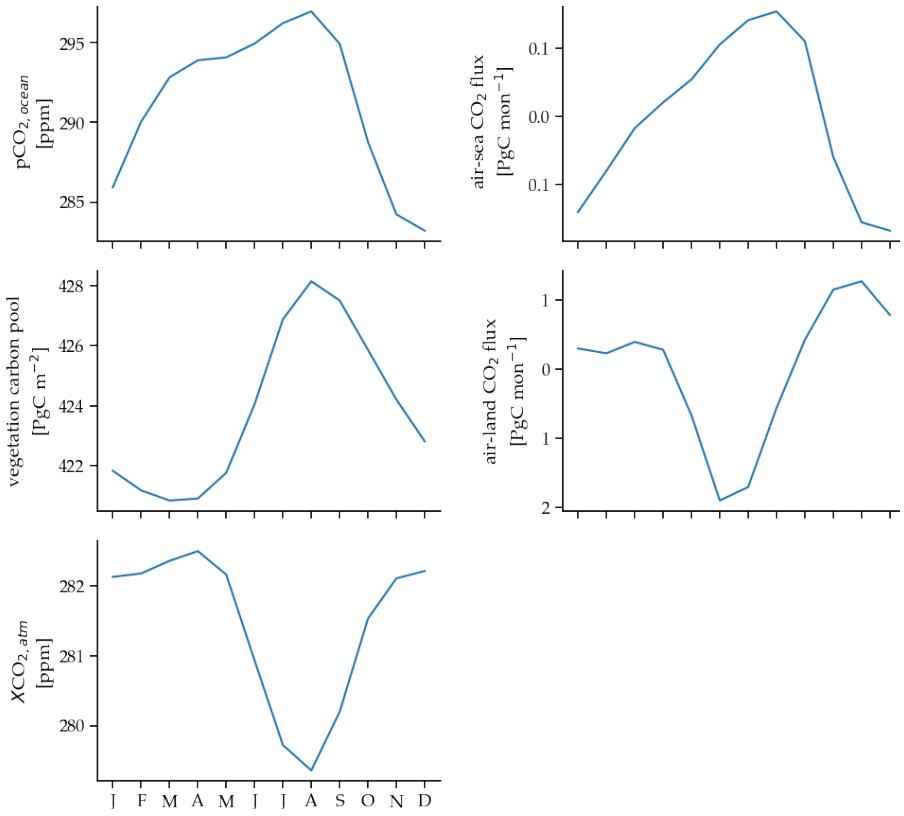


Figure III.SI.5: **Seasonality of the target simulation for global aggregated carbon cycle variables.**

III.SI.5 *Reproducibility*

Forecast verification was performed with the python package CLIMPRED [Brady and Spring, 2021] [https://github.com/pangeo-data/climpred], which was co-developed with Riley X. Brady from University of Colorado, Boulder. Scripts and data to reproduce this analysis are archived in http://hdl.handle.net/21.11116/0000-0007-A697-3.

CLIMPRED: VERIFICATION OF WEATHER AND CLIMATE FORECASTS

The attached article has been published in *Journal of Open Source Software*:

Brady, R. X., & Spring, A. [2021]. Climpred: Verification of weather and climate forecasts. *Journal of Open Source Software*, *6*[59], 2781. doi:10/gh9646

Author contributions: R.X.B. and A.S. jointly co-developed the software and drafted the manuscript.

CLIMPRED: verification of weather and climate forecasts

Riley X. Brady[1] **and Aaron Spring**[2,3]
[1] Boulder University, Boulder, Colorado, USA
[2] Max Planck Institute for Meteorology, Hamburg, Germany
[3] International Max Planck Research School of Earth System Modelling, IMPRS, Hamburg, Germany

Submitted: 16 October 2020, Published: 2 March 2021

IV.1 SUMMARY

Predicting extreme events and variations in weather and climate provides crucial information for economic, social, and environmental decision-making [Merryfield et al., 2020]. However, quantifying prediction skill for multi-dimensional geospatial model output is computationally expensive and a difficult coding challenge. The large datasets (order gigabytes to terabytes) require parallel and out-of-memory computing to be analyzed efficiently. Further, aligning the many forecast initializations with differing observational products is a straight-forward, but exhausting and error-prone exercise for researchers.

To simplify and standardize forecast verification across scales from hourly weather to decadal climate forecasts, we built CLIMPRED: a community-driven python package for computationally efficient and methodologically consistent verification of ensemble prediction models. The code base is maintained through open-source development. It leverages XARRAY [Hoyer and Hamman, 2017] to anticipate core prediction ensemble dimensions (ensemble member, initialization date and lead time) and DASK [Rocklin, 2015; Dask Development Team, 2016] to perform out-of-memory and parallelized computations on large datasets.

CLIMPRED aims to offer a comprehensive set of analysis tools for assessing the quality of dynamical forecasts relative to verification products (e.g., observations, reanalysis products, control simulations). The package includes a suite of deterministic and probabilistic verification metrics that are constantly expanded by the community and are generally organized in our companion package, XSKILLSCORE.

IV.2 STATEMENT OF NEED

While other climate verification packages exist (e.g., S2DVERIFICATION [Manubens et al., 2018] written in R and MURCSS [Illing et al., 2014] written with python-based CDO-bindings [Schulzweida, 2019]), CLIMPRED is unique for many reasons. (1) It spans broad temporal scales of prediction, supporting the weather, subseasonal-to-seasonal (S2S), and seasonal-to-decadal (S2D) communities. (2) CLIMPRED supports DASK [Dask Development Team, 2016; Rocklin, 2015] and thus works across all computational scales, from personal laptops to supercomputers (HPC). This leads to verification of a global 5° x 5° resolution climate prediction in a few seconds, compared to the 8 minutes required by MURCSS. This allows for a truly interactive analysis experience. However, note that benchmarking is inherently biased and MURCSS is valuable for their rigorous replication of decadal climate prediction metrics. (3) CLIMPRED is highly modular and supports the research process from end-to-end, from loading in model output, to interactive pre-processing and analysis, to visualization. (4) CLIMPRED is part of the wider scientific python community, PANGEO [Eynard-Bontemps et al., 2019]. A wide adoption of CLIMPRED could standardize prediction model evaluation and make verification reproducible [Irving, 2015]. (5) The CLIMPRED documentation serves as a repository of unified analysis methods through JUPYTER NOTEBOOK [Kluyver et al., 2016] examples and collects references and literature.

IV.3 PREDICTION SIMULATION TYPES

Weather and climate modeling institutions typically run so-called “hindcasts", where dynamical models are retrospectively initialized from many past observed climate states [Meehl et al., 2009]. Initializations are then slightly perturbed to generate an ensemble of forecasts that diverge solely due to their sensitive dependence on initial conditions [Lorenz, 1963]. Hindcasts are evaluated by using some statistical metric to score their performance against historical observations. “Skill" is established by comparing these results to the performance of some “reference" forecast, e.g. a persistence forecast [Jolliffe and Stephenson, 2011]. The main assumption is that the skill established relative to the past will propagate to forecasts of the future.

A more idealized approach is the so-called “perfect-model" framework, which is ideal for investigating processes leading to potentially exploitable predictability [Griffies and Bryan, 1997; Bushuk et al., 2018; Séférian et al., 2018; Spring and Ilyina, 2020]. Ensemble members are spun off an individual model (by slightly perturbing its state) to predict its own evolution. This avoids initialization shocks [Kröger et al., 2017], since the framework is self-contained. However, it cannot predict the real world. The perfect-model setup rather estimates the

theoretical upper limit timescale after which the value of dynamical initialization is lost due to chaos in the Earth system, assuming that the model perfectly replicates the dynamics of the real world. Skill quantification is accomplished by considering one ensemble member as the verification data and the remaining members as the forecasts [Griffies and Bryan, 1997].

IV.4 CLIMPRED CLASSES AND OBJECT-ORIENTED VERIFICATION

CLIMPRED supports both prediction system formats, offering *HindcastEnsemble* and *PerfectModelEnsemble* objects. *HindcastEnsemble* is instantiated with an initialized hindcast ensemble dataset and requires an observational dataset against which to verify. *PerfectModelEnsemble* is instantiated with an initialized perfect-model ensemble dataset and also accepts a control dataset against which to evaluate forecasts. Both objects can also track an uninitialized dataset, which represents a historical simulation that evolves solely due to random internal climate variability or can be used to isolate the influence of external forcing [e.g. Kay et al., 2015].

Assessing skill for *PredictionEnsemble* objects (the parent class to *HindcastEnsemble* and *PerfectModelEnsemble*) is standardized into a one-liner [Listing 1]:

Listing 1: **Example for predictive skill verification in CLIMPRED.**

```
PredictionEnsemble.verify(
    # Score forecast using the Anomaly Correlation Coefficient.
    metric="acc",
    # Compare the ensemble mean to observations.
    comparison="e2o",
    # Keep the same set of initializations at each lead time.
    alignment="same_inits",
    # Reduce the verification over the initialization dimension.
    dim="init",
    # Score performance of a persistence forecast as well.
    reference="persistence",
)
```

Each keyword argument allows flexibility from the user's end—one can select from a library of metrics, comparison types, alignment strategies, dimensional reductions, and reference forecasts. The most unique feature to CLIMPRED, however, is the ability for users to choose the alignment strategy to pair initialization dates with verification dates over numerous lead times. In other words, initialization dates need to be converted to target forecast dates by shifting them using the lead time coordinate. This is tedious, since one must remedy

disparities in calendar types between the model and observations and account for the time span of or gaps in observations relative to the time span of the model.

There is seemingly no unified approach to how hindcast initialization dates are aligned with observational dates in the academic literature. The authors of CLIMPRED thus identified three techniques, which can be selected by the user: (1) Maximize the degrees of freedom by selecting all initialization dates that verify with the available observations at each lead. In turn, initializations and verification dates are not held constant for each lead. (2) Use the identical set of initializations that can verify over the given observational window at all leads. However, the verification dates change at each lead. (3) Use the identical verification window at each lead, while allowing the set of initializations used at each lead to change. These strategies are shown graphically and explained in more detail in the documentation. Note that CLIMPRED offers extensive analysis functionality in addition to forecast verification, such as spatio-temporal smoothing [Goddard et al., 2013], bias removal [Boer et al., 2016], significance testing [Goddard et al., 2013; Boer et al., 2016; DelSole and Tippett, 2016], and a graphics library.

IV.5 USE IN ACADEMIC LITERATURE

CLIMPRED has been used to drive analysis in three academic papers so far. Brady et al. [2020] used the *HindcastEnsemble* class to highlight multi-year predictability of ocean acidification in the California Current; Spring and Ilyina [2020] used the *PerfectModelEnsemble* class to highlight predictability horizons in the global carbon cycle; and Krumhardt et al. [2020] used the *HindcastEnsemble* class to illuminate multi-year predictability in marine Net Primary Productivity.

ACKNOLWEDGEMENTS

We thank Andrew Huang for early stage refactoring and continued feedback on CLIMPRED. We also thank Kathy Pegion for pioneering the seasonal, monthly, and subseasonal time resolutions. Thanks in addition to Ray Bell for initiating and maintaining XSKILLSCORE[1], which serves to host the majority of metrics used in CLIMPRED.

1 Code: https://github.com/xarray-contrib/xskillscore, Documentation: https://xskillscore.readthedocs.io/en/stable/

BIBLIOGRAPHY

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., & Meinshausen, N. (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, *458*(7242), 1163–1166. doi:10/c8j665. (Cit. on p. 5)

Andrew, R. M. (2020). A comparison of estimates of global carbon dioxide emissions from fossil carbon sources. *Earth System Science Data*, *12*(2), 1437–1465. doi:10/ghxc7z. (Cit. on p. 34)

Archer, D. (2010). *The Global Carbon Cycle*. Princeton University Press. Retrieved November 14, 2018, from https://press.princeton.edu/books/paperback/9780691144146/the-global-carbon-cycle. (Cit. on p. 2)

Archer, D., Eby, M., Brovkin, V., Ridgwell, A., Cao, L., Mikolajewicz, U., … Tokos, K. (2009). Atmospheric Lifetime of Fossil Fuel Carbon Dioxide. *Annual Review of Earth and Planetary Sciences*, *37*(1), 117–134. doi:10/b963j9. (Cit. on p. 3)

Årthun, M., Bogstad, B., Daewel, U., Keenlyside, N. S., Sandø, A. B., Schrum, C., & Ottersen, G. (2018). Climate based multi-year predictions of the Barents Sea cod stock. *PLOS ONE*, *13*(10), e0206319. doi:10/gfg58b. (Cit. on p. 32)

Asayama, S., Bellamy, R., Geden, O., Pearce, W., & Hulme, M. (2019). Why setting a climate deadline is dangerous. *Nature Climate Change*, *9*(8), 570–572. doi:10/ggvwnq. (Cit. on p. 6)

Atger, F. (2003). Spatial and Interannual Variability of the Reliability of Ensemble-Based Probabilistic Forecasts: Consequences for Calibration. *Monthly Weather Review*, *131*(8), 1509–1523. doi:10/dn5knt. (Cit. on p. 26)

Aykut, S. C., Morena, E., & Foyer, J. (2020). 'Incantatory' governance: Global climate politics' performative turn and its wider significance for global politics. *International Politics*. doi:10/ghtnsg. (Cit. on p. 5)

Bacastow, R. B. (1976). Modulation of atmospheric carbon dioxide by the Southern Oscillation. *Nature*, *261*(5556), 116–118. doi:10/c9rrgm. (Cit. on pp. 5, 13, 63)

Bacastow, R. B., Adams, J. A., Keeling, C. D., Moss, D. J., Whorf, T. P., & Wong, C. S. (1980). Atmospheric Carbon Dioxide, the Southern Oscillation, and the Weak 1975 El Niño. *Science*, *210*(4465), 66–68. doi:10/dtg5gz. (Cit. on p. 13)

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., … Xu, S. (2016). A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT). *Earth System Science Data*, *8*(2), 383–413. doi:10/f3sgd6. (Cit. on p. 11)

Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P., & White, J. W. C. (2012). Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature*, *488*(7409), 70–72. doi:10/f35g9p. (Cit. on pp. 8, 13, 53, 67)

Balmaseda, M. A., Dee, D., Vidard, A., & Anderson, D. L. T. (2007). A multivariate treatment of bias for sequential data assimilation: Application to the tropical oceans. *Quarterly Journal of the Royal Meteorological Society*, *133*(622), 167–179. doi:10/czgj3m. (Cit. on pp. 109–111)

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161. doi:10/f5fbrq. (Cit. on p. 11)

Basu, S., Lehman, S. J., Miller, J. B., Andrews, A. E., Sweeney, C., Gurney, K. R., … Tans, P. P. (2020). Estimating US fossil fuel CO_2 emissions from measurements of ^{14}C in atmospheric CO_2. *Proceedings of the National Academy of Sciences*, *117*(24), 13300–13307. doi:10/ghzzqj. (Cit. on p. 33)

Bett, P. E., Williams, K. E., Burton, C., Scaife, A. A., Wiltshire, A. J., & Gilham, R. (2020). Skillful seasonal prediction of key carbon cycle components: NPP and fire risk. *Environmental Research Communications*, *2*(5), 055002. doi:10/ghtss9. (Cit. on p. 32)

Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., & Kennedy, J. J. (2016). El Niño and a record CO_2 rise. *Nature Climate Change*, *6*(9), 806–810. doi:10/gddfcr. (Cit. on pp. 22, 64, 76)

Betts, R. A., Jones Chris D., Knight Jeff. R., Keeling Ralph. F., Kennedy John. J., Wiltshire Andrew J., … Aragão Luiz 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. doi:10/gfthxx. (Cit. on pp. 4, 19, 22–24, 51, 64, 71, 76)

Bindoff, N. L., Scott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., … Zhang, X. (2013). Detection and Attribution of Climate Change: From Global to Regional. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, … P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 867–952). doi:10.1017/CBO9781107415324.022. (Cit. on p. 7)

Bloch, E. D. (2011). *Proofs and fundamentals: A first course in abstract mathematics* (2nd ed). New York: Springer. Retrieved from https://www.springer.com/gp/book/9781441971265. (Cit. on p. 18)

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., … Eade, R. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.* *9*(10), 3751–3777. doi:10/f89qdf. (Cit. on pp. 14, 125)

Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., & Mignot, J. (2020). Improved Decadal Predictions of North Atlantic Subpolar Gyre SST in CMIP6. *Geophysical Research Letters*, *48*(3), e2020GL091307. doi:10/ghssw5. (Cit. on p. 31)

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. doi:10/gdm28w. (Cit. on p. 15)

Brady, R. X., & Spring, A. (2021). Climpred: Verification of weather and climate forecasts. *Journal of Open Source Software*, *6*(59), 2781. doi:10/gh9646. (Cit. on pp. vii, 1, 15, 33, 111, 119, 121)

Brady, R. X., Lovenduski, N. S., Yeager, S. G., Long, M. C., & Lindsay, K. (2020). Skillful multiyear predictions of ocean acidification in the California Current System. *Nature Communications*, *11*(1), 1–9. doi:10/ggtpks. (Cit. on p. 125)

Branstator, G., & Selten, F. (2009). "Modes of Variability" and Climate Change. *Journal of Climate*, *22*(10), 2639–2658. doi:10/fgnx8j. (Cit. on p. 44)

Branstator, G., & Teng, H. (2010). Two Limits of Initial-Value Decadal Predictability in a CGCM. *Journal of Climate*, *23*(23), 6292–6311. doi:10/bwq92h. (Cit. on pp. 67, 69)

Brune, S., & Baehr, J. (2020). Preserving the coupled atmosphere–ocean feedback in initializations of decadal climate predictions. *WIREs Climate Change*, *11*(3), e637. doi:10/ghtnt8. (Cit. on pp. 32, 109)

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J. (2017). Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts. *Climate Dynamics*, 1–24. doi:10/gdsnf7. (Cit. on p. 32)

Buizza, R., & Leutbecher, M. (2015). The forecast skill horizon: The Forecast Skill Horizon. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 3366–3382. doi:10/gfx52f. (Cit. on p. 79)

Burke, E. J., Zhang, Y., & Krinner, G. (2020). Evaluating permafrost physics in the Coupled Model Intercomparison Project 6 (CMIP6) models and their sensitivity to climate change. *The Cryosphere*, *14*(9), 3155–3174. doi:10/ghn49c. (Cit. on p. 34)

Bushuk, M., Msadek, R., Winton, M., Vecchi, G., Yang, X., Rosati, A., & Gudgel, R. (2018). Regional Arctic sea–ice prediction: Potential versus operational seasonal forecast skill. *Climate Dynamics*. doi:10/gd7hfq. (Cit. on pp. 31, 78, 123)

CHE. (2017). CO_2 Human Emission: Separating Human Impact from the Natural Carbon Cycle, European Commission H2020. Retrieved January 2, 2021, from https://www.che-project.eu. (Cit. on p. 34)

Canadell, J. G., Quéré, C. L., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., ... Marland, G. (2007). Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the Na-*

tional Academy of Sciences, *104*(47), 18866–18870. doi:10/ddmj7h. (Cit. on pp. 3, 5, 7)

Carroll, D., Menemenlis, D., Adkins, J. F., Bowman, K. W., Brix, H., Dutkiewicz, S., … Zhang, H. (2020). The ECCO-Darwin Data-Assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multidecadal Surface Ocean pCO_2 and Air-Sea CO_2 Flux. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2019MS001888. doi:10/gg6t6t. (Cit. on p. 12)

Chevallier, F., Zheng, B., Broquet, G., Ciais, P., Liu, Z., Davis, S. J., … O'Dell, C. W. (2020). Local anomalies in the column-averaged dry air mole fractions of carbon dioxide across the globe during the first months of the coronavirus recession. *Geophysical Research Letters*, *47*(22), e2020GL090244. doi:10/ghhn39. (Cit. on p. 10)

Ciais, P., Tan, J., Wang, X., Roedenbeck, C., Chevallier, F., Piao, S.-L., … Tans, P. (2019). Five decades of northern land carbon uptake revealed by the interhemispheric CO_2 gradient. *Nature*, *568*(7751), 221–225. doi:10/gfxx8n. (Cit. on p. 6)

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*(6809), 184–187. doi:10/b2q48n. (Cit. on p. 13)

Dask Development Team. (2016). Dask: Library for dynamic task scheduling. Retrieved June 4, 2019, from https://docs.dask.org/en/latest/. (Cit. on pp. 122, 123)

DelSole, T., & Tippett, M. K. (2016). Forecast Comparison Based on Random Walks. *Monthly Weather Review*, *144*(2), 615–626. doi:10/f782pf. (Cit. on p. 125)

Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, *2*(11), 775–779. doi:10/gbpwxx. (Cit. on pp. 40, 50)

Dlugokencky, E. (2021). Annual CO_2 mole fraction increase (ppm) NOAA ESRL. Retrieved February 6, 2021, from https://www.esrl.noaa.gov/gmd/webdata/ccgg/trends/co2/co2_gr_gl.txt. (Cit. on p. 9)

Dlugokencky, E., & Tans, P. (2019). ESRL Global Monitoring Division - Global Greenhouse Gas Reference Network. Retrieved September 14, 2019, from ftp://aftp.cmdl.noaa.gov/products/trends/co2. (Cit. on pp. 46, 47, 56, 58)

Doney, S. C., Lindsay, K., Fung, I., & John, J. (2006). Natural Variability in a Stable, 1000-Yr Global Coupled Climate–Carbon Cycle Simulation. *Journal of Climate*, *19*(13), 3033–3054. doi:10/d54g8v. (Cit. on pp. 63, 71)

Eddebbar, Y. A., Rodgers, K. B., Long, M. C., Subramanian, A. C., Xie, S.-P., & Keeling, R. F. (2019). El Niño-like Physical and

Biogeochemical Ocean Response to Tropical Eruptions. *Journal of Climate*. doi:10/gfwdvz. (Cit. on p. 63)

Edenhofer, O., Flachsland, M., Jakob, M., & Lessmann, K. (2014). The Atmosphere as a Global Commons - Challenges for Challenges for International Cooperation and Governance. In L. Bernard & W. Semmler (Eds.), *The Oxford Handbook of the The Oxford Handbook of the Macroeconomics of Global Warming*. Oxford University Press. Retrieved January 13, 2021, from https://www.belfercenter.org/sites/default/files/files/publication/hpcadp58_edenhofer-flachsland-jakob-lessmann.pdf. (Cit. on p. 6)

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap* (1st edition). New York: Chapman and Hall/CRC. Retrieved from https://cds.cern.ch/record/526679/files/0412042312_TOC.pdf. (Cit. on pp. 66, 91)

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, *99*(C5), 10143–10162. doi:10/fpjxh8. (Cit. on pp. 109, 111)

Eynard-Bontemps, G., Abernathey, R., Hamman, J., Ponte, A., & Rath, W. (2019). The pangeo big data ecosystem and its use at CNES. In *Proc. of the 2019 conference on Big Data from Space (BiDS'19)*. doi:10.2760/848593. (Cit. on p. 123)

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* *9*(5), 1937–1958. doi:10/gbpwpt. (Cit. on pp. 11, 65, 88, 89)

Feldmann, H., g Pinto, J., Laube, N., Uhlig, M., Moemken, J., Pasternack, A., … Kottmeier, C. (2019). Skill and added value of the MiKlip regional decadal prediction system for temperature over Europe. *Tellus A: Dynamic Meteorology and Oceanography*, *71*(1), 1618678. doi:10/ghpz5q. (Cit. on p. 32)

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., … Rummukainen, M. (2013). Evalution of Climate Models. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, …P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Retrieved June 3, 2018, from https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter09_FINAL.pdf. (Cit. on pp. 7, 13)

Forster, P. M., Forster, H. I., Evans, M. J., Gidden, M. J., Jones, C. D., Keller, C. A., … Turnock, S. T. (2020). Current and future global

climate impacts resulting from COVID-19. *Nature Climate Change*, 1–7. doi:10/gg7s7w. (Cit. on pp. 8, 9)

Francey, R. J., Tans, P. P., Allison, C. E., Enting, I. G., White, J. W. C., & Trolier, M. (1995). Changes in oceanic and terrestrial carbon uptake since 1982. *Nature*, *373*(6512), 326–330. doi:10/cf2dgt. (Cit. on p. 13)

Frankignoul, C., Gastineau, G., & Kwon, Y.-O. (2017). Estimation of the SST Response to Anthropogenic and External Forcing and Its Impact on the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation. *Journal of Climate*, *30*(24), 9871–9895. doi:10/gcxk47. (Cit. on p. 44)

Fransner, F., Counillon, F., Bethke, I., Tjiputra, J., Samuelsen, A., Nummelin, A., & Olsen, A. (2020). Ocean Biogeochemical Predictions—Initialization and Limits of Predictability. *Frontiers in Marine Science*, *7*. doi:10/gg22rr. (Cit. on pp. 15, 25, 87, 88, 110, 111)

Friedlingstein, P., Andrew, R. M., Rogelj, J., Peters, G. P., Canadell, J. G., Knutti, R., … Le Quéré, C. (2014). Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nature Geoscience*, *7*(10), 709–715. doi:10/f6kt57. (Cit. on p. 34)

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2013). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *Journal of Climate*, *27*(2), 511–526. doi:10/f5nsgs. (Cit. on p. 34)

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., … Zaehle, S. (2019). Global Carbon Budget 2019. *Earth System Science Data*, *11*(4), 1783–1838. doi:10/ggd7p4. (Cit. on pp. 39, 55, 63, 67, 90)

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., … Zaehle, S. (2020). Global Carbon Budget 2020. *Earth System Science Data*, *12*(4), 3269–3340. doi:10/ghn75s. (Cit. on pp. 2–5, 8, 9, 22, 30, 33, 34, 88)

Frölicher, T. L., Joos, F., Raible, C. C., & Sarmiento, J. L. (2013). Atmospheric CO_2 response to volcanic eruptions: The role of ENSO, season, and variability. *Global Biogeochemical Cycles*, *27*(1), 239–251. doi:10/f4vwbx. (Cit. on p. 63)

Fyfe, J. C., Meehl, G. A., England, M. H., Mann, M. E., Santer, B. D., Flato, G. M., … Swart, N. C. (2016). Making sense of the early-2000s warming slowdown. *Nature Climate Change*, *6*, 224–228. doi:10/gbpwqj. (Cit. on p. 40)

Galbraith, E. D., Kwon, E. Y., Gnanadesikan, A., Rodgers, K. B., Griffies, S. M., Bianchi, D., … Held, I. M. (2011). Climate Variability and Radiocarbon in the CM2Mc Earth System Model. *Journal of Climate*, *24*(16), 4230–4254. doi:10/c4p5r5. (Cit. on p. 33)

Geden, O. (2016). An actionable climate target. *Nature Geoscience*, *9*. doi:10.1038/ngeo2699. (Cit. on p. 6)

Geden, O. (2018). Politically informed advice for climate action. *Nature Geoscience*, *11*, 380–383. doi:10.1038/s41561-018-0143-3. (Cit. on p. 6)

Geden, O., & Löschel, A. (2017). Define limits for temperature overshoot targets. *Nature Geoscience*, *10*(12), 881–882. doi:10/gcw9f2. (Cit. on p. 6)

Gier, B. K., Buchwitz, M., Reuter, M., Cox, P. M., Friedlingstein, P., & Eyring, V. (2020). Spatially resolved evaluation of Earth system models with satellite column-averaged CO_2. *Biogeosciences*, *17*(23), 6115–6144. doi:10/ghnthp. (Cit. on p. 32)

Giorgetta, M. A., Jungclaus Johann, Reick Christian H., Legutke Stephanie, Bader Jürgen, Böttinger Michael, … Stevens Bjorn. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597. doi:10/f5dzvh. (Cit. on p. 44)

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. doi:10/c6758w. (Cit. on p. 26)

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., … Delworth, T. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, *40*(1-2), 245–272. doi:10/f4jjvf. (Cit. on pp. 23, 66, 73, 125)

Griffies, S. M., & Bryan, K. (1997). A predictability study of simulated North Atlantic multidecadal variability. *Climate Dynamics*, *13*(7-8), 459–487. doi:10/ch4kc4. (Cit. on pp. 22, 65, 78, 89, 123, 124)

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., … Yeager, S. G. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development*, *9*(9), 3231–3296. doi:10/f85pjx. (Cit. on p. 90)

Gruber, N., Landschützer, P., & Lovenduski, N. S. (2019). The Variable Southern Ocean Carbon Sink. *Annual Review of Marine Science*, *11*(1), 159–186. doi:10/gf7sc9. (Cit. on p. 4)

Han, G., Zhu, J., & Zhou, G. (2004). Salinity estimation using the *T* - *S* relation in the context of variational data assimilation: SALINITY ESTIMATION. *Journal of Geophysical Research: Oceans*, *109*(C3). doi:10/b34qr8. (Cit. on pp. 109–111)

Haney, R. L. (1974). A Numerical Study of the Response of an Idealized Ocean to Large-Scale Surface Heat and Momentum Flux. *Journal of Physical Oceanography*, *4*(2), 145–167. doi:10/ch3kf3. (Cit. on p. 89)

Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P., & Ghil, M. (2016). Causal Counterfactual Theory for the Attribution of Weather and

Climate-Related Events. *Bulletin of the American Meteorological Society*, *97*(1), 99–110. doi:10/gf75gq. (Cit. on pp. 18, 39–41, 50)

Hausfather, Z., & Peters, G. P. (2020). Emissions – the 'business as usual' story is misleading. *Nature*, *577*(7792), 618–620. doi:10/ggjr4t. (Cit. on pp. 39, 43, 45)

Hawkins, E., Tietsche, S., Day, J. J., Melia, N., Haines, K., & Keeley, S. (2016). Aspects of designing and evaluating seasonal-to-interannual Arctic sea-ice prediction systems. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 672–683. doi:10/gfb3pn. (Cit. on pp. 31, 66)

Heinze, C., Meyer, S., Goris, N., Anderson, L., Steinfeldt, R., Chang, N., … Bakker, D. C. E. (2015). The ocean carbon sink – impacts, vulnerabilities and challenges. *Earth System Dynamics*, *6*(1), 327–358. doi:10/f3n8c6. (Cit. on pp. 2, 3, 63)

Hong, C., Burney, J. A., Pongratz, J., Nabel, J. E. M. S., Mueller, N. D., Jackson, R. B., & Davis, S. J. (2021). Global and regional drivers of land-use emissions in 1961–2017. *Nature*, *589*(7843), 554–561. doi:10/ghwp9c. (Cit. on p. 3)

Howe, L. C., MacInnis, B., Krosnick, J. A., Markowitz, E. M., & Socolow, R. (2019). Acknowledging uncertainty impacts public acceptance of climate scientists' predictions. *Nature Climate Change*, 1–5. doi:10/ggbj4j. (Cit. on p. 50)

Hoyer, S., & Hamman, J. (2017). Xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, *5*(1). doi:10/gdqdmw. (Cit. on pp. 15, 122)

Hua, W., Zhou, L., Nicholson, S. E., Chen, H., & Qin, M. (2019). Assessing reanalysis data for understanding rainfall climatology and variability over Central Equatorial Africa. *Climate Dynamics*, *53*(1), 651–669. doi:10/ghthbt. (Cit. on pp. 32, 111)

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, *9*(3), 90–95. doi:10/drbjhg. (Cit. on p. 75)

Hurrell, J. W. (1995). Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science*, *269*(5224), 676–679. doi:10/c34jtm. (Cit. on p. 4)

IEA. (2020). Global Energy Review 2020. Retrieved December 4, 2020, from https://www.iea.org/reports/global-energy-review-2020. (Cit. on pp. 8, 9)

Illing, S., Kadow, C., Oliver, K., & Cubasch, U. (2014). MurCSS: A Tool for Standardized Evaluation of Decadal Hindcast Systems. *Journal of Open Research Software*, *2*(1), e24. doi:10/gfxr7x. (Cit. on p. 123)

Ilyina, T., Li, H., Spring, A., Müller, W. A., Bopp, L., Chikamoto, M. O., … Yeager, S. (2021). Predictable Variations of the Carbon Sinks and Atmospheric CO2 Growth in a Multi-Model Framework. *Geophysical Research Letters*, *48*(6), e2020GL090695.

doi:10/ghsn7h. (Cit. on pp. vii, 10–12, 15, 25, 26, 29, 31, 86, 87, 111)

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, *5*(2), 287–315. doi:10/f5dz39. (Cit. on pp. 13, 44, 65, 88)

Irving, D. (2015). A Minimum Standard for Publishing Computational Results in the Weather and Climate Sciences. *Bulletin of the American Meteorological Society*, *97*(7), 1149–1158. doi:10/gf4wzh. (Cit. on pp. 75, 123)

Issa, S., Dahy, B., Ksiksi, T., & Saleous, N. (2020). A Review of Terrestrial Carbon Assessment Methods Using Geo-Spatial Technologies with Emphasis on Arid Lands. *Remote Sensing*, *12*(12), 2008. doi:10/ghssfb. (Cit. on p. 12)

Jacob, D. J. (1999). *Introduction to atmospheric chemistry*. Princeton, N.J: Princeton University Press. Retrieved from http://acmg.seas.harvard.edu/publications/jacobbook/index.html. (Cit. on pp. 2, 3, 6)

Jahn, A., Lindsay, K., Giraud, X., Gruber, N., Otto-Bliesner, B. L., Liu, Z., & Brady, E. C. (2015). Carbon isotopes in the ocean model of the Community Earth System Model (CESM1). *Geoscientific Model Development*, *8*(8), 2419–2434. doi:10/f7qhcb. (Cit. on p. 33)

Janssens-Maenhout, G., Pinty, B., Dowell, M., Zunker, H., Andersson, E., Balsamo, G., … Veefkind, P. (2020). Toward an Operational Anthropogenic CO_2 Emissions Monitoring and Verification Support Capacity. *Bulletin of the American Meteorological Society*, *101*(8), 1439–1451. doi:10/ggk2dx. (Cit. on pp. 6, 34)

Jeuken, A. B. M., Siegmund, P. C., Heijboer, L. C., Feichter, J., & Bengtsson, L. (1996). On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation. *Journal of Geophysical Research: Atmospheres*, *101*(D12), 16939–16950. doi:10/d64x9q. (Cit. on pp. 87, 94)

Jolliffe, I. T., & Stephenson, D. B. (2011). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. doi:10.1002/9781119960003. (Cit. on pp. 66, 78, 92, 112, 123)

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., … Zaehle, S. (2016). C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: Experimental protocol for CMIP6. *Geosci. Model Dev.* *9*(8), 2853–2880. doi:10/f84585. (Cit. on pp. 65, 89)

Jones, C. D., & Cox, P. M. (2005). On the significance of atmospheric $_2$ growth rate anomalies in 2002–2003. *Geophysical Research Letters*, *32*(14). doi:10/dj74zd. (Cit. on p. 64)

Jones, C. D., Collins, M., Cox, P. M., & Spall, S. A. (2001). The Carbon Cycle Response to ENSO: A Coupled Climate–Carbon Cycle Model Study. *Journal of Climate*, *14*(21), 4113–4129. doi:10/bwrw6c. (Cit. on pp. 4, 13, 19, 32, 39, 47, 63, 71)

Jones, C., Robertson, E., Arora, V., Friedlingstein, P., Shevliakova, E., Bopp, L., … Tjiputra, J. (2013). Twenty-First-Century Compatible CO_2 Emissions and Airborne Fraction Simulated by CMIP5 Earth System Models under Four Representative Concentration Pathways. *Journal of Climate*, *26*(13), 4398–4413. doi:10/f44bbn. (Cit. on pp. 45, 53, 58, 60)

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., … von Storch, J. S. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, *5*(2), 422–446. doi:10/f5d3x7. (Cit. on pp. 65, 88)

Karl, D. M., & Lukas, R. (1996). The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep Sea Research Part II: Topical Studies in Oceanography*, *43*(2), 129–156. doi:10/cj6kvq. (Cit. on p. 12)

Karoly, D. J. (1990). The role of transient eddies in low-frequency zonal variations of the Southern Hemisphere circulation. *Tellus A: Dynamic Meteorology and Oceanography*, *42*(1), 41–50. doi:10/fx3bcr. (Cit. on p. 4)

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., … Vertenstein, M. (2015). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. doi:10/f7r9st. (Cit. on pp. 44, 124)

Keeling, C. D., Whorf, T. P., Wahlen, M., & van der Plichtt, J. (1995). Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. *Nature*, *375*(6533), 666. doi:10/b69ptn. (Cit. on pp. 13, 17, 63, 71)

Keeling, C. D., Bacastow, R. B., & Whorf, T. P. (1982). Measurements of the concentration of carbon dioxide at Mauna Loa Observatory, Hawaii. In W. C. Clark (Ed.), *Carbon Dioxide Review*. New York: Oxford University Press. Retrieved February 3, 2021, from https://www.osti.gov/biblio/6438207. (Cit. on pp. 8, 13)

Keeling, C. D., Bacastow, R. B., Carter, A. F., Piper, S. C., Whorf, T. P., Heimann, M., … Roeloffzen, H. (1989a). A three-dimensional model of atmospheric CO_2 transport based on observed winds: 1. Analysis of observational data. In *Aspects of Climate Variability in the Pacific and the Western Americas* (pp. 165–236). doi:10.1029/GM055p0165. (Cit. on pp. 13, 17, 32)

Keeling, C. D., Piper, S. C., & Heimann, M. (1989b). A three-dimensional model of atmospheric CO_2 transport based on observed winds: 4. Mean annual gradients and interannual variations. In *Aspects of Climate Variability in the Pacific and the Western Americas* (pp. 305–363). doi:10.1029/GM055p0305. (Cit. on pp. 12, 32)

Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., & Meijer, H. A. (2005). Atmospheric CO_2 and $^{13}CO_2$ Exchange with the Terrestrial Biosphere and Oceans from 1978 to 2000: Observations and Carbon Cycle Implications. In I. Baldwin, M. Caldwell, G. Heldmaier, R. B. Jackson, O. Lange, H. Mooney, ... T. E. Cerling (Eds.), *A History of Atmospheric CO2 and Its Effects on Plants, Animals, and Ecosystems* (pp. 83–113). Ecological Studies. doi:10.1007/0-387-27048-5_5. (Cit. on pp. 72, 79)

Keeling, R. F. (2008). Recording Earth's Vital Signs. *Science*, *319*(5871), 1771–1772. doi:10/b4pz5b. (Cit. on p. 73)

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. doi:10/ffpmhd. (Cit. on pp. 10, 90)

Keppel-Aleks, G., Randerson, J. T., Lindsay, K., Stephens, B. B., Keith Moore, J., Doney, S. C., ... Wofsy, S. C. (2013). Atmospheric Carbon Dioxide Variability in the Community Earth System Model: Evaluation and Transient Dynamics during the Twentieth and Twenty-First Centuries. *Journal of Climate*, *26*(13), 4447–4475. doi:10/f439zf. (Cit. on pp. 32, 71)

Kirchmeier-Young, M. C., Zwiers, F. W., & Gillett, N. P. (2016). Attribution of Extreme Events in Arctic Sea Ice Extent. *Journal of Climate*, *30*(2), 553–571. doi:10/f9kp6p. (Cit. on p. 44)

Kirtman, B., Power, S. B., Adedoyin, A. J., Boer, G. J., Bojariu, R., Camilloni, I., ... Wang, H.-J. (2013). Near-term Climate Change: Projections and Predictability. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, ... P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Retrieved March 15, 2018, from https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter11_FINAL.pdf. (Cit. on pp. 7, 70)

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... al, e. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90). doi:10/gf48c9. (Cit. on p. 123)

Kobayashi, H., Shimota, A., Kondo, K., Okumura, E., Kameda, Y., Shimoda, H., & Ogawa, T. (1999). Development and evalua-

tion of the interferometric monitor for greenhouse gases: A high-throughput Fourier-transform infrared radiometer for nadir Earth observation. *Applied Optics*, *38*(33), 6801. doi:10/cjfzdw. (Cit. on p. 12)

Koul, V. (2020). *Decadal prediction of shelf-sea marine ecosystems in the eastern North Atlantic: The role of the Subpolar Gyre* (PhD Thesis, University of Hamburg, Hamburg). Retrieved from http://hdl.handle.net/21.11116/0000-0006-E1A6-0. (Cit. on p. 32)

Kröger, J., Pohlmann, H., Sienz, F., Marotzke, J., Baehr, J., Köhl, A., … Müller, W. A. (2017). Full-field initialized decadal predictions with the MPI earth system model: An initial shock in the North Atlantic. *Climate Dynamics*. doi:10/gdsnf8. (Cit. on pp. 65, 113, 123)

Krumhardt, K. M., Lovenduski, N. S., Long, M. C., Luo, J. Y., Lindsay, K., Yeager, S., & Harrison, C. (2020). Potential Predictability of Net Primary Production in the Ocean. *Global Biogeochemical Cycles*, *34*(6), e2020GB006531. doi:10/gg9ss8. (Cit. on pp. 32, 125)

Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., van Heuven, S., … Wanninkhof, R. (2015). The reinvigoration of the Southern Ocean carbon sink. *Science*, *349*(6253), 1221–1224. doi:10/f7q9th. (Cit. on pp. 12, 39)

Landschützer, P., Gruber Nicolas, & Bakker Dorothee C. E. (2016). Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, *30*(10), 1396–1417. doi:10/f9csrg. (Cit. on pp. 4, 23, 81)

Landschützer, P., Ilyina, T., & Lovenduski, N. S. (2019). Detecting Regional Modes of Variability in Observation-Based Surface Ocean pCO_2. *Geophysical Research Letters*, *46*(5), 2670–2679. doi:10/gfxkj6. (Cit. on pp. 4, 81)

Le Quéré, C., Jackson, R. B., Jones, M. W., Smith, A. J. P., Abernethy, S., Andrew, R. M., … Peters, G. P. (2020). Temporary reduction in daily global CO_2 emissions during the COVID-19 forced confinement. *Nature Climate Change*, *10*, 647–653. doi:10.1038/s41558-020-0797-x. (Cit. on pp. 8, 9)

Lee, D. E., & Biasutti, M. (2014). Climatology and Variability of Precipitation in the Twentieth-Century Reanalysis. *Journal of Climate*, *27*(15), 5964–5981. doi:10/f6b5jn. (Cit. on pp. 32, 111)

Lefèvre, N., Caniaux, G., Janicot, S., & Gueye, A. K. (2013). Increased CO_2 outgassing in February-May 2010 in the tropical Atlantic following the 2009 Pacific El Niño. *Journal of Geophysical Research: Oceans*, *118*(4), 1645–1657. doi:10/ghrngp. (Cit. on p. 95)

Levin, I., Kromer, B., Schmidt, M., & Sartorius, H. (2003). A novel approach for independent budgeting of fossil fuel CO_2 over Europe by $^{14}CO_2$ observations. *Geophysical Research Letters*, *30*(23). doi:10/ffh5mf. (Cit. on p. 33)

Lewandowsky, S., Risbey, J. S., & Oreskes, N. (2015). On the definition and identifiability of the alleged "hiatus" in global warming. *Scientific Reports*, *5*, 16784. doi:10/f7zc8r. (Cit. on p. 40)

Li, H., Ilyina, T., Müller, W. A., & Landschützer, P. (2019). Predicting the variable ocean carbon sink. *Science Advances*, *5*(4), eaav6471. doi:10/gf4fxm. (Cit. on pp. 10, 12, 15, 22, 23, 25, 29, 51, 64, 70, 73, 86, 87, 90, 110)

Li, H., & Ilyina, T. (2018). Current and Future Decadal Trends in the Oceanic Carbon Uptake Are Dominated by Internal Variability. *Geophysical Research Letters*, *45*(2), 916–925. doi:10/gc4fz9. (Cit. on pp. 44, 63)

Li, H., Ilyina, T., Müller, W. A., & Sienz, F. (2016). Decadal predictions of the North Atlantic CO_2 uptake. *Nature Communications*, *7*, 11076. doi:10/f8wkrs. (Cit. on pp. 15, 64, 70, 87)

Lin, S.-J., & Rood, R. B. (1996). Multidimensional Flux-Form Semi-Lagrangian Transport Schemes. *Monthly Weather Review*, *124*(9), 2046–2070. doi:10/fsh52z. (Cit. on pp. 65, 89)

Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S. J., Feng, S., … Schellnhuber, H. J. (2020). Near-real-time monitoring of global CO_2 emissions reveals the effects of the COVID-19 pandemic. *Nature Communications*, *11*(1), 5172. doi:10/ghfvb5. (Cit. on pp. 8, 9, 34)

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*(2), 130–141. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2. (Cit. on p. 123)

Lorenz, E. N. (1965). A study of the predictability of a 28-variable atmospheric model. *Tellus*, *17*(3), 321–333. doi:10/fgdw36. (Cit. on p. 28)

Lorenz, E. N. (1975). Climatic predictability. In *GARP Publications Series, 16. The Physical basis of climate and climate modelling : Report of the International Study Conference in Stockholm, 29 July - 10 August 1974, organised by WMO and ICSU and supported by UNEP*. Geneva: WMO. Retrieved from https://library.wmo.int/index.php?lvl=notice_display&id=6943. (Cit. on p. 10)

Lovenduski, N. S., Yeager, S. G., Lindsay, K., & Long, M. C. (2019a). Predicting near-term variability in ocean carbon uptake. *Earth System Dynamics*, *10*(1), 45–57. doi:10/gfvxkc. (Cit. on pp. 10, 15, 22, 23, 25, 51, 64, 70, 73, 86, 87, 94, 95, 110)

Lovenduski, N. S., Bonan, G. B., Yeager, S. G., Lindsay, K., & Lombardozzi, D. L. (2019b). High predictability of terrestrial carbon fluxes from an initialized decadal prediction system. *Environmental Research Letters*, *14*(12), 124074. doi:10/ggfv2w. (Cit. on pp. 11, 15, 22, 23, 64, 71, 73, 86, 87)

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., … Marotzke, J. (2019). The Max Planck Institute Grand Ensemble - Enabling the Exploration of Climate System

Variability. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2050–2069. doi:10/gf3kgt. (Cit. on pp. 18, 39, 43, 53)

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., & Francis, R. C. (1997). A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bulletin of the American Meteorological Society*, *78*(6), 1069–1080. doi:10/b9kgqq. (Cit. on p. 4)

Manubens, N., Caron, L.-P., Hunter, A., Bellprat, O., Exarchou, E., Fučkar, N. S., … Doblas-Reyes, F. J. (2018). An R package for climate forecast verification. *Environmental Modelling & Software*, *103*, 29–42. doi:10/gc9wzk. (Cit. on p. 123)

Marotzke, J. (2019). Quantifying the irreducible uncertainty in near-term climate projections. *Wiley Interdisciplinary Reviews: Climate Change*, *10*(1), e563. doi:10/gfq92h. (Cit. on pp. 6, 17, 18, 28, 33, 35, 37, 39–41, 45, 49–51)

Marotzke, J., Müller, W. A., Vamborg, F. S. E., Becker, P., Cubasch, U., Feldmann, H., … Ziese, M. (2016). MiKlip: A National Research Project on Decadal Climate Prediction. *Bulletin of the American Meteorological Society*, *97*(12), 2379–2394. doi:10/gddfck. (Cit. on pp. 10, 15, 23, 63, 73, 92)

Marotzke, J., Jakob, C., Bony, S., Dirmeyer, P. A., O'Gorman, P. A., Hawkins, E., … Tuma, M. (2017). Climate research must sharpen its view. *Nature Climate Change*, *7*(2), 89–91. doi:10.1038/nclimate3206. (Cit. on pp. 1, 30)

Mastrandrea, M. D. (2010). *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. IPCC. Jasper Ridge, CA, USA. Retrieved November 29, 2019, from https://archive.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf. (Cit. on p. 42)

Matthews, H. D., Gillett, N. P., Stott, P. A., & Zickfeld, K. (2009). The proportionality of global warming to cumulative carbon emissions. *Nature*, *459*(7248), 829–832. doi:10/d7pjtb. (Cit. on p. 5)

Matthews, H. D., Tokarska, K. B., Nicholls, Z. R. J., Rogelj, J., Canadell, J. G., Friedlingstein, P., … Zickfeld, K. (2020). Opportunities and challenges in using remaining carbon budgets to guide climate policy. *Nature Geoscience*, *13*(12), 769–779. doi:10/ghmt8r. (Cit. on p. 5)

Matthews, H. D., Tokarska, K. B., Rogelj, J., Smith, C. J., MacDougall, A. H., Haustein, K., … Knutti, R. (2021). An integrated approach to quantifying uncertainties in the remaining carbon budget. *Communications Earth & Environment*, *2*(1), 1–11. doi:10/ghvcdx. (Cit. on pp. 5, 6)

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., … Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to

Increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. doi:10/gftpps. (Cit. on pp. 23, 53, 65, 88)

McKenna, C. M., Maycock, A. C., Forster, P. M., Smith, C. J., & Tokarska, K. B. (2020). Stringent mitigation substantially reduces risk of unprecedented near-term warming rates. *Nature Climate Change*, 1–6. doi:10/ghntnm. (Cit. on p. 17)

McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., & Lovenduski, N. S. (2016). Timescales for detection of trends in the ocean carbon sink. *Nature*, *530*(7591), 469–472. doi:10/f8bhx4. (Cit. on p. 41)

McKinley, G. A., Fay, A. R., Lovenduski, N. S., & Pilcher, D. J. (2017). Natural Variability and Anthropogenic Trends in the Ocean Carbon Sink. *Annual Review of Marine Science*, *9*(1), 125–150. doi:10.1146/annurev-marine-010816-060529. (Cit. on pp. 4, 39, 63)

McLaren, D. P., Tyfield, D. P., Willis, R., Szerszynski, B., & Markusson, N. O. (2019). Beyond "Net-Zero": A Case for Separate Targets for Emissions Reduction and Negative Emissions. *Frontiers in Climate*, *1*. doi:10/ghvcf8. (Cit. on p. 6)

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., … Stockdale, T. (2009). Decadal Prediction: Can It Be Skillful? *Bulletin of the American Meteorological Society*, *90*(10), 1467–1486. doi:10/dpsjbp. (Cit. on pp. 10, 11, 63, 86, 87, 123)

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., … Yeager, S. (2013). Decadal Climate Prediction: An Update from the Trenches. *Bulletin of the American Meteorological Society*, *95*(2), 243–267. doi:10/f3cvw2. (Cit. on pp. 22, 63)

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., … van Vuuren, D. P. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1), 213. doi:10/fn3pw6. (Cit. on pp. 11, 43–46, 51, 55, 56, 60)

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., … Wang, H. J. (2019). The SSP greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development Discussions*, 1–77. doi:10/gf86j8. (Cit. on p. 11)

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., … Yeager, S. (2020). Current and Emerging Developments in Subseasonal to Decadal Prediction. *Bulletin of the American Meteorological Society*, *101*(6), 869–896. doi:10/ggvcqv. (Cit. on pp. vii, 22, 33, 86, 122)

Met Office. (2010). Cartopy: A cartographic python library with a Matplotlib interface. Exeter, Devon. Retrieved from http://scitools.org.uk/cartopy. (Cit. on p. 75)

Milinski, S., Bader, J., Haak, H., Siongco, A. C., & Jungclaus, J. H. (2016). High atmospheric horizontal resolution eliminates the wind-driven coastal warm bias in the southeastern tropical Atlantic.

Geophysical Research Letters, *43*(19), 10,455–10,462. doi:10/f89qf7. (Cit. on p. 94)

Millar, R. J., Fuglestvedt, J. S., Friedlingstein, P., Rogelj, J., Grubb, M. J., Matthews, H. D., … Allen, M. R. (2017). Emission budgets and pathways consistent with limiting warming to 1.5°C. *Nature Geoscience*, *10*(10), 741–747. doi:10/gb22sz. (Cit. on p. 6)

Moemken, J., Feldmann, H., Pinto, J. G., Buldmann, B., Laube, N., Kadow, C., … Marotzke, J. (2021). The regional MiKlip decadal prediction system for Europe: Hindcast skill for extremes and user-oriented variables. *International Journal of Climatology*, *41*(S1), E1944–E1958. doi:10/ghpgwc. (Cit. on p. 32)

Murphy, A. H. (1988). Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Monthly Weather Review*, *116*(12), 2417–2424. doi:10/fc7mxd. (Cit. on p. 66)

Murphy, A. H. (1993). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, *8*(2), 281–293. doi:10/bbpk8s. (Cit. on pp. 1, 14, 35)

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., … Yool, A. (2017). Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). *Geoscientific Model Development*, *10*(6), 2169–2199. doi:10/gcc5zk. (Cit. on p. 90)

Palmer, T. (2020). A Vision for Numerical Weather Prediction in 2030. *arXiv:2007.04830 [physics]*. arXiv: 2007.04830 [physics]. Retrieved November 25, 2020, from http://arxiv.org/abs/2007.04830. (Cit. on pp. 14, 33)

Park, J.-Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J., & Zhang, S. (2018). Modeling Global Ocean Biogeochemistry With Physical Data Assimilation: A Pragmatic Solution to the Equatorial Instability. *Journal of Advances in Modeling Earth Systems*, *10*(3), 891–906. doi:10/gddxmt. (Cit. on pp. 12, 65, 87, 110)

Park, J.-Y., Stock, C. A., Dunne, J. P., Yang, X., & Rosati, A. (2019). Seasonal to multiannual marine ecosystem prediction with a global Earth system model. *Science*, *365*(6450), 284–288. doi:10/gf7fbj. (Cit. on pp. 32, 109)

Pasternack, A., Bhend, J., Liniger, M. A., Rust, H. W., Müller, W. A., & Ulbrich, U. (2018). Parametric decadal climate forecast recalibration (DeFoReSt 1.0). *Geoscientific Model Development*, *11*(1), 351–368. doi:10/gcxtsx. (Cit. on p. 32)

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, *9*(1), 438–464. doi:10/f94tnr. (Cit. on pp. 65, 88)

Payne, M. R., Hobday, A. J., MacKenzie, B. R., Tommasi, D., Dempsey, D. P., Fässler, S. M. M., … Villarino, E. (2017). Lessons from

the First Generation of Marine Ecological Forecast Products. *Frontiers in Marine Science*, *4*. doi:10/gddfct. (Cit. on p. 32)

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York, NY: Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9780511803161. (Cit. on pp. 18, 39–41)

Peters, G. P., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Jackson, R. B., Korsbakken, J. I., … Peregon, A. (2019). Carbon dioxide emissions continue to grow amidst slowly emerging climate policies. *Nature Climate Change*, 1–4. doi:10/ggfv2z. (Cit. on p. 7)

Peters, G. P. (2016). The 'best available science' to inform 1.5°C policy choices. *Nature Climate Change*, *6*(7), 646–649. doi:10/bd6v. (Cit. on p. 6)

Peters, G. P. (2018). Beyond carbon budgets. *Nature Geoscience*, *11*, 378–383. doi:10.1038/s41561-018-0142-4. (Cit. on pp. 6, 34)

Peters, G. P., & Geden, O. (2017). Catalysing a political shift from low to negative carbon. *Nature Climate Change*, *7*(9), 619–621. doi:10/cbt4. (Cit. on p. 6)

Peters, G. P., Le Quéré, C., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Ilyina, T., … Tans, P. (2017). Towards real-time verification of CO_2 emissions. *Nature Climate Change*, *7*(12), 848–850. doi:10/cf7g. (Cit. on pp. 6, 7, 11, 17, 28, 30, 34, 39, 63)

Phillips, H. E., & Joyce, T. M. (2007). Bermuda's Tale of Two Time Series: Hydrostation S and BATS. *Journal of Physical Oceanography*, *37*(3), 554–571. doi:10/drffqc. (Cit. on p. 12)

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., & Tschuck, P. (2004). Estimating the Decadal Predictability of a Coupled AOGCM. *Journal of Climate*, *17*(22), 4463–4472. doi:10/d2qf62. (Cit. on pp. 15, 70, 78)

Pohlmann, H., Jungclaus, J. H., Köhl, A., Stammer, D., & Marotzke, J. (2009). Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic. *Journal of Climate*, *22*(14), 3926–3938. doi:10/cdvhcr. (Cit. on pp. 10, 15, 89, 90)

Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., & Marotzke, J. (2019). Realistic Quasi-Biennial Oscillation Variability in Historical and Decadal Hindcast Simulations Using CMIP6 Forcing. *Geophysical Research Letters*, *46*(23), 14118–14125. doi:10/ggfrh8. (Cit. on p. 90)

Rasmusson, E. M., & Wallace, J. M. (1983). Meteorological Aspects of the El Niño/Southern Oscillation. *Science*, *222*(4629), 1195–1202. doi:10/dp58sc. (Cit. on p. 4)

Rast, S., Brokopf, R., Esch, M., Gayler, V., Kirchner, I., Kornblueh, L., … Schulzweida, U. (2012). *User manual for ECHAM6*. Max Planck Institute for Meteorology. Hamburg. Retrieved from https://icdc.cen.uni-hamburg.de/fileadmin/user_upload/icdc_Dokumente/ECHAM/echam6_userguide.pdf. (Cit. on p. 90)

Rayner, P. J., Enting, I. G., Francey, R. J., & Langenfelds, R. (1999). Reconstructing the recent carbon cycle from atmospheric CO_2, $\Delta^{13}C$ and O_2/N_2 observations. *Tellus B: Chemical and Physical Meteorology*, *51*(2), 213–232. doi:10/ghs5dd. (Cit. on p. 12)

Reay, D. S., Dentener, F., Smith, P., Grace, J., & Feely, R. A. (2008). Global nitrogen deposition and carbon sinks. *Nature Geoscience*, *1*(7), 430–437. doi:10/b7xmcd. (Cit. on p. 3)

Rocklin, M. (2015). Dask: Parallel Computation with Blocked algorithms and Task Scheduling. In *Python in Science Conference* (pp. 126–132). doi:10/gfz6s5. (Cit. on pp. 15, 122, 123)

Rödenbeck, C., Houweling, S., Gloor, M., & Heimann, M. (2003). CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmospheric Chemistry and Physics*, *3*(6), 1919–1964. doi:10/cnmp77. (Cit. on p. 13)

Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., … Zeng, J. (2015). Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean pCO_2 Mapping intercomparison (SOCOM). *Biogeosciences*, *12*(23), 7251–7278. doi:10/gb9bnx. (Cit. on p. 81)

Rödenbeck, C., Zaehle, S., Keeling, R. F., & Heimann, M. (2018). History of El Niño impacts on the global carbon cycle 1957–2017: A quantification from atmospheric CO_2 data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1760), 20170303. doi:10/gf4j8j. (Cit. on p. 82)

Rodgers, K. B., Lin, J., & Frölicher, T. L. (2015). Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences*, *12*(11), 3301–3320. doi:10/f7gf94. (Cit. on p. 44)

Rogelj, J., Schaeffer, M., Friedlingstein, P., Gillett, N. P., van Vuuren, D. P., Riahi, K., … Knutti, R. (2016a). Differences between carbon budget estimates unravelled. *Nature Climate Change*, *6*(3), 245–252. doi:10/f8c6tq. (Cit. on p. 6)

Rogelj, J., den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., … Meinshausen, M. (2016b). Paris Agreement climate proposals need a boost to keep warming well below 2°C. *Nature*, *534*(7609), 631–639. doi:10/b4db. (Cit. on pp. 5, 18, 34, 39, 43)

Rogelj, J., Huppmann, D., Krey, V., Riahi, K., Clarke, L., Gidden, M., … Meinshausen, M. (2019). A new scenario logic for the Paris Agreement long-term temperature goal. *Nature*, *573*(7774), 357–363. doi:10/gf8g3m. (Cit. on p. 35)

Saito, M., Ito, A., & Maksyutov, S. (2011). Evaluation of Biases in JRA-25/JCDAS Precipitation and Their Impact on the Global Terrestrial Carbon Balance. *Journal of Climate*, *24*(15), 4109–4125. doi:10/fk6wrg. (Cit. on pp. 32, 111)

Samset, B. H., Fuglestvedt, J. S., & Lund, M. T. (2020). Delayed emergence of a global temperature response after emission mitiga-

tion. *Nature Communications*, *11*(1), 3261. doi:10/gg38gg. (Cit. on pp. 33, 34, 39, 40, 51)

Sarmiento, J. L., & Gruber, N. (2006). *Ocean biogeochemical dynamics*. Princeton, NJ: Princeton Univ. Press. Retrieved from https://press.princeton.edu/books/hardcover/9780691017075/ocean-biogeochemical-dynamics. (Cit. on p. 2)

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*(1). doi:10/gfk2tn. (Cit. on p. 15)

Schimel, D. S. (1995). Terrestrial ecosystems and the carbon cycle. *Global Change Biology*, *1*(1), 77–91. doi:10/dw5kbg. (Cit. on pp. 3, 4)

Schlesinger, M. E., & Ramankutty, N. (1994). An oscillation in the global climate system of period 65–70 years. *Nature*, *367*(6465), 723–726. doi:10/bsvctn. (Cit. on p. 4)

Schneck, R., Reick, C. H., & Raddatz, T. (2013). Land contribution to natural CO_2 variability on time scales of centuries. *Journal of Advances in Modeling Earth Systems*, *5*(2), 354–365. doi:10/f5dzt5. (Cit. on pp. 44, 65, 88)

Schneider, T., & Griffies, S. M. (1999). A Conceptual Framework for Predictability Studies. *Journal of Climate*, *12*(10), 3133–3155. doi:10/cf6zsg. (Cit. on p. 86)

Schulzweida, U. (2019). CDO: Climate Data Operators. Max Planck Institute for Meteorology. Hamburg. Retrieved from https://doi.org/10.5281/zenodo.2558193. (Cit. on pp. 75, 123)

Schuur, E. a. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., … Vonk, J. E. (2015). Climate change and the permafrost carbon feedback. *Nature*, *520*(7546), 171–179. doi:10/gb2fh4. (Cit. on p. 34)

Schwalm, C. R., Glendon, S., & Duffy, P. B. (2020). RCP8.5 tracks cumulative CO_2 emissions. *Proceedings of the National Academy of Sciences*. doi:10/gg6xw2. (Cit. on p. 45)

Schwartzman, A., & Keeling, R. F. (2020). Achieving atmospheric verification of CO_2 emissions. *Nature Climate Change*, *10*(5), 416–417. doi:10/ggwp73. (Cit. on pp. 7, 17, 28, 30, 31, 34, 39, 40, 50, 51)

Séférian, R., Bopp, L., Gehlen, M., Swingedouw, D., Mignot, J., Guilyardi, E., & Servonnat, J. (2014). Multiyear predictability of tropical marine productivity. *Proceedings of the National Academy of Sciences*, *111*(32), 11646–11651. doi:10/f6cgs3. (Cit. on p. 110)

Séférian, R., Berthet, S., & Chevallier, M. (2018). Assessing the Decadal Predictability of Land and Ocean Carbon Uptake. *Geophysical Research Letters*. doi:10/gdb424. (Cit. on pp. 12, 15, 22, 23, 25, 29, 64, 67, 70, 73, 78, 79, 87, 123)

Servonnat, J., Mignot, J., Guilyardi, E., Swingedouw, D., Séférian, R., & Labetoulle, S. (2015). Reconstructing the subsurface ocean

decadal variability using surface nudging in a perfect model framework. *Climate Dynamics*, *44*(1-2), 315–338. doi:10/f6v7kq. (Cit. on pp. 25, 87, 89)

Silver, N. (2012). *The signal and the noise: Why so many predictions fail–but some don't*. New York: Penguin Press. Retrieved from https://link.springer.com/article/10.1057%2Fbe.2012.33. (Cit. on p. 2)

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. doi:10/d36fnf. (Cit. on p. 10)

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., & Schmidt, A. (2016). Emergence of healing in the Antarctic ozone layer. *Science*, *353*(6296), 269–274. doi:10/bkn5. (Cit. on p. 31)

Spring, A., & Ilyina, T. (2020). Predictability Horizons in the Global Carbon Cycle Inferred From a Perfect-Model Framework. *Geophysical Research Letters*, *47*(9), e2019GL085311. doi:10/ggtbv2. (Cit. on pp. vii, 1, 11, 14, 15, 21, 23, 25, 29, 31, 45, 51, 53, 61, 86, 91, 92, 97, 106, 107, 123, 125)

Spring, A., Ilyina, T., & Marotzke, J. (2020). Inherent uncertainty disguises attribution of reduced atmospheric CO_2 growth to CO_2 emission reductions for up to a decade. *Environmental Research Letters*, *15*(11), 114058. doi:10/ghmjp8. (Cit. on pp. vii, 1, 7, 14, 19, 28, 31, 37, 86, 97)

Spring, A., Dunkl, I., Li, H., Brovkin, V., & Ilyina, T. (2021). Trivial improvements of predictive skill due to direct reconstruction of the global carbon cycle. *Earth System Dynamics Discussions*, 1–36. doi:10/gh3tn3. (Cit. on pp. vii, 1, 12, 26, 29, 32, 83)

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., … Roeckner, E. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*(2), 146–172. doi:10/f5dz5h. (Cit. on pp. 65, 89)

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., … Midgley, P. (Eds.). (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. Retrieved from https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_all_final.pdf. (Cit. on p. 2)

Stolpe, M. B., Medhaug, I., Sedláček, J., & Knutti, R. (2018). Multidecadal Variability in Global Surface Temperatures Related to the Atlantic Meridional Overturning Circulation. *Journal of Climate*, *31*(7), 2889–2906. doi:10/gdfdtr. (Cit. on p. 44)

Suarez-Gutierrez, L., Li, C., Müller, W. A., & Marotzke, J. (2018). Internal variability in European summer temperatures at 1.5°C and 2°C of global warming. *Environmental Research Letters*, *13*(6), 064026. doi:10/gf7792. (Cit. on p. 43)

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2011). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. doi:10/b7xgw2. (Cit. on pp. 11, 44)

Tebaldi, C., & Friedlingstein, P. (2013). Delayed detection of climate mitigation benefits due to climate inertia and variability. *Proceedings of the National Academy of Sciences*, 201300005. doi:10/n7p. (Cit. on pp. 6, 17, 28, 35, 39, 40, 49, 50)

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., … Edmonds, J. A. (2011). RCP4.5: A pathway for stabilization of radiative forcing by 2100. *Climatic Change*, *109*(1), 77. doi:10/bsxpcv. (Cit. on pp. 18, 43, 55)

Tietsche, S., Notz, D., Jungclaus, J. H., & Marotzke, J. (2013). Predictability of large interannual Arctic sea-ice anomalies. *Climate Dynamics*, *41*(9), 2511–2526. doi:10/f5f34j. (Cit. on p. 31)

Tjiputra, J. F., Schwinger, J., Bentsen, M., Morée, A. L., Gao, S., Bethke, I., … Schulz, M. (2020). Ocean biogeochemistry in the Norwegian Earth System Model version 2 (NorESM2). *Geoscientific Model Development*, *13*(5), 2393–2431. doi:10/ghz2t2. (Cit. on p. 33)

Toggweiler, J. R., Dixon, K., & Bryan, K. (1989). Simulations of radiocarbon in a coarse-resolution world ocean model: 1. Steady state prebomb distributions. *Journal of Geophysical Research: Oceans*, *94*(C6), 8217–8242. doi:10/ffvkfj. (Cit. on pp. 12, 65, 87)

Tohjima, Y., Patra, P. K., Niwa, Y., Mukai, H., Sasakawa, M., & Machida, T. (2020). Detection of fossil-fuel CO_2 plummet in China due to COVID-19 by observation at Hateruma. *Scientific Reports*, *10*(1), 18688. doi:10/ghhn4c. (Cit. on p. 10)

Tokarska, K. B., Arora, V. K., Gillett, N. P., Lehner, F., Rogelj, J., Schleussner, C.-F., … Knutti, R. (2020). Uncertainty in carbon budget estimates due to internal climate variability. *Environmental Research Letters*, *15*(10), 104064. doi:10/ghvcdj. (Cit. on p. 5)

Tommasi, D., Stock, C. A., Hobday, A. J., Methot, R., Kaplan, I. C., Eveson, J. P., … Werner, F. E. (2017a). Managing living marine resources in a dynamic environment: The role of seasonal to decadal climate forecasts. *Progress in Oceanography*, *152*, 15–49. doi:10/f96mg3. (Cit. on p. 32)

Tommasi, D., Stock, C. A., Alexander, M. A., Yang, X., Rosati, A., & Vecchi, G. A. (2017b). Multi-Annual Climate Predictions for Fisheries: An Assessment of Skill of Sea Surface Temperature Forecasts for Large Marine Ecosystems. *Frontiers in Marine Science*, *4*. doi:10/ggrwsk. (Cit. on p. 32)

Turetsky, M. R., Abbott, B. W., Jones, M. C., Anthony, K. W., Olefeldt, D., Schuur, E. A. G., … McGuire, A. D. (2020). Carbon release

through abrupt permafrost thaw. *Nature Geoscience*, *13*(2), 138–143. doi:10/ggj78x. (Cit. on p. 34)

Turnbull, J. C., Miller, J. B., Lehman, S. J., Tans, P. P., Sparks, R. J., & Southon, J. (2006). Comparison of $^{14}CO_2$, CO, and SF_6 as tracers for recently added fossil fuel CO_2 in the atmosphere and implications for biological CO_2 exchange. *Geophysical Research Letters*, *33*(1). doi:10/d98355. (Cit. on p. 33)

UNFCCC. (2015). Adoption of the Paris Agreement. In *United Nations Framework Convention on Climate Change*, Paris: United Nations. Retrieved August 12, 2018, from https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf. (Cit. on pp. 5, 17, 34, 39, 50, 51, 63, 74)

VERIFY. (2018). VERIFY - Verifying Greenhouse Gas Emissions, European Commission H2020. Retrieved February 2, 2020, from https://verify.lsce.ipsl.fr. (Cit. on p. 34)

Wang, J., Zeng, N., Wang, M., Jiang, F., Chen, J., Friedlingstein, P., … Wiltshire, A. J. (2018). Contrasting interannual atmospheric CO_2 variabilities and their terrestrial mechanisms for two types of El Niños. *Atmospheric Chemistry and Physics*, *18*(14), 10333–10345. doi:10/gdx3f2. (Cit. on p. 32)

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd ed). International Geophysics Series. Amsterdam; Boston: Academic Press. Retrieved from http://eamcweb4.usfs.msu.edu/study/2016-GCM/Statistical_Methods_Wilks.pdf. (Cit. on pp. 66, 92)

Winguth, A. M. E., Heinmann, M., Kurz, K. D., Maier-Reimer, E., Mikolajewicz, U., & Segschneider, J. (1994). El Niño-Southern Oscillation related fluctuations of the marine carbon cycle. *Global Biogeochemical Cycles*, *8*(1), 39–63. doi:10/ds7tc3. (Cit. on p. 13)

Winkler, A. (2020). *The response of global terrestrial photosynthesis to rising CO_2* (Doctoral dissertation, Universität Hamburg Hamburg). doi:10.17617/2.3190373. (Cit. on p. 3)

Yeager, S. G., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., … Lovenduski, N. S. (2018). Predicting near-term changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bulletin of the American Meteorological Society*. doi:10/gddfcs. (Cit. on pp. 23, 63, 73)

Zeng, N., Mariotti, A., & Wetzel, P. (2005). Terrestrial mechanisms of interannual CO_2 variability. *Global Biogeochemical Cycles*, *19*(1). doi:10/bjr69v. (Cit. on pp. 39, 47, 64, 71)

Zeng, N., Yoon, J.-H., Vintzileos, A., Collatz, G. J., Kalnay, E., Mariotti, A., … Lord, S. (2008). Dynamical prediction of terrestrial ecosystems and the global carbon cycle: A 25-year hindcast experiment. *Global Biogeochemical Cycles*, *22*(4). doi:10/cf445j. (Cit. on pp. 22, 64, 70, 71)

Zhang, L., Delworth, T. L., & Jia, L. (2017a). Diagnosis of Decadal Predictability of Southern Ocean Sea Surface Temperature in the GFDL CM2.1 Model. *Journal of Climate*, *30*(16), 6309–6328. doi:10/gbq6fz. (Cit. on p. 70)

Zhang, L., Delworth, T. L., Yang, X., Gudgel, R. G., Jia, L., Vecchi, G. A., & Zeng, F. (2017b). Estimating Decadal Predictability for the Southern Ocean Using the GFDL CM2.1 Model. *Journal of Climate*, *30*(14), 5187–5203. doi:10/gbktjk. (Cit. on p. 70)

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. (2007). System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies. *Monthly Weather Review*, *135*(10), 3541–3564. doi:10/dkvk78. (Cit. on pp. 109–111)

Zhu, J., & Kumar, A. (2018). Influence of surface nudging on climatological mean and ENSO feedbacks in a coupled model. *Climate Dynamics*, *50*(1), 571–586. doi:10/gcwrz4. (Cit. on pp. 87, 94, 109)

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., … Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*(1), 5. doi:10/brtf7r. (Cit. on p. 42)

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., … Sheffield, J. (2016). LS3MIP (v1.0) contribution to CMIP6: The Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome. *Geoscientific Model Development*, *9*(8), 2809–2832. doi:10/gcc538. (Cit. on p. 90)

ACKNOWLEDGMENTS

My PhD journey over three years was not a performance by an individual in isolation, but involved invaluable support and care of many people.

I am grateful for being supervised Tatiana, who nudged me on the topic of carbon cycle predictability, guided and supported me in the way I filled research gaps, promoting the topic and my work internally at the institute and externally at workshops and meetings.

I am thankful for my panel for the continuous supervision also outside of panel meetings: Inga, as my panel chair with a fresh and external perspective, and Peter for quickly reading so many abstracts as well as carbon variability and career advice.

I appreciate the contributions from co-authors in the presented papers: Jochem, for encouraging me to pursue the mitigation causation research question immediately and sharpening my writing skills; Hongmei, as my carbon variability and predictability sparring partner since my master's thesis; and Victor and István for helping an eager student to understand the land carbon cycle. I thank the proof-readers of this thesis Leo, Rike, Peter, Fabrice and Dan for valuable feedback and Yvonne Schrader for designing the cover picture. I am thankful for how the marine Biogeochemistry HAMOCC group adopted me as a physics master's student since 2016.

This whole journey would not have been possible without Ingo Harms forwarding my initial email inquiring about climate modeling in Hamburg to Johanna Baehr, which led me to a volunteering week in Hamburg in 2015, which eventually broad me to Tatiana. Thank you for giving me this opportunity. Since 2018 in the IMPRS program, Antje, Connie and Michaela facilitated all bureaucracy allowing me to focus on writing and coding.

I am indebted to the XARRAY and PANGEO community on GITHUB, which gave me the toolbox of scientific computation and software carpentry to work with code efficiently. I am happy for the coding journey with Riley, which led into CLIMPRED. Thanks Ray for initiating and maintaining XSKILLSCORE. Writing and layouting this thesis is based on Andre Miede's CLASSIC-THESIS template.

I am glad to have met and spent time with new friends, mainly Andreas, Fabrice, Sebastian and Clara and my cubicle neighbor Arjun.

I am grateful for my parents, letting me do and study whatever in whichever way I wanted. I am happy to have found best friend and life partner, Alisha.

VERSICHERUNG AN EIDES STATT - DECLARATION OF OATH

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel: „Internal Variability and Potential Predictability of the Global Carbon Cycle in a Perfect-Model Framework“ selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Dissertation oder Teile davon vorher weder im In- noch im Ausland in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Hamburg, Februar 2021

Aaron Spring

Hinweis / Reference

Die gesamten Veröffentlichungen in der Publikationsreihe des MPI-M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

sind über die Internetseiten des Max-Planck-Instituts für Meteorologie erhältlich:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html

All the publications in the series of the MPI -M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

are available on the website of the Max Planck Institute for Meteorology:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html