Report on the benchmarking and evaluation of simulated carbon budgets

Deliverable 1.9

Authors: Yuan Zhang, Philippe Peylin, Philippe Ciais, Laibao Liu, Sonia Seneviratne, Sönke Zaehle, Nicolas Mayot, Corrine LeQuere, Peter Llandschutzer, Camille Abadie, Cedric Bacour, Stephen Sitch, Nicolas Gruber

This project received funding from the Horizon 2020 programme under the grant agreement No. 821003.

Document Information

GRANT AGREEMENT	821003
PROJECT TITLE	Climate Carbon Interactions in the Current Century
PROJECT ACRONYM	4C
PROJECT START DATE	1/6/2019
RELATED WORK PACKAGE	WP1
RELATED TASK(S)	**T1.4.1 and T1.4.2**
LEAD ORGANIZATION	CEA
AUTHORS	Yuan Zhang, Philippe Peylin, Philippe Ciais, Laibao Liu, Sonia Seneviratne, Sönke Zaehle, Annelen Küchenmeister, Nicolas Mayot, Corine LeQuere, Peter Landschutzer, Camille Abadie, Cedric Bacour, Stephen Sitch, Mike O Sullivan, Nicolas Gruber
SUBMISSION DATE	31.01.2023
DISSEMINATION LEVEL	PU

History

DATE	SUBMITTED BY	REVIEWED BY	VISION (NOTES)
28/02/2023	P. Peylin		
01/03/2023		P. Friedlingstein	

Please cite this report as:. Y. Zhang, P. Peylin, P. Ciais, L. Liu, S. Seneviratne, S. Zaehle, A. Küchenmeister, N. Mayot, C. LeQuere, P. Llandschutzer, C. Abadie, C. Bacour, S. Sitch, N. Gruber (2020), Draft: Report on the benchmarking and evaluation of simulated carbon budgets D3.1 of the 4C project.

Table of Contents

1 Introduction 5

2 Evaluation of land & ocean carbon cycle models against established benchmark 5

2.1 Land models evaluation: LAI, evapotranspiration, GPP, biomass and soil C stocks 5

2.2 Ocean models evaluation: air-sea CO2 flux 6

3 Evaluation of land & ocean carbon cycle models against new observations 9

3.1 Net carbon budgets evaluated against O2 and APO global budget 9

3.2 Carbon budget models evaluated against 13C observations 10

3.3 Terrestrial water storage variability to constrain land carbon flux Interannual variability 11

3.4 Land GPP evaluation with atmospheric COS and Solar Induced Fluorescence data 15

3.4.1 Large scale atmospheric COS constraint on GPP 15

3.4.2 Satellite SIF data to constrain GPP 20

3.5 Land biomass evaluation with vegetation optical depth (L-VOD) satellite observations 22

3.5.1 Regional total live biomass changes 23

3.5.2 Evaluation of DGVMs used for the global carbon budget 24

3.6 Evaluation of land models against satellite based XCO2 data (MPI-B) 27

3.7 Ocean model evaluation with ocean interior data 30

4 Conclusions and Outlook 33

5 References 33

About 4C

Climate-Carbon Interactions in the Coming Century (4C) is an EU-funded H2020 project that addresses the crucial knowledge gap in the climate sensitivity to carbon dioxide emissions, by reducing the uncertainty in our quantitative understanding of carbon-climate interactions and feedback. This will be achieved through innovative integration of models and observations, providing new constraints on modeled carbon-climate interactions and climate projections, and supporting Intergovernmental Panel on Climate Change (IPCC) assessments and policy objectives.

Executive Summary

The first objective of Task 1.4 is to evaluate land and ocean models (mainly process-based models contributing to the global carbon budget or being used in earth system model simulations (i.e., CMIP6)) using established observations and benchmarks. According to these standard benchmarks, the performances of each land surface model vary between ecosystem processes and benchmark, but none of them performs well for all processes. Nevertheless, all land models capture the spatial patterns of above ground biomass and vegetation primary production relatively well. For the ocean models, the interannual variability of the air–sea CO_2 flux is strongly underestimated, which is mainly contributed by the southern extratropical oceans.

Besides existing observations, several new benchmarks have been developed, partly within the 4C project, to further evaluate land and ocean models. The benchmark based on O_2 and Atmospheric Potential Oxygen (APO) global budget shows that the current ocean models can reproduce the short-term interannual variation of southern ocean CO_2 sink, however missing the decadal variations. According to the observations of dissolved inorganic carbon, current ocean models also underestimated the cumulative ocean carbon storage in recent decades. For the evaluation of land models, the 4C project developed various benchmarks. Based on the stable carbon isotope ratio of atmospheric CO_2, we found a too strong decrease in the seasonal cycle of the ^{13}C ratio of atmospheric CO_2, suggesting a too weak influence of climate and land-use change on the seasonal amplitude of the net land carbon fluxes (using only one model). The terrestrial water storage variability indicates a too weak sensitivity of simulated tropical net land carbon fluxes. Atmospheric Carbonyl sulfide (COS) and Solar Induced Fluorescence (SIF) observations reveal contrasting model skills in terms of mean seasonal cycle of the Gross Primary Production (GPP), highlighting clear large-scale biases for some models (TRENDY collection and CMIP6 ensemble). The L-band vegetation optical depth derived from satellite observations indicate wrong biomass changes and biomass spatial distribution in current land models. Satellite XCO2 observations further highlight that some models (TRENDY collection) fail to correctly reproduce the interannual and seasonal variability of net land carbon uptake.

Overall, the work of Task 1.4 provides valuable insights in the strengths and deficiencies of current land and ocean models, highlighting possible directions of model improvements.

Keywords

Land carbon-cycle, Ocean carbon-cycle, soil moisture, soil carbon, ocean acidification

1 Introduction

This report consists of a comprehensive benchmarking and evaluation of simulated carbon budgets over land and ocean, from process-based land and ocean models, respectively. The main objective is to assess the strengths and weaknesses of coupled model simulation (CMIP6) as well as the latest land/ocean model versions (post CMIP6) that are run in an offline mode (i.e., with prescribed atmospheric forcing). To that end, we have thus used the standard output of the CMIP6 (and also CMIP5) model runs as well as the simulations performed in the context of the global carbon budget (GCB) performed each year (Friedlingstein et al., 2022).

The evaluation is divided into two sections. In the first section, we are using established observations and benchmarks to get a first order evaluation of all participating land and ocean models. In the second section, we use new/recent observations, including data-driven products, that were derived at least partly within the 4C project.

2 Evaluation of land & ocean carbon cycle models against established benchmark

2.1 Land models evaluation: LAI, evapotranspiration, GPP, biomass and soil C stocks

We conducted an initial benchmarking analysis of output from land carbon models used in GCB against global observation data sets using the ILAMB tool (Figure 2.1.1). This has so far focussed on four key main variables (biomass, land cover distribution, soil moisture, burnt area) that are provided at a global scale by ESA-CCI program. These model simulations correspond to the latest ensemble of runs done with the so-called TRENDY-V11 protocol (the latest one, performed during the summer 2022).

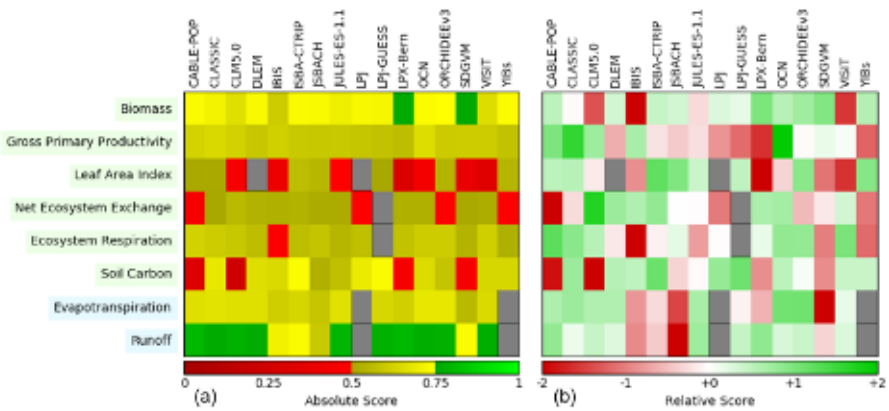

Figure 2.1.1 Carbon (green) and Water (blue) cycle variables benchmarked for the 16 DGVMs (including those from 4C) in GCB2022 (Friedlingstein et al. 2022). A) absolute score b) relative score across DGVMs.

The multi-model ensemble of DGVMs (MME-DGVM) used in GCB 2022 compare favorably against the ESA CCI biomass for the spatial pattern across the key global biomes (Figure 2.1.2), with the highest biomass estimates for the tropical regions (above 20 kgC m^{-2}) and more moderate in the temperate and boreal forests.

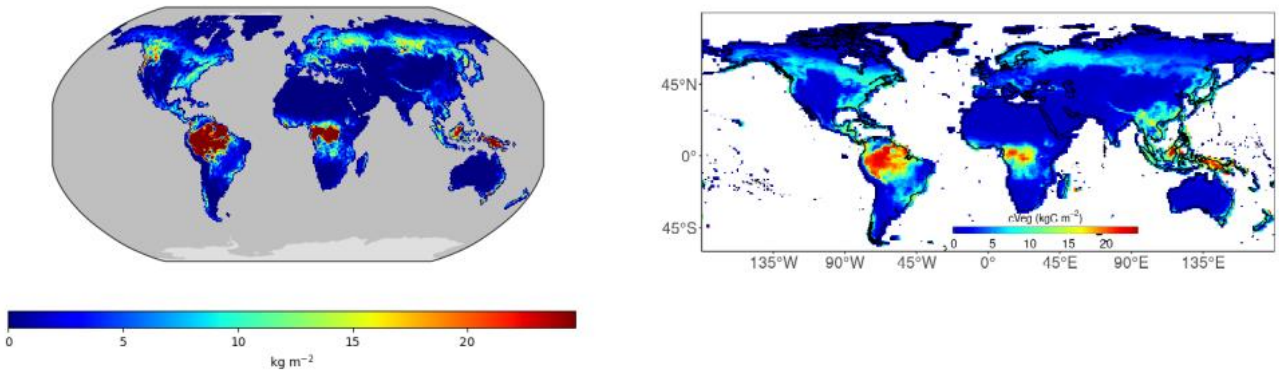

Figure 2.1.2 Simulated biomass from 16 GCB2022 DGVMs compared against the ESA CCI biomass product for 2010-11. Units are kgC m^{-2}.

The MME-DGVM was able to reproduce the observation-based spatial pattern of annual Gross Primary Productivity, GPP (Figure 2.1.3). Once again key features are captured, e.g. the highest GPP in the wet-tropics (above 8 gC m^{-2} day^{-1}) and more moderate in the temperate and boreal ecosystems.

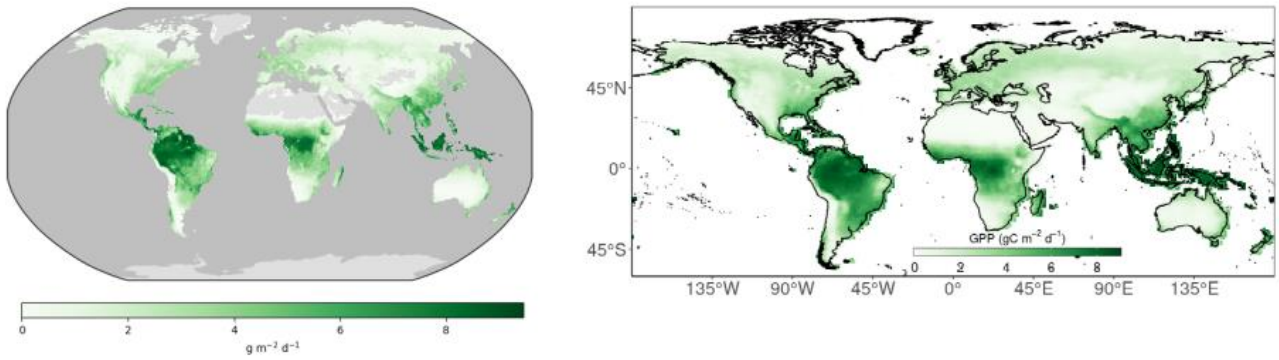

Figure 2.1.3 Simulated annual GPP from 16 GCB2022 DGVMs averaged over the period 1980-2014, compared against respective estimates derived from FLUXCOM. Units: gC m^{-2} day^{-1}.

2.2 Ocean models evaluation: air-sea CO_2 flux

The ocean models (as well as the data products) used within the Global Carbon Budget are evaluated with two metrics: the root-mean-squared-error (RMSE) between observed and simulated fCO_2, and the amplitude of the interannual variability of the flux (A-IAV) (Figure 2.2.1). The observed fCO_2 are from the SOCAT database

(Bakker et al. 2016). The RMSE is calculated from detrended, annually, and regionally averaged time series calculated from ocean models subsampled to SOCAT sampling points, in details: (i) subsample data points for which there are observations (ocean models and SOCAT), (ii) average spatially, (iii) calculate annual mean, (iv) detrend both time series (ocean models and SOCAT), and (v) calculate RMSE. The amplitude of the interannual variability is calculated as the temporal standard deviation of the detrended annual CO_2 flux time series. It is a direct measure of the variability of the oceanic CO_2 sink on interannual timescales.

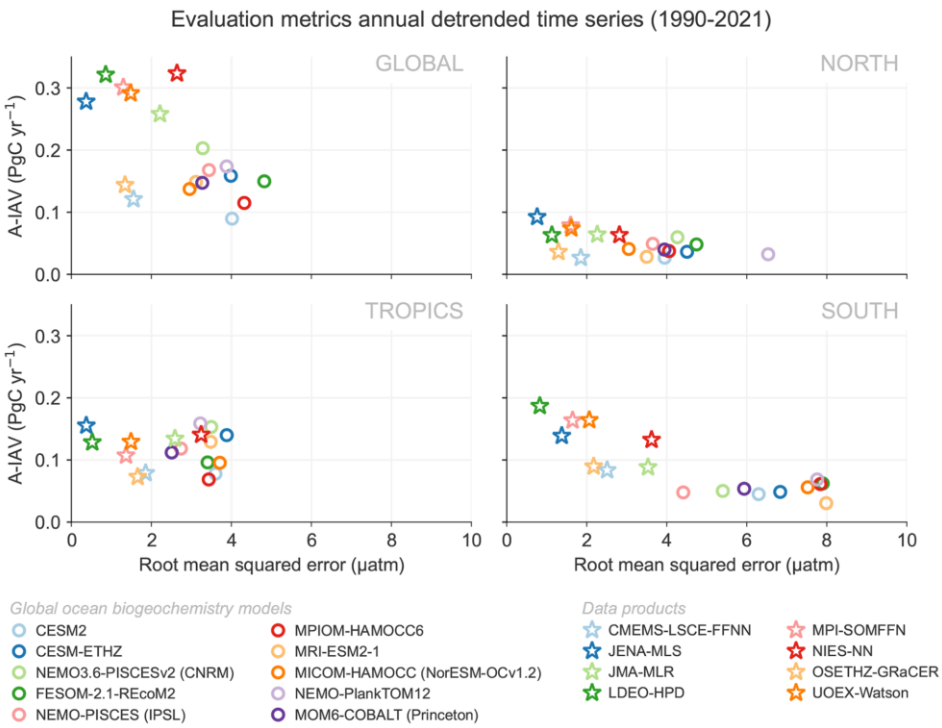

Figure 2.2.1 Evaluation of the ocean models and data products used in the Global Carbon Budget. The x axis shows the root-mean-squared error (RMSE) for the period 1990 to 2021 between the individual surface ocean fCO_2 mapping schemes and the SOCAT. The y axis shows the amplitude of the interannual variability of the air–sea CO_2 flux (A-IAV), taken as the standard deviation of the detrended annual time series. Results are presented for the globe, northern extratropics (> 30°N), tropics (30°S-30°N), and southern extratropics (< 30°S) for the ocean models (see legend, circles) and for the fCO_2-based data products (star symbols).

Over the globe, ocean models have a larger RMSE (3.0 to 4.8 µatm) than the data products (0.4 to 2.6 µatm) (Figure 2.2.1). However, the data products are based on the SOCAT database; hence, the SOCAT is not an independent dataset for the evaluation of the data products. For the ocean models, higher RMSEs occur in high latitude regions where stronger climate variability is observed. The interannual variability of the oceanic CO_2 flux is generally two times higher in the data products than in the ocean models. This difference in the amplitude is mostly driven by variations in the Southern Ocean (Figure 2.2.1).

As part of the 4C project and the RECCAP2 effort, an evaluation of the ocean models and data products was performed by focusing on the temporal variations of the Southern Ocean CO_2 sink (Mayot et al., in revision).

The interannual variations of this CO_2 sink could be decomposed into two frequencies: a short-term interannual component and a decadal component. These two frequencies should be clearly separated before being studied, because the environmental processes influencing the year-to-year changes might be different from the ones related to decadal variations.

The short-term interannual variations of the Southern Ocean CO_2 sink derived from data products and ocean models were consistent with each other. This result is consistent with other recent findings that, at global scale, ocean models can represent some of the interannual variabilities (e.g., Bennington et al., 2022; Mayot et al., in prep.). Therefore, we propose to improve the assessment of ocean models and data products in the Global Carbon Budget by decomposing the amplitude of the oceanic CO_2 sink variations into its short-term and decadal components (Figure 2.2.2). Such added information demonstrates that most discrepancies between ocean models and data products, and uncertainties in the oceanic CO_2 sink, are associated with the decadal variations of the oceanic CO_2 sink and not its year-to-year variations.

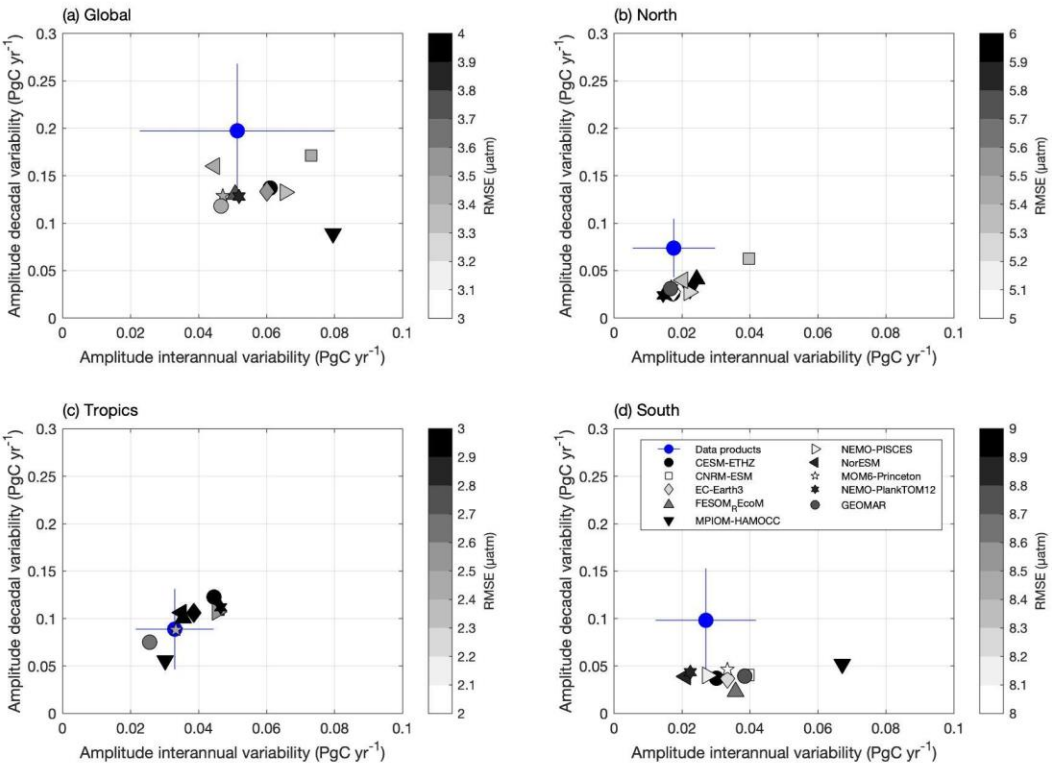

Figure 2.2.2: New evaluation of the ocean models for the period 1990 to 2018. The x axis shows the amplitude of the interannual variability of the air-sea CO_2 flux, and the y axis shows the amplitude of the decadal variability, both taken as the standard deviation of the temporarily decomposed monthly time series of air-sea CO_2 flux. The color scale shows the root-mean-squared error (RMSE) between the ocean models and SOCAT. Results are presented for the globe, northern extratropics (> 30°N), tropics (30°S-30°N), and southern extratropics

(< 30°S) for the ocean models (see legend) and for the data products (blue symbol, average value and standard deviation).

3 Evaluation of land & ocean carbon cycle models against new observations

3.1 Net carbon budgets evaluated against O_2 and APO global budget

The Atmospheric Potential Oxygen (APO) is a tracer that combines atmospheric concentrations of O_2 and CO_2 (i.e., APO = O_2 + 1.1 CO_2) in a manner that is conservative with regard to land biosphere exchanges (-O_2:CO_2 = 1.1, the molar exchange ratio for land uptake). The other processes that influence APO are the combustion of fossil fuels, which consumes O_2 and releases CO_2 (-O_2:CO_2 ≈ 1.4, depending on the fuel type), and air-sea exchanges of O_2 and CO_2. Therefore, APO measurements adjusted for the relatively well-known fossil fuel combustion are records of air-sea fluxes of O_2 and CO_2. It has been demonstrated that variabilities in APO concentrations are largely dominated by the variabilities in air-sea O_2 fluxes. In other terms, the variabilities in APO air-sea fluxes are approximately equal to the variabilities in air-sea exchanges of O_2. Thus, the variabilities in APO air-sea fluxes evaluated by the atmospheric inversion method provide observation-based estimates of the variations in O_2 air-sea fluxes. These estimates can be used as an independent dataset to evaluate the ability of ocean models to correctly simulate the oceanic components that influence air-sea exchanges of O_2, which impact air-sea CO_2 fluxes as well.

Air-sea O_2 fluxes from the ocean models used within the Global Carbon Budget were compared to observation-based estimates of air-sea O_2 fluxes from a global atmospheric inversion method (CarboScope APO inversion, https://www.bgc-jena.mpg.de/CarboScope) that is optimized to track the observed concentrations of APO. This comparison focused on the Southern Ocean, as large differences in the estimate of the oceanic CO_2 sink have been reported in this region between data products and ocean models (see previous section 2.2). We proposed that the degree of concordance between observed and simulated variabilities in air-sea fluxes of CO_2 and O_2, at different time scales, could be used to evaluate the ability of ocean models to simulate the climate-driven variability of the Southern Ocean CO_2 sink (Mayot et al., in revision).

The short-term interannual variations of the Southern Ocean CO_2 sink derived from data products and ocean models were consistent with the short-term interannual variations in air-sea O_2 flux derived from observations (Figure 3.1.1). Specifically, the current generation of ocean models can simulate the influence of stronger (weaker) winds that induce, in the Subpolar Zone, stronger (weaker) upwelling of deep waters and drive the short-term interannual variation of the Southern Ocean CO_2 sink. Regarding the decadal variation of the Southern Ocean CO_2 sink, suggested by several data products, it tends to be confirmed by the observed decadal variation of the air-sea O_2 flux. However, ocean models do not reproduce this decadal variation.

Although the climate-driven processes associated with these decadal variations remain unknown, data products suggest it originates in areas associated with the formation of the subantarctic mode water, a physical process that might be poorly represented in ocean models. However, more in-situ pCO_2 data are required to confirm the influence of these Southern Ocean areas, and other physical and biological components also need to be improved in the current generation of models.

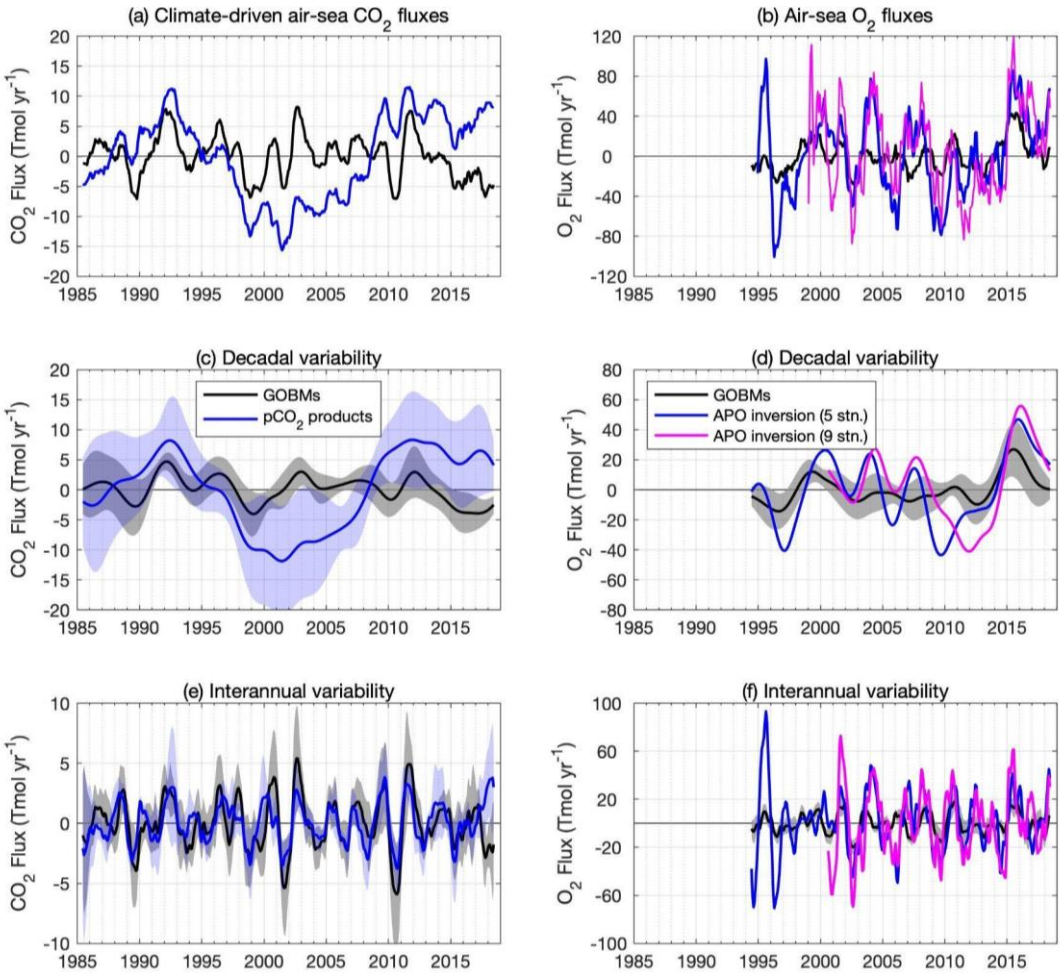

Figure 3.1.1 Comparison of the air-sea CO_2 and O_2 fluxes estimated by ocean models and observation-based products. (a-b) Air-sea CO_2 and O_2 fluxes from ocean models (black) and observation-based products (blue and magenta). By applying moving averages, time series from each model and observation-based product were decomposed into decadal and interannual components. All individual decomposed time series were averaged to provide estimates of (c-d) decadal and (e-f) interannual variability shown with ±1 standard deviation (shading).

3.2 Carbon budget models evaluated against ^{13}C observations

UBern submitted deliverable 1.4 on Isotopic-constrained atmospheric carbon budget and sink flux estimates (T1.1.2) per 31.05.2022 (Lienert, et al., 2022) and key results are summarized here again.

In situ measurements of the seasonal cycle of the stable carbon isotope ratio of atmospheric CO_2, $d^{13}C(CO_2)$, provide complementary information on the seasonality of the global carbon cycle but were thus far not exploited in the context of process-based carbon cycle models. UBern used isotope-enabled simulations of the Bern3D-LPX Earth System Model of Intermediate Complexity and fossil fuel emission estimates together with a model of atmospheric transport (TM3 transport matrices) to simulate local atmospheric $d^{13}C(CO_2)$. Good agreement is found between the measured and simulated seasonal cycle of atmospheric $d^{13}C(CO_2)$ (mean seasonal amplitude mismatch of 0.02 ‰ across 19 sites). Factorial simulations reveal that the seasonal cycle of $d^{13}C(CO_2)$ is primarily driven by land biosphere carbon exchange (Figure 3.2.1). Spatial and temporal fluxes of CO_2 and their signatures are analyzed to quantify the terrestrial drivers. The influence of climate and land-use change on seasonal amplitude is found to be small. Unlike the growth of seasonal amplitude of CO_2, no consistent change in seasonal amplitude of $d^{13}C(CO_2)$ is simulated over the historical period, nor evident in the available observations. We conclude that the seasonal cycle of $d^{13}C(CO_2)$ contains complementary information on the global carbon cycle, and should be employed as a novel atmospheric constraint in models used to project the carbon sink. A manuscript describing these results is in preparation.

UBern applied a Bayesian Monte Carlo Framework to establish the $d^{13}C$-based carbon budget over the industrial period. The $d^{13}C(CO_2)$ budget is well matched over the period of direct atmospheric measurements. However, the model simulates a too strong decrease in $d^{13}C(CO_2)$ compared to the ice core record. We attribute this mismatch, at least partly, to a too rapid expansion of C4 plants in the model under rising atmospheric CO_2.

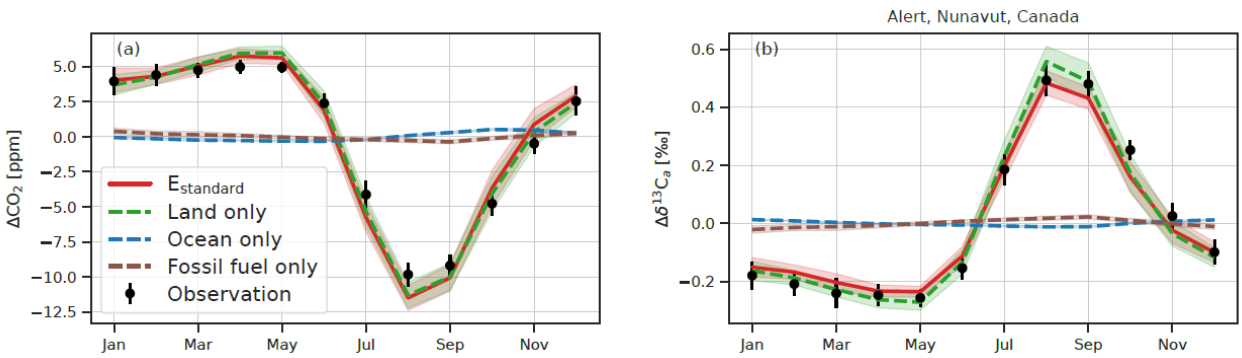

Figure 3.2.1 The seasonal cycle of CO_2 and $d^{13}C(CO_2)$ and contributions by different fluxes at Alert, Canada. The seasonal cycle of CO_2 (a) and its signature $d^{13}C(CO_2)$ (b) simulated by Bern3D-LPX and transported with TM3 (red), compared to observations (black dots). The results of only transporting fluxes of terrestrial (green, dashed), oceanic (blue, dashed), and from fossil sources (brown, dashed) are shown with dashed lines. Error bars and shading correspond to the interannual standard deviation.

3.3 Terrestrial water storage variability to constrain land carbon flux Interannual variability

Since 2002, the twin satellites of the Gravity Recovery and Climate Experiment (GRACE) enable the direct measurement of terrestrial water storage variability, and a subsequent analysis showed that it is tightly coupled to interannual variation (IAV) of atmospheric CO_2 growth rate (CGR) (Humphrey et al., 2018), in addition to the well-documented temperature effects. In the context of climate change, ETHZ further uses 59-year records of global atmospheric CO_2 to investigate the changes in the interannual relationship between tropical land climate conditions and CGR over the past decades (1960-2018). To complement the shorter observational record of the GRACE satellites, we additionally employ recently reconstructed long-term terrestrial water storage variability. Furthermore, the 6-month lagged yearly precipitation (LagP) can approximate aggregated tropical terrestrial water storage IAV well and correlates with CGR IAV, emerging as another efficient proxy for tropical terrestrial water availability IAV. In the end, ETHZ finds that the interannual relationship between tropical terrestrial water availability and CGR has become increasingly negative in the recent past (1989-2018) compared to prior climate conditions (1960-1989) (Figure 3.3.1). This could be related to spatiotemporal changes in tropical water availability anomalies driven by shifts in El Niño-Southern Oscillation teleconnections, including declining spatial compensatory water effects. We also demonstrate that most state-of-the-art coupled Earth System Models and offline Land Surface Models do not reproduce this intensifying water-carbon coupling (Figure 3.3.2). Our results suggest that tropical water availability is very likely increasingly controlling the interannual variability of terrestrial carbon cycle (Liu et al., under revision at Nature).

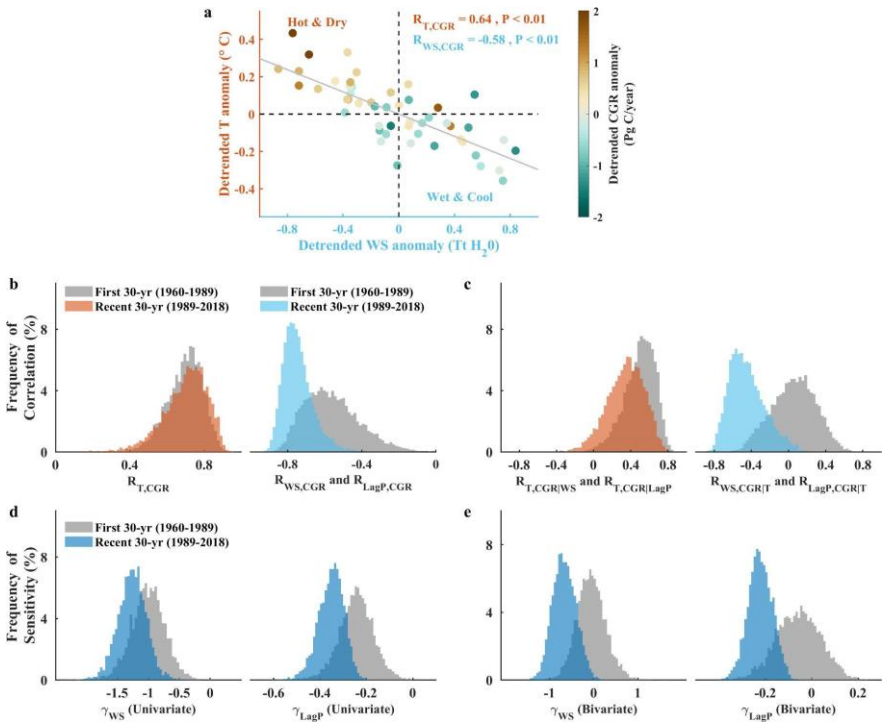

Figure 3.3.1 Tropical land climate-carbon interannual relationships. (a), Yearly tropical temperature versus tropical WS versus CGR in detrended anomalies. The values of CGR are indicated by the color-bar. (b), Histograms of climate-CGR interannual correlations in the first three decades (1960-1989) and in the recent three decades (1989-2018), derived using 5000 bootstrapping repeats. Both tropical water storage (WS) and

6-month lagged-precipitation (LagP) are used to represent tropical water availability. (c), same as (b) but showing histograms of the partial correlations of CGR to tropical temperature and tropical water after controlling tropical water and tropical temperature, respectively. (d), Histograms of the interannual sensitivity of CGR to tropical WS (Y_{WS}) and LagP (Y_{LagP}) in the univariate regression for the same two period, derived using 5000 bootstrapping repeats. Unlike correlations, Y_{WS} and Y_{LagP} differ a lot in magnitude due to the differences in WS and LagP IAV magnitude and are therefore shown separately. The unit of this sensitivity is PgC year^{-1} per Tt H_2O. (e), same as (d) but showing Y_{WS} and Y_{LagP} estimated using the bivariate regression with both tropical water and tropical temperature as predictors. Ridge regression is used here to reduce biases from high collinearity between water and temperature.

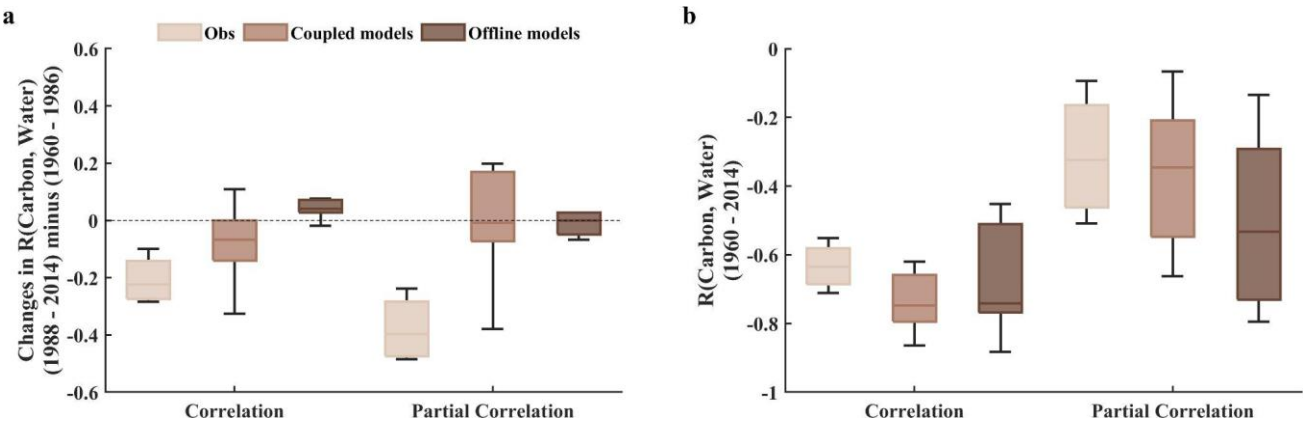

Figure 3.3.2 Historical tropical land water-carbon interannual correlations in observations and models. (a), Changes in interannual correlations and partial correlations between tropical water availability and land carbon fluxes from the prior 27-year (1960-1986) to the recent 27-year (1988-2014). For partial correlations, tropical temperature is controlled. All variables are detrended at yearly scale in each corresponding window. For observations, CGR/Residual land sink, reconstructed tropical WS/tropical LagP and tropical temperature are used for computations (n=4). For models, global net ecosystem exchange, tropical total soil moisture, and tropical temperature from each model are used for computations (coupled models, n=9; offline models, n=6). Box plots show the distribution of estimates for all models. (b), Same as (a), but showing the interannual correlations and partial correlations between tropical water availability and land carbon fluxes during 1960-2014.

In addition, ETHZ utilizes two generations of factorial Earth system model (ESM) experiments to show that the IAV of tropical land carbon uptake under both present and future climate are consistently dominated by soil water variations. The two generations of experiments (GLACE-CMIP5 and GLACE-CMIP6) share a similar protocol to investigate land-atmospheric feedbacks under present and future climate (*Seneviratne et al.*, 2013; *van den Hurk et al.*, 2016), which allows investigating impacts of soil moisture variability on the global carbon cycle. Model experiment results show that suppressing soil water IAV can reduce about 92 ± 8% and 84 ± 12% of present-to-future yearly tropical averaged net biome production (NBP) variances in CMIP5 (1970-2100) and CMIP6 models (1985-2099), respectively (Figure 3.3.3). Nonetheless, the magnitude of interannual sensitivity of tropical NBP to water variations ($Y_{IAV,W}$) under future climate shows a large spread across the latest 16 ESMs (2.3±1.5 PgCyr/Tt H_2O). Since year-to-year variation of tropical land carbon uptake is a relatively “fast” flux process, mainly including photosynthesis and respiration, the sensitivity of tropical NBP to water variations at interannual scale under future climate is expected to be related to that under present climate. Indeed, there is

a significant emergent relationship between $\Upsilon_{IAV,W}$ under future climate and that under present climate across 16 CMIP6 models (R^2=0.76) (Figure 3.3.4). Most importantly, this significant emergent relationship offers an opportunity on constraining $\Upsilon_{IAV,W}$ under a future warmer climate, if combined with observations. Given that only terrestrial water storage (TWS) rather than total soil water is available from observations, and given that not all ESMs have TWS as an output, we sum up all water bodies on land from the model's output (mainly total soil moisture and snow) as the proxy for TWS to keep consistency between models and observations. Combining observed CO_2 growth rate (the proxy for tropical land carbon sink) and TWS during 1995-2014, the resulting emergent constraint (EC) shows that, compared with the unconstrained ensemble of CMIP6 models (2.3 ± 1.5 PgC/yr/Tt H_2O), the mean and spread of future $\Upsilon_{IAV,W}$ are reduced by about 41% and 44%, respectively (1.3 ± 0.8 PgC/yr/Tt H_2O). Nonetheless, this result does not provide a process-based or bottom-up solution for model improvements. Calibrating model process representations is recommended to systematically improve modeled climate-carbon coupling and constrain long-term tropical carbon-climate feedbacks, thereby reducing uncertainty in climate change projections (Liu et al., In preparation).

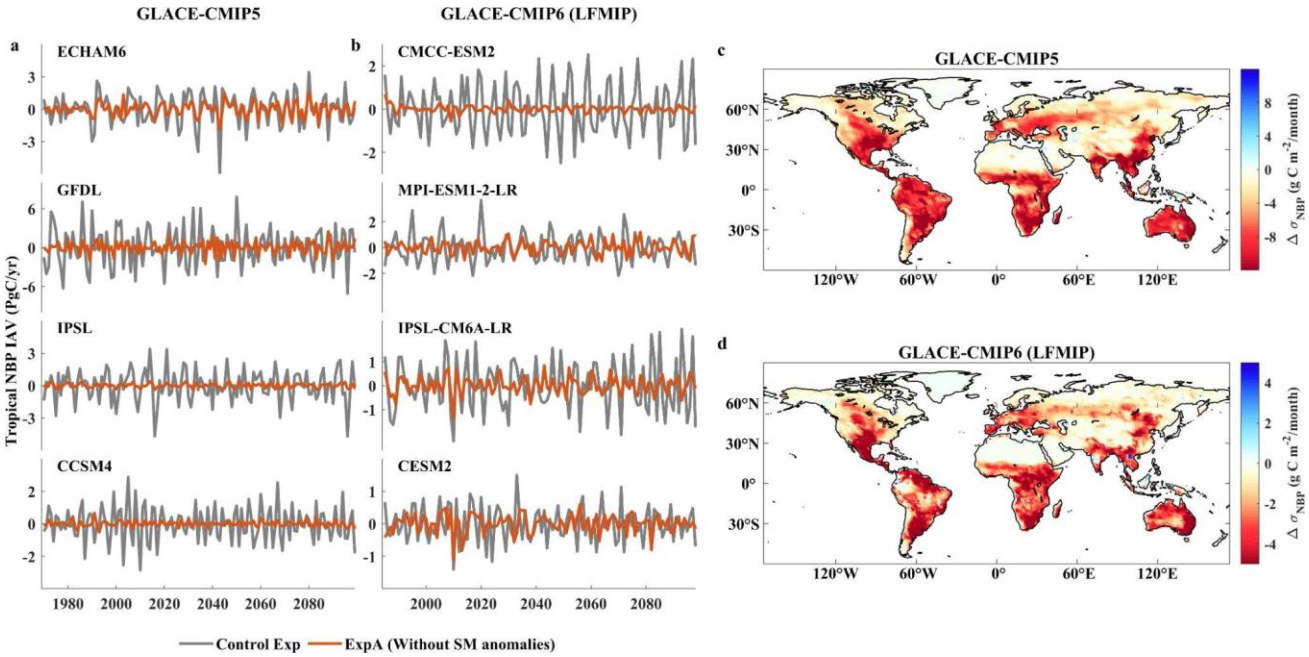

Figure 3.3.3 Present-to-future NBP IAV in model experiments. Detrended year-to-year variations in tropical NBP from present to future in experiments with and without SM anomalies, obtained from (a) GLACE-CMIP5 and (b) GLACE-CMIP6 (LFMIP). Spatial distributions of the reduction of tropical NBP standard deviation caused by suppressing SM anomalies in (c) GLACE-CMIP5 and (d) GLACE-CMIP6 (LFMIP).

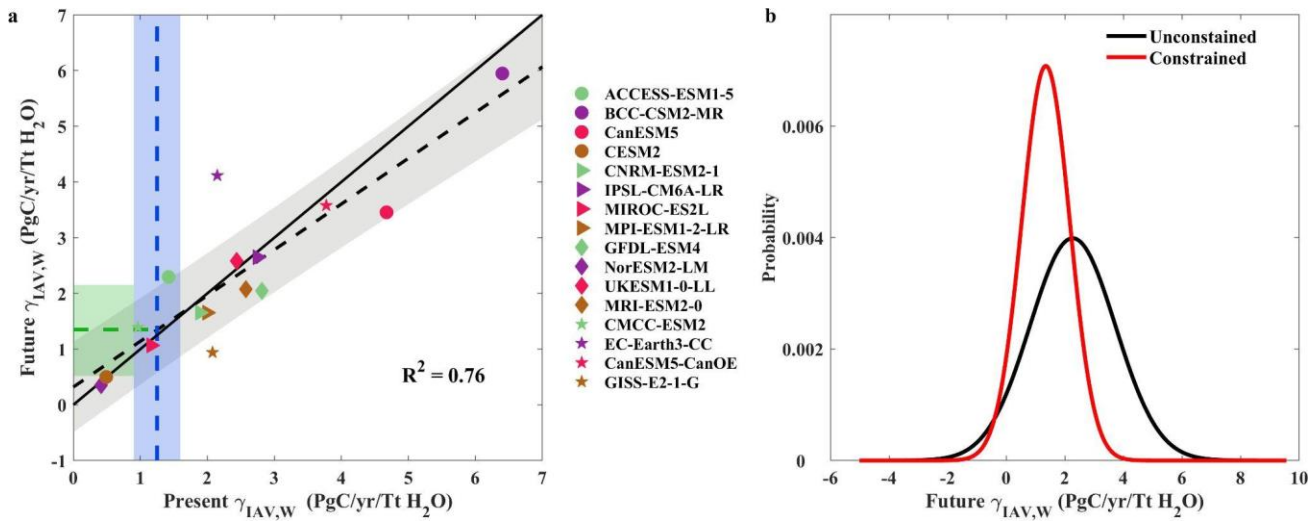

Figure 3.3.4 Emergent constraints on the interannual sensitivity of tropical land carbon to water variations under a future warmer climate in CMIP6. (a) Interannual sensitivity of tropical land carbon uptake to water variations under a future warmer climate (y-axis) versus present climate (x-axis). Each symbol represents a single ESM simulation, the grey bar represents the emergent relationship between the y variable and the x variable, the blue bar represents the observational estimate of the x variable, and the green bar represents the resulting emergent constraint on the y variable. The thicknesses represent ± one standard error. The solid black line indicates the 1:1 line. (b) Probability distributions of unconstrained and constrained interannual sensitivity of tropical land carbon uptake to water variations under a future warmer climate.

3.4 Land GPP evaluation with atmospheric COS and Solar Induced Fluorescence data

3.4.1 Large scale atmospheric COS constraint on GPP

Carbonyl sulfide (COS) has been recognized to be a promising surrogate for tracking the amount of CO_2 that is absorbed by terrestrial vegetation (see recent synthesis by Whealan et al. 2018). Indeed, COS follows the same diffusion pathway into the leaf chloroplasts as CO_2. While absorbed CO_2 following photosynthesis is re-emitted to the atmosphere through ecosystem respiration processes, COS is nearly inversely consumed by the enzyme carbonic anhydrase within the leaves. Therefore, the atmospheric drawdown of COS reflects to a large extent the plant uptake of COS, provided that the non-leaf COS fluxes have a much smaller temporal variability. In this respect, COS acts as a tracer of the CO_2 assimilation by plants (gross primary production, GPP). Launois et al. (2015) using atmospheric measurements of CO_2 and COS and a simple assimilation scheme of COS by terrestrial ecosystems have shown the potential of such an approach. Recently, Hilton et al. (2017) additionally showed that the variability of the COS temporal gradient is mainly driven by variation in GPP rather than other modeled COS flux.

We have used COS and CO_2 data from the National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA/ESRL) to evaluate the GPP and NEE from both CMIP5 and CMIP6 coupled model

simulation as well as from different versions of the TRENDY ensemble simulations (V7 and V10). Note that TRENDY-V7 corresponds to model runs made during the summer 2018 and V10 during the summer 2021. V10 corresponds to a post CMIP6 model development stage.

In order to relate the fluxes of COS and CO_2 to the atmospheric concentrations, we used the LMDZ atmospheric transport model. While CO_2 only provides a constraint on the net ecosystem exchange (NEE), COS data provide more directly a constraint on the GPP. The method has already been detailed in a previous deliverable: D1.7 - "New constraint on land carbon cycle".

We thus now analyze the results of the different ensemble of GPP and NEE simulations. The figure below illustrates the mean seasonal cycle of both COS and CO_2 concentrations at the Mauna Loa station (MLO, representative of the northern hemisphere) derived from the GPP and NEE fluxes of the different simulations. We recall that we transport these fluxes with the LMDZ model, using in addition flux estimates for the other components of the atmospheric COS budget (soil fluxes, air-sea fluxes, anthropogenic emissions, atmospheric sink). The same is done for CO_2 using reference fossil fuel emissions and reference ocean fluxes. From this diagnostic we can draw a first qualitative evaluation of the different model simulations and more specifically of their GPP (Figure 3.4.1):

- The CMIP5 model simulations provide a wide range of COS seasonal amplitudes at MLO with many models having a too large amplitude. This feature is also reflected in the CO_2 concentrations, but the COS diagnostic points toward biases in the model GPP rather than in the ecosystem respiration. This is also true and even enhanced at the Alert Station (ALT, Canada) which points towards model deficiencies for high latitude ecosystems (which are dominantly influencing the ALT station).
- In the most recent CMIP6 coupled simulations, the large biases in terms of amplitude have strongly decreased with most models having seasonal amplitudes comparable to that of the observations. This was mainly achieved with the optimization of the GPP in the coupled models. However, we still have several models that provide a too low amplitude for COS which is also apparent in CO_2 and thus points to too low GPP seasonal amplitude (mainly for high latitude ecosystems when using the same diagnostic at ALT). Note also that the COS drawdown occurs too early during the summer compared to the observations, which could indicate phase issues with the model GPP. However, this may be also partly related to biases in the ocean COS fluxes.
- Using prescribed climate in TRENDY-V7 simulations, allows investigating more specifically the biases linked to GPP and NEE and not linked to potential climate biases in the coupled simulations. The figure illustrates a reasonable agreement in terms of COS amplitude, with a few models having a too large amplitude. The CO_2 diagnostic further highlights that the respiration fluxes may also have some biases (i.e. model having a good COS amplitude but biases in the CO_2 amplitude).
- Using the TRENDY-V10 further illustrates the progress done in terms of model improvement between the year 2018 and 2021 (when the two sets of simulations were done). Although some individual models

have improved (according to the proposed diagnostic) the ensemble of TRENDY runs still show a similar spread with both too large and too low amplitudes induced by GPP biases.

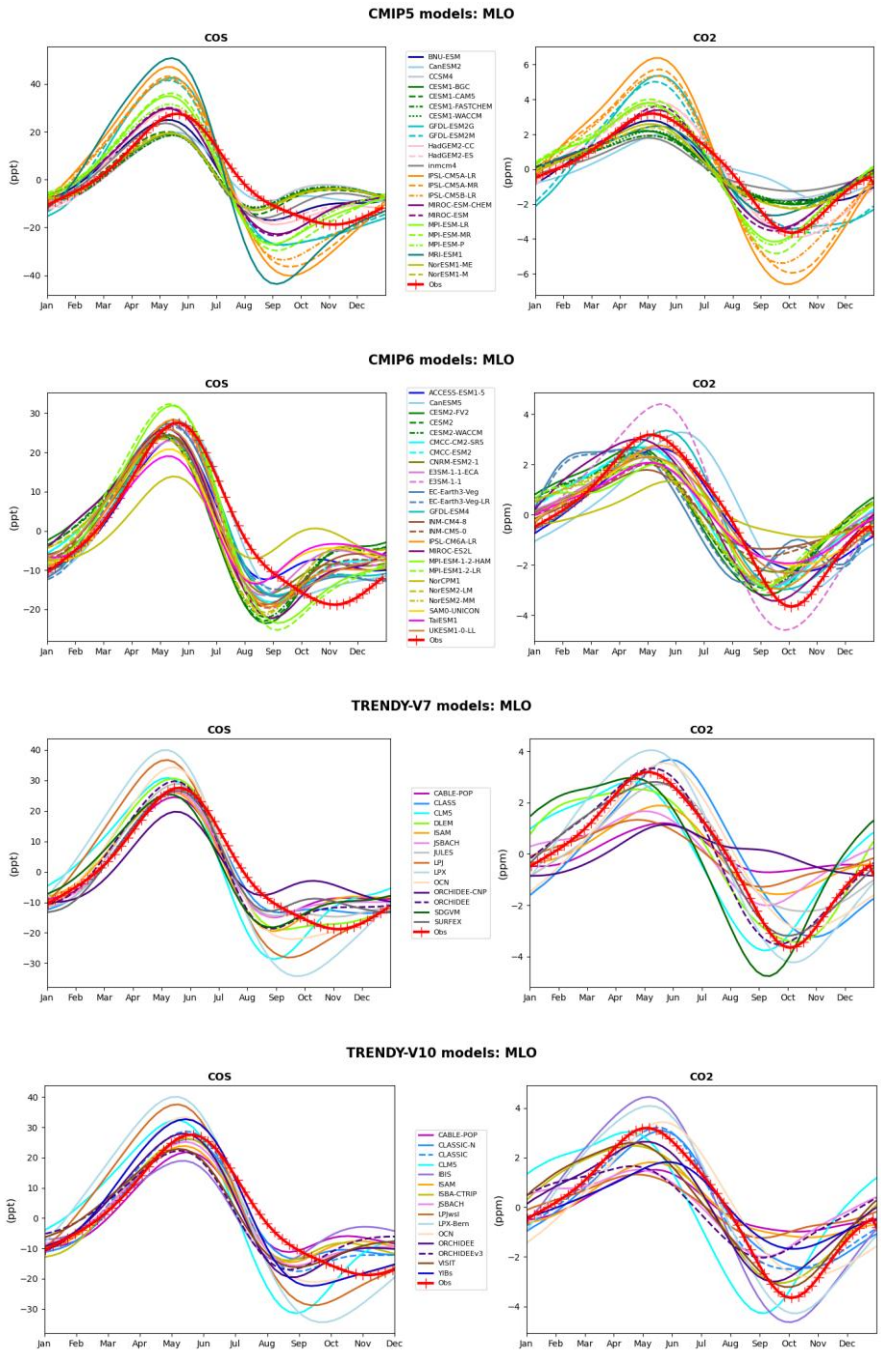

Figure 3.4.1: Smoothed seasonal cycles of atmospheric COS (left) and CO_2 (right) concentrations simulated at ALT station for CMIP5, CMIP6, TRENDY-V7 and TRENDY-V10 models. The observations are represented by red crosses.

We then further quantify the differences between the simulated COS / CO_2 concentrations and the respectives observations, focussing on the seasonal cycle and providing a specific metric for all stations. We analyzed the mean seasonal amplitude of the modeled atmospheric COS and CO_2 concentrations by normalizing the simulated seasonal amplitude with the observed one (see figure below). For each model, if the seasonal amplitudes of COS and CO_2 concentrations are equal to the observed ones, the model will be located on the red cross in Figure 3.4.2, which represents the non-biased situation.

From this figure we can draw the following conclusion for the different ensemble of runs:

- For the CMIP5 models, we clearly see at MLO (and to a large extent ALT) that the model ensemble falls on a line that crosses the “observation point” in the middle, with many models having too large COS and CO_2 amplitudes and a few having too low COS and CO_2 amplitudes. This simply quantifies the diagnostic made above with Figure 3.4.1. It further indicates that it is likely that our overall COS setup and methodology (in particular the so-called Leaf Relatif COS uptake (LRU)) may be adapted or that errors in some components are compensated by errors in others (LRU versus errors in soil COS uptake or atmospheric transport errors, etc). The fact that most models are in the lower-left or upper-right boxes of the plot, strongly suggest that the GPP of the model is key to explain these biases.
- For the CMIP6 models, we clearly see the improvement in the seasonal amplitudes that were noticed above. However, most models now tend to have a too low amplitude for both COS and CO_2, indicating possibly a too strong reduction of the GPP amplitude between the two CMIP phases. Overall the spread of the model is much lower for CMIP6, clearly indicating convergence.
- For the TRENDY-V7 models, using climate reanalysis to force the models does not lead to a smaller spread than for CMIP6. This is partly due to the fact that TRENDY comprises a larger ensemble of land surface models, while in CMIP6 some groups use the same land model (i.e. CLM is used in several earth system models).
- For the TRENDY-V10 models, we observe a similar range than for V7. Although some individual models have slightly improved at both stations, on average model performances remain similar. An ongoing analysis is checking whether the most recent TRENDY inter-comparison (V11) provides a similar feature.

Overall, the COS diagnostic provides valuable information on the model skis with respect to the mean seasonal cycle of GPP (combined with the NBP diagnostic using CO_2 data). All these results, including analysis at other atmospheric stations (more than 10 sites) are summarized into a paper that will be submitted early 2023 (*Peylin et al., Evaluation of CMIP/TRENDY model gross primary productivity using atmospheric COS and CO_2 data, Biogeosciences*).

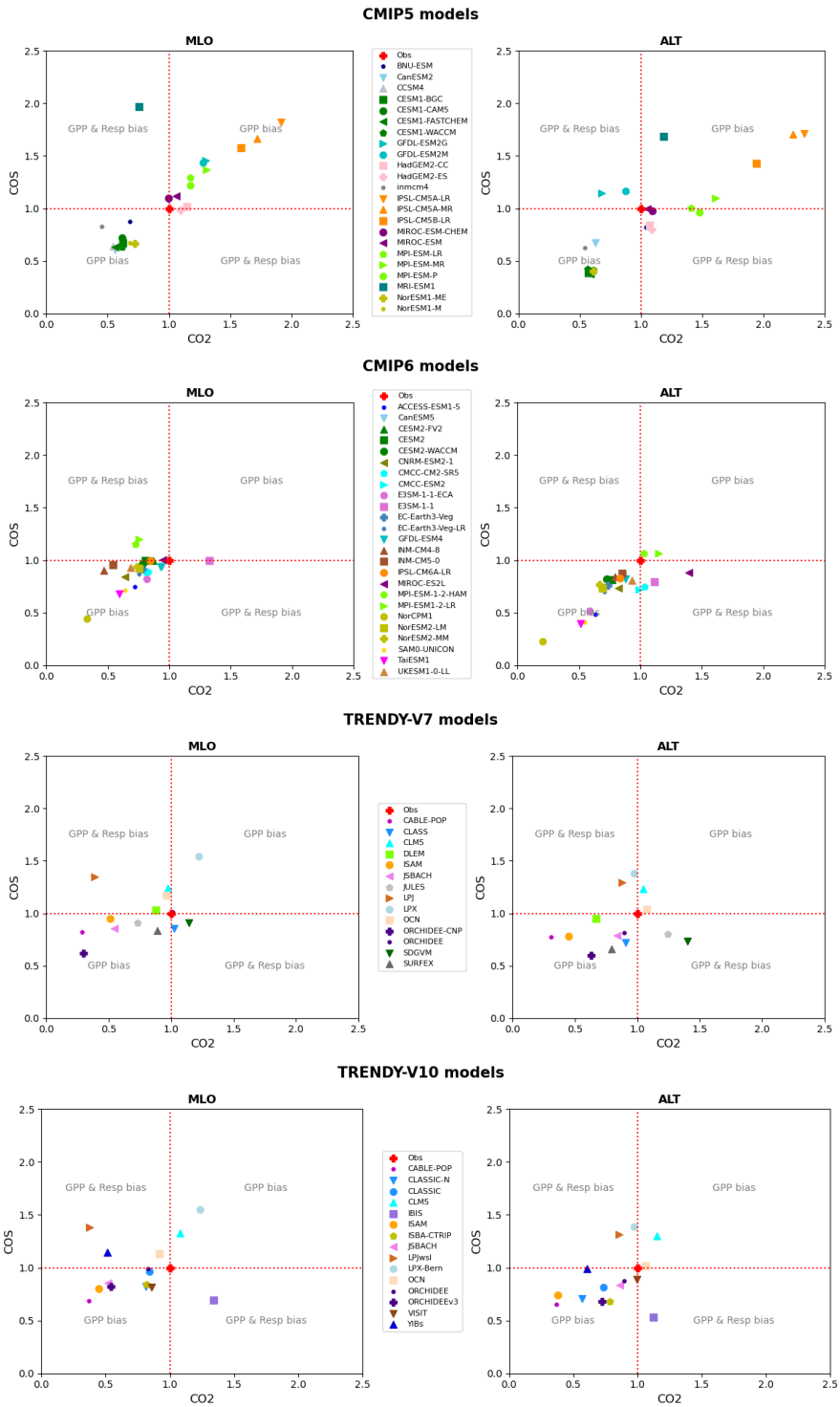

Figure 3.4.2: Scatter plots of CO_2 vs COS simulated amplitudes normalized by observed amplitudes of smoothed seasonal concentrations at MLO and ALT for CMIP5, CMIP6, TRENDY-V7 and TRENDY-V10 models. The observed phase is identified by the red cross in the middle.

Note finally that such analysis depends to a certain extent on the other COS fluxes, linked to soil, ocean, fire, industry and atmospheric chemistry. Soil COS fluxes with potentially a significant seasonality, as well as air-

sea exchanges, may bais to a certain extent the above diagnostics, but unlikely turning the extreme models in terms of amplitudes (lowest or highest amplitudes) into realistic models. In addition, in a recent paper (Remaud et al., 2022), we analyzed the impact of atmospheric transport errors on the simulated concentrations and especially the mean seasonal cycle amplitude. We compared seven atmospheric transport models providing simulations of COS mixing ratios in the troposphere over a 9-year period (2010–2018), using prescribed state-of-the-art surface fluxes for all components of the atmospheric COS budget: vegetation, soil, ocean, fire and industry. Models were run with the same fluxes to isolate transport differences and assess their skills with in-situ observations from surface stations and aircraft. The COS mixing ratios are underestimated by at least 50 parts per trillion (ppt) in the tropics, pointing to a missing tropical source. In contrast, in summer the mixing ratios are overestimated by at least 50 ppt above 40ºN, pointing to a likely missing sink in the high northern latitudes during this season. However, regarding the seasonal amplitude, the model spread can reach up to 50 ppt at 6 out of 15 sites, compared to an observed seasonal amplitude of 100 ppt, which points to the need to use different transport models in order to evaluate the GPP of land surface and draw more robust and quantitative conclusions. However, it does not change the main conclusions of the above study, given that for MLO and northern high latitudes stations the transport model spread is not as large.

3.4.2 Satellite SIF data to constrain GPP

We are now using recent observations of Solar Induced Fluorescence (SIF) to evaluate model GPP. These data should provide complementary information compared to the COS diagnostic about regional GPP gradients and temporal variations as well. The SIF dataset used for the evaluation comes from the TROPOSIF product (Guanter et al., 2021), which is derived from the observations of the TROPOMI (TROPOspheric Monitoring Instrument) instrument aboard Copernicus Sentinel-5P. The valid SIF retrievals at 740 nm are available at 3.5×5.5 km^2 spatial resolution in the form of daily files in the L1B product (https://s5p-troposif.noveltis.fr/data-access/).

The consistency of the TROPOSIF data (derived from the 743–758 nm fitting window) was assessed against the accurate SIF retrievals from the OCO-2 mission (Frankenberg et al., 2014; Sun et al., 2018) and with the well-established TROPOMI SIF product from Caltech (Köhler et al., 2020); see Figure 3.4.3. A very high similarity between the two TROPOMI SIF datasets (R usually greater than 0.93, 0.008 mW m^{-2} sr^{-1} nm^{-1} bias) was found, whereas the evaluation of precision errors over sand deserts and semi-desert sites shown slightly smaller random errors for the TROPOSIF product. The retrieval error is typically 0.5 mW m^{-2} sr^{-1} nm^{-1} while the mean bias for the raw SIF estimates is -0.080 mW m^{-2} sr^{-1} nm^{-1}.

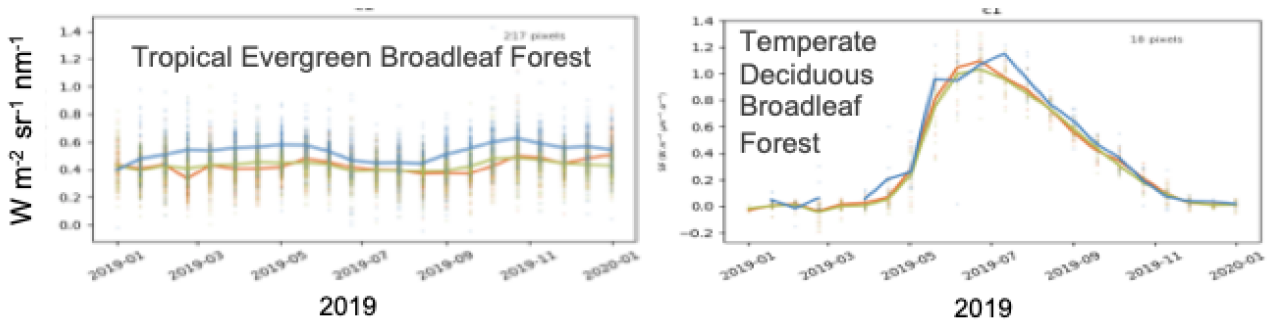

Figure 3.4.3: Comparison of three SIF satellite products for the year 2019 for two ecosystems: TROPOMI-ESA data in orange, TROPOMI-Caltech data in green and OCO-2 data in blue.

For the evaluation of the GPP products, we consider daily corrected TROPOSIF SIF data, removing retrievals associated with view zenith angles above 40° and cloud fraction larger than 0.5. The SIF estimates over the 2018-2021 period were further averaged at 0.5° / monthly resolutions. The two figures below (Figure 3.4.4 and Figure 3.4.5) compare for different regions / biomes the mean seasonal cycle of the model GPP from the CMIP6 or TRENDY - V10 collection with the mean SIF seasonal cycle from the TROPOMI - ESA product and also with the GPP from the FluxSAT product (a data-driven GPP product based also on SIF data).

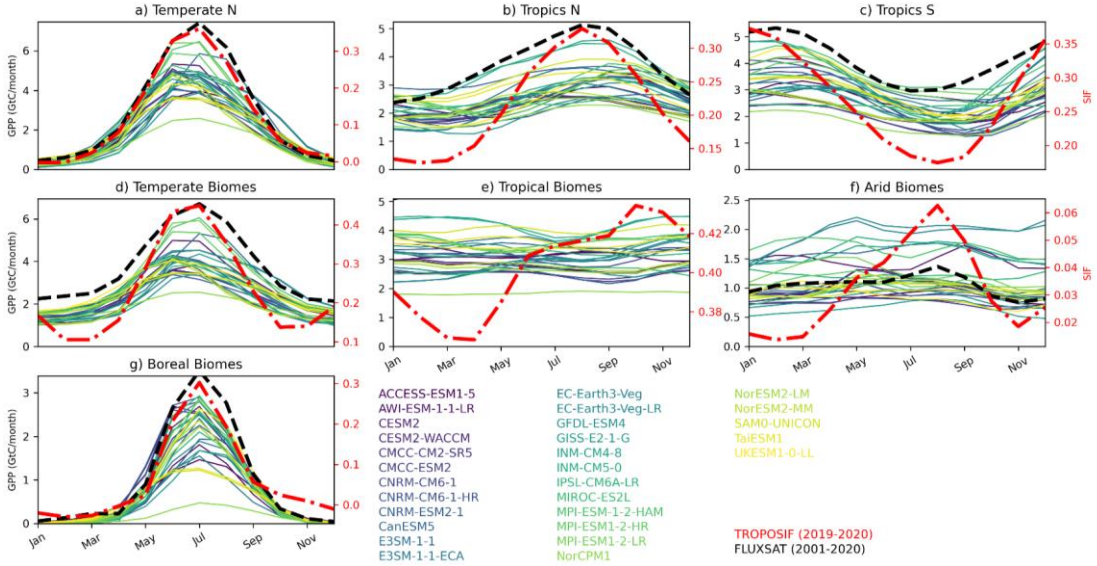

Figure 3.4.4 : Comparison of the mean seasonal cycle of the GPP from the CMIP6 ensemble simulations (left axis) with respect to the mean seasonal cycle of the SIF data from TROPOMI instrument (labeled TROPOSIF) (right axis). The GPP from the FluxSat product (data driven estimates using FluxNet CO_2 data and satellite SIF data) is also added.

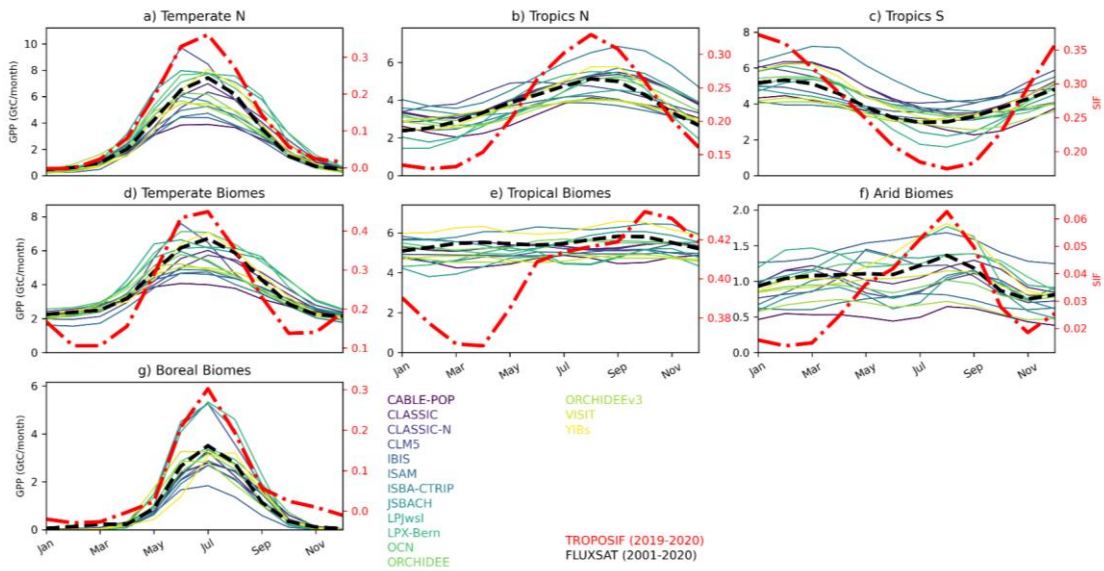

Figure 3.4.5: Same as Figure 3.4.4 but with the GPP simulated by the TRENDY - V10 model simulations.

From these comparisons, we can highlight specific features of the two land surface model ensembles:

- For the CMIP6 models (Figure 3.4.4), we see a clear seasonal cycle of SIF that is highly correlated with the seasonal cycle of the GPP for the extra-tropical ecosystems. However, it is difficult at this stage to evaluate the amplitude of the seasonal cycle given that we do not have a very robust and well-established relationship between the two quantities. We can only point toward potential model biases with respect to the temporality of the GPP. We clearly see that for some models, the growing season is too short compared to the SIF product. For tropical and arid biomes, on the other hand, the correlation between the modeled GPP and SIF data is much lower for most models. This clearly highlights the large uncertainty in these water limited regions associated with land surface models. However, the SIF signal should be taken also with some care as an increase of SIF may only partly relate to an increase of GPP.
- For the TRENDY - V10 models (Figure 3.4.5), we obtain similar diagnostics, for both extra-tropical and tropical/arid ecosystems. Even with prescribed climate forcing the tropical and arid ecosystems have a significantly different seasonality than the SIF signal, which calls for further investigation with respect to the GPP variation in these regions.

3.5 Land biomass evaluation with vegetation optical depth (L-VOD) satellite observations

Changes in terrestrial vegetation carbon storage due to environmental and land-use dynamics remain a critical uncertainty in carbon cycle science. Here, we generated space and time-continuous live vegetation biomass carbon maps using global L-band microwave vegetation optical depth (L-VOD) satellite observations. Carbon

stocks increased from 2010-2019 at a mean rate of 0.51 ±0.17 PgC yr^{-1}, the largest and second-largest contributors to which were boreal and temperate forests, respectively, while tropical forests were, due to deforestation and agricultural disturbances, small carbon sources. Young (<50 yrs) and middle-aged forests (50–140 yrs) dominated the biomass sink, whereas old-growth (>140 yrs) forests were approximately carbon neutral, opposite to the pattern predicted by vegetation models. Our findings improve attribution of long-term variations in terrestrial living biomass, highlighting the importance of forest age for predictions of carbon dynamics under a changing climate.

3.5.1 Regional total live biomass changes

Between 2010 and 2019, net L-VOD -derived live biomass carbon stocks increased at a rate of 0.51 ± 0.17 PgC yr^{-1} (mean ± s.d., s.d. represents the standard deviation of biomass changes estimated by 18 sets of L-VOD calibrations into biomass carbon, see Methods). Xu *et al.* (2021) reported a similar global annual net sink of +0.56 PgC yr^{-1} over the same period. The global patterning of carbon sinks and sources exhibited strong spatial and biome-scale heterogeneity (Figure 3.5.1). Boreal (+0.37 ± 0.02 PgC yr^{-1}) and temperate (+0.13 ± 0.05 PgC yr^{-1}) forests have acted as the largest and second-largest contributors, respectively, to the terrestrial live biomass carbon sink over the last decade. In contrast, wet tropical forests and arid biomes experienced net carbon losses (-0.07 ± 0.01 and -0.02 ± 0.03 PgC yr^{-1}, respectively), while dry tropical forests were carbon neutral (+0.0002 ± 0.0001 PgC yr^{-1}). These biome-scale results contrast with those of Xu *et al.* (2021), who reported larger net annual carbon sinks in the dry and wet tropics (+0.11 PgC yr^{-1} and +0.05 PgC yr^{-1}, respectively)

The spatial distribution of live biomass losses generally corresponds to that of global forest areal losses, as estimated by the Landsat-based Global Forest Change Data (Hansen et al., 2013). Specifically, where L-VOD net live biomass losses in the humid tropics are found in the eastern Amazon, the central Congo Basin and some southeast Asian rainforests (Figure 3.5.1a and b), cumulative forest area loss of 1.98×10^7 ha are reported from the Landsat -derived estimate (Figure 3.5.1b). On the other hand, in regions which experienced the largest carbon gains over the decade, L-VOD live biomass increased not only in zones of increasing forest area, but also where the latter remained stable (Figure 3.5.1c). This mismatch was apparent in West Africa and southern China, which both experienced large net carbon gains (Figure 3.5.1a and c), and could reflect the inability of Landsat imagery to detect stand growth and thickening, forest plantations, woody encroachment, and extra-forest woody systems establishment (Venter *et al.*, 2018).. This is especially the case in tree plantation areas in southern China, as well as in dryland Africa, where trees and shrubs grow in isolation without canopy closure (Brandt *et al.*, 2018).

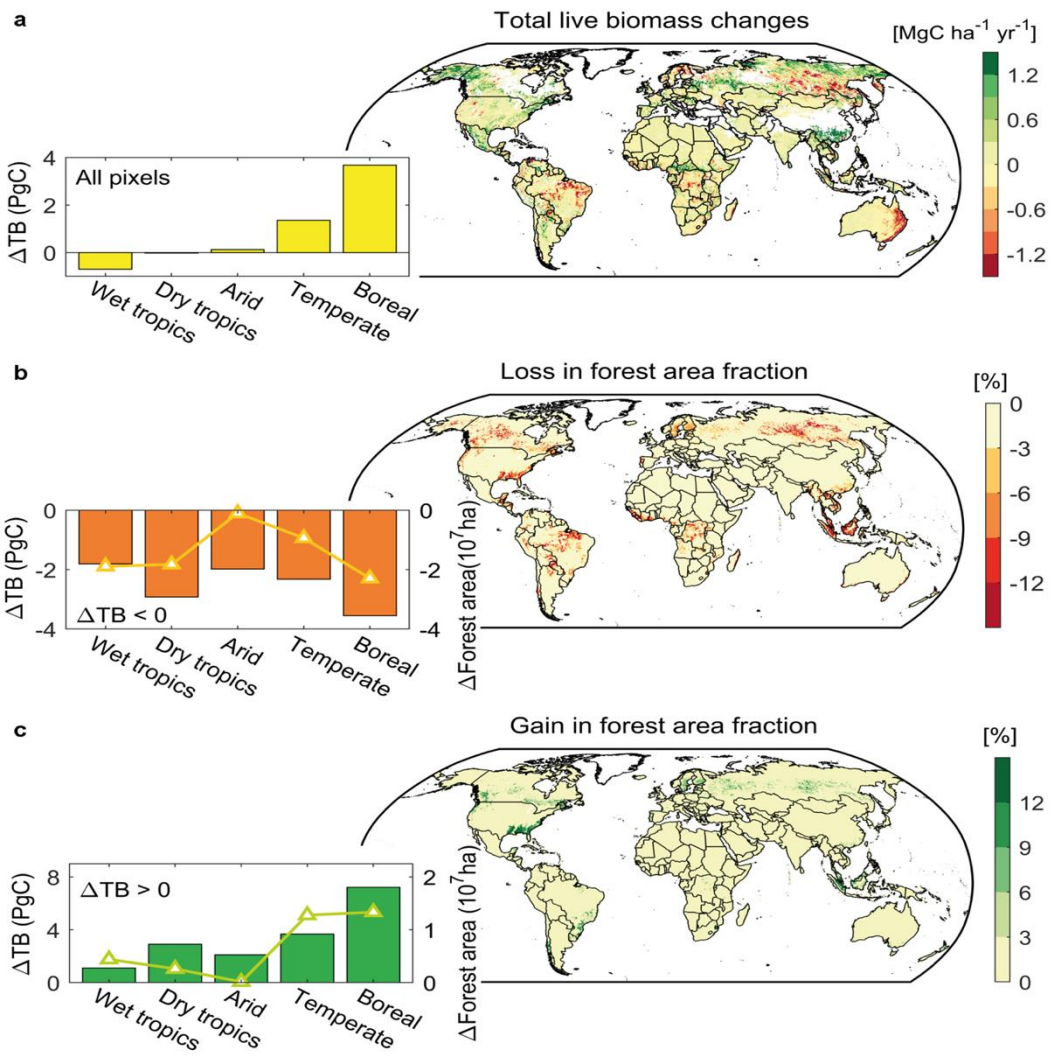

Figure 3.5.1. Spatial patterns of changes in total biomass carbon density and forest area fraction during 2010-2019. a, Mean annual changes in total live biomass (ΔTB) carbon density. The insert shows the total of ΔTB over all pixels (yellow bars) for five climatic biomes. b, The areal fraction of forest loss. The insert shows the total of ΔTB and forest area over the pixels with ΔTB < 0 (orange bars and curve). c, The areal fraction of forest gains. The insert shows the total changes of ΔTB and forest area over the pixels with ΔTB > 0 (green bars and curve).

3.5.2 Evaluation of DGVMs used for the global carbon budget

The TRENDY DGVMs show a very small net carbon sink in the temperate region (+0.03 ± 0.10 PgC yr^{-1}, ensemble mean ± s.d. across models) and a larger one in boreal forests (+0.28 ± 0.16 PgC yr^{-1}) (Figure 3.5.2). Their boreal forest carbon sink estimate approximates that of ours, but their temperate forest sink is less than half of the L-VOD derived value, perhaps due to an incomplete representation of plantations and forestry in most DGVMs. In the tropics, TRENDY models predict live biomass changes of an opposite sign to those of the L-VOD method, i.e., a carbon sink in the wet tropics (+0.07 ± 0.08 PgC yr^{-1}) and a large carbon source in the

dry tropics (-0.11 ± 0.10 PgC yr^{-1}). Modeled carbon gains in the wet tropics may result from the fact that DGVMs don't account for light competition between trees, and drought induced mortality which modulates the turnover of biomass, nor do they represent emerging phosphorus limitations on growth (Goll *et al.*, 2018), tropical forest degradation, and small fires (Pugh *et al.*, 2019; Yang *et al.*, 2020).

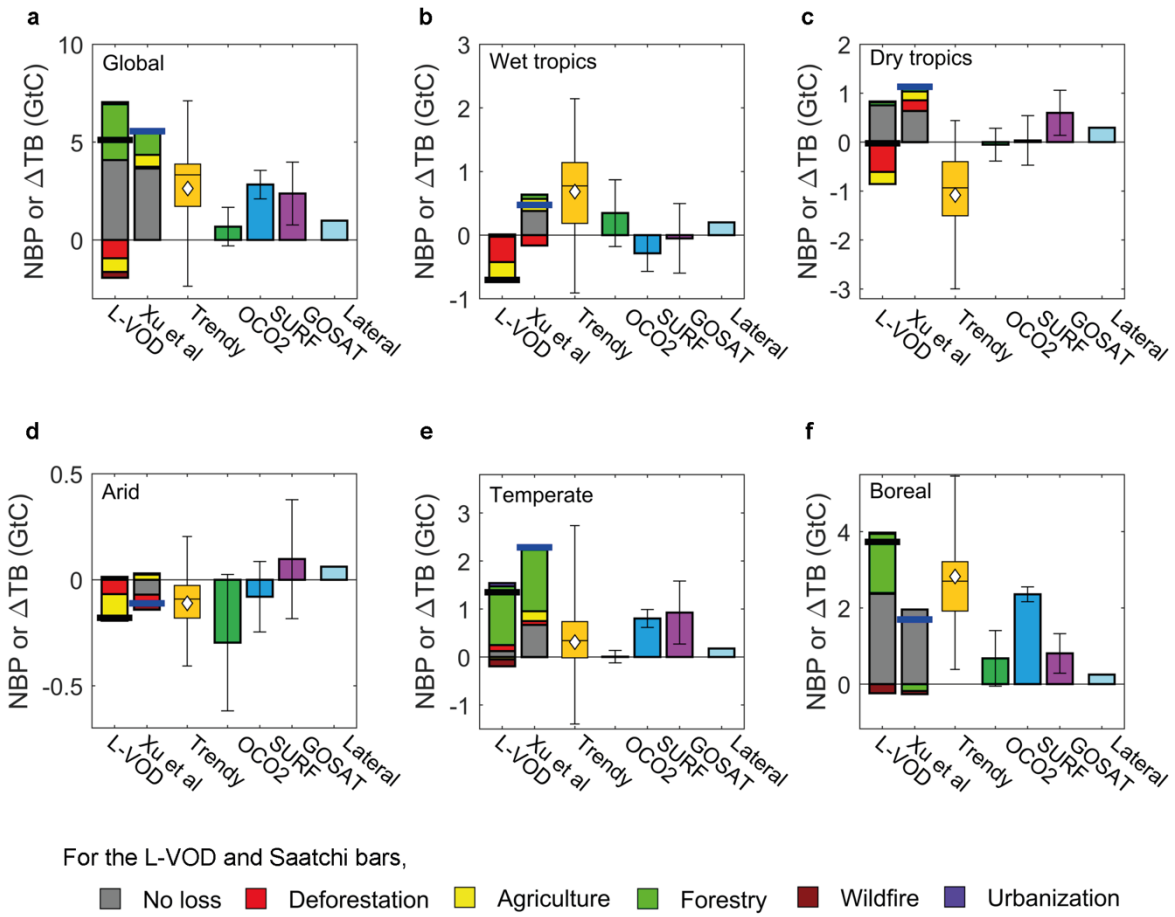

Figure 3.5.2. Changes in live biomass carbon density and mean net biome productivity for global and five climatic biomes. a-f, Changes in live biomass carbon density (ΔTB) were estimated using L-VOD, Xu et al. (2021), and Trendy DGVMs. Mean of net biome productivity (NBP) were estimated using three inversion models (OCO2, SURF, and GOSAT), and the lateral flux from river (light blue) using the estimates from Deng et al. (2021). ΔTB and NBP were calculated for global (a), and five climatic biomes: wet tropics (b), dry tropics (c), arid (d), temperate (e), and boreal (f). The stacked bars of L-VOD and Xu et al. (2021) show the subcomponents of ΔTB that are caused by no forest loss (grey), deforestation (red), agriculture (yellow), forestry (green), wildfire (brown) and urbanization (dark blue).

Perhaps a more important limitation of current DGVMs in this regard is their prediction of forest carbon uptake without consideration of differing carbon sequestration potentials across age classes. We binned the L-VOD data into 15 age class bins according to maps from Besnard *et al.* (2021) which derive the average forest age per pixel, and found that the average change in net live biomass density for young (< 50 yrs) and middle-aged

(50 – 140 yrs) forests correspond to areas acting as strong carbon sinks (Figure 3.5.3a). Conversely, old-growth forests (aged > 140 yrs), which represent approximately 10% of global forest area, were an atmospheric carbon source over the past decade, irrespective of whether they were intact or non-intact (Figure 3.5.3a). The absence of a carbon sink in old-growth forests is in accordance with the theoretical curves of forest growth and succession (Odum, 1969), but substantial debate remains from comparisons across studies based on in-situ data (Gundersen *et al.*, 2021; Luyssaert *et al.*, 2021).

In contrast to the L-VOD estimates, TRENDY DGVMs simulate diverse (even opposing) live biomass density trend directions over young and old-growth forest bins. The models predict very large carbon sinks over old-growth forests and small sources over young forests (Figure 3.5.3a). This is because these models predict a long residence time for living biomass in high biomass ecosystems (thus a higher sink potential in response to e.g. rising CO_2), and fail to consider the differences in CO_2 and climate effects on biomass accumulation across forest age classes. As illustration of the potential improvement of DGVMs in this regard, one model study that did incorporate forest age into a DGVM (LPJ-GUESS) predicted carbon sinks in young forests (Pugh *et al.*, 2019), in line with the results derived from L-VOD. Finally, since forest age is correlated with height, we used canopy height data derived from spaceborne Lidar measurements (Lang *et al.*, 2022) to assess the robustness of the differing biomass accumulation rates across forest ages found here. We find a response of live biomass density for forest canopy height bins similar to the one obtained for forest age bins (Figure 3.5.3b).

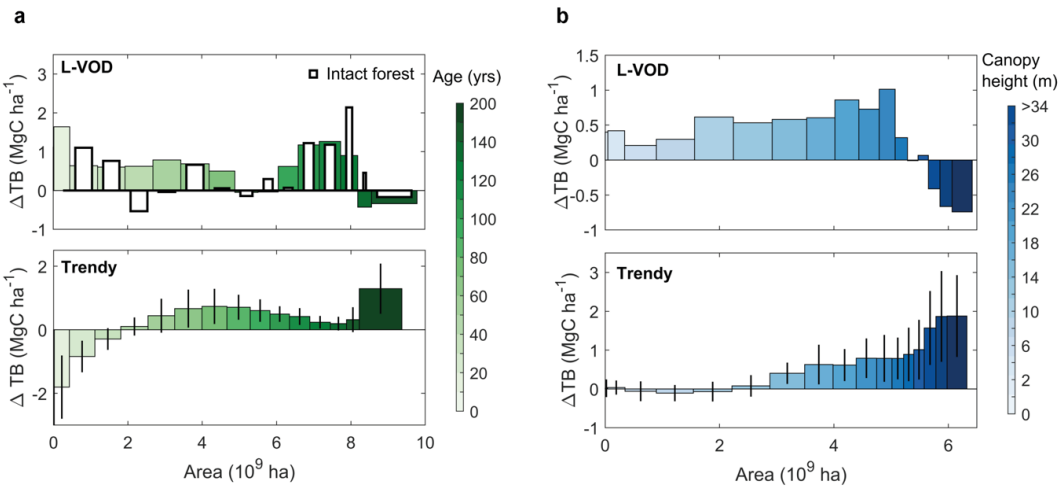

Figure 3.5.3. Relationship between total live biomass changes and forest age, canopy height. a, The averaged changes in total live biomass (ΔTB) density during 2010-2019 partitioned to different classes of forest ages (20-year intervals). b, changes in ΔTB density partitioned to different classes of canopy height (2m intervals). Total ΔTB density in each class were derived from L-VOD (upper panels) and Trendy DGVMs mean (bottom panels). The width of colored bars represents the forest area of different forest classes. The error-bars denote the uncertainty (standard deviation across Trendy models) of total ΔTB density in each forest class. In the panel a, the total ΔTB density of intact forests (Potapov et al., 2017) were also shown as white bars (the width of white bars does not represent the area of intact forests).

3.6 Evaluation of land models against satellite based XCO2 data (MPI-B)

Satellite observations of CO_2 offer a complimentary source of information compared to in situ observations of atmospheric CO_2, commonly used in atmospheric inversions as well as in land model evaluation using forward atmospheric transport. In this task, we used an updated version of the EMMA product (Reuter et al. 2013), provided by University of Bremen and accessed through the Copernicus Data Service.

Rather than comparing monthly mean XCO2 values, as discussed during the GA2022, we integrated the individual soundings of the different retrieval algorithms included in EMMA into the Jena CarboScope Inversion framework with the updated version v2022 (Rödenbeck, 2022). Next to the land fluxes (see below), ocean surface fluxes were prescribed from those provided by CarboScope v2022 (Friedlingstein et al., 2022) and fossil fuels were derived from GridFED v2022.2, which provides gridded monthly fossil CO_2 emissions from 1959 to 2021 including sources like oxidation of oil, coal or natural gases as well as cement production (Jones et al., 2021). Atmospheric transport of CO_2 was simulated with TM3 using 6-hourly reanalyzed wind data from the National Centers for Environmental Prediction (NCEP) and (Geels et al., 2007, updated).

For the evaluation, 13 out of 16 global vegetation models (DGVMs) from the TRENDYv11 dataset participating in the Global Carbon Project were examined using the S3 simulation including changes in atmospheric CO_2, N deposition, climate change and land-use changes (Friedlingstein et al. 2022).. These include CABLE-POP, CLASSIC, CLM5.0, IBIS, ISAM, JSBACH, JULES- ES, LPX-Bern, OCNv2, ORCHIDEEv3, SDGVM, VISIT and YIBs. The other 3 models did not provide the required output at monthly timescale.

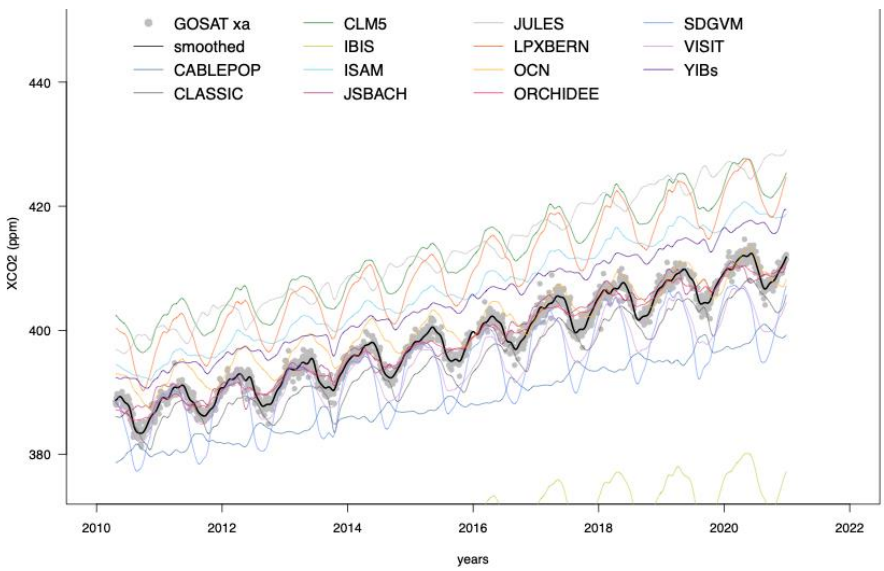

Figure 3.6.1 Observed and simulated XCO2 over North Africa ([15°W, 45°E], [10°N, 35°N]).

Figure 3.6.1 shows exemplary the simulated concentrations for the Sahel region by all 13 models in comparison to the satellite retrievals. Offsets between data and model are related to biases in the simulated (TRENDY) and assumed (combined ocean and fossil fluxes from CarboScope) net land carbon uptake, and should not be regarded as model error.

To compare model and data, we separated model information into trend, growth rate, detrended seasonal cycle and short-term variations using a common curve fitting method based on the method introduced by Thonning et al. (1989). This allows to investigate CO_2 distributions without the influences from possible sources or sinks and smoothes the data helping to analyze different signals on different timescales.

Since EMMA contains several retrieval products, it is necessary to determine the degree of consistency between different retrievals. For the example of the North Africa region (defined as a box with corners [15°W, 45°E], [10°N, 35°N]), Figure 3.6.2 shows the level of agreement between the different retrievals. While trends are robustly reproduced by all retrievals, there is a mismatch of high-frequency variations between the retrievals of about 1 ppm, with even somewhat larger deviations at the timescales of the detrended seasonal cycles. Notably, not all features of the interannual variations of the seasonal cycle are consistently captured in the retrievals. While the phasing of the annual growth rate is consistently represented in all retrievals, the magnitude of the interannual variations differs notably and up to 1 ppm relative to the long-term mean.

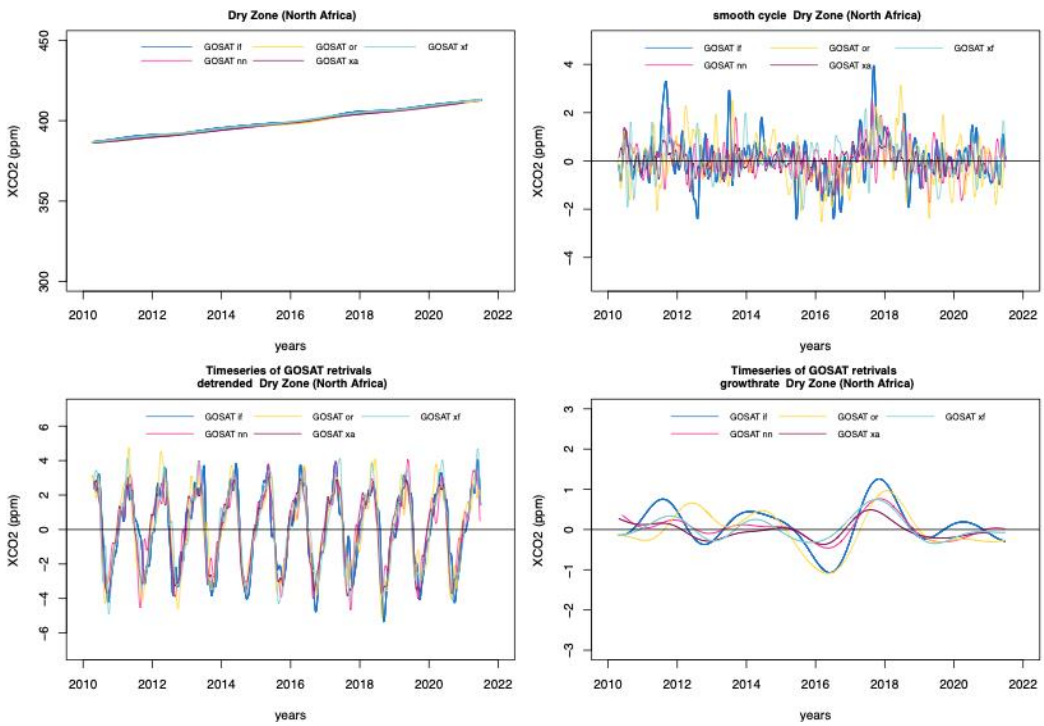

Figure 3.6.2: Comparison of different retrieval algorithms used to determine XCO2 from the GOSAT observations. Results are shown for the long-term trend, the short-timescale variations (<55 days), the seasonal cycle and its interannual variations ("detrended") and the growth rate.

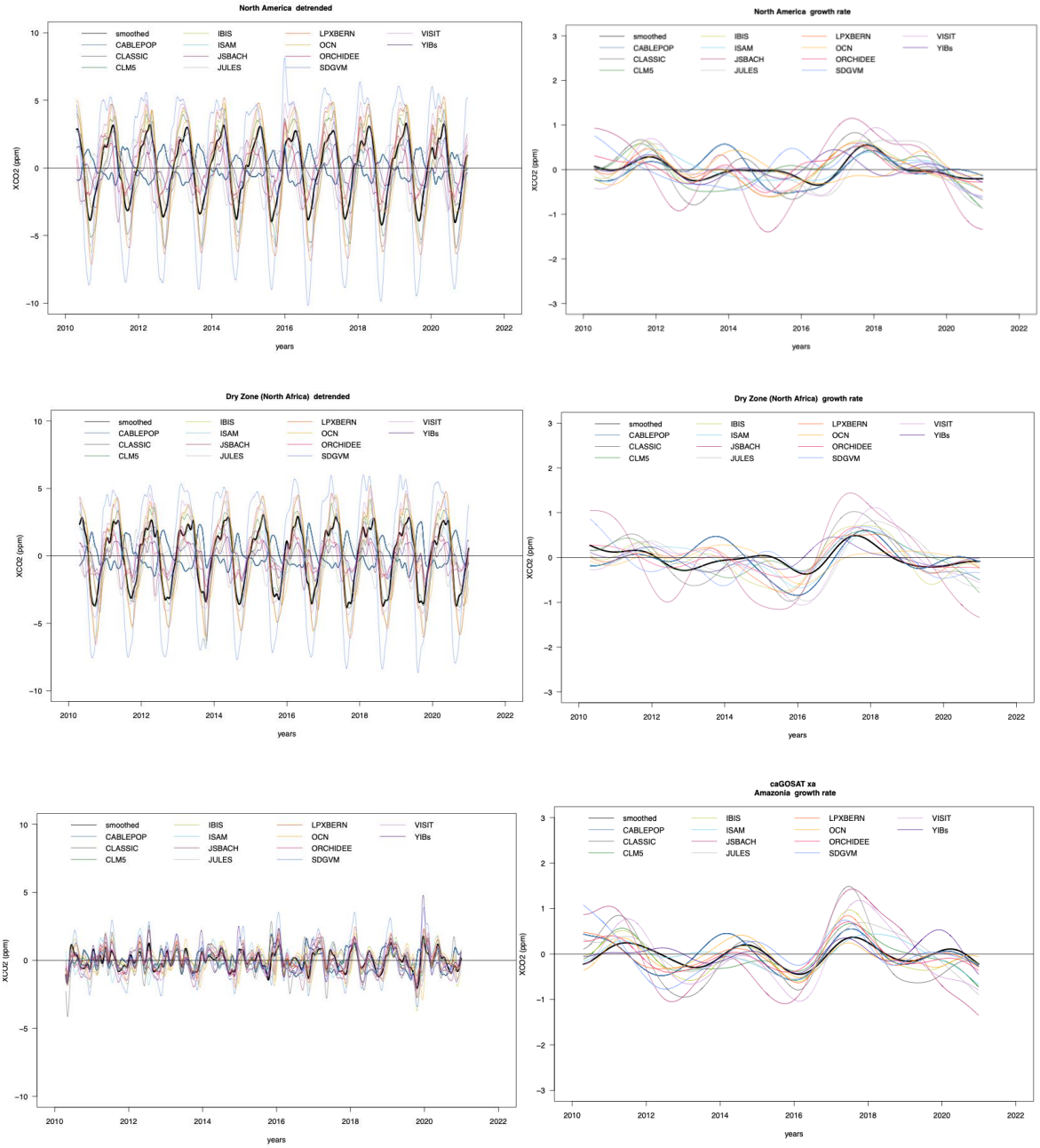

Figure 3.6.3 Evaluation of TRENDY models for the Northern America (top) Sahel zone (middle) and Amazonia (lower panels). For these comparisons, the GOSATxa retrieval from Figure 3.6.2 was used, as it showed the most consistent signal and had the most observations in this region.

Overall, comparing XCO2 with DGVMs shows important differences in modeled carbon uptake and release. While most of the models capture interannual seasonality well, others over- or underestimate the amplitude of CO_2. It is challenging to attribute model error to model assumptions or structures. However, the use of XCO2 allows to demonstrate that several models have significant issues not only in reproducing the interannual variability of net land carbon uptake, but also in the seasonal cycle. The ability of XCO2 to identify these biases across large regions suggests that there is a high potential of these data to benchmark and further evaluate land biosphere models. However, the large disagreement between individual retrievals at high frequency cautions against a too strict interpretation of the model-data mismatch in terms of the ability to capture short-term responses to climate variability (such as drought).

3.7 Ocean model evaluation with ocean interior data

Within 4C, at least two new products constraining the oceanic accumulation of DIC in the ocean interior have been created (see also D1.6) (Müller et al in review, Keppler et al in review). These form an independent constraint to evaluate the air-sea CO_2 flux across the air-sea interface and to investigate model biases that may cause discrepancies between observation-based and model estimates. The first product (Müller et al, in review) is based on the eMLR(C*) method (Clement and Gruber, 2018) and uses observations of dissolved inorganic carbon (DIC) from the late 1980s until the near present to estimate the accumulation of anthropogenic carbon (Cant) in the ocean interior for two decadal periods, i.e., from 1994 to 2004 and from 2004 to 2014. The authors find an uptake of anthropogenic CO_2 of 2.9±0.3 Pg yr^{-1} and 2.7±0.3 PgC yr^{-1} for the two decades, respectively. Although the two growth rates are not significantly different, they imply a reduction of the oceanic uptake fraction of the anthropogenic emissions from 36 ± 4 % to 27± 3 % from the first to the second decade. Müller et al. (in review) attributed this reduction to a decrease of the ocean buffer capacity and changes in ocean circulation, especially in the Atlantic Ocean.

The anthropogenic CO_2 uptake estimates tend to be on the high side compared to the ocean uptake estimates suggested by the Global Carbon Budget (Friedlingstein et al. 2022) on the basis of ocean models as well as surface ocean pCO_2-based air-sea CO_2 flux estimates. (Figure 3.7.1). The difference is particularly striking for the 1st decade, i.e., between 1994 and 2004 (Fig 3.7.1, panel b), while the difference is substantially smaller during the second decade (2004-2014) (Fig 3.7.1, panel c). Müller et al. (in review) interpreted the majority of this difference to be due to a loss of natural CO_2 and estimated this loss to amount to 8 ± 4 Pg C, all of which occurs during the first decade.

Even when accounting for this loss, the ocean interior uptake estimates still exceed those from the global ocean biogeochemical models (GOBM) that are being used for the Global Carbon Budget (yellow bars and lines in Figure 3.7.1). The ocean data suggest over the 20 year period from 1994 to 2014 an uptake of 56 Pg C of anthropogenic CO_2 and a loss of 8 for a total accumulation of of 48 Pg C. The models arrive, on average, at an accumulation of about 42 Pg C, with some models having taken up only 35 Pg C over these 20 years.

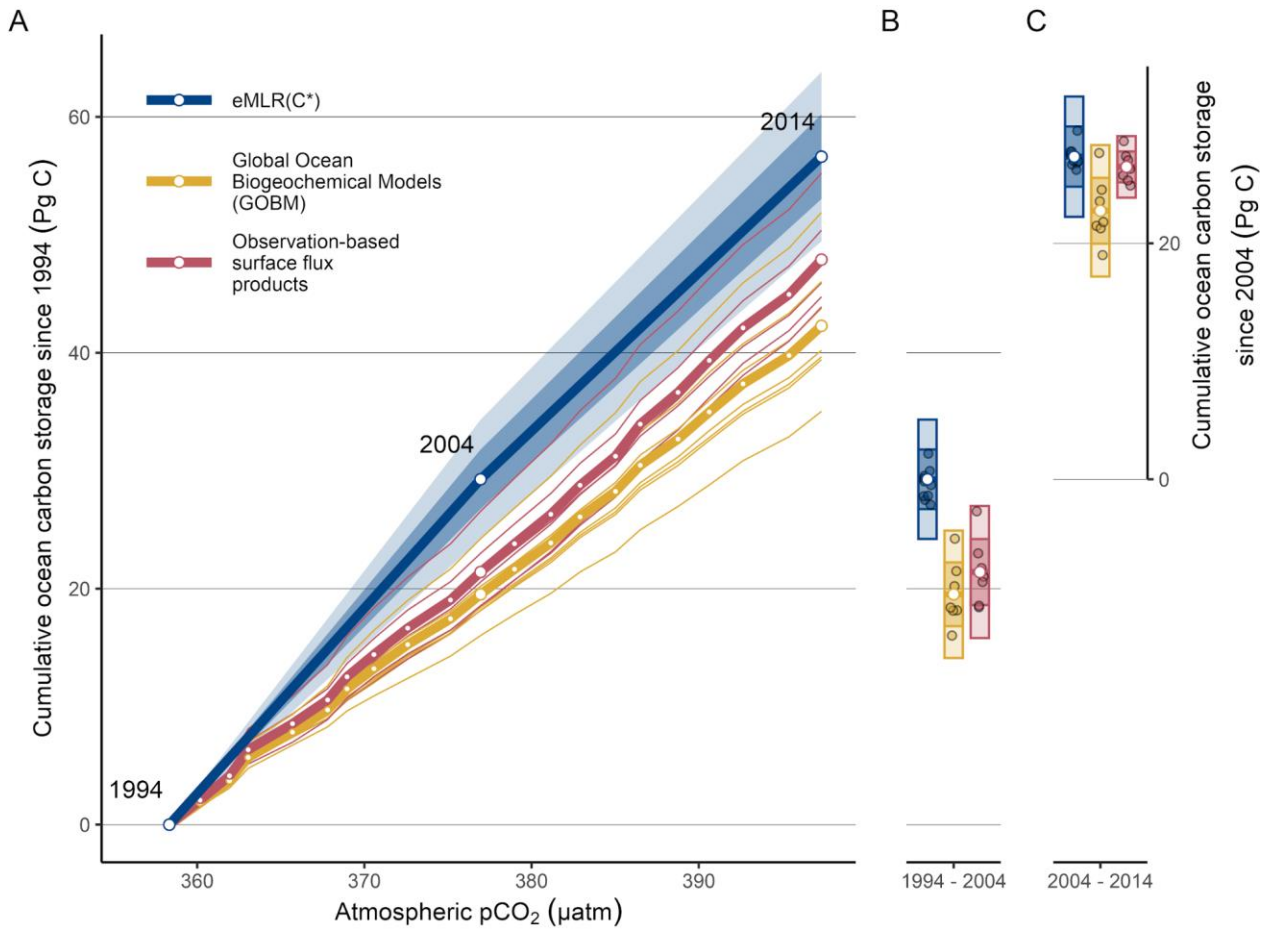

Figure 3.7.1 Ocean carbon storage from 1994 to 2014 according to the eMLR(C) estimates from Müller et al. (in review) (blue), in comparison to the cumulative fluxes from surface pCO2 observation-based air-sea CO_2 flux products (red) and Global Ocean Biogeochemical Models (yellow) from the Global Carbon Budget. The ocean carbon storage is displayed in (A) as a function of atmospheric pCO2 and in (B and C) as separate temporal integrals across the two considered decades. All cumulative estimates for the 2004–2014 period in (C) use the eMLR(C*) estimate for 2004 as the zero point. White points represent the ensemble mean for the GCB estimates and the standard case for the eMLR(C*) estimates. Bars in (B and C) indicate 1σ- and 2σ-uncertainty ranges. Note: The eMLR(C*) estimates represent storage changes of anthropogenic carbon only, while the GCB estimates include fluxes of natural and anthropogenic CO_2 (from Müller et al., in review).*

This conclusion is supported by the results of the second product, i.e., the reconstruction of the total DIC changes using the MOBO-DIC machine learning method of Keppler et al (in review). They suggest for the period between 2004 and 2020 a mean oceanic uptake rate of 3.2±0.7 PgC yr^{-1}. This is actually even larger than the scaled estimate based on the eMLR(C*) method, which would suggest for the same period an uptake of about 2.9 PgC yr^{-1} (see Table 3.7.1). This scaling is based on the recognition that the oceanic uptake of anthropogenic CO_2 scales with the increase in atmospheric CO_2 (Gruber et al., 2023) as is also evident from Figure 3.7.1. This high uptake estimate is even more remarkable since it is an estimate of the total change in the oceanic DIC content, accounting for the accumulation of both natural and anthropogenic CO_2. This is further evidence, that after ~2004, the ocean did not lose any natural CO_2. The high uptake in the MOBO-DIC estimate confirms the conclusion drawn from the eMLR(C*) method that the current models tend to underestimate the oceanic uptake of CO2 from the atmosphere (Table 3.7.1). This conclusion is also supported by the analyses of Terhaar et al. (2022) for the Earth System Models used in CMIP6 and by Terhaar et al. (in prep) for the GOBMs used by the Global carbon Budget.

Table 1 | Ocean CO_2 uptake from 1990 to 2019

Method	Components	1990–1999 (Pg C year^{-1})	2000–2009 (Pg C year^{-1})	2010–2019 (Pg C year^{-1})	Ref.
Atmospheric CO_2					
Change in atmospheric CO_2 (ppm)		15.0	18.7	23.6	155
Ocean CO_2 uptake					
Change in interior accumulation of anthropogenic CO_2 [a]	$F_{ant}^{ss}+F_{ant}^{ns}$	−2.1±0.2	−2.6±0.3	−3.3±0.3	4
Ocean inverse model (Green's function)	F_{ant}^{ss}	−2.0±0.6	−2.3±0.6	NA	3
Ocean inverse model (adjoint method)	F_{ant}^{ss}	−2.2±0.1	−2.5±0.1	−2.9±0.2	136
Ocean inverse model (adjoint method)[b]	$F_{ant}^{ss}+F_{nat}^{ns}$	−2.0±0.1	−2.3±0.1	−2.7±0.2	136
Ocean forward models	$F_{ant}^{ss}+F_{ant}^{ns}+F_{nat}^{ns}$	−2.0±0.2	−2.1±0.3	−2.5±0.3	25
Surface pCO_2 products[c]	$F_{ant}^{ss}+F_{ant}^{ns}+F_{nat}^{ns}$	−2.1±0.4	−2.3±0.2	−3.1±0.2	103

F_{ant}^{ns}, non-steady-state uptake component of anthropogenic CO_2 (part driven by variations in ocean circulation and other physical drivers); F_{ant}^{ss}, steady-state uptake flux component of anthropogenic CO_2 (part driven solely by the increase in atmospheric CO_2); F_{nat}^{ns}, non-steady-state exchange component of natural CO_2 (part driven by variations in ocean circulation and other physical drivers) (see Box 1); NA, not available; pCO_2, partial pressure of CO_2. [a]Scaled using sensitivity β=1.39 ± 0.16 Pg C per ppm CO_2 and the change in atmospheric CO_2 indicated in the first row. [b]Non-steady-state component is only due to variability in sea-surface temperature. [c]Adjusted for the steady-state outgassing of river-derived CO_2.

Table 3.7.1 Ocean CO_2 uptake estimates for the three decades from 1990 until 2019 based on various constraints. From Gruber et al. (2023).

Additionally, the time varying nature of the MOBO-DIC products allows to independently investigate the observed deviations between ocean models and pCO_2 products from 2002 onwards in the Global Carbon Budget (Friedlinstein et al 2022) and particularly in the Southern Ocean, where the largest differences in the decadal trends emerge. Figure 3.7.2 shows the time evolution of DIC accumulation of the Southern Ocean from the MOBO-DIC estimate (Keppler et al in review) in the Southern Ocean.

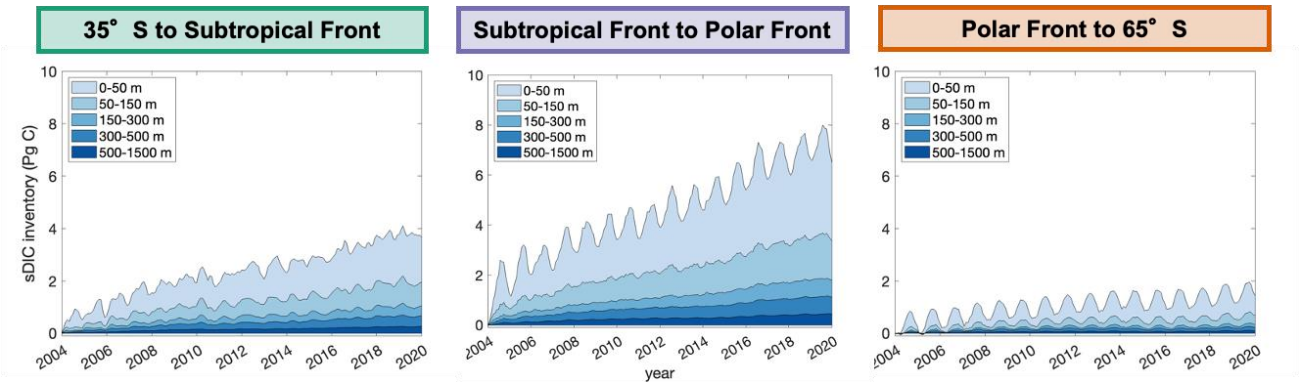

Figure 3.7.2 Temporal evolution of the salinity normalized interior dissolved inorganic carbon (DIC) pool in the Southern Ocean from a temporal extension of the MOBO-DIC estimate, developed within the 4C project.

Additionally, within the Regional Carbon Cycle Assessment and Processes Phase II (RECCAP2) project, MOBO-DIC is currently used to assess biases in ocean model simulations. Preliminary results (Rodgers et al in prep) suggest that a too weak seasonal DIC cycle is responsible for discrepancies in the high latitude northern and southern hemispheres in ocean models.

4 Conclusions and Outlook

The different sections of this report provided a comprehensive assessment of land and ocean carbon cycle model simulations, through an in-depth evaluation of various model outputs with respect to standard and new benchmarks. The land and ocean CO_2 fluxes were evaluated both in terms of seasonal cycle and inter-annual variations. In addition the land forest and soil carbon stocks as well as the ocean carbon content were also evaluated. Main model strengths and deficiencies were pointed out and discussed in the various sections, highlighting possible directions of model improvements.

Next steps will be to analyze the output of the post CMIP6 coupled earth system model (ESM) simulations, to see how recent model developments, including those performed within the 4C project, further decrease ESM model biases and with respect to the carbon budgets. These model simulations are currently on-going within 4C and complementary European projects (such as ESM2025). In addition some of the new benchmarks proposed in this report will integrate the ESMValTool (Task 1.4.3 of project).

5 References

Bakker, D. C. E., et al. (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.

Bennington, V., Gloege, L., & McKinley, G. A. (2022) Variability in the global ocean carbon sink from 1959 to 2020 by correcting models with observations. Geophysical Research Letters, 49, e2022GL098632

Besnard, S. et al. (2021). Mapping global forest age from forest inventories, biomass and climate data. Earth System Science Data Discussions, 1-22.

Brandt, M. et al. (2018). Reduction of tree cover in West African woodlands and promotion in semi-arid farmlands. Nature Geoscience, 11(5), 328-333.

Clement, D., & Gruber, N. (2018). The eMLR(C*) Method to Determine Decadal Changes in the Global Ocean Storage of Anthropogenic CO 2. Global Biogeochemical Cycles, 32(4), 654–679. https://doi.org/10.1002/2017GB005819

Friedlingstein, P., et al. (2022) Global Carbon Budget 2022, Earth System Science Data, 14(11), 4811-4900.

Frankenberg, C., et al. (2014) Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2, Remote Sensing of Environment, 147, 1–12, 2014.

Geels, C., et al. (2007) Comparing atmospheric transport models for future regional inversions over Europe – Part 1: mapping the atmospheric CO_2 signals, Atmospheric Chemistry and Physics, 7(13), 3461-3479.

Goll, D. S., Joetzjer, E., Huang, M., & Ciais, P. (2018) Low phosphorus availability decreases susceptibility of tropical primary productivity to droughts. Geophysical Research Letters, 45(16), 8231-8240

Gruber, N., Bakker, D. C. E., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., et al. (2023). Trends and variability in the ocean carbon sink. Nature Reviews Earth & Environment. https://doi.org/10.1038/s43017-022-00381-x

Guanter, L.. et al. (2021) The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission. Earth System Science Data, 13(11), pp.5423-5440.

Gundersen, P., Thybring, E. E., Nord-Larsen, T., Vesterdal, L., Nadelhoffer, K. J., & Johannsen, V. K. (2021) Old-growth forest carbon sinks overestimated. Nature, 591(7851), E21-E23.

Hansen, M. C. et al. (2013) High-resolution global maps of 21st-century forest cover change. Science, 342(6160), 850-853.

Hilton, T.W. et al. (2017) Peak growing season gross uptake of carbon in North America is largest in the Midwest USA. Nature Climate Change, 7(6), pp.450-454.

Humphrey, V., Zscheischler, J., Ciais, P., Gudmundsson, L., Sitch, S., & Seneviratne, S. I. (2018). Sensitivity of atmospheric CO_2 growth rate to observed changes in terrestrial water storage. *Nature*, *560*(7720), 628-631.

Jones, M. W., R. M. Andrew, G. P. Peters, G. Janssens-Maenhout, A. J. De-Gol, P. Ciais, P. K. Patra, F. Chevallier, and C. Le Quéré (2021), Gridded fossil CO_2 emissions and related O2 combustion consistent with national inventories 1959–2018, Scientific Data, 8(1), 2.

Keppler, L., Landschützer, P., Lavset, s. and Gruber, N.: Recent trends and variability in the oceanic storage of dissolved inorganic carbon, Global Biogeochemical Cycles, in review, 2023

Köhler, P., Behrenfeld, M. J., Landgraf, J., Joiner, J., Magney, T. S., & Frankenberg, C. (2020) Global Retrievals of Solar-Induced Chlorophyll Fluorescence at Red Wavelengths With TROPOMI, Geophysical Research Letters, 47, e2020GL087 541,https://doi.org/https://doi.org/10.1029/2020GL087541, e2020GL087541 10.1029/2020GL087541.

Launois, T., Peylin, P., Belviso, S., & Poulter, B. (2015) A new model for the global biogeochemical cycle of carbonyl sulfide – Part 2: Use of carbonyl sulfide to constrain gross primary productivity in current vegetation models. Atmospheric Chemistry and Physics 15: 9285–9312.

Lienert, S., Frölicher, T. L., Joos, F., (2022), Isotopic-constrained atmospheric carbon budget and sink flux estimates, D1.4 of the 4C project

Liu, L., et al. (in revision at Nature). Increasing negative coupling between tropical water and interannual CO_2 growth rate.

Liu, L., et al. (in preparation). Could interannual variability directly constrain long-term tropical land carbon-climate feedbacks?

Luyssaert, S., Schulze, E., Knohl, A., Law, B. E., Ciais, P., & Grace, J. (2021). Reply to: Old-growth forest carbon sinks overestimated. Nature, 591(7851), E24-E25.

Mayot, N., et al. (in revision). Climate-driven variability of the Southern Ocean CO_2 sink. Philosophical Transactions of the Royal Society A.

Mayot, N., et al. (in preparation). Constraints on the variability of the oceanic CO_2 sink from observations and theory. To be submitted to: Biogeosciences.

Müller, J. D. et al. (in review at AGU Advance) Decadal Trends in the Oceanic Storage of Anthropogenic Carbon from1994 to 2014.

Odum, E. P. (1969). The strategy of ecosystem development. Science (New York, NY), 164(3877), 262-270.

Peylin et al. (in prep). Evaluation of CMIP/TRENDY model gross primary productivity using atmospheric COS and CO_2 data. To be submitted to: Biogeosciences

Potapov, P., et al. (2017). The last frontiers of wilderness: Tracking loss of intact forest landscapes from 2000 to 2013. Science advances, 3(1), e1600821.

Pugh, T. A., Arneth, A., Kautz, M., Poulter, B., & Smith, B. (2019). Important role of forest disturbances in the global biomass turnover and carbon sinks. Nature geoscience, 12(9), 730-735

Remaud, M., Ma, J., Krol, M., Abadie, C., Cartwright, M., Patra, P.K., Niwa, Y., Rödenbeck, C., Belviso, S., Kooijmans, L. and Lennartz, S., 2022. Intercomparison of atmospheric Carbonyl Sulfide (TransCom-COS; Part one): Evaluating the impact of transport and emissions on tropospheric variability using ground-based and aircraft data; in review.

Reuter, M., et al. (2013), A joint effort to deliver satellite retrieved atmospheric CO_2 concentrations for surface flux inversions: the ensemble median algorithm EMMA, Atmospheric Chemistry and Physics, 13(4), 1771-1780.

Rödenbeck, C., DeVries, T., Hauck, J., Le Quéré, C., & Keeling, R. F. (2022), Data-based estimates of interannual sea–air CO_2 flux variations 1957–2020 and their relation to environmental drivers, Biogeosciences, 19(10), 2627-2652.

Rodgers K. B., Jörg Schwinger, Andrea J. Fassbender, Peter Landschützer, Ryohei Yamaguchi, Hartmut Frenzel, Seth Bushinsky, Thi-Tuyet-Trang Chau, Marion Gehlen, M. Angeles Gallego, Luke Gloege, Nadine Goris, Luke Gregor, Nicolas Gruber, Judith Hauck, Yosuke Iida, Masao Ishii Lydia Keppler, Ji-Eun Kim, Jens Daniel Müller, Sarah Schlunegger, Sahil Sharma, Karl Stein, Jerry Tjiputra, Katsuya Toyama, Pradeebane Vaittinada Ayar: Seasonal variability of the surface ocean carbon cycle: a synthesis, in prep, 2023

Seneviratne, S. I. et al.. (2013). Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. *Geophysical Research Letters*, *40*(19), 5212-5217.

Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., & Magney, T. (2018) Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP, Remote Sensing of Environment, 209, 808–823, https://doi.org/https://doi.org/10.1016/j.rse.2018.02.016.

Terhaar, J., Frölicher, T. L., & Joos, F. (2022). Observation-constrained estimates of the global ocean carbon sink from Earth system models. Biogeosciences, 19(18), 4431–4457. https://doi.org/10.5194/bg-19-4431-2022

Terhaar, J., N. Goris, J.D. Müller, T. DeVries, N. Gruber, J. Hauck, F.F. Perez, R. Seferian et al. Observation-based assessment of the regional and global carbon sink from global ocean biogeochemical models used in RECCAP2, in preparation.

Van den Hurk, et al. (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.

Whelan, M. E., et al. (2018), Reviews and syntheses: Carbonyl sulfide as a multi-scale tracer for carbon and water cycles, Biogeosciences, 15(12), 3625-3657.

Xu, L., et al. (2021). Changes in global terrestrial live biomass over the 21st century. Science Advances, 7(27), eabe9829.

Yang, H.,et al. (2020). Comparison of forest above-ground biomass from dynamic global vegetation models with spatially explicit remotely sensed observation-based estimates. Global Change Biology, 26(7), 3997-4012.