

Enter authors here: Morgan Loechli¹, Britton B. Stephens², Roisin Commane³, Frédéric Chevallier⁴, Kathryn McKain^{5,6}, Ralph F. Keeling⁷, Eric J. Morgan⁷, Prabir K. Patra⁸, Maryann R. Sargent⁹, Colm Sweeney¹⁰, and Gretchen Keppel-Aleks¹

¹University of Michigan, Ann Arbor, MI, USA

²National Center for Atmospheric Research, Boulder, CO, USA

³Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10064

⁴LSCE Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette Cedex, France

⁵Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA

⁶Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

⁷Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA

⁸Research Institute for Global Change, JAMSTEC, Yokohama 2360001, Japan

⁹Harvard University, Cambridge, MA, United States

¹⁰NOAA Earth System Research Laboratory Global Monitoring Division, Boulder, CO, USA

Corresponding author: Morgan Loechli (morgamic@umich.edu)

Key Points:

- Aircraft observations of atmospheric carbon dioxide concentrations are used to infer the northern extratropical growing season net flux.
- The observations suggest a larger net flux and shorter growing season than simulated in Earth system models.
- An emergent constraint approach is used to estimate productivity and respiration fluxes.

Abstract

Understanding terrestrial ecosystems and their response to anthropogenic climate change requires quantification of land-atmosphere carbon exchange. However, top-down and bottom-up estimates of large-scale land-atmosphere fluxes, including the northern extratropical growing season net flux (GSNF), show significant discrepancies. We develop a data-driven metric for the GSNF using atmospheric carbon dioxide concentration observations collected during the High-Performance Instrumented Airborne Platform for Environmental Research (HIAPER) Pole-to-Pole Observations (HIPPO) and Atmospheric Tomography

Mission (ATom) flight campaigns. This aircraft-derived metric is bias corrected using three independent atmospheric inversion systems. We estimate the northern extratropical GSNF to be 5.7 ± 0.2 Pg C and use it to evaluate net biosphere productivity from the Coupled Model Intercomparison Project phase 5 and 6 (CMIP5 and CMIP6) models. While the model-to-model spread in the GSNF has decreased in CMIP6 models relative to that of the CMIP5 models, there is still disagreement on the magnitude and timing of seasonal carbon uptake with most models underestimating the GSNF and overestimating the length of the growing season relative to the observations. We also use an emergent constraint approach to estimate annual northern extratropical gross primary productivity to be 56 ± 15 Pg C, heterotrophic respiration to be 25 ± 11 Pg C, and net primary productivity to be 28 ± 10 Pg C. The flux inferred from these aircraft observations provides an additional constraint on large-scale, gross fluxes in prognostic Earth system models that may ultimately improve our ability to accurately predict carbon-climate feedbacks.

1 Introduction

Approximately half of the carbon dioxide (CO_2) released annually by the combustion of fossil fuels stays in the atmosphere (Keeling et al., 1976; Schimel et al., 2001; Friedlingstein et al., 2021). The remaining CO_2 is taken up by the terrestrial biosphere and ocean in roughly equal proportion (Khatiwala et al., 2009; Keeling et al., 2014; Sabine et al., 2004). The efficiency of the ocean and land sinks varies with both climate and atmospheric CO_2 , representing an important feedback in the climate system (e.g. Ballantyne et al. 2012; Fung et al., 2005; Fernandez-Martinez et al., 2019). The strength of the land sink may be related to the amplitude of the seasonal cycle of atmospheric CO_2 (e.g. Keeling et al., 1996; Randerson et al., 1997) via annual and seasonal imbalances between photosynthesis and respiration. However, the magnitude, and spatial and temporal distributions of gross primary productivity (GPP) and net primary productivity (NPP) vary noticeably among Earth system models (ESMs) (e.g. Hu et al., 2022). Furthermore, models typically underestimate the change in amplitude of seasonal CO_2 exchange in northern land ecosystems over time (e.g. Graven et al., 2013) or underestimate CO_2 uptake in the Northern Hemisphere mid-high latitudes. (e.g. Canadell et al., 2021, fig 5.24).

Multi-model ensembles of coupled carbon-climate models show large differences in their land sink projections, especially for terrestrial carbon uptake (e.g. Arora et al., 2020; Cadule et al., 2010). For example, Friedlingstein et al. (2014) showed that the Coupled Model Intercomparison Project phase 5 (CMIP5) models range between -173 and 758 Pg C in simulations of cumulative land carbon uptake for 1850 to 2100 when forced by RCP8.5. This uncertainty exists in historical simulations where models both overestimate and underestimate the historical atmospheric CO_2 increase by over 20%. These differences are mainly due to uncertainties in the land carbon cycle response, with differences in their cumulative land flux estimates of 214 Pg C for 1850-2005, more than double the differences in their cumulative ocean flux estimates (Friedlingstein et al., 2014).

Quantifying the exchange of carbon between the atmosphere and the land surface at hemispheric and global scales is challenging because the heterogeneity of Earth’s surface makes it difficult to upscale local flux measurements (e.g. Friend et al., 2007; Kumar et al., 2016). Atmospheric inversion, wherein carbon fluxes are estimated from atmospheric CO₂ observations using atmospheric tracer transport models, provides a method to infer large-scale carbon fluxes (e.g. Tans et al., 1990; Ciais et al., 2010; Thompson et al., 2016); however, this method has been shown to be sensitive to uncertainty due to the simulation of vertical transport (Schuh et al., 2019; Stephens et al., 2007; Verma et al., 2017). Atmospheric inversions that rely only on surface observations must accurately represent vertical mixing to estimate CO₂ concentrations aloft. Uncertainty in atmospheric inversion flux estimates can be characterized through the use of observations of the vertical profile of atmospheric CO₂ (e.g. Peiro et al., 2022; Stephens et al., 2007).

Global-scale aircraft observations, such as those made during the High-Performance Instrumented Airborne Platform for Environmental Research (HIAPER) Pole-to-Pole Observations project (HIPPO, 2009-11) and the Atmospheric Tomography Mission (ATom, 2016-18), are representative of large regions and capture the vertical profile of atmospheric CO₂ (Wofsy et al., 2011; Thompson et al., 2021). These campaigns measured the vertical structure of CO₂ in the atmosphere across a range of latitudes and over the full seasonal cycle, and allow for analysis of seasonal changes in hemispheric-scale atmospheric CO₂ (e.g. Jin et al., 2021), which are dominated by land exchange. We use the seasonal cycle of atmospheric CO₂ concentrations measured during the HIPPO and ATom flight campaigns to develop a metric for evaluating the simulation of terrestrial CO₂ exchange in prognostic ESMs.

We derive estimates of the northern hemisphere net land flux integrated over the growing season, or growing season net flux (GSNF), as a benchmark for model evaluation (e.g. Collier et al., 2018). The creation of flux benchmarks allows for a direct comparison of observations and model simulations at the flux level rather than at the concentration level (e.g. Keppel-Aleks et al., 2013), which requires either using an atmospheric transport model or emulator (Liptak et al., 2017) to translate fluxes into atmospheric mole fraction variations. This research explores an alternative approach to formal inverse modeling to constrain net land-atmosphere carbon fluxes at hemispheric scale. We use CO₂ measurements from the HIPPO and ATom flight campaigns to infer the GSNF with only minimal reliance on atmospheric transport models. Thus, our estimated flux is less sensitive to errors in transport simulation and gives more robust insight into prognostic model inconsistencies.

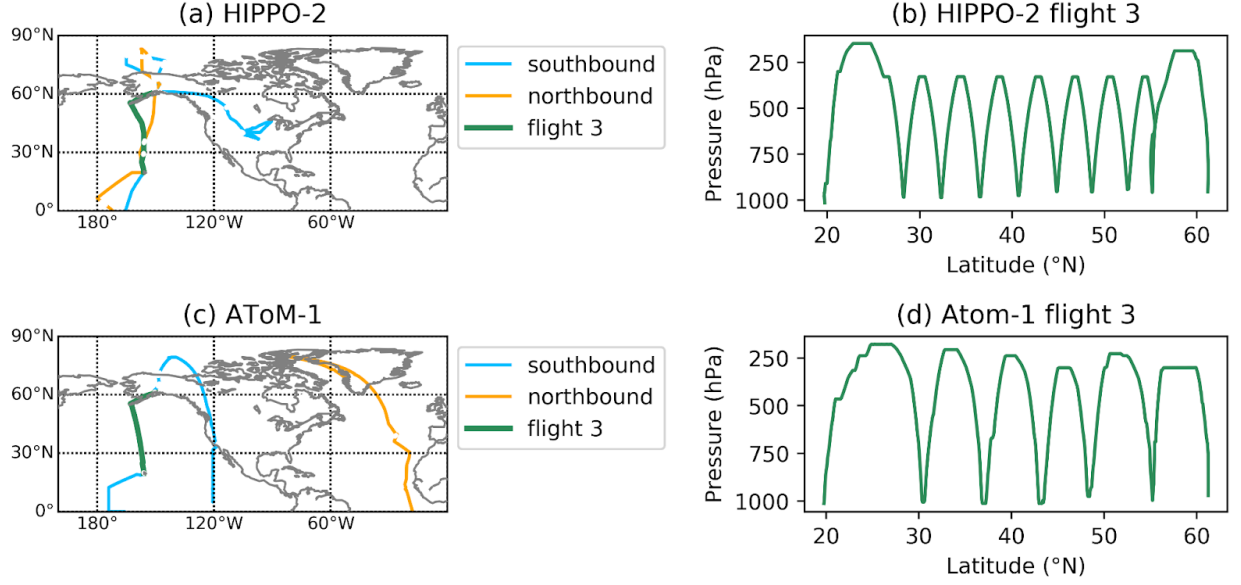
We describe the data sets and methods used to derive GSNF in Section 2. We discuss the GSNF and compare to ESM estimates of net biosphere productivity (NBP), GPP, heterotrophic respiration (RH), and net primary productivity (NPP), using output from the Coupled Model Intercomparison Project phase 5 and 6 (CMIP5 and CMIP6) in Section 3. This is followed by a discussion of

those results in Section 4 and conclusions in Section 5.

2 Data and Methods

2.1 Aircraft Observations

We used dry air mole fractions of CO_2 measured in the free troposphere during the HIPPO and ATom aircraft campaigns. HIPPO (Wofsy, 2011; Wofsy et al. 2017) used the NSF/NCAR GV aircraft to make measurements primarily over the remote Pacific from 87° N to 67° S (Fig. 1a) during five campaigns that spanned all four seasons between 2009 and 2011 (Table 1). The aircraft flew vertical profiles from near the surface to an altitude of 14 km; typically, a full profile was completed over $\sim 2.2^\circ$ of latitude (Fig. 1b). During these flights, measurements were made of greenhouse gasses and related tracers. CO_2 mole fractions were measured using three different in situ instruments and two whole air samplers: the Harvard Quantum Cascade Laser System (QCLS, Santoni et al., 2014), the Harvard Observations of the Middle Stratosphere (OMS, Daube et al., 2002) instrument, the National Center for Atmospheric Research (NCAR) Airborne Oxygen Instrument (AO2, Stephens et al., 2021), the National Oceanic and Atmospheric Administration (NOAA) Portable Flask Packages (PFP, Sweeney et al., 2015), and the NCAR/Scripps Medusa Whole Air Sampler (Stephens et al., 2021). For our analysis, we used the recommended CO_2X variable, which is derived primarily from QCLS measurements with calibration periods gap-filled using OMS measurements, reported as part per million dry air mole fraction (Wofsy et al., 2017). We use the 10-second merge data product and all CO_2 measurements are reported to be within 0.2 ppm with respect to the WMO X2007 scale (Santoni et al., 2014). The mean bias between QCLS and NOAA flask measurements across all five HIPPO campaigns is 0.11 ppm (Santoni et al., 2014). We use comparisons to the other 4 systems as a measure of analytical uncertainty. We also use observations of N_2O made by QCLS to identify stratospheric samples.



1 Flight paths for (a) HIPPO-2, which flew over the remote Pacific in November 2009, and (c) ATom-1, which flew over the Pacific and Atlantic in August 2016. All other campaigns followed similar flight paths. Flight path with continuous vertical profiling for flight 3, which flew from Anchorage, AK to Kona, HI for (b) HIPPO-2 and (d) ATom-1. All other flights flew a similar path.

ATom (Wofsy et al., 2021) is a more recent series of flight campaigns that used the NASA DC-8 aircraft to measure atmospheric trace gas concentrations by traveling south over the Pacific and north over the Atlantic (Fig. 1c) and which included a much larger scientific payload. As with HIPPO, a full annual cycle was measured, with flights that occurred in each of the four seasons over a three-year period from 2016 to 2018 (Table 1). Flights spanned 83°N to 86°S and sampled vertical profiles from 0.2 to 12 km in altitude (Fig. 1d). ATom measured CO₂ using the QCLS, AO2, Medusa, and PFPs similarly to HIPPO but also included a NOAA Picarro instrument. For our analysis, we used the CO2.X variable, which consists of NOAA Picarro measurements gap-filled using QCLS measurements. However during the first two flights of ATom-1, the NOAA Picarro measurements were not reported due to an inlet problem. Similar to HIPPO, we use the 10-second merge data product (Wofsy et al., 2021). To identify and remove stratospheric samples, we use observations of N₂O from QCLS and the NOAA PAN and Trace Hydrohalocarbon Experiment (PANTHER, ATom-1 only). While the WMO CO₂ scale has been recently updated, both the HIPPO and ATom observations used here have been calibrated with respect to the previous, WMO X2007, scale.

Table 1 Aircraft data used in this study

Deployment	Northern Hemisphere (Southbound) Dates	Northern Hemisphere (Northbound) Dates
HIPPO-1	1/8/09-1/16/09	1/28/09-1/30/09
HIPPO-2	10/31/09-11/7/09	11/16/09-11/22/09
HIPPO-3	3/24/10-3/31/10	4/10/10-4/16/10
HIPPO-4	6/14/11-6/22/11	7/4/11-7/11/11
HIPPO-5	8/9/11-8/24/11	9/3/11-9/8/11
ATom-1	7/29/16-8/6/16	8/17/16-8/23/16
ATom-2	1/26/17-2/3/17	2/15/16-2/21/17
ATom-3	9/28/17-10/6/17	10/19/17-10/28/17
ATom-4	4/24/18-5/1/18	5/14/18-5/21/18

2.2 Curtain Averages from Atmospheric Concentrations

The CO₂ observations from all flight campaigns are combined to estimate the average northern extratropical tropospheric CO₂ seasonal cycle (Bent, 2014). We then use a set of transport models to convert the time derivative of this cycle into estimates of northern extratropical terrestrial CO₂ flux. We refer to this process as “bias correction.”

To isolate tropospheric CO₂ signals, we define an upper cutoff of 300 hPa and remove any remaining observations with detectable stratospheric influence using the measured concentration of nitrous oxide (N₂O) and a cutoff value of 319 parts per billion (ppb) after detrending the data to 2009; samples whose N₂O concentration falls below this threshold are removed from the observations (Bent, 2014). We also manually remove outlying samples primarily obtained during takeoffs and landings, to avoid strong local influences from biospheric exchange or fossil emissions. The flights and times filtered are identified in Data Sets S1 and S2. We filter output at the same locations and times for the transport model CO₂ mole fractions Simulated along the flight tracks, discussed in section 2.3. This stratospheric and local influence filtering removes 2.8% of the observations within the defined domain from the HIPPO and ATom datasets. We do not use observations from the northbound leg of HIPPO-1 because it only extended to 40°N, and both QCLS and OMS have been filtered for altitude-dependent biases on these flights.

We then detrend the filtered data by removing the long-term trend in the NOAA Mauna Loa in situ CO₂ mole fraction record (Thoning et al., 2022), found by Seasonal-Trend decomposition using locally estimated scatterplot smoothing (STL, Cleveland et al., 1990) with a 2-year smoothing window. The filtered and detrended data primarily reflect the seasonal variations in tropospheric CO₂ concentrations that are of interest.

We calculate the extratropical mean drawdown by first aggregating the detrended data in latitude and pressure bins. We discretize the atmosphere into bins of 5° in latitude and 50 hPa in pressure, for the latitude range 20°N to 90°N and the pressure range 300 hPa to 1000 hPa. Observations at latitudes

south of 20°N are excluded because of the differences in the phasing of the tropical seasonal cycle to that north of 20°N, and observations at pressures below 300 hPa were excluded because measurements were sparse and frequently in the stratosphere. Within each bin, we average all data collected for a given day of the year and then fit a second-order harmonic with an offset due to the difference in the annual mean relative to Mauna Loa (Fig. S1). We then generate seasonal time series at daily resolution from the harmonic fits and take the pressure-weighted average of these values for each latitude bin. These partial columns are then integrated over latitude from 20°N to 90°N (Eq. 1), using $\cos(\text{lat})$ weighting to reflect the influence on the full zonal volume. We call the result of this integration the “curtain average” concentration of atmospheric CO₂ (Bent, 2014).

$$\text{CurtainAvg} = \frac{\int_{20^\circ\text{N}}^{90^\circ\text{N}} \int_{300\text{ hPa}}^{1000\text{ hPa}} \text{aveCO}_2(\varphi, P) \, dP \cos \varphi \, d\varphi}{\int_{20^\circ\text{N}}^{90^\circ\text{N}} \int_{300\text{ hPa}}^{1000\text{ hPa}} dP \cos \varphi \, d\varphi} \quad (1)$$

The curtain average is shown in black in Figure 2, and is compared to northbound and southbound legs of each HIPPO and ATom mission where each point is found by filtering, detrending, interpolating and extrapolating to get a full altitude and latitude slice, then taking a pressure and cosine of latitude weighted average (Akima, 1978).

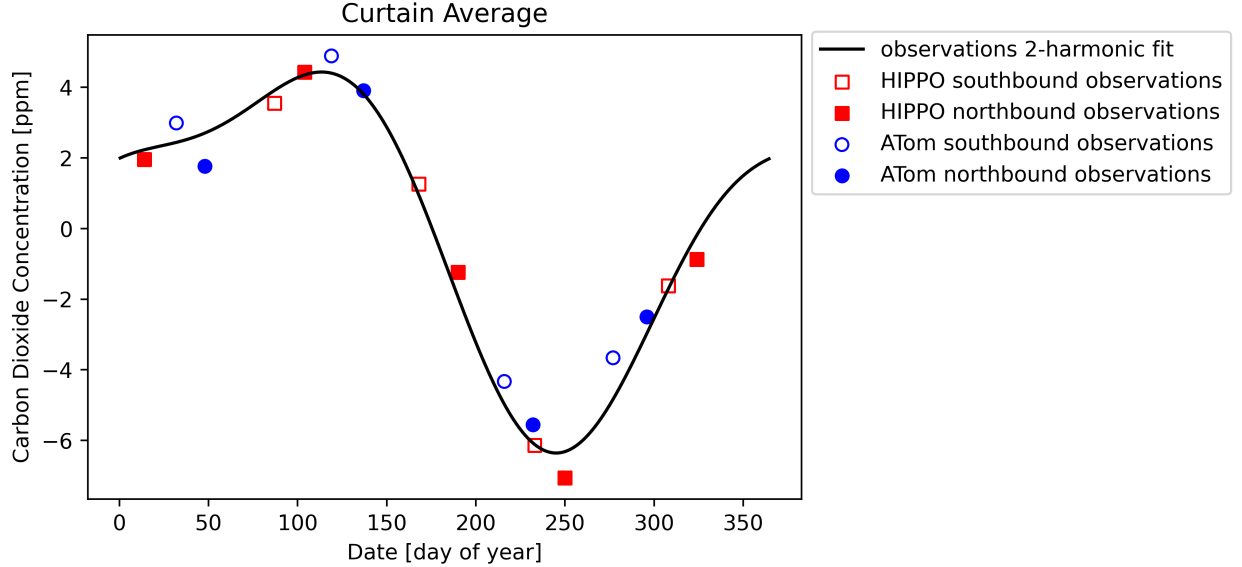


Figure 2 Two-harmonic fit to detrended average carbon dioxide concentration in ppm as a function of day of year for HIPPO and ATom flight campaigns in the atmospheric curtain between 20°N and 90°N in latitude and between 1000 hPa and 300 hPa in pressure. The points are found by filtering, detrending, interpolating and extrapolating to get a full altitude and latitude slice, then taking a pressure and cosine

of latitude weighted average. The black line is the average of all fits to individual latitude-pressure bins with the annual mean removed.

The derivative, found as a finite-difference, of the curtain average concentration fit line with respect to time, then gives the rate of change of CO_2 of this atmospheric volume as a function of the day of the year.

We convert from a rate of change in concentration to a rate of change in mass balance (MB) by multiplying the concentration by the mass of dry air within the domain from the ERA5 reanalysis fields used by Tracer Model 5 (TM5) within CarbonTracker 2019 (Krol et al., 2005; Jacobson et al., 2020). We use the time mean mass within the domain across the HIPPO and ATom time periods, neglecting annual and seasonal variations, which are less than 0.2%.

2.3 Flux Estimates Using Atmospheric Transport Models

Although the HIPPO and ATom observations over the remote ocean provide representative estimates of background values, the sampling and discretization method, zonal gradients, and mixing out of the domain result in differences between our MB time derivatives and zonal fluxes. Also, fossil fuel emissions and air-sea gas exchange make small contributions to the observed cycles. We use atmospheric transport models to account for the cumulative effect of: 1) atmospheric mixing across the southern boundary and above the pressure boundary; 2) spatial sampling biases associated with specific flight tracks; 3) zonal sampling bias; 4) temporal sampling biases associated with synoptic variability, subseasonal sample distribution, and interannual variability; and 5) contributions from fossil-fuel emissions and ocean uptake.

Atmospheric inversions provide optimal estimates of surface-atmosphere CO_2 exchange derived from both atmospheric CO_2 mole fraction data and initial estimates for land-atmosphere and ocean-atmosphere exchange in the context of an atmospheric transport model. We used posterior concentrations from three different inversions (Table 2) to reduce the uncertainty that may arise due to biases present in the choice of transport model as differences in transport have previously been shown to lead to large differences in optimized fluxes (Gurney et al., 2002; Schuh et al., 2019; Stephens et al., 2007).

CarbonTracker is a data assimilation system consisting of the TM5 atmospheric transport model coupled to an ensemble Kalman filter (Jacobson et al., 2020; Peters et al., 2007). TM5 is a global two-way nested transport model driven by 3-h meteorological forcing from the ERA5 operational forecast model (Krol et al., 2005). We used output from the most recent Carbon Tracker release (CT2019B, Jacobson et al., 2020), which includes optimized carbon fluxes through the HIPPO and ATom time period, and CO_2 mole fractions simulated along the flight paths for the HIPPO and ATom campaigns, which match the dates, times, and locations for the HIPPO and ATom data included in the GLOBALVIEW-plus v5.0 ObsPack product (Cooperative Global Atmospheric Data Integration Project, 2019).

The Model for Interdisciplinary Research on Climate version 4 atmospheric general circulation model based chemistry transport model (MIROC4-ACTM) provides posterior 4-D CO₂ fields and optimized surface fluxes through the HIPPO and ATom period (Chandra et al., 2022). Atmospheric CO₂ transport in MIROC4-ACTM is simulated by the Japan Agency for Marine-Earth Science and Technology’s ACTM, a transport model driven by meteorological parameters from the Japanese 55-year Reanalysis (JRA55, Patra et al., 2018). We use MIROC4-ACTM output from 2020 and MIROC CO₂ mole fractions simulated along the flight paths for the HIPPO and ATom campaigns matching the dates, times, and locations for the HIPPO and ATom data included in the GLOBALVIEWplus v5.0 ObsPack product..

We use a third set of inverse modeling output from the Copernicus Atmosphere Monitoring Service (CAMS; Chevallier et al., 2005). Within CAMS, transport of atmospheric CO₂ is simulated by the global climate model of the Laboratoire de Météorologie Dynamique, zoom capacity (LMDZ) driven by meteorological parameters from ECMWF (Chevallier et al., 2005). We used CO₂ mole fractions simulated along the HIPPO and ATom flight paths matching the dates, times, and locations for the HIPPO and ATom data included in the GLOBALVIEWplus v5.0 ObsPack product, and posterior carbon fluxes from CAMS v20r1, which contains output through the HIPPO and ATom period.

Table 2 Inverse models used in this study.

	CT2019B	MIROC4-ACTM	CAMS
Years Available	2000-2018	1996-2018	1979-2020
Years Used	2009-2018	2009-2018	2009-2018
Transport	TM5	ACTM	LMDZ
Meteorology	ERA5	JRA55	ERA5
Resolution (lat x lon in degrees)	Glb2x3, N America 1x1	Glb2.8x2.8	Glb1.9x3.75
Fossil Fuels	Miller and ODIAC	EDGARv432	GCP-GridFEDv2021.2
Reference	(Jacobson et al., 2020)	(Chandra et al., 2022)	(Chevallier et al., 2005)

For each model, we calculate the annual cycle of the northern extratropical net land flux by removing the long-term annual mean of the posterior land flux, which excludes fossil fuel emissions and ocean fluxes, from 2000 - 2018 at each grid cell and then taking an area-weighted average north of 20°N. We linearly interpolate between monthly means to get an annual cycle at daily resolution to allow for direct comparison to the aircraft-observation-derived MB.

To correct for bias, we match the model ObsPack output date, time, and location to the 10-second merge files for HIPPO and ATom, then repeat the analysis described in Section 2.2 using posterior CO₂ mole fractions simulated along the HIPPO and ATom flight tracks to calculate the curtain average and MB for each atmospheric inversion system. The MB, akin to the observations, is then subtracted from the area-weighted averaged posterior land flux to de-

rive a seasonal correction. The MB found using posterior CO_2 mole fractions simulated along the HIPPO and ATom flight tracks is the MB that would be observed if transport in our world perfectly matched transport in the model. Thus, in each model, we assume the difference between the MB for each model (solid lines in Fig. 3a) and northern extratropical average posterior land flux (dashed lines in Fig. 3a) is due primarily to mixing outside the domain, with additional influences listed above. We determine the correction for each model and subtract it from the observationally derived MB (dotted black line in Fig. 3b) resulting in transport-model specific flux estimates (solid color lines in Fig. 3b). The average difference across the three models (dashed black line in Fig. 3b) is subtracted from the observationally derived MB to estimate the seasonal cycle of the average net flux, hereafter referred to as the “flux cycle”, into the atmosphere (solid black line in Fig. 3b).

We find that the contribution of atmospheric transport uncertainty for the large spatial scale over which we average is small. In particular, noted variations in representations of vertical mixing (Stephens et al., 2007; Schuh et al. 2019) may change the distribution of CO_2 within our averaging domain but not the domain average.

2.4 Growing Season Net Flux from Seasonal Flux Cycles

We then calculate the net atmospheric carbon exchange during the growing season, or GSNF, as the integral of the flux cycle during the growing season, defined to be when the bias corrected flux cycle is negative, which is nominally equivalent to the

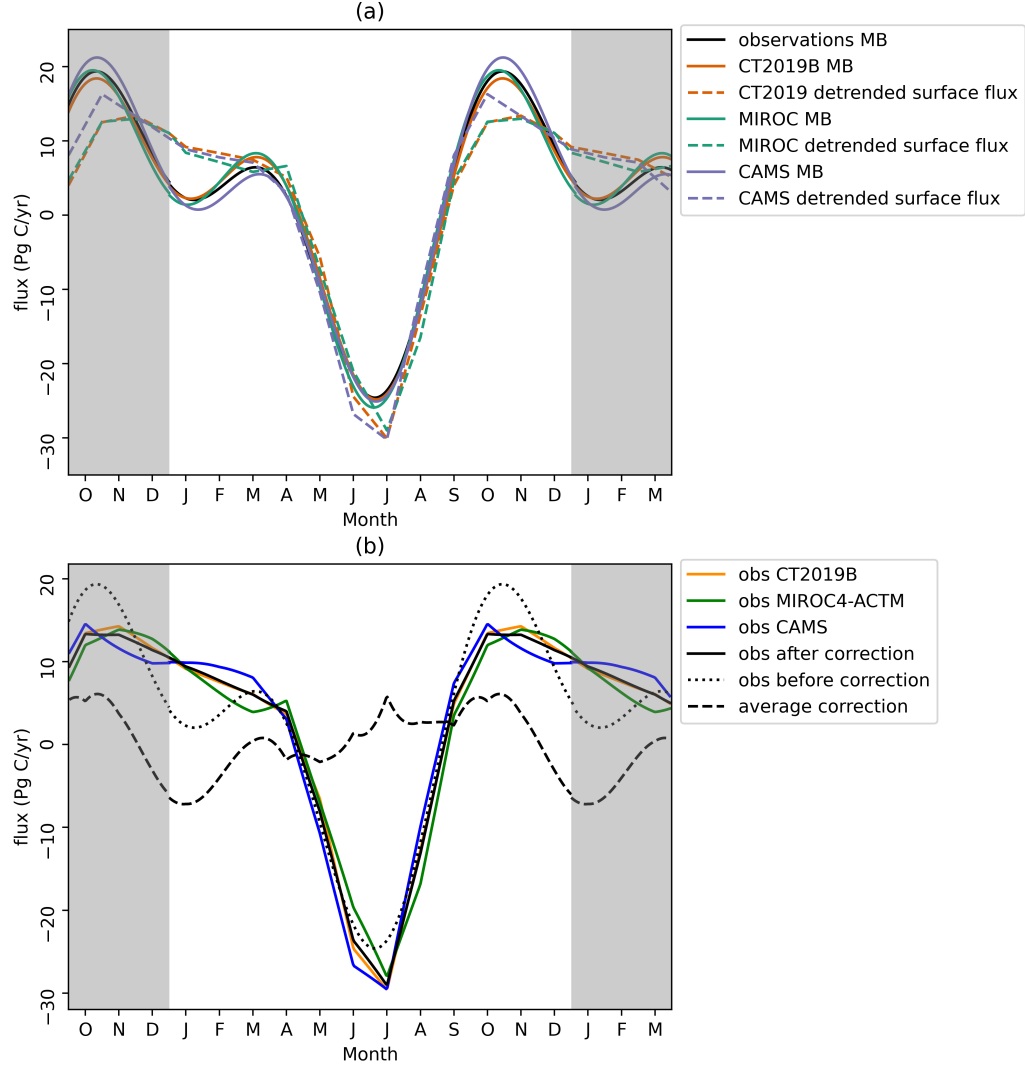


Figure 3 (a) Time derivative of concentrations from observations and inverse models, along with model fluxes. The dashed lines the area weighted average of posterior land fluxes from each inversion system in the domain 20°N-90°N. The solid lines are found by using the carbon dioxide mole-fractions along the flight track for each model to calculate the MB as described in Section 2.2. The solid black line is the time derivative of concentrations using the HIPPO and ATom observations. The estimated flux for the observations is bias corrected by finding the difference between the dotted and solid lines for a given model and applying that difference to the time derivative of the concentration. (b) Estimated flux after bias correction. The

colored lines are found by calibrating using only the model indicated and the solid black line is found using the average correction. The dotted black line is the time derivative of concentration before the correction.

period when detrended atmospheric CO₂ is declining, primarily due to additional uptake by the biosphere as GPP outpaces respiration.

By detrending the observations, this estimate excludes the annual mean flux of CO₂, which itself includes fossil fuel emissions and terrestrial and oceanic sinks. Thus, our estimate of GSNF reflects the seasonal-only component of terrestrial exchange; the actual net uptake by the terrestrial biosphere during the growing season is larger when the annual component (long-term sink) is included. This approach is consistent with prior use of GSNF (e.g. Fung et al., 1983; Yang et al., 2007). Seasonal variations in fossil-fuel emissions and air-sea exchange contribute to seasonal variations in atmospheric CO₂, but these influences are small at about 3% and 5% of land exchange respectively on average for the three inversion systems, and have been removed by our use of the land flux in our model-based bias correction.

2.5 Earth System Models (CMIP5 and CMIP6)

The Coupled Model Intercomparison Project (CMIP) is an international, multi-model research intercomparison project whose purpose is to compare a coordinated set of simulations from ESMs in order to gain a better understanding of our ability to model climate change and associated feedbacks (Friedlingstein et al., 2006). The ESMs that participate simulate relevant physical, chemical, and biological processes within the coupled Earth system (Eyring et al., 2016) using models developed by individual modeling centers worldwide, with the goal of including the most important processes that feedback into the climate system.

Here, we analyze the historical simulations for CMIP5 and CMIP6 (Table 3) that span the period from 1850 to 2005 for CMIP5 and 2014 for CMIP6. We analyze the CO₂-concentration driven historical simulations in which environmental forcing, such as greenhouse gas concentrations and solar forcing, are prescribed. Land and ocean fluxes are allowed to evolve prognostically in response to greenhouse gasses and other forcings (Eyring et al., 2016).

Table 3 Earth system models used in this study. Models in bold are included in the subset used to analyze GPP, RH, and NPP.

Model	Generation	Land Component	Ocean Component	Institution ID
ACCESS-ESM1-5	CMIP6	CABLE2.4	ACCESS-OM2 (MOM5)	CSIRO
CanESM5	CMIP6	CLASS3.6/CTEM1.2	NEMO3.4.1	CCCma
CESM2	CMIP6	CLM5	POP2	NCAR
CESM2-FV2	CMIP6	CLM5	POP2	NCAR
CESM2-WACCM	CMIP6	CLM5	POP2	NCAR
CESM2-WACCM-FV2	CMIP6	CLM5	POP2	NCAR

Model	Generation	Land Component	Ocean Component	Institution ID
CMCC-CM2-SR5	CMIP6	CLM4.5	NEMO3.6	CMCC
CMCC-ESM2	CMIP6	CLM4.5	NEMO3.6	CMCC
GISS-E2-1-G	CMIP6	GISS LSM	GISS	NASA-GISS
GISS-E2-1-H	CMIP6	GISS LSM	HYCOM	NASA-GISS
IPSL-CM6A-LR	CMIP6	ORCHIDEE	NEMO-OPA	IPSL
MPI-ESM-1-2-HAM	CMIP6	JSBACH3.2	MPIOM1.63	HAMMOZ-Co
MPI-ESM1-2-LR	CMIP6	JSBACH3.2	MPIOM1.63	MPI-M
NorCPM1	CMIP6	CLM4	MICOM1.1	NCC
NorESM2-LM	CMIP6	CLM	MICOM	NCC
NorESM2-MM	CMIP6	CLM	MICOM	NCC
TaiESM1	CMIP6	CLM4	POP2	AS-RCEC
CanESM2	CMIP5	CLASS2.7 and CTEM1	CanOM4 and CMOC1.2	CCCma
CCSM4	CMIP5	CLM4	POP2	NCAR
CESM1-BGC	CMIP5	CLM	BEC	NSF-DOE-NO
GFDL-ESM2G	CMIP5	LM3	TOPAZ	NOAA GFDL
HadGEM2-CC	CMIP5	MOSES2 and TRIFFID	HadGOM2	MOHC
HadGEM2-ES	CMIP5	MOSES2 and TRIFFID	HadGOM2	MOHC
INM-CM4.0	CMIP5	-	-	INM
IPSL-CM5A-LR	CMIP5	ORCHIDEE	ORCA2	IPSL
IPSL-CM5A-MR	CMIP5	ORCHIDEE	ORCA2	IPSL
MIROC-ESM	CMIP5	MATSIRO	COCO	MIROC
MIROC-ESM-CHEM	CMIP5	MATSIRO	COCO	MIROC
NorESM1-M	CMIP5	CLM	MICOM	NCC

The northern extratropical seasonal land flux for CMIP5 and CMIP6 models is calculated by removing the long-term annual mean and taking the area-weighted average of each model’s NBP output north of 20°N, linearly interpolating between monthly mean values, similar to the method used on the inversion posterior fluxes (Section 2.3). In addition, we fit a second-order harmonic to the NBP seasonal cycle to find seasonal timing within the models. We define the growing season in each model as the period for which the model simulates net terrestrial uptake, which allows us to evaluate each model’s growing season start and end dates against those inferred from the aircraft observations. In ESMs, NBP reflects the balance of gross photosynthetic uptake, ecosystem respiration, and disturbance and harvest fluxes, and corresponds to the land-atmosphere carbon exchange, making it comparable with our observationally derived flux. We note that ESMs generally do not represent lateral carbon fluxes in rivers, but we expect these to have a relatively minor contribution to our observed seasonal variations. We average multiple years of NBP output from the CMIP models to derive a climatology; for CMIP6, we average over 2009 - 2014, and for CMIP5, we average over 2000-2005 because the output is not available through the HIPPO and ATom timeframe. We evaluate the ensembles as a whole by taking the multi-model mean as it is commonly used. We also evaluate area-

weighted averages of three major component fluxes: GPP, RH, and NPP. Instead of analyzing the fluxes integrated over the growing season, we analyze the fluxes integrated over the entire year for GPP, RH, and NPP, which still correlate with the GSNF and are more useful than the seasonal fluxes when analyzing the carbon budget (e.g. Ballantyne et al., 2015; Tans et al., 1990). For the flux analysis, we use a subset of the models (9 of 12 CMIP5 models and 13 of 17 CMIP6 models, Table 3) for which historical NBP, GPP, RH, and NPP are available.

3 Results

Observations of atmospheric carbon dioxide from the HIPPO and ATom aircraft campaigns were used to estimate a GSNF of 5.7 ± 0.2 Pg C out of the atmosphere north of 20°N averaged over the period 2009-2018 (Fig. 5). This value is equivalent to the net CO_2 exchange between the land and atmosphere during the growing season after removing the annual mean. The growing season is defined to be the period when seasonal fluxes are negative (net uptake by land greater than the annual mean) and occurs between day 117 in late April and day 248 in early September. This corresponds to the day when the curtain average is maximum to the day when the curtain average is minimum (Fig. 2). The flux cycle shows maximum uptake on day 196 (Fig. 3b).

We conducted sensitivity tests to ensure that our choices for the latitudinal and vertical boundaries of our domain and bin size did not have a large influence on the calculated GSNF. Expanding the region north (south) of 20°N by 5°N resulted in a decrease (increase) in the strength of the GSNF of just under 1%. The GSNF was also generally robust to the choice of pressure ceiling for the aircraft observations, increasing by just over 1% when we instead used 350 hPa as a ceiling. The relative standard deviation for all boundary combinations tested (all possible combinations of pressure cutoffs of 300 hPa, 325 hPa, 350 hPa, 375 hPa, and 400 hPa with latitude cutoffs of 20°N and 25°N) was 2%. Similarly, we saw just over a 1% increase in GSNF magnitude for a doubling of latitude or pressure bin size. The relative standard deviation for all bin sizes tested (all possible combinations of pressure bin sizes of 25 hPa, 50 hPa, and 100 hPa with latitude bin sizes of 5° and 10°) was 1%.

Given the small differences that the choice of boundary and bin size make on the magnitude of the calculated GSNF, most of the uncertainty in GSNF results from the transport model bias correction process. We see a spread of 0.2 Pg C or 4% when using CT2019B alone to bias correct vs. using CAMS alone to bias correct. When adding the uncertainty from bin size, boundary choice, and bias correction in quadrature, assuming uncorrelated errors, the 1-sigma error on GSNF is 0.2 Pg C and on the start and end of the growing season is 2 days.

Considering the extensive altitude-latitude coverage of the aircraft observations, the inferred flux represents a unique and robust hemispheric estimate of terrestrial biosphere exchange and its seasonal phasing. We used the observationally inferred GSNF metric to evaluate NBP and its seasonal phasing from the CMIP5

and CMIP6 ensembles (Fig. 5). CMIP6 models (mean GSNF: 5.0 ± 1.6 Pg C, range: 1.5 - 6.9 Pg C) on the whole have less spread than CMIP5 models (mean GSNF: 5.8 ± 2.4 Pg C, range: 2.5 - 10.0 Pg C) (Fig 5). Two of the 17 CMIP6 ensemble members evaluated, CESM2 and CESM2-WACCM, were within 0.2 Pg C of the observed value, and most of the models with a large bias underestimated the GSNF.

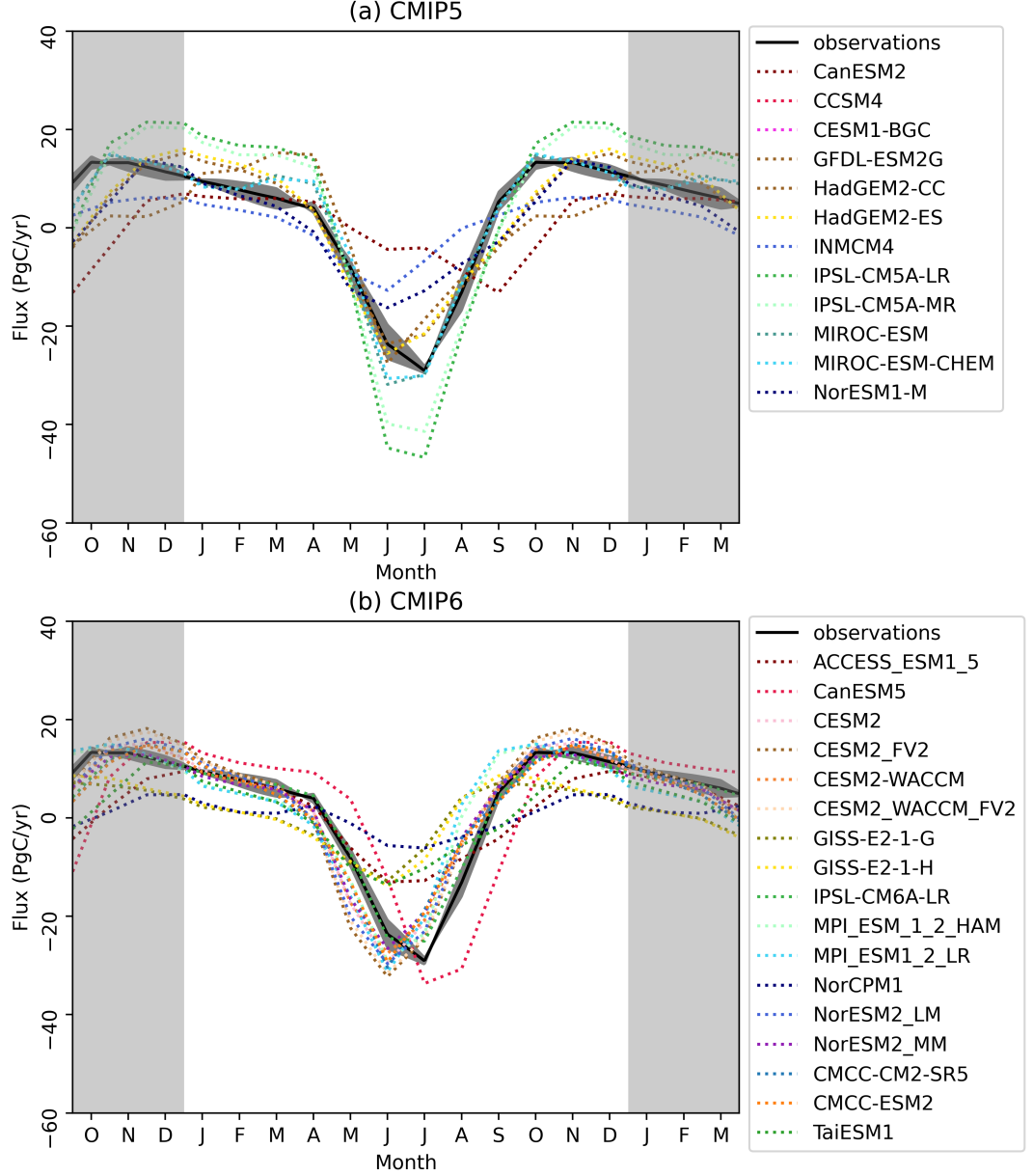


Figure 4 Corrected flux estimated from the HIPPO and ATom campaigns in comparison to area-weighted average NBP in the same domain from the (a) CMIP5 and (b) CMIP6 models. The bias corrected observation error is the standard deviation between correction using the three different inverse models. While the spread in magnitude and timing of the flux in CMIP6 models is smaller than that of CMIP5 models, there is still disagreement between models.

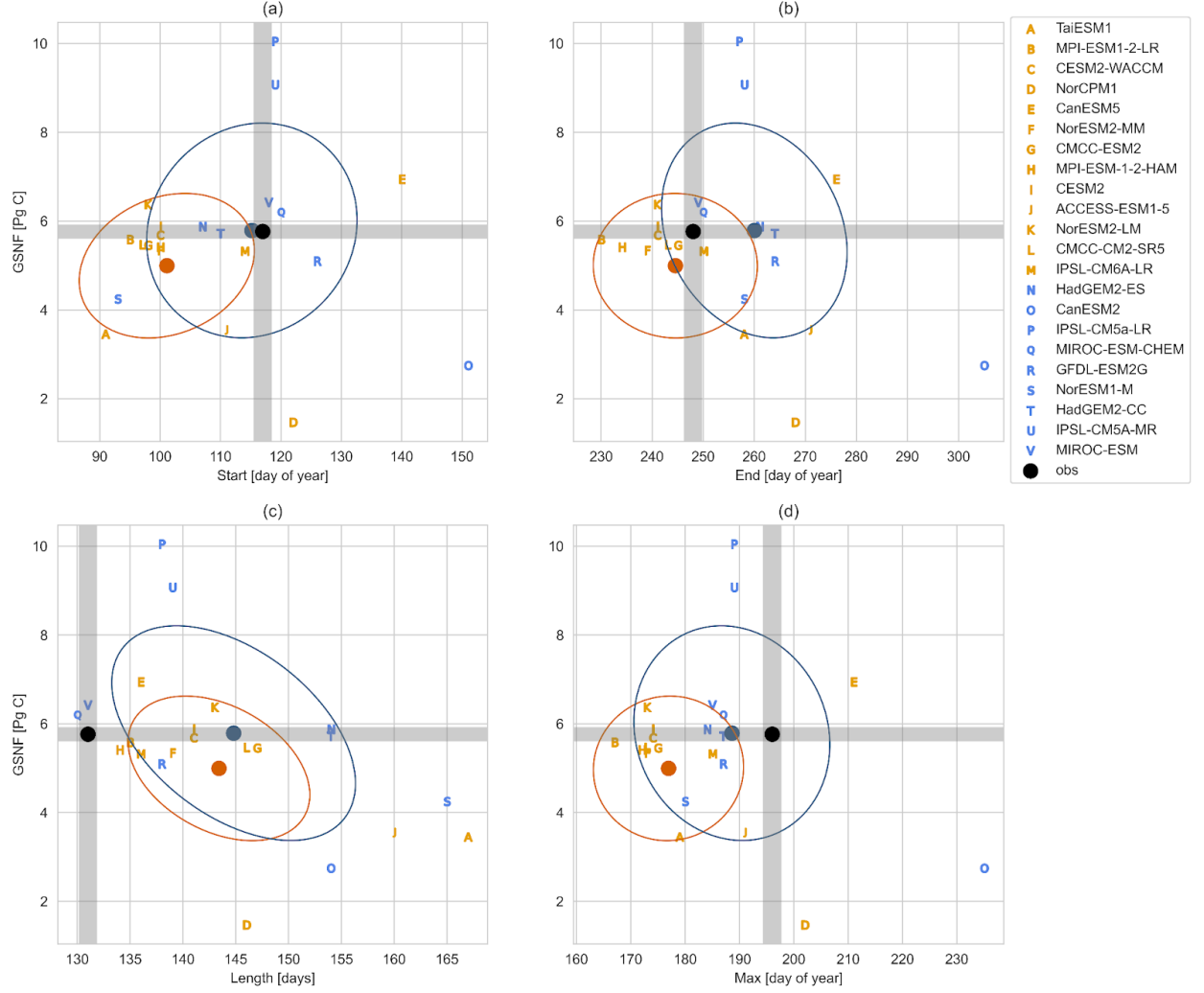


Figure 5 GSNF plotted against the (a) start of the season, defined to be the first day when the seasonal component of atmospheric CO_2 is decreasing (seasonal component of flux changes from positive to negative), (b) end of the season, defined to be the last day when the seasonal component of atmospheric CO_2 is decreasing (seasonal component of flux changes from negative to positive),

(c) length of the season, and (d) max of season, defined to be the day when flux is most negative. The black point is the number inferred from the observations with the gray lines showing uncertainty. CMIP5 models are shown in blue and CMIP6 models are shown in orange. The blue and orange points are the multi-model mean for the CMIP5 and CMIP6 ensembles respectively. The surrounding ellipses show the covariance to one standard deviation. Only models where GPP, RH, and NPP output was available are included.

We note that some modeling centers showed substantial improvement in capturing GSNF between CMIP5 and CMIP6 (Fig. 6) by decreasing the absolute value of their z-score, which indicates how many standard deviations away from the observed value a model falls. For example, the two versions of the IPSL model overestimate GSNF by more than 4 Pg C in CMIP5, which improved substantially in CMIP6 to underestimate the GSNF by less than 1 Pg C. These overestimates are a large reason why the CMIP5 mean is closer to the observed value than the CMIP6 mean. When the IPSL models are excluded from the mean, the CMIP5 ensemble mean decreases to 4.8 Pg C, even smaller than the CMIP6 ensemble mean. In both CMIP5 and CMIP6, the majority of ensemble members underestimate the seasonal flux, falling below the horizontal gray bar in Fig. 5.

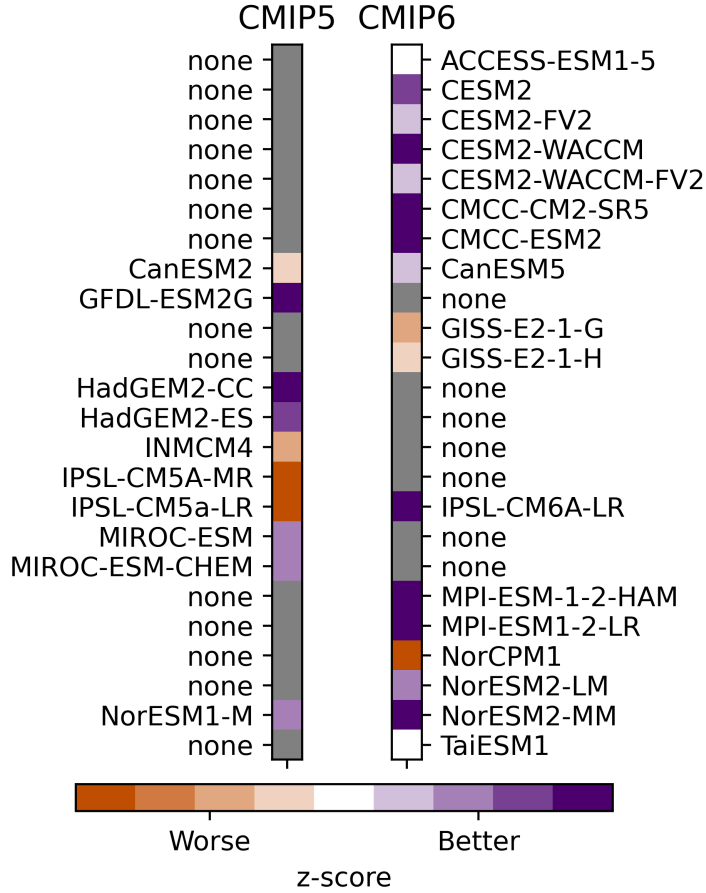


Figure 6 CMIP5 and CMIP6 model absolute value of z-score calculated for all models. The color gray and the label none has been used when one generation of a model is not used or not existent.

A decrease in GSNF model spread in the newer generation has not necessarily led to an improved agreement between models and observations on the phasing of the seasonal cycle. For example, The CMIP5 ensemble mean start day (Julian day 115 ± 17) is closer to the observed start day (Julian day 117 ± 2) than the CMIP6 ensemble mean start day (Julian day 101 ± 14) (Fig. 5a). None of the 17 CMIP6 models evaluated fell within 2 days of the observed start day while three of the 12 CMIP5 models evaluated fell within the 2-day uncertainty range (Fig. 5a). The observations suggest that the growing season onset has become more biased in CMIP6. However, the CMIP6 ensemble did show a smaller bias than CMIP5 for the model-mean end day, which was Julian day 245 ± 16 for CMIP6 and 260 ± 18 for CMIP5, in contrast to day 248 ± 2 in the observations (Fig. 5b). At 143 ± 9 days, the CMIP6 ensemble mean growing season length

is 12 days longer than the observed length of 131 ± 2 . In comparison, the CMIP5 ensemble mean growing season length is 14 days longer at 145 ± 12 days (Fig. 5c). While the CMIP6 ensemble mean end day and growing season length compare more favorably with the observed end day and growing season length than does the CMIP5 ensemble mean values, in both cases, the simulated growing season is longer than what is observed (Fig. 5c).

Seasonal phasing in general does not appear to be correlated with GSNF in the model ensembles, suggesting that phasing is not a dominant driver of GSNF spread among models. No correlation was seen ($r^2 < 0.2$, $p > 0.05$) between GSNF magnitude and the start day, end day, length, or max day across models (Fig. 5). This suggests that factors other than the phasing of the growing season may explain inter-model differences.

We analyzed GPP, RH, and NPP for the subset of CMIP5 and CMIP6 ensemble members which include these outputs to see if these component fluxes might explain model disagreement on GSNF (Fig. 7). We found that in both the CMIP5 and CMIP6 ensembles, GSNF was correlated ($r^2 = 0.68$, $p < 0.05$) with GPP, where models with larger GPP generally had larger GSNF. Models with large GPP also tend toward higher respiration values with an r^2 value of 0.77 between GPP and RH, as GPP provides the inputs to support RH. RH showed a weaker correlation with GSNF than did GPP, however, the correlation is still moderately strong ($r^2 = 0.52$, $p < 0.05$). As expected, models with higher GPP values also tend toward higher NPP values, and the correlation between NPP and GSNF is moderately strong with an r^2 value of 0.66 and a p-value less than 0.05.

We see a large range for GPP, RH, and NPP across the CMIP5 and CMIP6 ensembles. The range in GPP is smaller for the CMIP6 ensemble (mean = 50 Pg C, spread = 37 Pg C) than for the CMIP5 ensemble (mean = 60 Pg C, spread = 47 Pg C). However, this decrease in the model range between ensemble generations is not seen for RH or NPP. The spreads for RH were 23 Pg C (mean = 29 Pg C) for CMIP5 and 27 Pg C (mean = 22 Pg C) for CMIP6. NPP from CMIP5 and CMIP6 had mean values of 32 Pg C and 24 Pg C respectively and a spread of 28 Pg C for both CMIP5 and CMIP6. We note that two models in the CMIP6 ensemble have GSNF values consistent with the observational constraint, MPI-ESM1-2-LR and IPSL-CM6A-LR, but the GPP values spanned by these models are over 10 Pg C, the RH values spanned by these models are 10 Pg C, and the NPP spanned by these models are 9 Pg C.

The strong correlations and large ensemble spread enabled us to indirectly constrain northern extratropical GPP, RH, and NPP through an “emergent constraint” (EC) approach (e.g. Eyring et al., 2019, Williamson et al., 2021; Simpson et al., 2021). ECs are correlations between some observable element X that varies across the ESM ensemble and some important variable Y assuming a physically meaningful relationship exists between X and Y. Here, we assumed GSNF to be X and assumed a physically meaningful relationship between the net flux and its component fluxes GPP, RH, and NPP (Fig. 7). The CMIP5 and

CMIP6 ensembles can be analyzed separately to estimate GPP, RH, and NPP. The CMIP5 ensemble EC estimate for GPP is 58 ± 21 Pg C, for RH is 28 ± 11 Pg C, and for NPP is 31 ± 11 Pg C where the error is the 95% prediction interval from a hypothetical sample generated using a Monte Carlo simulation of $n = 10000$ and assuming a Gaussian distribution. Similarly, the CMIP6 ensemble EC estimate for GPP is 54 ± 11 Pg C, for RH is 24 ± 11 Pg C, and for NPP is 26 ± 9 Pg C. However, we can also combine the two ensembles to create a larger sample size as the estimates agree within error and the modeled relationship does not change substantially between the two generations. The combined EC estimate for GPP is 56 ± 15 Pg C, for RH is 25 ± 11 Pg C, and for NPP is 28 ± 10 Pg C. Although we did not constrain autotrophic respiration (RA) explicitly, the constraint on GPP and NPP imply RA is 28 ± 18 Pg C where the error has been propagated by summing the errors in quadrature. Applying an emergent constraint to RA explicitly would likely give smaller errors.

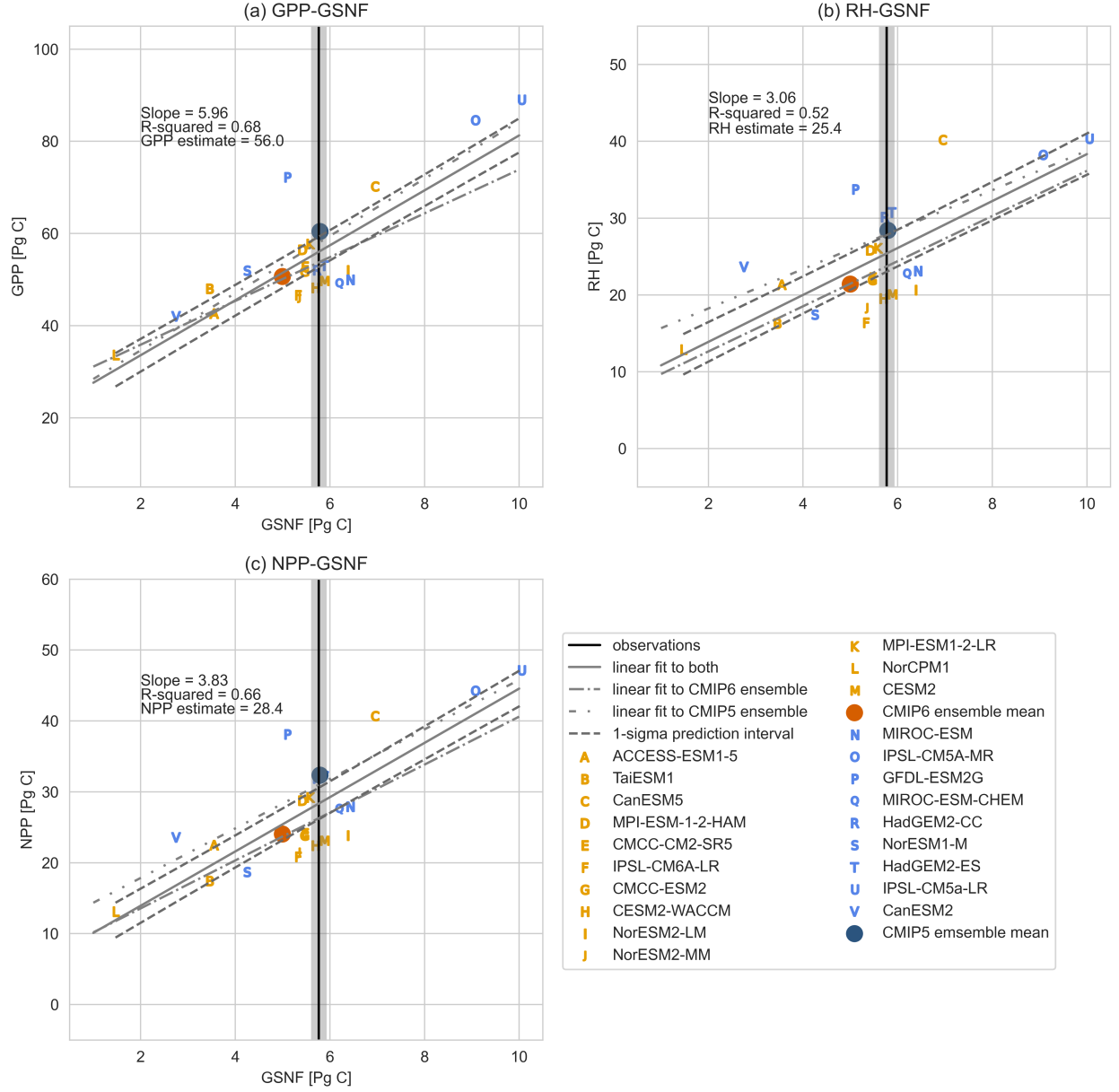


Figure 7 GSNF plotted as a function of (a) integrated GPP and (b) integrated RH, and (c) integrated NPP for the CMIP5 and CMIP6 models. The estimated GSNF from the HIPPO and ATom observations is shown in gray, CMIP5 models are shown in blue, and CMIP6 models are shown in orange. Only models where GPP, RH, and NPP data was available are included.

4 Discussion

Here we derive an observational constraint on northern extratropical GSNF from two novel aircraft campaigns that measured the atmospheric CO_2 curtain over remote oceans. Our estimated GSNF of $5.7 \pm 0.2 \text{ Pg C}$ is significantly smaller in magnitude than the GSNF of 7.9 Pg C yr^{-1} out of the atmosphere estimated north of 30°N by Yang et al. (2007) based on total column observations and spatially sparse aircraft profiles in North America. These differences are most likely due to differences in methodology, as Yang et al. used a combination of total column and aircraft measurements at 8 locations between 30°N and 70°N to scale fluxes from a terrestrial ecosystem model (TEM), while we used aircraft data that sampled nearly continuously between our cutoff of 20°N and roughly 87°N to estimate GSNF without using a specific TEM. When our cutoff is changed from 20°N to 30°N , our estimate remains as 5.8 Pg C . Fung et al. (1983) quantified GSNF in the Northern Hemisphere using a three-dimensional tracer transport model. When our cutoff is changed to the Equator, our estimate increases to 6.1 Pg C , which falls within the $3.4 - 10.7 \text{ Pg C}$ range given by Fung et al. (1983). However, the range given by Fung et al. (1983) is large. The uncertainty on northern hemisphere GSNF could likely be reduced with application of multiple transport models to existing records.

We may expect to see an increase in GSNF over time due to the observed increase in the CO_2 seasonal cycle amplitude (SCA). Graven et al. (2013) saw an increase of 32-59 % in the NH seasonal cycle amplitude in the 50-year period from 1958-1963 to 2009-2011. This equates to an increase of $0.56 - 0.93 \text{ \% yr}^{-1}$. If this trend continued through the HIPPO and ATom time period, we would expect to see a 4.0 - 6.7% increase across the 7 years between these two missions. Applying the methodology to HIPPO alone, which covers 2009-2011, gives a value of 11.2 ppm for the SCA of the curtain average versus 10.7 ppm when using ATom alone, which covers 2016-2018. This corresponds to a more than 1% decrease, contrary to the increase expected from Graven. However, this may not be accurate as there may not be enough data in either mission alone to fully constrain the seasonal cycle (Fig. S3) and account for interannual variability in CO_2 fluxes. Jin et al. (2021) showed that interannual variability in the hemispheric carbon inventory is not negligible. Here, we assume the bias-correction process accounts for interannual variability, and further note that the interannual variability, calculated as the standard deviation of the GSNF estimated from the average of the three inversion fluxes for each year over the period 2009-2018, is less than 0.2 Pg C or just over 3%.

Despite any interannual variability, the expected increase in GSNF is not apparent when comparing the flight-based estimates to previous estimates. Yang et al. (2007) estimate a larger GSNF despite covering an earlier time period. The large range given by Fung et al. (1983) means that the increase in GSNF may be seen if the true GSNF in 1982 fell within the range of $4.6 - 5.2 \text{ Pg C}$. Looking at surface records, Keeling and Graven (2021) saw an increase in the amplitudes at Mauna Loa and Barrow between 2010 and 2017 equating to 10.9% at Mauna Loa and 1.1% at Barrow using 5-year running means, suggesting possibly a shift to less SCA growth at high latitudes.

The HIPPO and ATom observations reflect the atmospheric mass of carbon integrated over a large latitudinal and altitudinal extent, thus, our estimate is less sensitive to specific representations of atmospheric transport than are other estimates. For example, the GSNF estimated over our domain bias correcting with the three inverse models individually only varied by just 0.2 Pg C or less than 4% (Fig. 3b), even while previous research has reported the relative spread of 10 simulations for the posterior annual mean northern extratropical land flux to be 13 % (Gaubert et al., 2019). Additionally, inverse models seem to be converging on the land flux as the GSNF estimated from the posterior inversion land fluxes averaged 5.8 Pg C, slightly larger than the observationally based estimate of 5.7 Pg C, and varied by just 0.2 Pg C. The small differences among the inverse models at the hemispheric scale mean that even though it represented the largest source of uncertainty, the bias correction process imparted minimal uncertainty on our ultimate GSNF value.

To further reduce uncertainty, use of the transformed coordinate, M_e , introduced by Jin et al. (2021) when integrating to find the curtain average may largely eliminate the contribution of sparse sampling and synoptic variability in calculating curtain averages.

The low uncertainty on the hemispheric integral makes GSNF and its phasing robust targets for evaluating TEMs and the land components of ESMs. Although direct comparison with atmospheric CO_2 mole fraction has been used to evaluate ESMs previously, these comparisons rely on simulation of the three-dimensional atmospheric CO_2 mole fraction within the ESM (e.g., Keppel-Aleks et al., 2013) or require estimating CO_2 using an offline transport model or operator (e.g., Liptak et al, 2017). In contrast, our GSNF constraint can be used to evaluate ESMs at the flux level rather than relying on comparisons at the concentration level. Compared to our metric, the CMIP6 ensemble has shown some improvements relative to the CMIP5 ensemble, namely a reduction in the spread between models (5.5 Pg C vs. 7.3 Pg C) and a more favorable simulation of the timing of the end of the growing season (3 days early vs. 12 days late) when considering the ensemble mean values (Fig. 5 b, c). However, there are still large disagreements, and in some ways, CMIP5 models perform better in relation to the observationally inferred flux than do CMIP6 models. For example, CMIP5 models outperform CMIP6 models in simulating the start of the growing season (2 days early vs. 16 days early) and the GSNF magnitude (5.8 ± 2.4 Pg C vs 5.0 ± 1.6 Pg C compared to observational value of 5.7 ± 0.2 Pg C) when considering the ensemble mean values (Fig. 5). The models tend to underestimate the magnitude of GSNF on the whole with 4 of the 10 CMIP5 models and 10 of the 17 CMIP6 models underestimating the flux (falling below the horizontal gray lines in Figure 5), 2 CMIP5 and 3 CMIP6 models falling within the uncertainty range, and only 4 CMIP5 models and 4 CMIP6 models overestimating the net flux. This is consistent with previous results from Keppel-Aleks et al. (2013) suggesting the Community Land Model version 4 (CLM4), a version of which is used as the land model for over one-third of the ESMs evaluated, underestimated GSNF. Three of the models with the largest underestimate

of GSNF (NorCPM1, NorESM1-M, and TaiESM1) utilize CLM4 as their land model, however, we note that two of the best performing models (CESM2 and CESM2-WACCM) both utilize the Community Land Model version 5 (CLM5) as their land component, reflecting major improvement in CLM5 compared to CLM4 in simulating CO₂ seasonality (Lawrence et al., 2019). Correlations between GSNF and seasonal phasing were weak or nonexistent, meaning that the improvements in CLM5 are more likely due to improvements in simulations of the overall magnitude of photosynthesis and respiration than to improvements in simulations of the timing of the growing season.

The chosen methods to define the growing season have some effect on model-observation comparisons. We could have chosen to prescribe the growing season dates for calculating model metrics from the observations rather than allowing the dates to vary from model to model. The primary difference in the results using this approach is a slight reduction in the magnitude of the GSNF for most of the CMIP5 and CMIP6 models, a less strong correlation between GSNF and GPP, RH, and NPP, and a slight increase in the inferred values of northern GPP, RH, and NPP (Fig S2). Another choice that affects model-observation comparisons is the use of the ensemble mean rather than the ensemble median as representative of the ensemble. Use of the median may be preferable because outliers exist in both the CMIP5 and CMIP6 ensembles, however, the mean was used as it is a common approach used in model evaluation and evidence exists that the multi-model mean provides the best comparison to observations (Lambert and Boer, 2001). For the CMIP6 ensemble, the median is 6.0 Pg C and for the CMIP5 ensemble, the median is 8.0 Pg C, indicating that CMIP6 models may better capture the GSNF than CMIP5 models do.

Although the aircraft CO₂ data provide a constraint on GSNF, this benchmark is also useful for constraining component fluxes that are difficult to infer observationally at large spatial scales. We examine the GPP, RH, and NPP of the ESM ensembles in relation to GSNF. GPP, RH, and NPP are components of the net flux, and we see moderate correlations between simulated GSNF and simulated productivity (both GPP and NPP) and RH. It is noteworthy that correlations between GSNF and the various annual fluxes are stronger than those between GSNF and seasonal timing, which may indicate that the magnitudes of photosynthesis and respiration are more dominant drivers of GSNF than seasonal phasing (e.g., Valentini et al., 2000, Baldocchi et al., 2017). Cadule et al. (2010) analyzed three CMIP models and concluded that models generally underestimate the seasonal amplitude due to shortcomings in vegetation phenology and heterotrophic respiration response to climate. Our results generally support this role for discrepancies in component fluxes as driving a discrepancy in the resulting GSNF.

These correlations between GSNF and its component fluxes provide an opportunity to constrain GPP, RH, and NPP using an emergent constraint approach. Previous estimates of northern extratropical GPP are highly uncertain, with large disagreements between estimates. For example, Mao et al., (2012) used

the Moderate Resolution Imaging Spectroradiometer (MODIS) GPP product to estimate Northern Hemisphere GPP averaged over 2000-2009 as 64.75 ± 0.97 Pg C yr⁻¹. In contrast, the FluxCom product, based on upscaled observations from FLUXNET sites using various machine learning approaches (Jung et al., 2019), estimated GPP north of 20°N averaged over 2009-2014 to be 48 Pg C yr⁻¹. Our GSNF-constrained value of 56 ± 15 Pg C for GPP is consistent with both of these estimates given the larger error bars on the emergent constraint approach and indicates that GPP falls between these two estimates. The large value for GPP relative to Fluxcom is interesting in light of the uncertainty in global GPP. The range in global mean GPP magnitudes for 2008–2010 from FLUXCOM members is 106 - 130 Pg C yr⁻¹ (Jung et al., 2020). This range covers the observation-based estimate of global mean GPP of 123 ± 8 Pg C yr⁻¹ found using eddy covariance flux data and various diagnostic models (Beer et al., 2010) but is smaller than the GPP magnitudes of 150–175 Pg C yr⁻¹ derived from an isotope-based study (Welp et al., 2011). Our results may indicate a higher global mean GPP than the flux tower upscaling yields; however, we are constraining extratropical GPP which may be affected by different factors than tropical GPP. These results therefore need to be interpreted with caution.

5 Conclusions

Earth system models disagree in their simulation of large-scale carbon fluxes, making it crucial to evaluate models to contextualize their climate predictions. We have presented an approach to constrain the seasonal land flux using aircraft data from the HIPPO and ATom flight campaigns. The northern extratropical GSNF of 5.7 ± 0.2 Pg C derived from these observations can be used to evaluate prognostic model fluxes for net flux, directly, and for component fluxes via an emergent constraint approach. We note that this constraint is tied to atmospheric transport models because the flux is bias corrected by comparing posterior mole fractions from inverse models with their fluxes. We found, however, that at the hemispheric scale, the constraint is robust to the choice of atmospheric transport model since common transport errors tend to cancel out at this scale.

When compared to CMIP5 and CMIP6 models, the inferred GSNF suggests a larger net flux and shorter growing season than simulated in both model ensembles. This gives modelers an additional observation to target during model development and could be added to a benchmarking system such as the International Land Model Benchmarking (ILAMB) System (Collier et al., 2018). While there is decreased model spread between CMIP6 models, this benchmark also highlights some of the ways in which CMIP6 models have not improved from CMIP5, such as in simulating the start of the growing season.

Correlations within the CMIP5 and CMIP6 ensembles allowed us to apply an emergent constraint approach to estimate northern hemisphere GPP, RH, and NPP. We found that the GSNF-constrained value for GPP is at the higher end of a commonly used upscaling product.

Overall, the HIPPO and ATom inferred GSNF provides a robust metric as it allows for the evaluation of large-scale fluxes in flux space and sheds light on component fluxes, filling a need highlighted by Collier et al., (2018) for land model benchmarking. More regular global scale airborne tomography could resolve GSNF at higher time resolution and leverage carbon cycle interannual variability for improved tests of ESM process representations.

Acknowledgments

This research was partially supported by the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Science Focus Area, which is sponsored by the Regional and Global Model Analysis (RGMA) activity of the Earth & Environmental Systems Modeling (EESM) Program in the Earth and Environmental Systems Sciences Division (EESSD) of the Office of Biological and Environmental Research (BER) in the US Department of Energy Office of Science.

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. The HIPPO data were collected using NSF’s Lower Atmosphere Observing Facilities, which are managed and operated by NCAR’s Earth Observing Laboratory.

We thank the HIPPO science team and the flight crew and support staff of the NCAR Research Aviation Facility. We thank the ATom science team and the flight crew and support staff of the NASA DC-8, which is supported by the NASA Airborne Science Program and Earth Science Project Office. For additional support of HIPPO and ATom airborne CO₂ and N₂O measurements, we thank S. Afshar, J. Bent, J. Budney, B. Daube, G. Diskin, J. Elkins, Y. Gonzalo Ramos, J. Higgs, E. Hintsa, R. Jimenez, E. Kort, P. Lang, T. Legard, E. Moglia, S. Montzka, F. Moore, D. Neff, T. Newberger, B. Paplawsky, S. Park, J. Pittman, G. Santoni, A. Watt, S. Wofsy, S. Wolter, and B. Xiang.

We would like to thank the climate modeling groups (listed in Table 2 and 3 of this paper) for producing and making freely available their model output.

PKP is partly supported by the Arctic Challenge for Sustainability II grant (JPMXD1420318865) of the Ministry of Education, Science, Culture and Sports (MEXT), Japan.

Open Research

HIPPO data are available at the Earth Observing Laboratory (EOL) HIPPO data archive at <http://www.eol.ucar.edu/projects/hippo/>. (http://doi.org/10.3334/CDIAC/HIPPO_010).

ATom data are available at the Oak Ridge National Laboratory DAAC at <https://daa.ornl.gov/ATOM/campaign/>. (<https://doi.org/10.3334/ORNLDAAAC/1925>).

CarbonTracker CT2019B results provided by NOAA ESRL, Boulder, Colorado, USA from the website at <http://carbontracker.noaa.gov>. (<https://doi.org/10.25925/20201008>)

CMIP5 and CMIP6 model output files are freely available from the Earth System Grid Federation data replication centers and used under creative commons license CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0/>.

References

Hiroshi Akima, "A Method of Bivariate Interpolation and Smooth Surface Fitting for Irregularly Distributed Data Points", ACM Transactions on Mathematical Software, Vol. 4, No. 2, June 1978, pp. 148-159. Copyright 1978, Association for Computing Machinery, Inc., reprinted by permission.

Alexandrov, G. A. (2014). Explaining the seasonal cycle of the globally averaged CO₂ with a carbon-cycle model. *Earth System Dynamics*, 5(2). <https://doi.org/10.5194/esd-5-345-2014>

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzer, E., Kawamiya, M., Koven, C. D., ... Ziehn, T. (2020). Carbon-concentration and carbon-climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, 17(16), 4173–4222. <https://doi.org/10.5194/bg-17-4173-2020>

Baldocchi, D., Chu, H., & Reichstein, M. (2018). Inter-annual variability of net and gross ecosystem carbon fluxes: A review. *Agricultural and Forest Meteorology*, 249, 520–533. <https://doi.org/10.1016/j.agrformet.2017.05.015>

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*. <https://doi.org/10.1038/nature11299>

Ballantyne, A. P., Andres, R., Houghton, R., Stocker, B. D., Wanninkhof, R., Anderegg, W., Cooper, L. A., Degrandpre, M., Tans, P. P., Miller, J. B., Alden, C., & White, J. W. C. (2015). Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty. *Biogeosciences*, 12, 2565–2584. <https://doi.org/10.5194/bg-12-2565-2015>

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Rouspard, O., ... Papale, D. (2010). Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science*, 329(5993), 834–838. <https://doi.org/10.1126/science.1184984>

Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., Hinton, T. J., et al. (2011). The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, 4(3), 723–757. <https://doi.org/10.5194/gmd-4-723-2011>

Bent, J.: Airborne oxygen measurements over the Southern Ocean as an integrated constraint of seasonal biogeochemical processes, PhD thesis, University of California, San Diego, USA, 2014.

Bentsen, Mats; Olivieri, Dirk Jan Leo; Seland, Øyvind; Toniazzi, Thomas; Gjermundsen, Ada; Graff, Lise Seland; Debernard, Jens Boldingh; Gupta, Alok Kumar; He, Yanchun; Kirkevåg, Alf; Schwinger, Jörg; Tjiputra, Jerry; Aas, Kjetil Schanke; Bethke, Ingo; Fan, Yuanchao; Griesfeller, Jan; Grini, Alf; Guo, Chungheng; Ilicak, Mehmet; Karset, Inger Helene Hafsahl; Landgren, Oskar Andreas; Liakka, Johan; Moseid, Kine Onsum; Nummelin, Aleks; Spensberger, Clemens; Tang, Hui; Zhang, Zhongshi; Heinze, Christoph; Iversen, Trond; Schulz, Michael (2019). NCC NorESM2-MM model output prepared for CMIP6 CMIP historical. *Version 20191126* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.8040> .

Bethke, Ingo; Wang, Yiguo; Counillon, François; Kimmritz, Madlen; Fransner, Filippa; Samuelsen, Annette; Langehaug, Helene Reinertsen; Chiu, Ping-Gin; Bentsen, Mats; Guo, Chungheng; Tjiputra, Jerry; Kirkevåg, Alf; Olivieri, Dirk Jan Leo; Seland, Øyvind; Fan, Yuanchao; Lawrence, Peter; Eldevik, Tor; Keenlyside, Noel (2019). NCC NorCPM1 model output prepared for CMIP6 CMIP historical. *Version 20190921* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.10894> .

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, 12(7). <https://doi.org/10.1029/2019MS002010>

Boucher, Olivier; Denvil, Sébastien; Levvasseur, Guillaume; Cozic, Anne; Caubel, Arnaud; Foujols, Marie-Alice; Meurdesoif, Yann; Cadule, Patricia; Devilliers, Marion; Ghattas, Josefine; Lebas, Nicolas; Lurton, Thibaut; Mellul, Lidia; Musat, Ionela; Mignot, Juliette; Cheruy, Frédérique (2018). IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical. *Version 20180711* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.5195> .

Cadule, P., Friedlingstein, P., Bopp, L., Sitch, S., Jones, C. D., Ciais, P., et al. (2010). Benchmarking coupled climate-carbon models against long-term atmospheric CO₂ measurements. *Global Biogeochemical Cycles*, 24(2). <https://doi.org/10.1029/2009GB003556>

Josep G. Canadell, Pedro M.S. Monteiro, Marcos H. Costa, Leticia Cotrim Da Cunha, Peter M. Cox, et al.. Global Carbon and other Biogeochemical Cycles and Feedbacks. IPCC AR6 WGI, Final Government Distribution, chapter 5, 2021. hal-03336145

Chandra, N., Patra, P. K., Niwa, Y., Ito, A., Iida, Y., Goto, D., Morimoto, S., Kondo, M., Takigawa, M., Hajima, T., and Watanabe, M.: Estimated regional CO₂ flux and uncertainty based on an ensemble of atmospheric CO₂ inversions,

Atmos. Chem. Phys., in press, 2022.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., et al. (2018). Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *Journal of Advances in Modeling Earth Systems*, 2018MS001369. <https://doi.org/10.1029/2018MS001369>

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F. M., et al. (2005). Inferring CO₂ sources and sinks from satellite observations: Method and application to TOVS data. *Journal of Geophysical Research Atmospheres*, 110(24), 1–13. <https://doi.org/10.1029/2005JD006390>

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., & Ramonet, M. (2010). Atmospheric inversions for estimating CO₂ fluxes: Methods and perspectives. *Climatic Change*. <https://doi.org/10.1007/s10584-010-9909-3>

Cleveland, R.B. and Cleveland, W.S. (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6, 3-33.

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al. (2018). The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation. *Journal of Advances in Modeling Earth Systems*. <https://doi.org/10.1029/2018MS001354>

Cooperative Global Atmospheric Data Integration Project. (2019). Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2018; obspack_co2_1_GLOBALVIEWplus_v5.0_2019_08_12 [Data set]. NOAA Earth System R

Danabasoglu, G., Lamarque, J. -F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2). <https://doi.org/10.1029/2019MS001916>

Danabasoglu, Gokhan (2019). NCAR CESM2 model output prepared for CMIP6 CMIP historical. *Version 20190116* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.7627>

Danabasoglu, Gokhan (2019). NCAR CESM2-FV2 model output prepared for CMIP6 CMIP historical. *Version 20190731* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.11297> .

Danabasoglu, Gokhan (2019). NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical. *Version 20190131* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.10071> .

Danabasoglu, Gokhan (2019). NCAR CESM2-WACCM-FV2 model output prepared for CMIP6 CMIP historical. *Version 20190831* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.11298> .

Daube, B. C., Boering, K. A., Andrews, A. E., & Wofsy, S. C. (2002). A High-Precision Fast-Response Airborne CO₂ Analyzer for In Situ

Sampling from the Surface to the Middle Stratosphere. *Journal of Atmospheric and Oceanic Technology*, 19(10). [https://doi.org/10.1175/1520-0426\(2002\)019<1532:AHPFRA>2.0.CO;2](https://doi.org/10.1175/1520-0426(2002)019<1532:AHPFRA>2.0.CO;2)

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics*, 40(9–10), 2123–2165. <https://doi.org/10.1007/s00382-012-1636-1>

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., et al. (2013). GFDL’s ESM2 Global Coupled Climate–Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics*. *Journal of Climate*, 26(7), 2247–2267. <https://doi.org/10.1175/JCLI-D-12-00150.1>

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. *Geoscientific Model Development*. <https://doi.org/10.5194/gmd-9-1937-2016>

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., ... Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, 9(2), 102–110. <https://doi.org/10.1038/s41558-018-0355-y>

Fernández-Martínez, M., Sardans, J., Chevallier, F., Ciais, P., Obersteiner, M., Vicca, S., et al. (2019). Global trends in carbon sinks and their relationships with CO₂ and temperature. *Nature Climate Change*, 9(1). <https://doi.org/10.1038/s41558-018-0367-7>

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., et al. (2006). Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *Journal of Climate*, 19(14), 3337–3353. <https://doi.org/10.1175/JCLI3800.1>

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27(2), 511–526. <https://doi.org/10.1175/JCLI-D-12-00579.1>

Friedlingstein, P., Jones, M. W., O’Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., ... Zeng, J. (2022). Global Carbon Budget 2021. *Earth System Science Data*, 14(4), 1917–2005. <https://doi.org/10.5194/essd-14-1917-2022>

Friend, A. D., Arneth, A., Kiang, N. Y., Lomas, M., Ogée, J., Rödenbeck, C., et

- al. (2007). FLUXNET and modelling the global carbon cycle. *Global Change Biology*. <https://doi.org/10.1111/j.1365-2486.2006.01223.x>
- Fung, I. Y., Doney, S. C., Lindsay, K., & John, J. (2005). Evolution of carbon sinks in a changing climate. *Proceedings of the National Academy of Sciences*, 102(32). <https://doi.org/10.1073/pnas.0504949102>
- Fung, I., Prentice, K., & Matthews, E. (1983). Three-dimensional tracer model study of atmospheric CO₂: Response to seasonal exchanges with the terrestrial biosphere. *Journal of Geophysical Research*, 88(C2), 1281–1294. <https://doi.org/10.1029/JC088iC02p01281>
- Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., Patra, P. K., Peters, W., Rödenbeck, C., Saeki, T., Schimel, D., van der Laan-Luijkx, I., Wofsy, S., & Yin, Y. (2019). Global atmospheric CO₂ inverse models converging on neutral tropical land exchange but diverging on fossil fuel and atmospheric growth rate. *Biogeosciences Discussions*, 1–25. <https://doi.org/10.5194/bg-2018-384>
- Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., et al. (2011). The Community Climate System Model Version 4. *Journal of Climate*, 24(19), 4973–4991. <https://doi.org/10.1175/2011JCLI4083.1>
- Graven, H. D., Keeling, R. F., Piper, S. C., Patra, P. K., Stephens, B. B., Wofsy, S. C., et al. (2013). Enhanced Seasonal Exchange of CO₂ by Northern Ecosystems Since 1960. *Science*, 341(6150). <https://doi.org/10.1126/science.1239207>
- Guo, C., Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., et al. (2019). Description and evaluation of NorESM1-F: a fast version of the Norwegian Earth System Model (NorESM). *Geoscientific Model Development*, 12(1), 343–362. <https://doi.org/10.5194/gmd-12-343-2019>
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., et al. (2002). Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models. *Nature*. <https://doi.org/10.1038/415626a>
- Hu, Q., Li, T., Deng, X., Wu, T., Zhai, P., Huang, D., Fan, X., Zhu, Y., Lin, Y., Xiao, X., Chen, X., Zhao, X., Wang, L., & Qin, Z. (2022). Intercomparison of global terrestrial carbon fluxes estimated by MODIS and Earth system models. *Science of The Total Environment*, 810, 152231. <https://doi.org/10.1016/j.scitotenv.2021.152231>
- Jacobson, A. R., Schuldt, K. N., Miller, J. B., Oda, T., Tans, P., Arlyn Andrews, Mund, J., Ott, L., Collatz, G. J., Aalto, T., Afshar, S., Aikin, K., Aoki, S., Apadula, F., Baier, B., Bergamaschi, P., Beyersdorf, A., Biraud, S. C., Bollenbacher, A., ... Mirosław Zimnoch. (2020). CarbonTracker CT2019B. NOAA Global Monitoring Laboratory. <https://doi.org/10.25925/20201008>
- Jin, Y., Keeling, R. F., Morgan, E. J., Ray, E., Parazoo, N. C., and Stephens, B. B. (2021). A mass-weighted isentropic coordinate for mapping chemical tracers

- and computing atmospheric inventories. *Atmospheric Chemistry and Physics*, 21(1), 217–238. <https://doi.org/10.5194/acp-21-217-202>
- Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., et al. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. [Dataset] *Scientific Data*, 6(1). <https://doi.org/10.1038/s41597-019-0076-8>
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler, P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., ... Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, 17(5), 1343–1365. <https://doi.org/10.5194/bg-17-1343-2020>
- Keeling, C. D., Chin, J. F. S., & Whorf, T. P. (1996). Increased activity of northern vegetation inferred from atmospheric CO₂ measurements. *Nature*, 382(6587). <https://doi.org/10.1038/382146a0>
- Keeling, C. D., Adams, J. A., Ekdahl, C. A., & Guenther, P. R. (1976). Atmospheric carbon dioxide variations at the South Pole. *Tellus*, 28(6), 552–564. <https://doi.org/10.1111/j.2153-3490.1976.tb00702.x>
- Keeling, R. F., & Manning, A. C. (2014). Studies of Recent Changes in Atmospheric O₂ Content. In *Treatise on Geochemistry*. Elsevier. <https://doi.org/10.1016/B978-0-08-095975-7.00420-4>
- Keeling, R. F., & Graven, H. D. (2021). Insights from Time Series of Atmospheric Carbon Dioxide and Related Tracers. *Annual Review of Environment and Resources*, 46(1), 85–110. <https://doi.org/10.1146/annurev-environ-012220-125406>
- Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., et al. (2020). GISS-E2.1: Configurations and Climatology. *Journal of Advances in Modeling Earth Systems*, 12(8). <https://doi.org/10.1029/2019MS002025>
- Keppel-Aleks, G., Randerson, J. T., Lindsay, K., Stephens, B. B., Keith Moore, J., Doney, S. C., et al. (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*. <https://doi.org/10.1175/JCLI-D-12-00589.1>
- Khatiwala, S., Primeau, F., & Hall, T. (2009). Reconstruction of the history of anthropogenic CO₂ concentrations in the ocean. *Nature*, 462(7271). <https://doi.org/10.1038/nature08526>
- Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., & Bergamaschi, P. (2005). The two-way nested global chemistry-transport zoom model TM5: algorithm and applications. *Atmospheric Chemistry and Physics*, 5(2), 417–432. <https://doi.org/10.5194/acp-5-417-2005>

- Kumar, J., Hoffman, F., Hargrove, W., & Collier, N. (2016). Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements. *Earth System Science Data Discussions*, 1–25. <https://doi.org/10.5194/essd-2016-36>
- Lambert, S., Boer, G. CMIP1 evaluation and intercomparison of coupled climate models. *Climate Dynamics* 17, 83–106 (2001). <https://doi.org/10.1007/PL00013736>
- Lee, W.-L., Wang, Y.-C., Shiu, C.-J., Tsai, I., Tu, C.-Y., Lan, Y.-Y., et al. (2020). Taiwan Earth System Model Version 1: description and evaluation of mean state. *Geoscientific Model Development*, 13(9), 3887–3904. <https://doi.org/10.5194/gmd-13-3887-2020>
- Lee, Wei-Liang; Liang, Hsin-Chien (2020). AS-RCEC TaiESM1.0 model output prepared for CMIP6 CMIP historical. *Version 20200616* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.9755> .
- Liptak, J., Keppel-Aleks, G., & Lindsay, K. (2017). Drivers of multi-century trends in the atmospheric CO₂ mean annual cycle in a prognostic ESM. *Biogeosciences*, 14(6), 1383–1401. <https://doi.org/10.5194/bg-14-1383-2017>
- Long, M. C., Lindsay, K., Peacock, S., Moore, J. K., & Doney, S. C. (2013). Twentieth-Century Oceanic Carbon Uptake and Storage in CESM1(BGC)*. *Journal of Climate*, 26(18), 6775–6800. <https://doi.org/10.1175/JCLI-D-12-00184.1>
- Lovato, Tomas; Peano, Daniele (2020). CMCC CMCC-CM2-SR5 model output prepared for CMIP6 CMIP historical. *Version 20200527* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.3825> .
- Lovato, Tomas; Peano, Daniele; Butenschön, Momme (2021). CMCC CMCC-ESM2 model output prepared for CMIP6 CMIP historical. *Version 20201225* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.13195> .
- Mao, J., Thornton, P. E., Shi, X., Zhao, M., & Post, W. M. (2012). Remote sensing evaluation of CLM4 GPP for the period 2000–09. *Journal of Climate*, 25(15), 5327–5342. <https://doi.org/10.1175/JCLI-D-11-00401.1>
- Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO₂. *Journal of Advances in Modeling Earth Systems*, 11(4), 998–1038. <https://doi.org/10.1029/2018MS001400>
- NASA Goddard Institute for Space Studies (NASA/GISS) (2018). NASA-GISS GISS-E2.1G model output prepared for CMIP6 CMIP historical. *Version 20181015* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.7127> .
- NASA Goddard Institute for Space Studies (NASA/GISS) (2019). NASA-GISS

GISS-E2.1H model output prepared for CMIP6 CMIP historical. *Version 20190328* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.7128>

Neubauer, David; Ferrachat, Sylvaine; Siegenthaler-Le Drian, Colombe; Stoll, Jens; Folini, Doris Sylvia; Tegen, Ina; Wieners, Karl-Hermann; Mauritsen, Thorsten; Stemmler, Irene; Barthel, Stefan; Bey, Isabelle; Daskalakis, Nikos; Heinold, Bernd; Kokkola, Harri; Partridge, Daniel; Rast, Sebastian; Schmidt, Hauke; Schutgens, Nick; Stanelle, Tanja; Stier, Philip; Watson-Parris, Duncan; Lohmann, Ulrike (2019). HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 CMIP historical. *Version 20191124* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.5016>

Patra, P. K., Takigawa, M., Watanabe, S., Chandra, N., Ishijima, K., & Yamashita, Y. (2018). Improved chemical tracer simulation by MIROC4.0-based atmospheric chemistry-transport model (MIROC4-CTM). *Scientific Online Letters on the Atmosphere*. <https://doi.org/10.2151/SOLA.2018-016>

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., et al. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0708986104>

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp, D., Deng, F., Weir, B., Basu, S., Johnson, M. S., Philip, S., & Baker, I. (2022). Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7. *Atmospheric Chemistry and Physics*, 22(2), 1097–1130. <https://doi.org/10.5194/acp-22-1097-2022>

Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y., & Field, C. B. (1997). The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Global Biogeochemical Cycles*, 11(4), 535–560. <https://doi.org/10.1029/97GB02268>

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., et al. (2004). The oceanic sink for anthropogenic CO₂. *Science*, 305(5682), 367–371. <https://doi.org/10.1126/science.1097403>

Santoni, G. W., Daube, B. C., Kort, E. A., Jiménez, R., Park, S., Pittman, J. V., et al. (2014). Evaluation of the airborne quantum cascade laser spectrometer (QCLS) measurements of the carbon and greenhouse gas suite-CO₂, CH₄, N₂O, and CO-during the CalNex and HIPPO campaigns. *Atmos. Meas. Tech*, 7, 1509–1526. <https://doi.org/10.5194/amt-7-1509-2014>

Schimel, D. S., House, J. I., Hibbard, K. A., Bousquet, P., Ciais, P., Peylin, P., et al. (2001). Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature*, 414(6860). <https://doi.org/10.1038/35102500>

Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., et

al. (2019). Quantifying the Impact of Atmospheric Transport Uncertainty on CO₂ Surface Flux Estimates. *Global Biogeochemical Cycles*, 33(4), 484–500. <https://doi.org/10.1029/2018GB006086>

Seland, Ø., Bentsen, M., Olivié, D., Toniazzi, T., Gjermundsen, A., Graff, L. S., et al. (2020). Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geoscientific Model Development*, 13(12), 6165–6200. <https://doi.org/10.5194/gmd-13-6165-2020>

Seland, Øyvind; Bentsen, Mats; Olivié, Dirk Jan Leo; Toniazzi, Thomas; Gjermundsen, Ada; Graff, Lise Seland; Debernard, Jens Boldingh; Gupta, Alok Kumar; He, Yanchun; Kirkevåg, Alf; Schwinger, Jörg; Tjiputra, Jerry; Aas, Kjetil Schanke; Bethke, Ingo; Fan, Yuanchao; Griesfeller, Jan; Grini, Alf; Guo, Chungheng; Ilıcak, Mehmet; Karset, Inger Helene Hafsahl; Landgren, Oskar Andreas; Liakka, Johan; Moseid, Kine Onsum; Nummelin, Aleks; Spensberger, Clemens; Tang, Hui; Zhang, Zhongshi; Heinze, Christoph; Iversen, Trond; Schulz, Michael (2019). NCC NorESM2-LM model output prepared for CMIP6 CMIP historical. *Version20190917* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.8036> .

Simpson, I. R., McKinnon, K. A., Davenport, F. v., Tingley, M., Lehner, F., al Fahad, A., & Chen, D. (2021). Emergent constraints on the large scale atmospheric circulation and regional hydroclimate: do they still work in CMIP6 and how much can they actually constrain the future? *Journal of Climate*, 1–62. <https://doi.org/10.1175/JCLI-D-21-0055.1>

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., et al. (2007). Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂. *Science*. <https://doi.org/10.1126/science.1137004>

Stephens, B. B., Morgan, E. J., Bent, J. D., Keeling, R. F., Watt, A. S., Shertz, S. R., & Daube, B. C. (2021). Airborne measurements of oxygen concentration from the surface to the lower stratosphere and pole to pole. *Atmospheric Measurement Techniques*, 14(3), 2543–2574. <https://doi.org/10.5194/AMT-14-2543-2021>

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12(11), 4823–4873. <https://doi.org/10.5194/gmd-12-4823-2019>

Swart, Neil Cameron; Cole, Jason N.S.; Kharin, Viatcheslav V.; Lazare, Mike; Scinocca, John F.; Gillett, Nathan P.; Anstey, James; Arora, Vivek; Christian, James R.; Jiao, Yanjun; Lee, Warren G.; Majaess, Fouad; Saenko, Oleg A.; Seiler, Christian; Seinen, Clint; Shao, Andrew; Solheim, Larry; von Salzen, Knut; Yang, Duo; Winter, Barbara; Sigmond, Michael (2019). CCCma CanESM5 model output prepared for

CMIP6 CMIP historical. *Version 20190501* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.3610> .

Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., et al. (2015). Seasonal climatology of CO₂ across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network. *Journal of Geophysical Research: Atmospheres*, 120(10). <https://doi.org/10.1002/2014JD022591>

Tans, P. P., Fung, I. Y., & Takahashi, T. (1990). Observational constraints on the global atmospheric CO₂ budget. *Science*. <https://doi.org/10.1126/science.247.4949.1431>

Thompson, R. L., Patra, P. K., Chevallier, F., Maksyutov, S., Law, R. M., Ziehn, T., van der Laan-Luijkx, I. T., Peters, W., Ganshin, A., Zhuravlev, R., Maki, T., Nakamura, T., Shirai, T., Ishizawa, M., Saeki, T., Machida, T., Poulter, B., Canadell, J. G., & Ciais, P. (2016). Top-down assessment of the Asian carbon budget since the mid 1990s. *Nature Communications*, 7(1), 10724. <https://doi.org/10.1038/ncomms10724>

Thompson, C. R., Wofsy, S. C., Prather, M. J., Newman, P. A., Hanisco, T. F., Ryerson, T. B., Fahey, D. W., Apel, E. C., Brock, C. A., Brune, W. H., Froyd, K., Katich, J. M., Nicely, J. M., Peischl, J., Ray, E., Veres, P. R., Wang, S., Allen, H. M., Asher, E., ... Zeng, L. (2021). THE NASA ATMOSPHERIC TOMOGRAPHY (ATom) MISSION: Imaging the Chemistry of the Global Atmosphere. *Bulletin of the American Meteorological Society*, 1–53. <https://doi.org/10.1175/BAMS-D-20-0315.1>

K.W. Thoning, A.M. Crotnell, and J.W. Mund (2022), Atmospheric Carbon Dioxide Dry Air Mole Fractions from continuous measurements at Mauna Loa, Hawaii, Barrow, Alaska, American Samoa and South Pole. 1973-2021, Version 2022-05 National Oceanic and Atmospheric Administration (NOAA), Global Monitoring Laboratory (GML), Boulder, Colorado, USA <https://doi.org/10.15138/yaf1-bk21>

Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., et al. (2013). Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). *Geoscientific Model Development*, 6(2), 301–325. <https://doi.org/10.5194/gmd-6-301-2013>

Valentini, R., Matteucci, G., Dolman, A. J., Schulze, E. D., Rebmann, C., Moors, E. J., et al. (2000). Respiration as the main determinant of carbon balance in European forests. *Nature*, 404(6780), 861–865. <https://doi.org/10.1038/35009084>

Verma, S., Marshall, J., Gerbig, C., Rödenbeck, C., & Uwe Totsche, K. (2017). The constraint of CO₂ measurements made onboard passenger aircraft on surface-atmosphere fluxes: The impact of transport model errors in vertical mixing. *Atmospheric Chemistry and Physics*, 17(9), 5665–5675. <https://doi.org/10.5194/acp-17-5665-2017>

- Volodin, E. M., Dianskii, N. A., & Gusev, A. V. (2010). Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya, Atmospheric and Oceanic Physics*, 46(4), 414–431. <https://doi.org/10.1134/S000143381004002X>
- Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, 4(4), 845–872. <https://doi.org/10.5194/gmd-4-845-2011>
- Welp, L. R., Keeling, R. F., Meijer, H. A. J., Bollenbacher, A. F., Piper, S. C., Yoshimura, K., et al. (2011). Interannual variability in the oxygen isotopes of atmospheric CO₂ driven by El Niño. *Nature*, 477(7366), 579–582. <https://doi.org/10.1038/nature10421>
- Williamson, M. S., Thackeray, C. W., Cox, P. M., Hall, A., Huntingford, C., & Nijse, F. J. M. M. (2021). Emergent constraints on climate sensitivities. *Reviews of Modern Physics*, 93(2), 025004. <https://doi.org/10.1103/RevModPhys.93.025004>
- Wieners, Karl-Hermann; Giorgetta, Marco; Jungclaus, Johann; Reick, Christian; Esch, Monika; Bittner, Matthias; Legutke, Stephanie; Schupfner, Martin; Wachsmann, Fabian; Gayler, Veronika; Haak, Helmuth; de Vrese, Philipp; Raddatz, Thomas; Mauritsen, Thorsten; von Storch, Jin-Song; Behrens, Jörg; Brovkin, Victor; Claussen, Martin; Crueger, Traute; Fast, Irina; Fiedler, Stephanie; Hagemann, Stefan; Hohenegger, Cathy; Jahns, Thomas; Kloster, Silvia; Kinne, Stefan; Lasslop, Gitta; Kornblueh, Luis; Marotzke, Jochem; Matei, Daniela; Meraner, Katharina; Mikolajewicz, Uwe; Modali, Kameswarrao; Müller, Wolfgang; Nabel, Julia; Notz, Dirk; Peters-von Gehlen, Karsten; Pincus, Robert; Pohlmann, Holger; Pongratz, Julia; Rast, Sebastian; Schmidt, Hauke; Schnur, Reiner; Schulzweida, Uwe; Six, Katharina; Stevens, Bjorn; Voigt, Aiko; Roeckner, Erich (2019). MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 CMIP historical. *Version 20190929* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.6595> .
- Wofsy, S. C., The HIPPO Science Team and Cooperating Modellers and Satellite Teams, (2011). HIAPER Pole-to-Pole Observations (HIPPO): Fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. <https://doi.org/10.1098/rsta.2010.0313>
- Wofsy, S. C., Daube, B. C., Jimenez, R., Kort, E., Pittman, J. V., Park, S., et al. (2017). HIPPO Merged 10-second Meteorology, Atmospheric Chemistry, and Aerosol Data. Version 1.0. [Dataset] UCAR/NCAR - Earth Observing Laboratory. Retrieved from http://doi.org/10.3334/CDIAC/HIPPO_010
- Wofsy, S.C., S. Afshar, H.M. Allen, E.C. Apel, E.C. Asher, B. Barletta, J. Bent, H. Bian, B.C. Biggs, D.R. Blake, N. Blake, I. Bourgeois, C.A. Brock, W.H. Brune, J.W. Budney, T.P. Bui, A. Butler, P. Campuzano-Jost, C.S. Chang, M. Chin, R. Commane, G. Correa, J.D. Crounse, P. D. Cullis, B.C. Daube,

D.A. Day, J.M. Dean-Day, J.E. Dibb, J.P. DiGangi, G.S. Diskin, M. Dollner, J.W. Elkins, F. Erdesz, A.M. Fiore, C.M. Flynn, K.D. Froyd, D.W. Gesler, S.R. Hall, T.F. Hanisco, R.A. Hannun, A.J. Hills, E.J. Hints, A. Hoffman, R.S. Hornbrook, L.G. Huey, S. Hughes, J.L. Jimenez, B.J. Johnson, J.M. Katich, R.F. Keeling, M.J. Kim, A. Kupc, L.R. Lait, K. McKain, R.J. McLaughlin, S. Meinardi, D.O. Miller, S.A. Montzka, F.L. Moore, E.J. Morgan, D.M. Murphy, L.T. Murray, B.A. Nault, J.A. Neuman, P.A. Newman, J.M. Nicely, X. Pan, W. Paplawsky, J. Peischl, M.J. Prather, D.J. Price, E.A. Ray, J.M. Reeves, M. Richardson, A.W. Rollins, K.H. Rosenlof, T.B. Ryerson, E. Scheuer, G.P. Schill, J.C. Schroder, J.P. Schwarz, J.M. St.Clair, S.D. Steenrod, B.B. Stephens, S.A. Strode, C. Sweeney, D. Tanner, A.P. Teng, A.B. Thames, C.R. Thompson, K. Ullmann, P.R. Veres, N.L. Wagner, A. Watt, R. Weber, B.B. Weinzierl, P.O. Wennberg, C.J. Williamson, J.C. Wilson, G.M. Wolfe, C.T. Woods, L.H. Zeng, and N. Vieznor. 2021. ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, Version 2. [Dataset] ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/1925>

Yang, Z., Washenfelder, R. A., Keppel-Aleks, G., Krakauer, N. Y., Randerson, J. T., Tans, P. P., et al. (2007). New constraints on Northern Hemisphere growing season net flux. *Geophysical Research Letters*. <https://doi.org/10.1029/2007GL029742>

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., et al. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, 70(1), 193. <https://doi.org/10.1071/ES19035>

Ziehn, Tilo; Chamberlain, Matthew; Lenton, Andrew; Law, Rachel; Bodman, Roger; Dix, Martin; Wang, Yingping; Dobrohotoff, Peter; Srbinovsky, Jhan; Stevens, Lauren; Vohralik, Peter; Mackallah, Chloe; Sullivan, Arnold; O'Farrell, Siobhan; Druken, Kelsey (2019). CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP historical. *Version 20191115* Earth System Grid Federation. doi:<https://doi.org/10.22033/ESGF/CMIP6.4272>