

Global atmospheric CO₂ inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate

Benjamin Gaubert¹, Britton B. Stephens¹, Sourish Basu², Frédéric Chevallier³, Feng Deng⁴, Eric A. Kort⁵, Prabir K. Patra⁶, Wouter Peters⁷, Christian Rödenbeck⁸, Tazu Saeki⁹, David Schimel¹⁰, Ingrid Van der Laan-Luijkx⁷, Steven Wofsy¹¹, and Yi Yin¹²

¹National Center for Atmospheric Research, Boulder, CO, USA

²University of Colorado Boulder and NOAA Earth System Research Laboratory Boulder, CO, USA

³Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, CEA-CNRS-UVSQ, Gif sur Yvette, 91191 CEDEX, France

⁴Department of Physics, University of Toronto, Canada

⁵University of Michigan, Ann Arbor, MI, USA

⁶RGGC/IACE/ACMPT, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama 236 0001, Japan

⁷Meteorology and Air Quality, Wageningen University, Wageningen, Netherlands

⁸Max Planck Institute for Biogeochemistry, 07745 Jena, Germany

⁹Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan

¹⁰Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

¹¹Harvard University, Cambridge, MA, USA

¹²California Institute of Technology, Pasadena, CA, USA

Correspondence: Benjamin Gaubert (gaubert@ucar.edu)

Abstract. We have compared a suite of recent global CO₂ atmospheric inversion results to independent airborne observations and to each other, to assess their dependence on differences in northern extratropical vertical transport and to identify some of the drivers of model spread. We evaluate posterior CO₂ concentration profiles against observations from the High-Performance Instrumented Airborne Platform for Environmental Research (HIAPER) Pole-to-Pole Observations (HIPPO) aircraft campaigns over the mid Pacific in 2009-2011. Although the models differ in inverse approaches, assimilated observations, prior fluxes, and transport models, their broad latitudinal separation of land fluxes has converged significantly since the Atmospheric Carbon Cycle Inversion Intercomparison (TransCom3) and the REgional Carbon Cycle Assessment and Processes (RECCAP) projects, with model spread reduced by 80 % since TransCom3 and 70 % since RECCAP. Most modelled CO₂ fields agree reasonably well with the HIPPO observations, specifically for the annual mean vertical gradients in the northern hemisphere. Northern hemisphere vertical mixing no longer appears to be a dominant driver of northern versus tropical annual flux differences. Our newer suite of models still gives northern extratropical land uptake that is modest relative to previous estimates (Gurney et al., 2002; Peylin et al., 2013) and near neutral tropical land uptake for 2009-2011. Given estimates of emissions from deforestation, this implies a continued uptake in intact tropical forests that is strong relative to historical estimates (Gurney et al., 2002; Peylin et al., 2013). The results from these models for other time periods (2004-2014, 2001-2004, 1992-1996), and re-evaluation of the TransCom3 Level 2 and RECCAP results confirms that tropical land carbon fluxes including deforestation

have been near neutral for several decades. However, models still have large disagreements on ocean-land partitioning. The fossil fuel and the atmospheric growth rate terms have been thought to be the best-known terms in the global carbon budget, but we show that they currently limit our ability to assess regional scale terrestrial fluxes and ocean-land partitioning from the model ensemble.

5 1 Introduction

Current appraisals of the global atmospheric carbon budget are informed by surface fluxes computed by inverse transport models (e.g., Newsam and Enting, 1988; Tans et al., 1990; Rayner et al., 1999; Gurney et al., 2002, 2003, 2004; Peylin et al., 2013). Net carbon flux to the atmosphere is derived from temporal and spatial CO₂ gradients given by atmospheric observations and prior estimates of component fluxes and their uncertainties. This assessment of atmospheric sources and sinks relies on 1) atmospheric tracer transport models that link fluxes to atmospheric CO₂ fields, 2) prior emissions and sinks (e.g., from process model flux estimates), 3) the spatial and temporal representativeness and coverage of the observational network, and 4) error statistics associated with each information piece. Since the problem is underdetermined, it is essential to quantify the uncertainty and biases of posterior fluxes and CO₂ concentrations with independent observations and cross-model comparisons. The most prominent community-wide inverse result intercomparison that included comparisons of posterior concentrations to independent observations was the TransCom 3 study (Gurney et al., 2002, 2004), which studied fluxes for the 1992-1996 period. This comparison could focus on the impact of transport model differences by optimizing the fluxes using a common method over the same regions (11 land and 11 ocean). One particular feature of the seasonally resolved (Level 2) TransCom 3 inversions (hereafter denoted as T3L2) was the direct dependence of flux estimates on vertical gradients of CO₂ (Stephens et al., 2007), leading to a different partitioning between Northern Extratropical (NET) versus Tropical (T) land sinks.

A more recent community-wide CO₂ inverse model intercomparison was carried out as part of the REgional Carbon Cycle Assessment and Processes Project (RECCAP, <https://www.globalcarbonproject.org/reccap>; Canadell et al. 2011). The atmospheric inversion component of RECCAP was a comprehensive intercomparison that analyzed long-term mean, long-term trend, interannual variations and mean seasonal variations of CO₂ fluxes using common post-processing (Peylin et al., 2013). In RECCAP, the 11 inverse models used different inversion techniques, atmospheric models, and observational datasets. When the fluxes were analyzed for the years 2001 to 2004, Peylin et al. (2013) found an overall improved consistency between inversions at large scale and over specific regions compared to T3L2 when the network of atmospheric sites was less dense. RECCAP inversions showed a general agreement on the total natural land carbon flux long-term mean and its interannual variability over 1991-2010. The total ocean plus land sink estimates were more robust over the NET than for the Tropics and in the Southern Extratropics (SET). The remaining spread led to a disagreement on the NET/T/SET land partitioning, with some models simulating a stronger tropical source compensated by larger NET and SET sinks. Peylin et al. (2013) also noted that the group of models that assimilated observations at their corresponding times rather than using monthly means had more consistent, weaker tropical sources, and weaker northern sink land fluxes.

Several additional inverse modelling intercomparison studies have more recently involved satellite, surface and joint surface-

satellite inversion (e.g., Chevallier et al., 2014; Houweling et al., 2015). In these studies, the inversion systems used space borne retrievals of column-average dry air-mole fraction of CO₂ (XCO₂) from the Orbiting Carbon Observatory-2 (OCO-2) satellite since July 2014 (Eldering et al., 2017) and from the Greenhouse Gas Observing SATellite (GOSAT; Kuze et al. 2009) instrument since January 2009. Those inverse exercises, however, are still sensitive to satellite retrieval algorithms and the
5 inversions' prior assumptions. In particular, the results are sensitive to systematic errors from transport and satellite retrievals (Houweling et al., 2010; Chevallier, 2015).

Schimel et al. (2015) investigated the NET versus T+SET land flux partitioning as indicated by atmospheric inversions, biosphere process model simulations, and forest inventory estimates, and estimated a large land uptake over the tropics by intact forests due to a significant CO₂ fertilization effect. This study argued for the importance of comparing posterior CO₂ fields to
10 observations, which was not done in RECCAP, in order to fully understand and predict terrestrial land sinks, and their variation due to CO₂ and climate feedbacks. A follow-up inversion intercomparison focused on East Asia and found that large flux adjustments were possible even though models well simulated the observed gradient in vertical profiles measured by aircraft, because the uncertainties from model transport and fossil fuel (FF) prior emissions were compensated by the flux adjustments (Thompson et al., 2016).

15 The HIPPO campaign (Wofsy, 2011), spanned large latitudinal, vertical and temporal coverage (2009 to 2011) and provides a useful atmospheric trace gas data set for investigating the consistency of inverse fluxes and posterior concentration results. Graven et al. (2013) found an increase of the CO₂ seasonal amplitude by up to 50 % at mid to high latitudes of the northern hemisphere and at altitudes ranging between 3 and 6 km between the HIPPO period and the 1950s. Deng et al. (2015) compared posterior CO₂ and O₃ fields from GEOS-Chem to the HIPPO observations to diagnose the impact of the upper troposphere
20 and lower stratosphere (UTLS) definition on retrieved fluxes. These results indicate a significant impact of transport errors on retrieved fluxes. Frankenberg et al. (2016) evaluated the CarbonTracker CT2013B and Monitoring Atmospheric Composition and Climate MACC v13r1 atmospheric inverse models, and satellite retrievals from GOSAT, TES (Tropospheric Emission Spectrometer) and AIRS (Atmospheric Infrared Sounder) in comparison to HIPPO measurements. They found that despite an overall agreement between inversions and HIPPO measurements, systematic model transport errors remain important.
25 After years of continuous model development, the goal of this study is to investigate whether global inverse models are still highly dependent on northern hemisphere vertical transport errors and on prior flux estimates and their uncertainties used in the inversions.

Our two main approaches to answer this question are:

- First, we compare modelled CO₂ after flux optimization to independent aircraft in-situ CO₂ observations from the HIPPO
30 campaign (2009-2011).
- Second, we compare the observationally-constrained fluxes for latitudinal bands and on a global scale both across models and to budget estimates provided by the Global Carbon budget 2016 (hereafter denoted GCP2016; Le Quéré et al. 2016).

Measurements and inversion systems are described in Sect. 2. In Sect. 3.1, we present the results of the comparison of modeled posterior CO₂ vertical gradients with HIPPO measurements. In Sect. 3.2, we analyze the differences in the meridional

distribution of land sinks and global carbon estimates for the years 2009 to 2011 from inverse modelling of atmospheric in-situ observations together and with GCP2016. In Sect. 3.3, we compare inverse model and GCP2016 estimates at the global scale, including prescribed fossil and retrieved atmospheric growth rate terms. Conclusions and a summary of the findings are given in Sect. 4.

5 2 Methods

2.1 Participating models

The list of participating inverse models is shown in Table 1 and more details are available in the supplement. These include 10 different inverse modeling systems or system variants. The inversion systems differ in many aspects such as transport models, wind fields, analysis procedures, and subset of assimilated observations. The ACTM system performed 2 inversions with different prescribed fossil fuel (FF) emissions (Saeki and Patra, 2017), one based on totals from the Carbon Dioxide Information Analysis Center (CDIAC; Boden et al. 2016) and another based on the International Energy Agency (IEA 2016), which allows us to assess sensitivity to the FF prior only. This is also the case for the 2 Carbon Tracker Europe versions, CTE2016-FT (Fast Track) and CTE2017-FT, where only the subset of observations and the FF prior are different (van der Laan-Luijkx et al., 2017).

It is worth noting that some inverse models are constructed in a similar framework. Some share the same transport model, such as TM5 that is used in 4 inversions, and some use the same meteorological fields. Five inverse systems nudge their forecast field to the ERA-Interim reanalysis (Dee et al., 2011). The two longest flux estimates, from CAMS (v16r1) and Jena (s85_v4.1), are used to reproduce the comparison with observations as in (Stephens et al., 2007) over the T3L2 period (1992 to 1996). The Jena s85_v4.1 and s04_v4.1 inversions differ in their calculation periods and station sets used: Jena s85_v4.1 starts in 1985 using only 23 stations that cover this entire period, while s04_v4.1 uses many more sites (59) and starts in 2004. This also allows us to separate the impact of the number of sites assimilated over the most recent period.

2.2 The Global Carbon Budget 2016

The Global Carbon Project (GCP) gathers observational and model-based flux estimates from multiple organizations and research groups around the world to yearly report a global budget of atmospheric CO₂ (Le Quéré et al., 2016). GCP2016 is the most recent version with flux estimates forced to balance globally. The most recent version (GCP2017; Le Quéré et al. 2018) separated an explicit unknown ocean or land flux term, which prevents simple comparisons of the type presented here. Specifically, the land/ocean partitioning in GCP2016 is based on multiple observational constraints on the ocean flux for the 1990s, extrapolated forward with a suite of 7 global ocean models. As pointed out in Le Quéré et al. (2018), there are considerable uncertainties in this extrapolation, with the estimated ocean/land partitioning for later decades dependent on the models.

Acronym	References	Grid Spacing	Fossil Fuel priors	Transport Model	Number of vertical layers	Meteorological fields	Available Period
CAMS (v16r1)	Chevallier et al. (2005, 2010)*	3.75° x 1.875°	CDIAC/GCP2016	LMDZ	39	ERA interim	1979 to 2016
Jena (s04_v4.1)	Rödenbeck et al. (2003) Rödenbeck (2005)	4°x5°	CDIAC	TM3	19	NCEP	2004 to 2016
Jena (s85_v4.1)	//	4°x5°	CDIAC	TM3	19	NCEP	2004 to 2016
CTE2016-FT	van der Laan-Luijkx et al. (2017)	1°x1°	CDIAC	TM5	25	ERA interim	2001 to 2015
CTE2017-FT	//	1°x1°	CDIAC	TM5	25	ERA interim	2000 to 2016
CT2016	Peters et al. (2007)**	1°x1°	ODIAC v2016 and "Miller"	TM5	25	NCEP	2001 to 2015
ACTM-IEA	Saeki and Patra (2017) Patra et al. (2011)	Inversion (2.8°x2.8°)	IEA	ACTM	32	NCEP2 (for inversion)	2003 to 2011
ACTM-CDIAC	//	and forward (1.1°x1.1°)	CDIAC	ACTM	32	JRA25 for forward	2003 to 2011
TM5-4DVar	Basu et al. (2013)	3°x2°	EDGAR +CDIAC	TM5	25	ERA interim	2007 to 2012
GEOS-Chem	Deng et al. (2014)	4°x5°	CDIAC, ICOADS and 3-D aviation	GEOS	47	GEOS5	2009 to 2011

Table 1. List of the inverse modelling systems used in this study and general characteristics. *with updates documented at <https://atmosphere.copernicus.eu/>. **with updates documented at <http://carbontracker.noaa.gov>.

The GCP2016 atmospheric growth rate is derived from atmospheric CO₂ measurements at marine boundary layer (MBL) sites made by the US National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA/ESRL; Masarie and Tans 1995; Dlugokencky and Tans 2018). CO₂ emissions from land-use-change (ELUC) are the net sum of all anthropogenic activities: deforestation, afforestation, logging and shifting cultivation. Total emissions are estimated, following the bookkeeping method (Houghton, 2003; 2012), with complementary inter-annual variability calculated from satellite data when available (van der Werf et al., 2010; Giglio et al., 2013). The average ELUC for the year 2009 to 2011 included here is estimated to be 0.85 PgC yr⁻¹ with an uncertainty of 0.5 PgC yr⁻¹. These emissions are added to the GCP2016 land sink for comparison to atmospheric inversion estimates.

Finally, the land sink is estimated in GCP2016 as a residual from all other components of the carbon budget. The GCP2016 method treats the riverine flux of carbon from land to ocean to atmosphere as separate components of the total air-land and air-sea fluxes and subtracts an estimate of this flux (0.45 PgC yr⁻¹; Jacobson et al. 2007) from the pCO₂-based sea-to-air

flux estimates to match estimates of the anthropogenic ocean sink alone. Because the land sink is a residual, this increase in the magnitude of the ocean sink results in a corresponding reduction by 0.45 PgC yr^{-1} in the magnitude of the land sink in GCP2016. To compare to atmospheric inverse flux estimates, which represent the total air-sea and air-land fluxes, we have adjusted the GCP2016 ocean and land flux estimates by this same 0.45 PgC yr^{-1} , decreasing the ocean sink and increasing the

5 land sink.

Note that we do not show GCP2016 estimates here as a truth metric against which to evaluate the models, but rather as one estimate of an internally consistent global budget that provides a useful reference for exploring axes of variability in our models and comparing to other community estimates.

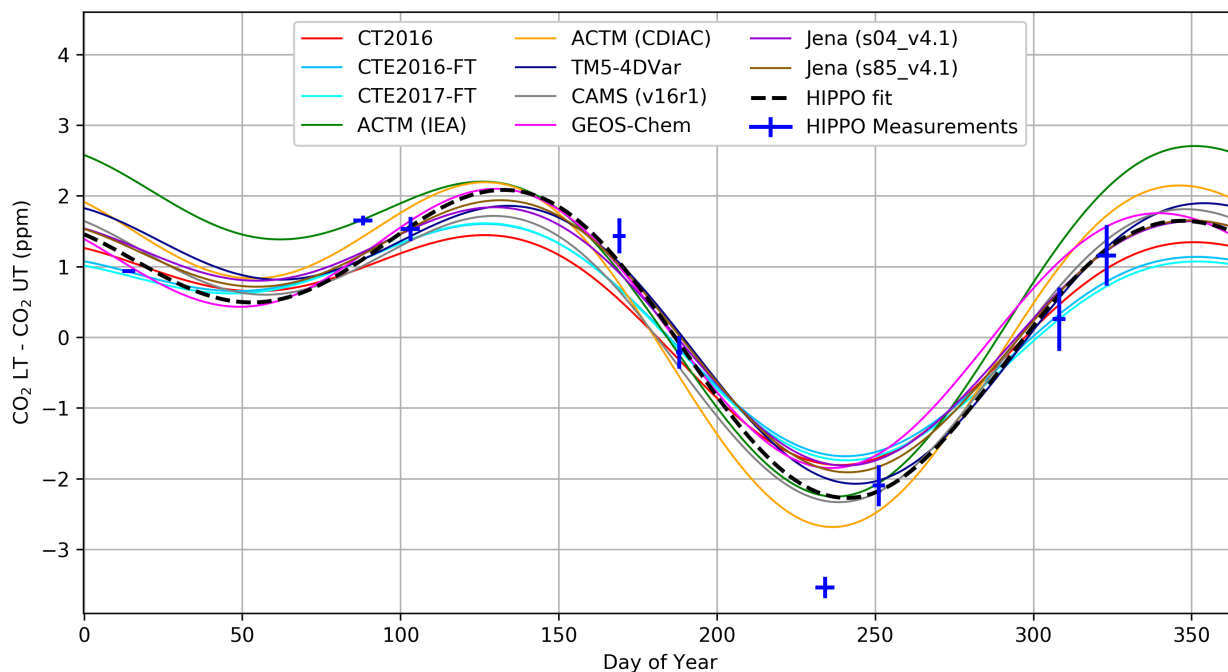


Figure 1. Reconstructed annual cycle in northern extratropical vertical CO_2 gradients, obtained from 2-harmonic fits of the HIPPO data and correspondingly sampled model outputs, averaged over 20°N to 90°N (1000 hPa to 800 hPa minus 800 hPa to 400 hPa). The CO_2 average curtain observations for each of 9 atmospheric transects have been added on the graph to illustrate the data uncertainties and temporal coverage, the y-axis errorbar is derived from the range of disagreement among the three in situ instruments on board (QCLS, OMS, and AO2; see supplement), the line average is derived from the $\text{CO}_2.X$ merged dataset. The horizontal whiskers represent the time span of the flights contributing to each average. The observed line shown here is not a direct fit to the observation points, but rather comes from an average of fits to individual 100 hPa by 5 degree latitude bins as described in the text.

2.3 HIPPO observations and fitting procedures

The HIPPO project (Wofsy, 2011) used the NSF/NCAR Gulfstream V aircraft (GV) to conduct 5 month-long campaigns in different seasons over three years (2009-2011, see supplement), that consisted of vertical profiling along north-south Pacific transects between 87°N and 67°S. The 5 campaigns included 9 transects of the NET Pacific. We exclude observations over North America conducted between Colorado and Alaska (Fig. S1). HIPPO flew three different in situ CO₂ instruments and 2 whole air samplers with laboratory CO₂ measurements. We use the recommended CO2.X variable which comes primarily from the Harvard Quantum Cascade Laser Spectrometer (QCLS), gap filled during calibration sequences, and compare to the other systems to constrain potential systematic biases (see supplement). We calculate the NET vertical gradient as the difference between the average from 1000 hPa to 800 hPa for the lower troposphere (LT), and the average from 800 hPa to 400 hPa for the upper troposphere (UT), spanning the latitude range from 20°N to 90°N. To do this, we first detrend the observations and model sampled along the flight-track output by subtracting a deseasonalized and smoothed long-term trend record from the fit of the Mauna Loa Observatory in-situ measurement time series to provide a common reference for both observations and models, and bin the observations by 100 hPa in pressure and 5 degrees in latitude bins. We then fit each bin with a 2-harmonic curve and constant offset, and average the resulting fits across boxes and pressure levels, with latitude weighting (see supplement). Fig. 1 shows the resulting daily fit of the annual cycle for the HIPPO observations and model simulations of the NET vertical gradient. Qualitatively, it shows that most models reproduce the CO₂ cycle well, with positive gradients in winter over a broad peak and negative gradients in summer over a narrower trough. The three CarbonTracker inversions (CT2016, CTE2016-FT, and CTE2017-FT) have somewhat lower seasonal gradient amplitude, while the two ACTM inversions (ACTM-IEA and ACTM-CDIAC) show larger amplitude. More quantitative details are given in Section 3.1. To illustrate the temporal coverage of the observations, we plot the measurements of the 9 HIPPO transects on Fig. 1 as simple differences of the latitude weighted average concentrations within the LT and UT boxes for each transect, while an example of a fit to an individual bin is shown in Fig. S1.

The QCLS instrument has a 1- σ precision of 20 ppb (Santoni et al., 2014), and for all 5 CO₂ systems on the GV the instrumental precision is negligible for the large-scale average metrics we present here. More relevant sources of uncertainty are associated with the potential for altitude dependent biases that might result from inlet or cabin-pressure effects, and misrepresentation of synoptic transport in the models. We estimate uncertainty in the annual-mean NET vertical gradient metric by comparison of the 5 independent instruments and whole air samplers to be ± 0.15 ppm (see supplement), and uncertainty on the individual HIPPO transect values to range from 0.02 ppm to 0.48 ppm as shown by the vertical bars in Fig. 1 These values are derived from the maximum absolute differences between the sensors, which we conservatively treat as best-guess 1 σ uncertainty estimates. These uncertainty estimates correspond to the vertical gradient as observed by the HIPPO flight tracks and calculated with the fitting procedure used here. Because we use model output along the flight tracks and treat model output and observations identically in our calculations, we do not include an estimate of potential spatial sampling bias, but we do use model output to assess the spatial representativeness of our calculated metrics with respect to full 150°W transect and full zonal means in Sect. 4 of supplement (Fig. S5, S6). Also, because the models are driven by reanalysis winds, they should capture the position of

synoptic systems and associated transport. However, the wind fields and model transport may be biased which could result in different vertical gradients for reasons unrelated to the fluxes of interest. We have estimated synoptic variability in the vertical gradient metric and find a worst-case potential model synoptic sampling bias of ± 0.06 ppm for the annual mean, ± 0.14 ppm for JFM, ± 0.15 ppm for JAS (1σ ; see supplement).

5

3 Results

3.1 Fluxes and posterior CO₂ comparisons with HIPPO

Each individual inversion system adjusts fluxes to fit the concentration fields with its given transport scheme and a priori source/sink information. Biases can appear in the retrieved posterior CO₂ resulting from errors in the estimated fluxes or from specific biases in transport to the location of the independent data (here in particular vertical transport to the upper atmosphere). We first evaluate if the spread of retrieved land fluxes over different zonal bands is correlated with NET vertical CO₂ gradients and if the modelled gradients match observations, as was previously done for the T3L2 models by Stephens et al. (2007). Fig. 2A presents the results for the HIPPO and model vertical gradients and model fluxes, broken into NET and T+SET regions for the years 2009-2011. The mean and relative spread of 10 simulations for the posterior annual mean NE land flux is -2.24 PgC yr⁻¹ ± 0.29 PgC yr⁻¹ (13 %, 1σ). Aside from the ACTM-IEA simulation, all models are within the uncertainty range of 0.15 ppm or 50 % of the measured vertical gradient. This contrasts to the TransCom3 Level 2 simulations which had an annual mean of -2.42 PgC yr⁻¹ ± 1.05 (43 %) PgC yr⁻¹ for NE land flux and disagreed with the observed vertical gradient by $\tilde{0}.5$ ppm on average and as much as 1.3 ppm (186 %). As listed in Table 1, the inversions have significant differences in transport model, resolution, and driving meteorology and are converging despite these differences. In addition, there are no apparent relationships between vertical gradients and NET nor T+SET land fluxes. The standard deviation across 10 simulations on the difference between NET Land and T+SET is 0.4 PgC yr⁻¹ while it was 2.1 PgC yr⁻¹ in T3L2 (Gurney et al., 2004; Gurney and Denning, 2013) and 1.28 PgC yr⁻¹ in RECCAP (Peylin et al., 2013), representing a steady and dramatic convergence of model estimates over the past 15 years. We reproduce the Stephens et al. (2007) annual mean figure in Fig. 2B, with the exception of showing T+SET instead of T, to highlight those differences. It is important to note that these results correspond to a different period and different models, with a smaller network of assimilated in-situ network measurements, and assimilation of monthly-mean rather than discrete measurements. We took advantage of the 2 models that span the 1992-1996 period, CAMS (v16r1) and Jena (s85_v4.1) to further investigate differences from the T3L2 period. Those two models are quite close to the 2009-2011 vertical gradient observations (Fig. 2A), but they both overestimate the 1992-1996 vertical gradients (Fig. 2B). Notably, they fall along the lines fit to the T3L2 models in Fig. 2B, which could be a coincidence, but might also suggest that despite agreeing with the other models on the latitudinal flux distribution for 2009-2011 these models overestimate tropical sources and northern sinks during 1992-1996. This would require that these models be more dependent on vertical mixing biases in the earlier period. The different number of assimilated sites is one potential factor that might explain different biases in retrieved fluxes for these two periods, but this is not seen for the comparison of the 2 versions of the Jena model assimilating different numbers of

sites during 2009-2011. It is worth noting that reanalyses of meteorological observations have noticeably improved thanks to a better representation of unresolved processes in global models, improved data assimilation methods, and the increasing availability of satellite data, which makes the reanalyses perform better in the 2000s than for the 1990s and earlier (e.g., Gelaro et al., 2017; Bauer et al., 2015). As an example, the assimilation of new observations from the constellation of COSMIC global positioning system radio occultation (GPSRO) satellites has led to a significant improvement in meteorological analyses and forecasts (e.g., Healy, 2008).

One concern is the spatial representativeness of the HIPPO measurements which were made over the Pacific Ocean while the light aircraft observations used by Stephens et al. (2007) were mostly measuring profiles over land. We discuss this issue in the supplement and show that across models HIPPO vertical gradients are significantly representative of the zonal mean for the three-year mean and every year individually (Fig. S5). Seasonally (Fig. S6), it appears that the vertical gradients are representative of the parallel 150W for winter (JFM), spring (AMJ) and fall (OND) seasons, representative of the zonal mean for winter (JFM) and fall (OND), and representative of the zonal average over land only in boreal summer (JAS). We did find a significant correlation between vertical gradients defined by the HIPPO flight tracks and land zonal means during summer (JAS), when vertical gradients are weak.

Fig. 2C and Fig. 2D show the vertical gradients and fluxes for 2009-2011 winter (JFM) and summer (JAS). The agreement between the models and HIPPO observations is not as strong as for annual means. The vertical gradient in the NET winter is reasonably well reproduced by 9 models with differences lower than 0.36 ppm. The ACTM-IEA inversion is an outlier and overestimates by 0.94 ppm the winter season average vertical gradient. For ACTM, the global annual IEA emissions are less than CDIAC (Fig. 4C and D) which results in a weaker northern extratropical sink (Fig. 2A and 3A) which corresponds with a more positive LT-UT northern extratropical gradient (Fig. 2A and S2) and a more positive N-S gradient (Figure S2), comparing just the 2 ACTM versions. Differences across inversion systems in Figure S2 also depend on transport and inversion scheme and the resulting spatial distribution of sources and sinks.

There are generally larger differences between observed and modelled vertical gradients in northern hemisphere summer (JAS), with only 2 models (ACTM-IEA and CAMS) within observation error bars, but the whole range of values is only 0.75 ppm. In this case a linear relationship ($r^2=0.4$) is found between the modelled vertical gradient and the retrieved T+SET fluxes, but not for the NET flux. There is a significant relationship between HIPPO and the land only zonal average vertical gradient and both are correlated with the T+SET fluxes (Figure S7), but with a slope of 2.16 ppm/PgC yr⁻¹ for HIPPO while it is 0.93 ppm/PgC yr⁻¹ over land where the vertical gradients are bigger. This suggests that transport errors may be more critical in the summer season or that other factors compensate to obscure the relationship for these relatively coarse time averages in other seasons and for the annual means. While additional insights into model behavior could be gained from more detailed comparisons to individual models or in more controlled inversion ensembles, the varied nature of these inversion systems makes detailed analyses more challenging and beyond the scope of our current study.

For the annual means and winter there are no statistical relationships between the vertical gradients and the retrieved fluxes. This suggests that northern hemisphere vertical mixing errors do not play a major role in biasing the flux estimation across these models. However, the retrieved fluxes can still be biased because of the transport errors.

One potential limitation in our analysis could be the use of similar meteorological fields from the ECMWF base analysis and forecast cycle, which is the case for 5 out of 10 simulations. A careful comparison of model transport suggests that nudging to a particular reanalysis product does not imply identical tracer transport between the models (e.g., Prather et al., 2008; Locatelli et al., 2015; Orbe et al., 2017). The transport errors arise from resolved advection and heavily parameterized transport schemes such as convection and boundary layer mixing (Locatelli et al., 2015; Orbe et al., 2017; Krol et al., 2018). Qualitatively, we cannot distinguish the CO₂ vertical gradient from models using ERA-Interim winds from the 5 other models.

3.2 The latitudinal distribution of retrieved land fluxes

In this section, we present the retrieved land flux partitioning between the NET and the T+SET, as shown in Fig. 3 and Fig. 2. Because the total sink is the sum of T+SET and NET, these lines have a slope of -1 and any deviation perpendicular to the lines indicates disagreement on the total land sink from the GCP2016 estimate. As noted in the previous section, inverse modelling results for the HIPPO period (2009-2011) are remarkably close to one another (Fig. 3A). These results converge on a NET land sink value slightly larger than 2 PgC yr⁻¹ (-2.24 ± 0.29 PgC yr⁻¹) and a T+SET land sink of -0.38 ± 0.31 PgC yr⁻¹. In Fig. 3, multi model means are represented by blue diamonds and associated error bars are estimated by the standard deviation across models. The 2009-2011 period is marked by a large tropical land sink because of the strong La Niña event of 2011 (Bastos et al., 2013; Poulter et al., 2014). For these 3 years, the models clearly indicate a negative flux over the Tropics and SET land. There are also increasing lines of evidence that the rate of deforestation and climate stress over tropics have been moderated in recent decades (e.g., 2000s), compared to the 1990s (Kondo et al., 2018), with a reduced change in tropical forest cover because the decrease in the South American deforestation has been compensated by an increased South East Asian deforestation (Hansen et al., 2013).

In order to place these recent fluxes estimates in the context of previous studies, we show the flux estimates by the new models that also estimate fluxes for the earlier periods; 2 models have available outputs for the T3L2 period (1992-1996) and 4 for the RECCAP period (2001-2004), as shown in Fig. 3B and 3C. For Jena, one inversion (s85_v4.1) starts in 1985 and is constrained by only 23 atmospheric sites while the other (s04_V4.1) starts in 2004 and uses 59 sites. Interestingly, the difference between s85_v4.1 and s04_V4.1 for 2009-2011 is rather small (Fig. 3a), less than 0.15 PgC yr⁻¹.

According to GCP2016, the total land sink in 2009-2011 was around twice as large (around 3 PgC yr⁻¹) than for 1992-1996 (around 1.7 PgC yr⁻¹) and 2001-2004 (around 1.3 PgC yr⁻¹). This is due to the combined effect of natural inter-annual variability as well as a long-term trend (Ballantyne et al., 2012). The retrieved total land fluxes for all study periods appear to be close to the corresponding GCP estimates with most models falling within the GCP2016 1 σ uncertainty range. For the 2001-2004 period, the newer simulations move fluxes parallel to the GCP line in the direction of a weaker tropical source and a weaker NE sink relative to the original RECCAP estimates. For the 1992-1996 period, 1 of the 2 newer simulations shifts fluxes in that same direction, but not as far as suggested by Stephens et al. (2007).

However, we have revisited the Stephens et al. (2007) estimates, by considering the intercept of the regression lines with the aircraft observations rather than the mean of the 3 models nearest the annual mean observations, and evaluating the error using

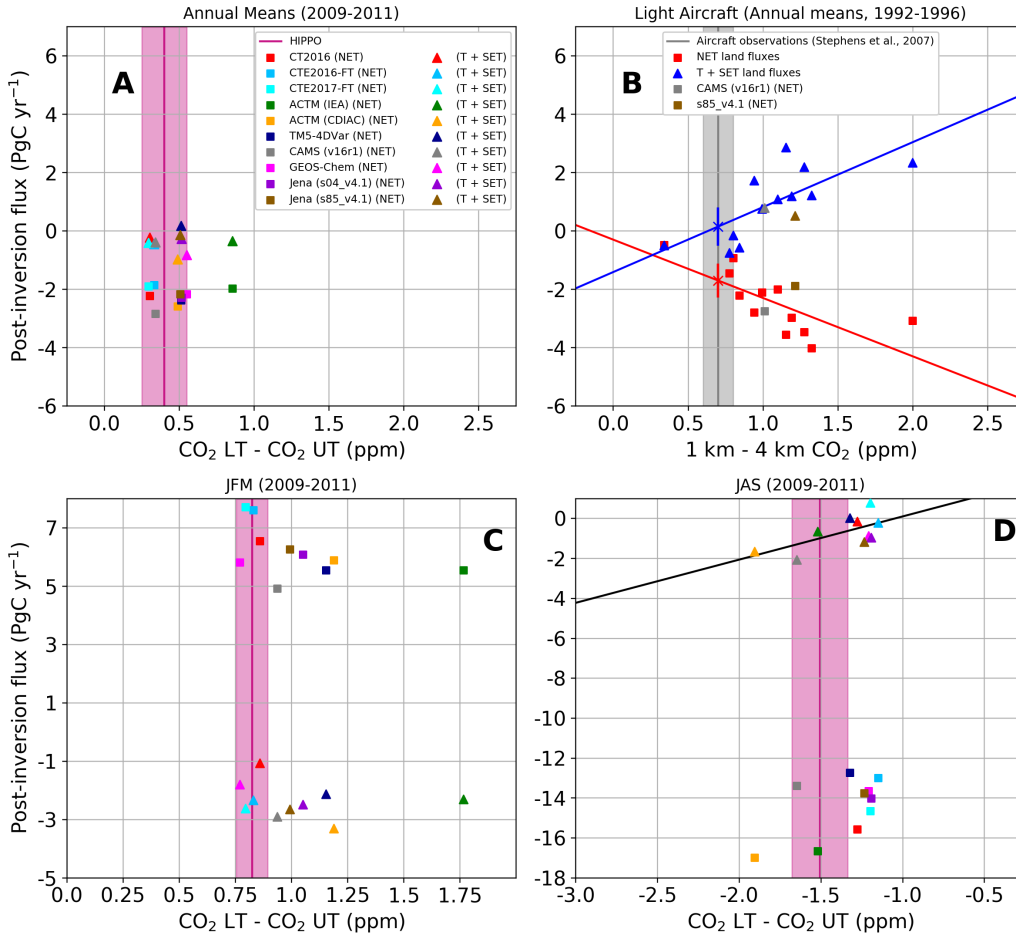


Figure 2. Retrieved fluxes versus NE vertical gradients. A) Annual mean NET land and T+SET land fluxes versus posterior NE vertical gradients (lower minus upper troposphere) from model output along HIPPO flight tracks and HIPPO observations (pink line) for the period 2009-2011. The shaded area represents an estimate of measurement uncertainty of ± 0.15 ppm for the annual mean, as estimated in the Sect. 2 of the supplement. Inverse model posterior concentration gradients and fluxes are shown as points (squares = NET, triangles = T+SET). The vertical axis represents the integrated annual mean land fluxes (PgC yr^{-1}). B) Same as (A) but for 1992-1996 and showing TransCom3 Level 2 models and our 2 current models that span this time period, showing dependence of posterior fluxes on transport and a large range of transport biases. Annual mean NE (red squares) and T+SET (blue triangles) land carbon fluxes for 1992-1996 estimated by the 12 T3L2 models plotted as a function of the models' post-inversion predicted mean vertical CO₂ gradients at 10 light aircraft profiling sites (adapted from Stephens et al. 2007) with fluxes partitioned by TransCom region. The Jena (s85_v4.1) and the CAMS (v16r1) simulations have also been sampled at the same light aircraft locations but their fluxes are partitioned at 20°N and 20°S. The crosses show our new best estimate of the fluxes estimated by the regression of all T3L2 models. The error bars on these points are estimated using the standard error of the regressions. C) same as panel A for January-February-March (JFM), and D) same as panel A for July-August-September (JAS). For the seasonal plots, the width of the pink bar is of 0.07 ppm for JFM and 0.17 for JAS. On panel D, the black line represents the regression line, shown because the relationship is statistically significant at a 95% confidence interval.

the standard error of the linear regressions. The selection of 3 models by Stephens et al. (2007) was somewhat arbitrary as they did not directly overlap the observations and did not agree as well as other models seasonally. This new approach relying on the correlated signal from all models leads to a NET flux of $-1.7 \pm 0.59 \text{ PgC yr}^{-1}$ and a T+SET flux of $0.15 \pm 0.66 \text{ PgC yr}^{-1}$, a similar shift in NE fluxes but only 2/3 of the shift in T+SET fluxes using the Stephens et al. (2007) subset of models, as shown in Fig. 3B.

For the RECCAP period, we used their Group 1 simulations (JENA, LSCE, MACC-II, CT2011_oi, CTE2013) identified in Peylin et al. (2013), four of which assimilated the observations at the sample time as opposed to using monthly means and all of which solved for fluxes at the resolution of the transport model or for small ecoregions over land. The T+SET flux estimate averaged over the RECCAP Group 1 models is $0.34 \pm 0.27 \text{ PgC yr}^{-1}$. This is nearly identical to the average of the new models from this study ($0.34 \pm 0.27 \text{ PgC yr}^{-1}$; using CTE2016-FT, CTE2017-FT, CT2016, CAMSv16r1, and Jena s85_v4.1). Both estimate slightly positive T+SET fluxes that are only half of the RECCAP all-model average ($0.93 \pm 0.90 \text{ PgC yr}^{-1}$). Our NET land sink estimates using newer models are less than the previous estimates in the original T3L2 and RECCAP studies for the 1992-1996 and 2001-2004 periods. Conversely, our new estimates suggest a change in the T+SET flux towards greater uptake / less emission for these periods; we found a decrease of the T+SET land flux by 0.71 PgC yr^{-1} from $0.56 \pm 0.32 \text{ PgC yr}^{-1}$ for the 1994-2004 period compare to $-0.15 \pm 0.43 \text{ PgC yr}^{-1}$ for the 2004-2014 period (Figure S9). Then, to obtain a flux estimate less sensitive to year to year variability we calculate the fluxes for the full 11-year 2004-2014 period (Fig. 3D), for which we have 5 model estimates. For this longer period, the model spread is largely reduced, in particular for the NET land fluxes, and again we find near neutral T+SET land fluxes. Taking all 4 of the estimation periods together (Table 2) all of our central estimates for T+SET are within 0.4 PgC yr^{-1} of zero. The tropical land fluxes are $-0.2 \pm 0.3 \text{ PgC yr}^{-1}$ for 2009-2011 and $0.0 \pm 0.12 \text{ PgC yr}^{-1}$ for 2004-2014. This implies a consistent uptake of carbon by intact tropical forests over several decades.

3.3 Variation in retrieved global carbon budgets

The global carbon budget partitioning for 2009-2011 is shown for our suite of models and for GCP2016 (river adjusted) on Fig. 4 with the model mean and GCP2016 reported in Table 3. On every panel of Fig. 4, the light pink error band shows the constraint imposed by fixing the values to those of GCP2016, and the associated equation is shown on the graph. The pink diamond represents the GCP2016 estimate while the cross and the gray shaded area show the model mean and 1 standard deviation in darker and 2 standard deviations in lighter gray. For the models, the total flux is calculated as the subtraction of the ocean and land sink from the FF emissions. Note that by mass conservation the total flux equals the whole-atmosphere growth rate (WAGR), but that WAGR may differ from the MBL atmosphere growth rate (AGR) defined by surface stations, because of sampling biases or interannual variability in tropospheric mixing or stratosphere-troposphere exchange. GCP uses the MBL AGR (Dlugokencky and Tans, 2018) as an estimate of total flux, and assigns uncertainty of $\pm 0.19 \text{ PgC yr}^{-1}$ (Le Quéré et al., 2016) for recent decades, with speculation that the relative uncertainty should decrease when averaging multiple years. Note that even though the CAMS results systematically align with the GCP2016 estimates in Fig. 4, the two are independent, except for the FF and for the atmospheric data that serve to estimate the total flux in GCP2016. By mass balance, the total annual flux

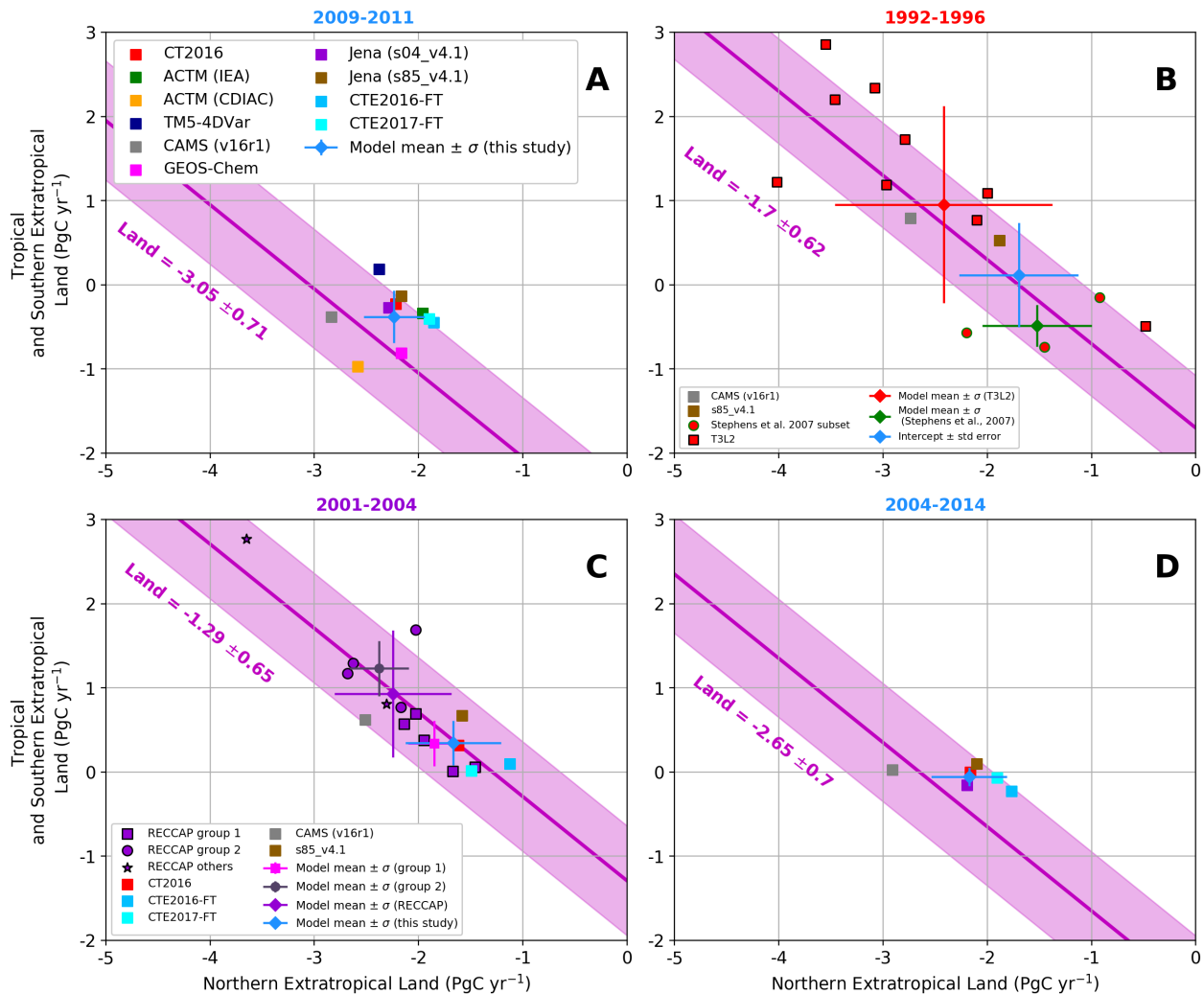


Figure 3. Tropical and Southern Extratropical (T+SET) versus Northern Extratropical (NET) land fluxes for the periods A) 2009-2011, B) 1992-1996, C) 2001-2004 and D) 2004-2014. The new models used in this study are represented by squares and the average of the available or selected simulations is shown in blue with 1 standard deviation error bars. The pink line and shaded area represents the GCP2016 (river adjusted) estimates of the total land sink for the given period. A) results for the HIPPO period 2009-2011, B) results for the T3L2 period 1992-1996. The TransCom 3 Level 2 outputs (Gurney et al., 2004) are shown in red, with the vertical gradient selected models from (Stephens et al., 2007) as circles outlined in green and the rest as red squares outlined in black. The intercept of the regression line with the observed vertical gradient (Fig. 2) is used to define our best flux estimate with error bars estimated by the standard error of the linear regression. C) results for the RECCAP period 2001-2004. Also, from Peylin et al. (2013), model means and standard deviations are shown in pink for the subgroup 1 (Jena, LSCEa, MACC-II, CTE2013, CT2011_oi) and in gray for the subgroup 2 (MATCH, CCAM, TrC, NICAM). Panel D shows the results from our new set of models for the period 2004-2014.

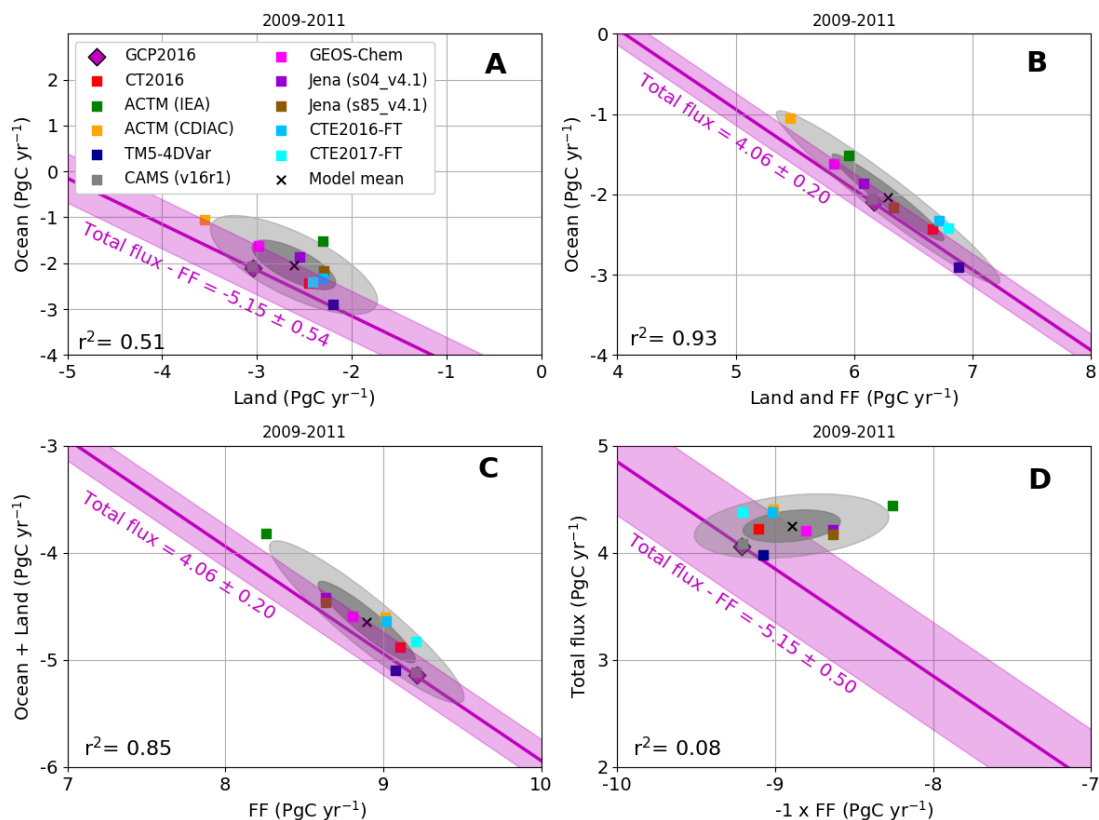


Figure 4. Synthesis of globally integrated fluxes for the 2009-2011 period, in PgC yr^{-1} . Each inversion is represented by a square and the model mean by a \times . The GCP2016 estimates are a pink diamond, sometimes hard to see because it is superimposed in each panel by the gray CAMS point. We have adjusted the GCP2016 ocean and land flux estimates by the riverine flux of carbon from land to ocean to atmosphere (0.45 PgC yr^{-1} ; Jacobson et al. 2007; Le Quéré et al. 2018), decreasing the ocean sink and increasing the land sink. The magenta line and light pink shaded area shows the corresponding mass balance estimates from GCP2016. In each panel the line and equation shown represent the sum of the x and y variables, and thus has a slope of -1 and any deviation perpendicular to the line indicates disagreement on the sum. Here we use the total flux which by mass balance is the whole-atmosphere growth rate (see text), and for panels A and D, the total flux - FF line also equals $O + L$, while for panels B and C, the total flux line equals $O + L + \text{FF}$. Ellipses denote the variability around the model means of 1σ (darker gray) and 2σ (lighter gray). A) ocean versus land; B) ocean versus land + FF; C) ocean + land versus FF; D) total flux versus $-1 \times \text{FF}$.

Time period	Source	Number of models	NET Land	T+SET Land
1992-1996	T3L2	12	-2.42 ± 1.05	0.95 ± 1.17
	Stephens et al. 2007	3	-1.52 ± 0.53	-0.49 ± 0.25
	T3L2 (Intercept)	12	-1.70 ± 0.59	0.15 ± 0.66
2001-2004	RECCAP All models	11	-2.25 ± 0.58	0.93 ± 0.90
	RECCAP Group 1	5	-1.85 ± 0.25	0.34 ± 0.27
	This study	5	-1.67 ± 0.46	0.34 ± 0.27
2009-2011	This study	10	-2.24 ± 0.29	-0.38 ± 0.31
2004-2014	This study	6	-2.17 ± 0.36	-0.05 ± 0.11

Table 2. Previous and our new best estimates (in bold) of the latitudinal partitioning of land fluxes over 4 time periods. All values are in PgC yr^{-1} . Values are indicated by the model mean ± 1 standard deviation or 1σ error uncertainties. Regarding the T3L2 period (Gurney et al., 2004), our new estimate for the 1992-1996 period comes from the intercept of the fit lines with the observations in Fig. 2B, and the uncertainties on these values from the standard error on these metrics from the fits. Regarding the RECCAP period (Peylin et al., 2013), our new estimate for the 2001-2004 period is the average of the 5 new models from this study.

must equal the total growth rate integrated over the entire atmosphere, and this is what we refer to as the total flux.

The integrated ocean versus land fluxes are presented in Fig. 4A. The equation for the range of ocean and land fluxes that would match FF and the total flux estimates from GCP2016 is also shown on Fig. 4A. The models and GCP2016 agree well on the ocean flux with a mean of $-2.04 \pm 0.51 \text{ PgC yr}^{-1}$ for the over the three years 2009-2011. The multi model mean of the land flux is $-2.61 \pm 0.42 \text{ PgC yr}^{-1}$. The GCP2016 land flux is $-3.04 \pm 0.5 \text{ PgC yr}^{-1}$ and thus overestimates the model mean. The cloud of model ocean versus land flux estimates are rather scattered around the model mean with a correlation coefficient of only 0.51.

To better understand the reasons for these discrepancies, and specifically to investigate how much of the land spread in Fig.

	FF	Land	Ocean	total flux
GCP2016	9.21 ± 0.46	-3.04 ± 0.50	-2.05 ± 0.50	4.06 ± 0.20
Multi-model	8.9 ± 0.29	-2.61 ± 0.42	-2.04 ± 0.51	4.25 ± 0.14

Table 3. Global Carbon budget for 2009 to 2011 inclusive, estimated by the Global Carbon Project 2016 (first row, with river adjustment) and by the suite of models from this study (second row), all values are in PgC yr^{-1} . Values are indicated by the model mean $\pm 1 \sigma$ error uncertainties, provided by GCP2016 or by the model standard deviation.

4a is a result of differences in fossil fuel priors, we plotted the ocean flux versus the sum of land and FF emissions in Fig. 4B.

This figure shows a tight correlation across models for these 2 parameters ($r^2 = 0.93$). Given that prior uncertainties specified in

the inversions for ocean fluxes are typically smaller than those for land, and fossil emissions are fixed, this implies for a given ocean and FF flux combination, the models are adjusting the land fluxes while matching CO₂ observations. While combining land and fossil fluxes together reduces the random scatter, it does not reduce the range of the continental fluxes, illustrating the fact that models do not simply compensate for biases in fossil priors with land fluxes, but rather that ocean fluxes are affected too (Saeki and Patra, 2017). Conversely, we plot the sum of ocean and land fluxes against FF emissions in Fig. 4C. This figure shows that the ocean + land total sink is largely controlled by the prescribed FF emissions. In general, the models use smaller fossil fuel sources than reported in GCP2016.

Fig. 4D compares the opposite of FF emissions versus the total flux, again defined by subtraction of the land and ocean fluxes from FF. The spread in models is not parallel to the line defined by the GCP2016 budget closure. We hypothesize that models that overestimate fossil emissions prioritize matching the spatial distribution of CO₂ and thus estimate overcompensating sinks. The spatial patterns of the different FF priors must also play a role, as well as the strength of the atmospheric constraint on annual time scales imposed by the inversion systems.

Taking the 2 extreme models the ACTM-CDIAC and TM5-4DVar estimates provide very different distributions of fluxes. ACTM-CDIAC suggests stronger land sinks, both over the NET and the T+SET regions, and a lower ocean sink while TM5-4DVar suggests the opposite. This leads to a range of around 2 PgC yr⁻¹ on the model ocean sink. Because of an intentionally different FF source, but with the same inversion system, the ACTM-CDIAC and ACTM-IEA retrieved land fluxes differ by slightly less than 1 PgC yr⁻¹ and ocean fluxes differ by half a PgC yr⁻¹. Overall, this analysis suggests that errors in FF priors are larger than the uncertainty prescribed to them, or more specifically, the range of FF estimates used by leading inversions exceeds the uncertainty that GCP2016 places on the CDIAC estimates. This implies that uncertainties in FF emissions do not adequately consider potential regional biases (Peylin et al., 2011; Thompson et al., 2016; Saeki and Patra, 2017). The large spread of model results along the mass balance line in 4C highlights the need to reduce uncertainty in estimates of FF emissions, and to develop modelling systems that relax rigid FF prior constraints and observational systems that can support optimizing FF emission estimates. For the period 1980-2015, the total flux estimates from GCP2016 are estimated by the MBL AGR of Dlugokencky and Tans (2018). Only background sites that are located in the MBL are used in this calculation. Ballantyne et al. (2012) calculated a sampling error of 0.38 PgC yr⁻¹ (2 σ) among the 40 sites, and GCP2017 estimate uncertainty of ± 0.19 PgC yr⁻¹ (1 σ) for the period 1980-2015 with respect to the total flux. We show the model retrieved WAGR (equal to total flux) for each individual year on Fig. 5 along with the GCP2016 estimate and error bars. The total spread in the total flux from the inverse models over the three years 2009-2011 equates to 1.38 PgC as shown in Fig. 5b. This is well outside of the uncertainty range estimated for the extrapolation of MBL measurements, implying several inversions are not rigidly constrained to match observed MBL AGR, even over periods of 3 years. Because CO₂ is variably mixed in different years and by different models in the troposphere and between the troposphere and the stratosphere, some inconsistency between the MBL-defined AGR and the total flux of CO₂ in the models might be expected. However, using CT2017 as a test case, the annual difference between the model total surface flux and the observed MBL growth rate over 2000-2016 has a standard deviation of 0.29 PgC yr⁻¹, and for three-year averages within this period a standard deviation of only 0.10 PgC yr⁻¹, much smaller than the discrepancies shown in Fig. 5. Buchwitz et al. (2018) made a similar AGR comparison using CAMS output of total column and surface data

and also found good agreement with differences of only $\pm 0.2 \text{ PgC yr}^{-1}$ (1σ) on an annual basis. Another potential challenge to inversions having a consistent total flux during this time may be due to large interannual variability in natural fluxes, with rapid changes resulting from different climatic conditions from the moderate El Niño of 2009 to the strong La Niña of 2011 (Bastos et al., 2013; Poulter et al., 2014). This period has also been marked by rapid changes in emissions, related to lower emissions in 2009 during the financial crisis and rapid increase in 2010 (Peters et al., 2011). However, Fig. 5 does not indicate that the model total flux estimates for the years 2009-2011 are more divergent than other years. Further work investigating these differences is needed but beyond the scope of this study. In particular, the length of assimilation window may have an impact. It may also be possible to force the inverse systems to agree, at least within the MBL, with the observationally defined AGR, and this may help to reduce model spread elsewhere.

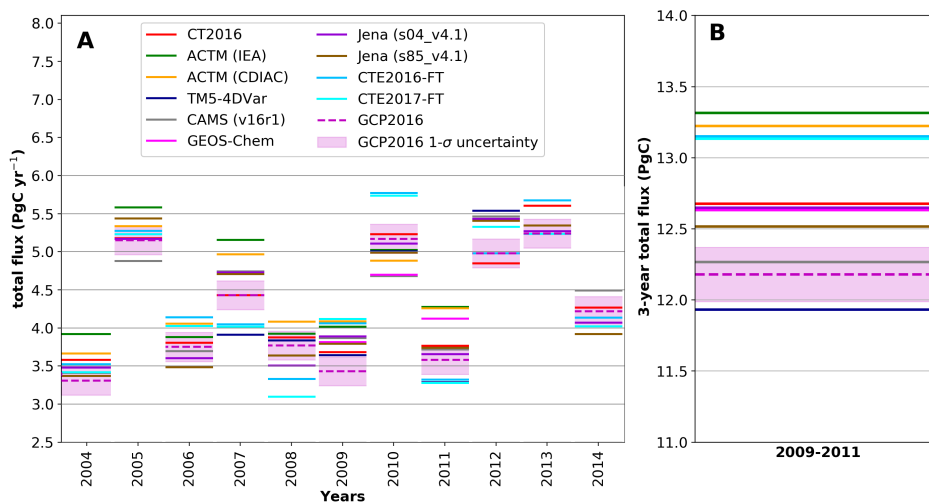


Figure 5. Modelled total flux (lines), equal to whole-atmosphere growth rate, that is the difference between the global FF emissions and the land and ocean fluxes. Atmospheric growth rate from GCP2016, derived from atmospheric CO_2 measurements made in the marine boundary layer by the NOAA/ESRL flask network (Masarie and Tans, 1995; Dlugokencky and Tans, 2018) and GCP2016 assigned uncertainty (pink bands). Panel B shows the sum of the total flux for the three years (2009 to 2011).

10 4 Summary and future work

Atmospheric transport has long been a major contributor to top-down atmospheric inverse model flux uncertainty. We applied the technique of Stephens et al. (2007) to a suite of state-of-the-art inversion systems assimilating primarily surface observations to take advantage of the unique HIPPO global airborne data set for independent validation in assessing fluxes. We also compared the models to each other and to the GCP2016 carbon budget synthesis. The major findings of these comparisons can be summarized as follows:

1. Model estimates of the latitudinal distribution of land fluxes are remarkably consistent across models and this represents a convergence over the past 15 years of inverse model development. The standard deviation across our 10 simulations of the difference between northern extratropical land and tropical land fluxes is 0.4 PgC yr^{-1} for the period 2009-2011 and 0.43 PgC yr^{-1} for the period 2004-2014 across 5 models. These are considerable reductions from 2.1 PgC yr^{-1} for 12 simulations in T3L2 (differing only in transport modelling) for the period 1992-1996 and 1.28 PgC yr^{-1} for 11 simulations in RECCAP study for the period 2001-2004.
2. Our suite of 10 inversions gives a NET land uptake of $-2.22 \pm 0.27 \text{ PgC yr}^{-1}$ (1σ) and a net T+SET uptake of $-0.37 \pm 0.31 \text{ PgC yr}^{-1}$ for 2009-2011 ($-0.2 \pm 0.3 \text{ PgC yr}^{-1}$ for the tropics only). For 2004-2014, a subset of 6 models gives NE land uptake of $-2.17 \pm 0.36 \text{ PgC yr}^{-1}$, T+SET uptake of $-0.06 \pm 0.11 \text{ PgC yr}^{-1}$, and T of $0.0 \pm 0.12 \text{ PgC yr}^{-1}$, thus allowing for deforestation implying a strong uptake in intact tropical forests, in line with forest inventories (Pan et al., 2011).
3. The group of RECCAP models that primarily assimilated discrete rather than monthly mean observations agree with estimates from our subset of 5 newer models regarding the lack of strong net emissions from tropical land. This is not too surprising because most of our models, with the exception of LSCEa, are the updated versions of the same models in the RECCAP Group 1 (Peylin et al., 2013). Those 5 models estimated a net NET land sink of $-1.85 \pm 0.25 \text{ PgC yr}^{-1}$ and our subset of 4 models covering the RECCAP period estimate $-1.71 \pm 0.5 \text{ PgC yr}^{-1}$. Regarding T+SET, the newer model estimate is a source of $0.34 \pm 0.27 \text{ PgC yr}^{-1}$, while it is $0.34 \pm 0.27 \text{ PgC yr}^{-1}$ in RECCAP's Group 1.
4. For the 1992-1996 period, we define an update to the Stephens et al. (2007) result, using the intercept of the model output linear regression with the observed annual mean vertical gradient of 0.7 ppm , leading to a NE land uptake of $-1.7 \pm 0.57 \text{ PgC yr}^{-1}$ and a T+SET flux of $0.12 \pm 0.62 \text{ PgC yr}^{-1}$ for 1992-1996. Our results for the more recent decadal period, the 11 years from 2004 to 2014 indicate a somewhat larger NET sink of $2.21 \pm 0.34 \text{ PgC yr}^{-1}$ and a neutral tropical land flux of $0.04 \pm 0.13 \text{ PgC yr}^{-1}$, in line with a trend of a larger land sink (Sarmiento et al., 2010; Keenan et al., 2016) if shared across both latitudinal bands.
5. We present our best estimates of the latitudinal land flux partitioning for the 4 periods 1992-1996, 2001-2004, 2009-2011, and 2004-2014 in Table 2. We present in Fig. 6 the time series of the NET and T+SET land fluxes from 1979 to 2016, using all simulations available in this study. This figure shows a decrease of the T+SET land flux by 0.71 PgC yr^{-1} , from $+0.56 \text{ PgC yr}^{-1}$ to $-0.15 \text{ PgC yr}^{-1}$ between the decades 1994-2004 and 2004-2014, respectively. The land-use change flux estimated by GCP2017 was nearly identical for these two time periods ($+1.31$ and $+1.29 \text{ PgC yr}^{-1}$, respectively), and assuming these numbers primarily reflect tropical land-use change emissions this implies an increase in the intact tropical forest sink on decadal timescales. Our re-evaluations of the T3L2 and RECCAP study results (Table 2) confirms that the sum of the tropics and southern extratropics have been near neutral for several decades, despite large scale tropical deforestation, and in accordance with the recent literature on the tropical land carbon budget (Hansen et al., 2013; Keenan et al., 2016; Mitchard, 2018).

6. At global scale, we find in agreement with earlier studies that our model results are strongly dependent on the prescribed FF emissions. While the total of global land and ocean uptake adjusts to match differences in FF emissions, this compensation is not perfect.
7. Our suite of 10 simulations also retrieve surprisingly different three-year whole atmospheric growth rates, as defined by the total fluxes. The model range is 1.38 PgC over 3 years, compared to an estimated uncertainty of ± 0.10 PgC in CT2017 matching between MBL CO₂ concentration trends and total flux over three years and a 0.2 PgC yr⁻¹ uncertainty assigned by GCP2017. The yearly ranges of up to 1 PgC yr⁻¹ in the model total flux estimates imply 0.5 ppm disagreements in whole-atmosphere CO₂ concentrations, and the 1.4 PgC yr⁻¹ range for the 3-year period implies disagreements of 0.7 ppm in the whole-atmosphere CO₂ concentration change over that time period.

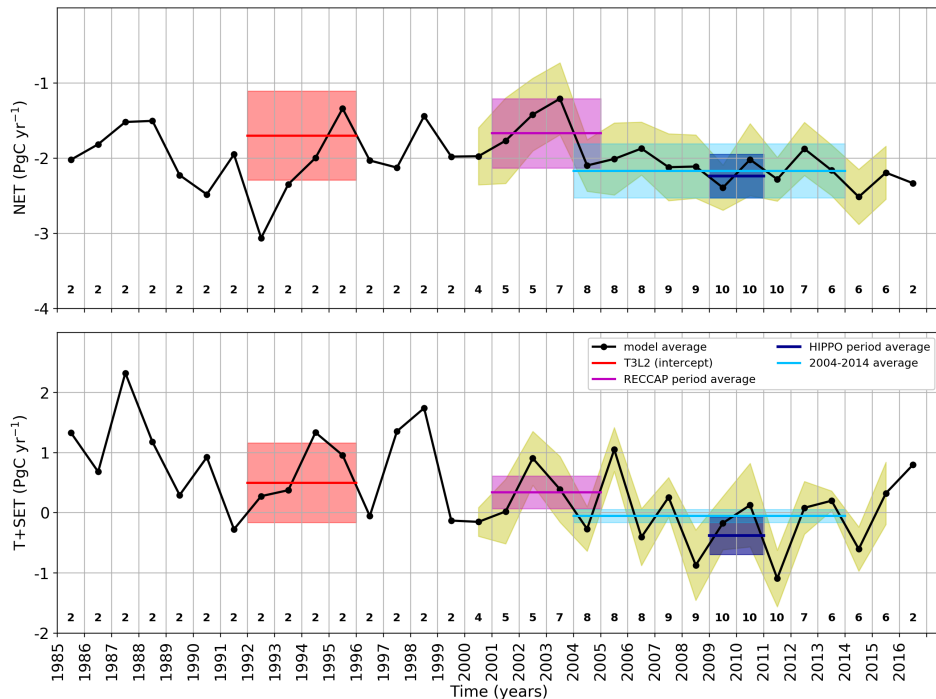


Figure 6. Time series of annual land fluxes for the NET (top panel) and the T+SET (bottom panel). The black line represents the model mean and standard deviation derived from available simulations, the number of simulations is shown by the numbers below the curve. The standard deviation is shown only if there are more than 2 model simulations available. Estimates from the specific period (Table 2) are added as multi-year average and standard deviation (shaded area).

- 10 Across 7 state-of-the-art systems running 10 inversions, there does not appear to be a correlation between posterior NET vertical gradients and the retrieved latitudinal distribution of land fluxes in winter and for the annual mean. This is suggesting that northern hemisphere vertical mixing, albeit significant in summer, is not currently the major driver of tropical versus

northern extratropical land flux spread. However, transport errors can still contribute significantly to the flux estimates. Repeating the experiment in T3L2 where transport was the only thing that varied across models would be required to rule out other factors masking a vertical mixing effect, but given the diversity of modern inversion systems this is not practical. Having a common FF prior, eventually also optimized and with known uncertainties, would improve our ability to retrieve the natural
5 fluxes. Other components, such as other features of the transport models, the prior fluxes or the inversion method also drive the discrepancies in the global atmospheric budget. The ocean, land, and ocean-land partitioning appear to be a function of the FF prior. This also results in large differences in retrieved total flux or WAGR. The increase of the absolute error in fossil fuel emissions and the large sensitivity of the carbon uptake estimates to those errors (Ballantyne et al., 2015), suggests that despite being thought to be the best-known term in the global carbon budget, systematic errors in fossil fuel emission estimates limit
10 our assessment of the natural fluxes and the ocean-land partitioning from this inversion ensemble.

Our ability to isolate transport effects in this study is limited in comparison to T3L2 in that many other features of the inversion systems also vary; however, this variability allows us to assess the state of the art FF inventories and their importance in the retrieved flux estimates. Gurney et al. (2005), Peylin et al. (2011) and Saeki and Patra (2017) already demonstrated the importance of FF emission uncertainties in inverse modelling studies, suggesting the importance of temporally defined emis-
15 sion inventory. However, Peylin et al. (2011) found that transport errors were still the main source of uncertainty in regional inversions. With the aim of quantifying CO₂ fluxes at regional scales, it is more than ever necessary to assess systematic errors of inverse modelling results with independent in-situ observations. FF emissions could be optimized with the addition of additional species (e.g., Turnbull et al., 2011; Nathan et al., 2018), such as carbon monoxide (Liu et al., 2017; Bowman et al., 2017), although it can be challenging at the most local and urban scales (Ammoura et al., 2016).

20 There is a significant correlation between NET vertical gradients and the T+SET retrieved fluxes in summer only. This study reaffirms that systematic evaluation of posterior concentrations against independent measurements is essential to assess the biases and accuracy of inverse modelling systems. Future work will naturally involve comparison against CO₂ observations from the more recent NASA Atmospheric Tomography (ATom) project, which is similar to HIPPO, but is augmented with additional flights over the Atlantic Ocean (see e.g. Prather et al. 2017) and an extensive atmospheric chemistry payload, and will
25 involve the inclusion of models assimilating satellite total column CO₂ measurements. It is possible that the larger observation coverage from satellite observations, expanded ¹⁴CO₂ measurements, and urban and power-plant scale observations will help to narrow down the FF emissions, which in turn will allow us to better evaluate inverse model global and regional land and ocean CO₂ flux estimates.

Acknowledgements. We thank the entire HIPPO team for making these measurements possible and available. We gratefully acknowledge
30 all the hard work that went into collecting and cross-validating the HIPPO CO₂ data set: for the Harvard QCLS and OMS instruments by Bruce Daube, Greg Santoni, Jasna Pittman, Rodrigo Jimènez-Pizarro, Roisin Commane, Bin Xiang, and Sunyoung Park; for the NCAR AO2 instrument and NCAR/Scripps Medusa flask sampler by Ralph Keeling, Jonathan Bent, Andrew Watt, and Stephen Shertz; and for the NOAA PFP sampler by Colm Sweeney, Fred Moore, Ben Miller, Steve Montzka, Eric Hintsa, Jim Elkins, and Sonja Wolter. We would like to thank

Andrew Jacobson for helping in the analysis of the Carbon Tracker results and the entire CarbonTracker team. CarbonTracker CT2016 and CarbonTracker CT2017 results are provided by NOAA ESRL, Boulder, Colorado, USA from the website at <http://carbontracker.noaa.gov>. The National Center for Atmospheric Research is sponsored by the National Science Foundation. This work was supported by NASA ROSES grant NNX15AG87G. F. Chevallier is funded by the Copernicus Atmosphere Monitoring Service, implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Commission, and was granted access to the HPC resources of TGCC under the allocation A0030102201. P. K. Patra and T. Saeki are partly supported by Japan Environment Research and Technology Development Fund of the Ministry of the Environment (grant no. 2-1701, 2-1401). F. Deng was supported by Environment and Climate Change Canada (ECCC) and Canada Space Agency (CSA).

References

- Ammoura, L., Xueref-Remy, I., Vogel, F., Gros, V., Baudic, A., Bonsang, B., Delmotte, M., Té, Y., and Chevallier, F.: Exploiting stagnant conditions to derive robust emission ratio estimates for CO₂, CO and volatile organic compounds in Paris, *Atmospheric Chemistry and Physics*, 16, 15 653–15 664, <https://doi.org/10.5194/acp-16-15653-2016>, 2016.
- 5 Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P., and White, J. W. C.: Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years, *Nature*, 488, 70–72, <https://doi.org/10.1038/nature11299>, 2012.
- 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., and White, J. W. C.: 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>, 2015.
- 10 Bastos, A., Running, S. W., Gouveia, C., and Trigo, R. M.: The global NPP dependence on ENSO: La Niña and the extraordinary year of 2011, *Geophys. Res. Biogeosci.*, 118, 1247–1255, <https://doi.org/10.1002/jgrg.20100>, 2013.
- Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO₂ fluxes estimated from GOSAT retrievals of total column CO₂, *Atmospheric Chemistry & Physics*, 13, 8695–8717, <https://doi.org/10.5194/acp-13-8695-2013>, 2013.
- 15 Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, *Nature*, 525, 47–55, <https://doi.org/10.1038/nature14956>, 2015.
- Boden, T., Andres, R., and Marland, G.: Regional, and National Fossil-Fuel CO₂ Emissions, https://doi.org/10.3334/CDIAC/00001_V2016, 2016.
- Bowman, K. W., Liu, J., Bloom, A. A., Parazoo, N. C., Lee, M., Jiang, Z., Menemenlis, D., Gierach, M. M., Collatz, G. J., Gurney, K. R., and Wunch, D.: Global and Brazilian Carbon Response to El Niño Modoki 2011-2010, *Earth and Space Science*, 4, 637–660, <https://doi.org/10.1002/2016EA000204>, 2017.
- Buchwitz, M., Reuter, M., Schneising, O., Noël, S., Gier, B., Bovensmann, H., Burrows, J. P., Boesch, H., Anand, J., Parker, R. J., Somkuti, P., Detmers, R. G., Hasekamp, O. P., Aben, I., Butz, A., Kuze, A., Suto, H., Yoshida, Y., Crisp, D., and O’Dell, C.: Computation and analysis of atmospheric carbon dioxide annual mean growth rates from satellite observations during 2003–2016, *Atmospheric Chemistry and Physics*, 18, 17 355–17 370, <https://doi.org/10.5194/acp-18-17355-2018>, 2018.
- 25 Canadell, J. G., Ciais, P., Gurney, K., Quéré, C. L., Piao, S., Raupach, M. R., and Sabine, C. L.: An International Effort to Quantify Regional Carbon Fluxes, *Eos, Transactions American Geophysical Union*, 92, 81, <https://doi.org/10.1029/2011eo100001>, 2011.
- Chevallier, F.: On the statistical optimality of CO₂ atmospheric inversions assimilating CO₂ column retrievals, *Atmospheric Chemistry and Physics*, 15, 11 133–11 145, <https://doi.org/10.5194/acp-15-11133-2015>, 2015.
- 30 Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO₂ sources and sinks from satellite observations: Method and application to TOVS data, *Journal of Geophysical Research (Atmospheres)*, 110, D24309, <https://doi.org/10.1029/2005JD006390>, 2005.
- Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguá, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *Journal of Geophysical Research*, 115, <https://doi.org/10.1029/2010jd013887>, 2010.

- Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional CO₂ flux estimates from in situ and spaceborne measurements of atmospheric CO₂, *Geophysical Research Letters*, 41, 1065–1070, <https://doi.org/10.1002/2013gl058772>, 2014.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597, <https://doi.org/10.1002/qj.828>, 2011.
- 5 Deng, F., Jones, D. B. A., Henze, D. K., Bousserez, N., Bowman, K. W., Fisher, J. B., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P. O., Kort, E. A., Wofsy, S. C., Blumenstock, T., Deutscher, N. M., Griffith, D. W. T., Hase, F., Heikkinen, P., Sherlock, V., Strong, K., Sussmann, R., and Warneke, T.: Inferring regional sources and sinks of atmospheric CO₂ from GOSAT XCO₂ data, *Atmospheric Chemistry & Physics*, 14, 3703–3727, <https://doi.org/10.5194/acp-14-3703-2014>, 2014.
- 10 Deng, F., Jones, D. B. A., Walker, T. W., Keller, M., Bowman, K. W., Henze, D. K., Nassar, R., Kort, E. A., Wofsy, S. C., Walker, K. A., Bourassa, A. E., and Degenstein, D. A.: Sensitivity analysis of the potential impact of discrepancies in stratosphere troposphere exchange on inferred sources and sinks of CO₂, *Atmospheric Chemistry and Physics*, 15, 11 773–11 788, <https://doi.org/10.5194/acp-15-11773-2015>, 2015.
- 15 Dlugokencky, E. and Tans, P.: Trends in atmospheric carbon dioxide, National Oceanic & Atmospheric Administration, Earth System Research Laboratory (NOAA/ESRL), available at: <http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html>, 2018.
- 20 Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A., Liu, J., Schwandner, F. M., Sun, Y., O'Dell, C. W., Frankenberg, C., Taylor, T., Fisher, B., Osterman, G. B., Wunch, D., Hakkarainen, J., Tamminen, J., and Weir, B.: The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes, *Science*, 358, eaam5745, <https://doi.org/10.1126/science.aam5745>, 2017.
- Frankenberg, C., Kulawik, S. S., Wofsy, S. C., Chevallier, F., Daube, B., Kort, E. A., O'Dell, C., Olsen, E. T., and Osterman, G.: Using airborne HIAPER Pole-to-Pole Observations (HIPPO) to evaluate model and remote sensing estimates of atmospheric carbon dioxide, *Atmospheric Chemistry & Physics*, 16, 7867–7878, <https://doi.org/10.5194/acp-16-7867-2016>, 2016.
- 25 Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *Journal of Climate*, 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- 30 Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), *Journal of Geophysical Research: Biogeosciences*, 118, 317–328, <https://doi.org/10.1002/jgrg.20042>, 2013.
- 35 Graven, H. D., Keeling, R. F., Piper, S. C., Patra, P. K., Stephens, B. B., Wofsy, S. C., Welp, L. R., Sweeney, C., Tans, P. P., Kelley, J. J., Daube, B. C., Kort, E. A., Santoni, G. W., and Bent, J. D.: Enhanced Seasonal Exchange of CO₂ by Northern Ecosystems Since 1960, *Science*, 341, B44A-04, <https://doi.org/10.1126/science.1239207>, 2013.

- Gurney, K. R. and Denning, A. S.: TransCom 3: Seasonal CO₂ Flux Estimates from Atmospheric Inversions (Level 2), <https://doi.org/10.3334/ORNLDAAC/1198>, 2013.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.-W.: Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models, *Nature*, 415, 626–630, <https://doi.org/10.1038/415626a>, 2002.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Kowalczyk, E., Maki, T., Maksyutov, S., Peylin, P., Prather, M., Pak, B. C., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.-W.: TransCom 3 CO₂ inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, *Tellus Series B Chemical and Physical Meteorology B*, 55, 555–579, <https://doi.org/10.1029/J09-94479>, 2003.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M., and Taguchi, S.: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, *Global Biogeochemical Cycles*, 18, n/a–n/a, <https://doi.org/10.1029/2003gb002111>, <https://doi.org/10.1029/2003GB002111>, 2004.
- Gurney, K. R., Chen, Y.-H., Maki, T., Kawa, S. R., Andrews, A., and Zhu, Z.: Sensitivity of atmospheric CO₂ inversions to seasonal and interannual variations in fossil fuel emissions, *Journal of Geophysical Research*, 110, D10308, <https://doi.org/10.1029/2004JD005373>, 2005.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G.: High-Resolution Global Maps of 21st-Century Forest Cover Change, *Science*, 342, 850–853, <https://doi.org/10.1126/science.1244693>, 2013.
- Healy, S. B.: Forecast impact experiment with a constellation of GPS radio occultation receivers, *Atmospheric Science Letters*, 9, 111–118, <https://doi.org/10.1002/asl.169>, 2008.
- Houweling, S., Aben, I., Breon, F.-M., Chevallier, F., Deutscher, N., Engelen, R., Gerbig, C., Griffith, D., Hungershofer, K., Macatangay, R., Marshall, J., Notholt, J., Peters, W., and Serrar, S.: The importance of transport model uncertainties for the estimation of CO₂ sources and sinks using satellite measurements, *Atmospheric Chemistry and Physics*, 10, 9981–9992, <https://doi.org/10.5194/acp-10-9981-2010>, 2010.
- Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., Deng, F., Dlugokencky, E. J., Feng, L., Ganshin, A., Hasekamp, O., Jones, D., Maksyutov, S., Marshall, J., Oda, T., O'Dell, C. W., Oshchepkov, S., Palmer, P. I., Peylin, P., Poussi, Z., Reum, F., Takagi, H., Yoshida, Y., and Zhuravlev, R.: An intercomparison of inverse models for estimating sources and sinks of CO₂ using GOSAT measurements, *Journal of Geophysical Research: Atmospheres*, 120, 5253–5266, <https://doi.org/10.1002/2014jd022962>, 2015.
- Jacobson, A. R., Fletcher, S. E. M., Gruber, N., Sarmiento, J. L., and Gloor, M.: A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, *Global Biogeochemical Cycles*, 21, <https://doi.org/10.1029/2005gb002556>, 2007.
- Keenan, T. F., Prentice, I. C., Canadell, J. G., Williams, C. A., Wang, H., Raupach, M., and Collatz, G. J.: Recent pause in the growth rate of atmospheric CO₂ due to enhanced terrestrial carbon uptake, *Nature Communications*, 7, 13 428, <https://doi.org/10.1038/ncomms13428>, 2016.
- Kondo, M., Ichii, K., Patra, P. K., Canadell, J. G., Poulter, B., Sitch, S., Calle, L., Liu, Y. Y., van Dijk, A. I. J. M., Saeki, T., Saigusa, N., Friedlingstein, P., Arneth, A., Harper, A., Jain, A. K., Kato, E., Koven, C., Li, F., Pugh, T. A. M., Zaehle, S., Wiltshire, A., Chevallier, F.,

- Maki, T., Nakamura, T., Niwa, Y., and Rödenbeck, C.: Land use change and El Niño-Southern Oscillation drive decadal carbon balance shifts in Southeast Asia, *Nature Communications*, 9, 074011, <https://doi.org/10.1038/s41467-018-03374-x>, 2018.
- Krol, M., de Bruine, M., Killaars, L., Ouwersloot, H., Pozzer, A., Yin, Y., Chevallier, F., Bousquet, P., Patra, P., Belikov, D., Maksyutov, S., Dhomse, S., Feng, W., and Chipperfield, M. P.: Age of air as a diagnostic for transport timescales in global models, *Geoscientific Model Development*, 11, 3109–3130, <https://doi.org/10.5194/gmd-11-3109-2018>, 2018.
- Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, *Applied Optics*, 48, 6716, <https://doi.org/10.1364/ao.48.006716>, 2009.
- Le Quéré, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., Manning, A. C., Boden, T. A., Tans, P. P., Houghton, R. A., Keeling, R. F., Alin, S., Andrews, O. D., Anthoni, P., Barbero, L., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Currie, K., Delire, C., Doney, S. C., Friedlingstein, P., Gkritzalis, T., Harris, I., Hauck, J., Haverd, V., Hoppema, M., Klein Goldewijk, K., Jain, A. K., Kato, E., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lombardozi, D., Melton, J. R., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-i., O'Brien, K., Olsen, A., Omar, A. M., Ono, T., Pierrot, D., Poulter, B., Rödenbeck, C., Salisbury, J., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Sutton, A. J., Takahashi, T., Tian, H., Tilbrook, B., van der Laan-Luijckx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2016, *Earth System Science Data*, 8, 605–649, <https://doi.org/10.5194/essd-8-605-2016>, 2016.
- Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Jackson, R. B., Boden, T. A., Tans, P. P., Andrews, O. D., Arora, V. K., Bakker, D. C. E., Barbero, L., Becker, M., Betts, R. A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Cosca, C. E., Cross, J., Currie, K., Gasser, T., Harris, I., Hauck, J., Haverd, V., Houghton, R. A., Hunt, C. W., Hurtt, G., Ilyina, T., Jain, A. K., Kato, E., Kautz, M., Keeling, R. F., Goldewijk, K. K., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lima, I., Lombardozi, D., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., ichiro Nakaoka, S., Nojiri, Y., Padin, X. A., Pregon, A., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Reimer, J., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijckx, I. T., van der Werf, G. R., van Heuven, S., Viovy, N., Vuichard, N., Walker, A. P., Watson, A. J., Wiltshire, A. J., Zaehle, S., and Zhu, D.: Global Carbon Budget 2017, *Earth System Science Data*, 10, 405–448, <https://doi.org/10.5194/essd-10-405-2018>, 2018.
- Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., Bloom, A. A., Wunch, D., Frankenberg, C., Sun, Y., O'Dell, C. W., Gurney, K. R., Menemenlis, D., Gierach, M., Crisp, D., and Eldering, A.: Contrasting carbon cycle responses of the tropical continents to the 2015–2016 El Niño, *Science*, 358, eaam5690, <https://doi.org/10.1126/science.aam5690>, 2017.
- Locatelli, R., Bousquet, P., Saunois, M., Chevallier, F., and Cressot, C.: Sensitivity of the recent methane budget to LMDz sub-grid-scale physical parameterizations, *Atmospheric Chemistry & Physics*, 15, 9765–9780, <https://doi.org/10.5194/acp-15-9765-2015>, 2015.
- Masarie, K. A. and Tans, P. P.: Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record, *Journal of Geophysical Research*, 100, 11 593, <https://doi.org/10.1029/95jd00859>, 1995.
- Mitchard, E. T. A.: The tropical forest carbon cycle and climate change, *Nature*, 559, 527–534, <https://doi.org/10.1038/s41586-018-0300-2>, 2018.
- Nathan, B., Lauvaux, T., Turnbull, J., and Gurney, K.: Investigations into the use of multi-species measurements for source apportionment of the Indianapolis fossil fuel CO₂ signal, *Elem Sci Anth*, 6, A51K-0240, <https://doi.org/10.1525/elementa.131>, <http://doi.org/10.1525/elementa.131>, 2018.

- Newsam, G. N. and Enting, I. G.: Inverse problems in atmospheric constituent studies. I. Determination of surface sources under a diffusive transport approximation, *Inverse Probl*, 4, 1037–1054, <https://doi.org/10.1088/0266-5611/4/4/008>, 1988.
- Orbe, C., Waugh, D. W., Yang, H., Lamarque, J.-F., Tilmes, S., and Kinnison, D. E.: Tropospheric transport differences between models using the same large-scale meteorological fields, *Geophysical Research Letters*, 44, 1068–1078, <https://doi.org/10.1002/2016gl071339>, 2017.
- Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., and Hayes, D.: A Large and Persistent Carbon Sink in the World's Forests, *Science*, 333, 988–993, <https://doi.org/10.1126/science.1201609>, 2011.
- Patra, P. K., Niwa, Y., Schuck, T. J., Brenninkmeijer, C. A. M., Machida, T., Matsueda, H., and Sawa, Y.: Carbon balance of South Asia constrained by passenger aircraft CO₂ measurements, *Atmospheric Chemistry and Physics*, 11, 4163–4175, <https://doi.org/https://doi.org/10.5194/acp-11-4163-2011>, 2011.
- Peters, G. P., Marland, G., Quéré, C. L., Boden, T., Canadell, J. G., and Raupach, M. R.: Rapid growth in CO₂ emissions after the 2008-2009 global financial crisis, *Nature Climate Change*, 2, 2–4, <https://doi.org/10.1038/nclimate1332>, 2011.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proceedings of the National Academy of Sciences*, 104, 18 925–18 930, <https://doi.org/10.1073/pnas.0708986104>, 2007.
- Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., Vermeulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G., Ciais, P., and Heimann, M.: Importance of fossil fuel emission uncertainties over Europe for CO₂ modeling: model intercomparison, *Atmospheric Chemistry and Physics*, 11, 6607–6622, <https://doi.org/10.5194/acp-11-6607-2011>, 2011.
- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions, *Biogeosciences*, 10, 6699–6720, <https://doi.org/10.5194/bg-10-6699-2013>, 2013.
- Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., Running, S. W., Sitch, S., and van der Werf, G. R.: Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, *Nature*, 509, 600–603, <https://doi.org/10.1038/nature13376>, 2014.
- Prather, M. J., Zhu, X., Strahan, S. E., Steenrod, S. D., and Rodriguez, J. M.: Quantifying errors in trace species transport modeling, *Proceedings of the National Academy of Sciences*, 105, 19 617–19 621, <https://doi.org/10.1073/pnas.0806541106>, 2008.
- Prather, M. J., Zhu, X., Flynn, C. M., Strode, S. A., Rodriguez, J. M., Steenrod, S. D., Liu, J., Lamarque, J.-F., Fiore, A. M., Horowitz, L. W., Mao, J., Murray, L. T., Shindell, D. T., and Wofsy, S. C.: Global atmospheric chemistry - which air matters, *Atmospheric Chemistry & Physics*, 17, 9081–9102, <https://doi.org/10.5194/acp-17-9081-2017>, 2017.
- Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R.: Reconstructing the recent carbon cycle from atmospheric CO₂, δ¹³C and O₂/N₂ observations, *Tellus Series B Chemical and Physical Meteorology B*, 51, <https://doi.org/J99-48306>, 1999.
- Rödenbeck, C.: Estimating CO₂ sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport, Tech. rep., Technical Report 6, Max Planck Institute for Biogeochemistry, Jena., 2005.
- Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO₂ flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmospheric Chemistry and Physics*, 3, 1919–1964, <https://doi.org/10.5194/acp-3-1919-2003>, 2003.

- Saeki, T. and Patra, P. K.: Implications of overestimated anthropogenic CO₂ emissions on East Asian and global land CO₂ flux inversion, *Geoscience Letters*, 4, <https://doi.org/10.1186/s40562-017-0074-7>, 2017.
- Santoni, G. W., Daube, B. C., Kort, E. A., Jiménez, R., Park, S., Pittman, J. V., Gottlieb, E., Xiang, B., Zahniser, M. S., Nelson, D. D., McManus, J. B., Peischl, J., Ryerson, T. B., Holloway, J. S., Andrews, A. E., Sweeney, C., Hall, B., Hints, E. J., Moore, F. L., Elkins, J. W., Hurst, D. F., Stephens, B. B., Bent, J., and Wofsy, S. C.: 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, *Atmospheric Measurement Techniques*, 7, 1509–1526, <https://doi.org/10.5194/amt-7-1509-2014>, 2014.
- Sarmiento, J. L., Gloor, M., Gruber, N., Beaulieu, C., Jacobson, A. R., Fletcher, S. E. M., Pacala, S., and Rodgers, K.: Trends and regional distributions of land and ocean carbon sinks, *Biogeosciences*, 7, 2351–2367, <https://doi.org/10.5194/bg-7-2351-2010>, 2010.
- Schimel, D., Stephens, B. B., and Fisher, J. B.: Effect of increasing CO₂ on the terrestrial carbon cycle, *Proceedings of the National Academy of Sciences*, 112, 436–441, <https://doi.org/https://doi.org/10.1073/pnas.1407302112>, 2015.
- Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R. L., Steele, L. P., Francey, R. J., and Denning, A. S.: Weak Northern and Strong Tropical Land Carbon Uptake from Vertical Profiles of Atmospheric CO₂, *Science*, 316, 1732–1735, <https://doi.org/10.1126/science.1137004>, 2007.
- Tans, P. P., Fung, I. Y., and Takahashi, T.: Observational constraints on the global atmospheric CO₂ budget, *Science*, 247, 1431–1438, <https://doi.org/10.1126/science.247.4949.1431>, 1990.
- 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., and Ciais, P.: Top down assessment of the Asian carbon budget since the mid 1990s, *Nature Communications*, 7, 10724, <https://doi.org/10.1038/ncomms10724>, <http://dx.doi.org/10.1038/ncomms10724>, 2016.
- Turnbull, J. C., Karion, A., Fischer, M. L., Faloona, I., Guilderson, T., Lehman, S. J., Miller, B. R., Miller, J. B., Montzka, S., Sherwood, T., Saripalli, S., Sweeney, C., and Tans, P. P.: Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over Sacramento, California in spring 2009, *Atmospheric Chemistry & Physics*, 11, 705–721, <https://doi.org/10.5194/acp-11-705-2011>, 2011.
- van der Laan-Luijkx, I. T., van der Velde, I. R., van der Veen, E., Tsuruta, A., Stanislawski, K., Babenhauerheide, A., Zhang, H. F., Liu, Y., He, W., Chen, H., Masarie, K. A., Krol, M. C., and Peters, W.: The CarbonTracker Data Assimilation Shell (CTDAS) v1.0: implementation and global carbon balance 2001-2015, *Geoscientific Model Development*, 10, 2785–2800, <https://doi.org/10.5194/gmd-10-2785-2017>, 2017.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009), *Atmospheric Chemistry and Physics*, 10, 11707–11735, <https://doi.org/10.5194/acp-10-11707-2010>, 2010.
- Wofsy, S. C.: 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*, 369, 2073–2086, <https://doi.org/10.1098/rsta.2010.0313>, 2011.