Sources of Uncertainty in Regional and Global Terrestrial CO₂ Exchange Estimates

To cite this version:

HAL Id: hal-02904064
https://hal.inrae.fr/hal-02904064
Submitted on 21 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License
Sources of Uncertainty in Regional and Global Terrestrial CO₂ Exchange Estimates


1Department of Geography, Ludwig-Maximilians Universität, München, Germany, 2College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK, 3Laboratoire des Sciences du Climat et de l’Environnement, CEA-CNRS-UVSQ, UMR8212, Gif-sur-Yvette, France, 4University Paris-Saclay, AgroParisTech, INRAE, UMR 211, Thiverval-Grignon, France, 5CIRED, Nogent-sur-Marne, France, 6College of Life and Environmental Sciences, University of Exeter, Exeter, UK, 7Max Planck Institute for Biogeochemistry, Jena, Germany, 8Max Planck Institute for Meteorology, Hamburg, Germany, 9Department of Meteorology and Air Quality, Wageningen University and Research, Wageningen, Netherlands, 10Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan, 11CSIRO Oceans and Atmosphere, Canberra, ACT, Australia, 12Department of Geoscience, Environment and Society, Université Libre de Bruxelles, Bruxelles, Belgium, 13Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China, 14Centre for Isotope Research, University of Groningen, Groningen, Netherlands, 15Department of Geography, University of Augsburg, Augsburg, Germany, 16Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 17Institute of Applied Energy, Minato, Japan, 18Climate and Environmental Physics, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland, 19Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, 20CNRM, Université de Toulouse, Mto-France, CNRS, Toulouse, France, 21Biospheric Sciences Lab, NASA, Greenbelt, MD, USA, 22International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL, USA, 23Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Abstract The Global Carbon Budget 2018 (GCB2018) estimated by the atmospheric CO₂ growth rate, fossil fuel emissions, and modeled (bottom-up) land and ocean fluxes cannot be fully closed, leading to a “budget imbalance,” highlighting uncertainties in GCB components. However, no systematic analysis has been performed on which regions or processes contribute to this term. To obtain deeper insight on the sources of uncertainty in global and regional carbon budgets, we analyzed differences in Net Biome Productivity (NBP) for all possible combinations of bottom-up and top-down data sets in GCB2018: (i) 16 dynamic global vegetation models (DGVMs), and (ii) 5 atmospheric inversions that match the atmospheric CO₂ growth rate. We find that the global mismatch between the two ensembles matches well the GCB2018 budget imbalance, with Brazil, Southeast Asia, and Oceania as the largest contributors. Differences between DGVMs dominate global mismatches, while at regional scale differences between inversions contribute the most to uncertainty. At both global and regional scales, disagreement on NBP interannual variability between the two approaches explains a large fraction of differences. We attribute this mismatch to distinct responses to El Niño–Southern Oscillation variability between DGVMs and inversions and to uncertainties in land use change emissions, especially in South America and Southeast Asia. We identify key needs to reduce uncertainty in carbon budgets: reducing uncertainty in atmospheric inversions (e.g., through more observations in the tropics) and in land use change fluxes, including more land use processes and evaluating land use transitions (e.g., using high-resolution remote-sensing), and, finally, improving tropical hydroecological processes and fire representation within DGVMs.

1. Introduction

The United Nations Framework Convention on Climate Change Paris Agreement from 2015 (UNFCCC, 2015) has the goal to limit the increase in global average temperature well below 2 °C above preindustrial
levels, and pursue efforts not to exceed the target of 1.5 °C above preindustrial levels. To achieve this overarching goal, the agreement calls for "a balance between anthropogenic emissions by sources and removals by sinks of greenhouse gases in the second half of this century.”

On the policy side, the problem is expected to be tackled at the national level through Nationally Determined Contributions, that is, mitigation policies defined by countries (UNFCCC, 2015). The collective progress toward the overall goal of the agreement is to be assessed regularly in the “global stocktake” process to support the update of Nationally Determined Contributions according to the best available science. However, the ultimate goal of stabilizing global mean surface temperature requires evaluating common progress at the global scale, as well as the evolution of natural sinks. It is therefore crucial that bottom-up estimates of CO₂ fluxes (e.g., inventories and models) provide accurate estimates that are consistent with the global atmospheric greenhouse gas mole fractions.

Global carbon (C) budgets for the anthropogenic perturbation have been estimated in the successive Intergovernmental Panel on Climate Change (IPCC) assessment reports (IPCC, 2001, 2007, 2013) and have been revised and updated in the Global Carbon Budget (GCB) by the Global Carbon Project almost every year since 2005 (Le Quéré et al., 2009; Le Quéré, Andrew, Friedlingstein, Sitch, Pongratz, et al., 2018; Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b).

Fossil fuel and cement production emissions ($E_{FF}$) can be estimated from historical energy statistics (Boden et al., 2017; UNFCCC, 2018) with a 1 standard deviation (1σ) uncertainty of 5–11% (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b; Quilcaille et al., 2018). The fluxes from changes in land use and management ($F_{LUC}$) cannot be directly measured and are estimated in Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b) by bookkeeping models (Hansis et al., 2015; Houghton & Nassikas, 2017), with a reported uncertainty of 0.7 Pg C yr⁻¹ (1σ) for decadal average $F_{LUC}$ over the industrial era. Global carbon uptake by the ocean and land can also not be directly measured, and are estimated in the GCB by process-based models for the land, and process- and data-driven models for the ocean (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b; Sitch et al., 2013).

In the Global Carbon Budget 2018 (GCB2018; Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b), the net annual balance between the anthropogenic sources and the sinks of CO₂ from process-based models does not exactly match the accurately observed atmospheric CO₂ growth rate (Gₐ₅₈; Dlugokencky & Tans, 2018). The residual flux resulting from this gap, the “budget imbalance,” can be interpreted as a measure of the limitations of the data sets used and of the imperfect process understanding by the modeling community (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b). Since the budget imbalance does not show a significant trend since the 1960s but high year-to-year variability, Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b) proposed that errors in the land and ocean sinks explain most of the imbalance term, but uncertainty in $F_{LUC}$ is also known to be high (Arneth et al., 2017; Piao et al., 2018). In GCB2018, estimates of ocean and land net CO₂ fluxes from process-based models were compared with those from four observation-based atmospheric inversions (Chevallier et al., 2005; Rödenbeck et al., 2003; Saeki & Patra, 2017; van der Laan-Luijkx et al., 2017) for the globe and over three latitudinal bands, where large differences between process-based models and inversions, but also between inversions were found. A recent study (Gaubert et al., 2019) further compared 10 different inversion systems, including the four mentioned above, showing an overall fair match with aircraft CO₂ measurements and varying consistency (<10% on 3-year mean varying with the inversion) with Gₐ₅₈. However, their results showed that differences in fossil fuel emission priors used affected the ocean-land partitioning, setting limits to the accuracy of inversions for quantifying regional surface-atmosphere CO₂ fluxes. Even though inversions were adjusted for differences in $E_{FF}$ priors in Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b), large disparities were still found for regional fluxes estimated by the four atmospheric inversions, especially the balance between the Northern Hemisphere and the tropics.

Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b) proposed that regional-level analyses may uncover sources of errors in the inferred fluxes. For example, comparisons of net CO₂ surface flux estimates from atmospheric inversions and dynamic global vegetation models (DGVMs) allowed assessing in detail the response of the tropical net CO₂ exchange to the 2015–2016 El Niño event, and evaluating the DGVMs’ ability to simulate the anomalies in carbon fluxes in 2015/16 (Bastos et al., 2018; Gloor et al., 2018; Rödenbeck et al., 2018; van Schaik et al., 2018). Likewise, the studies produced in the framework of the REgional Carbon Cycle Assessment and Processes (RECCAP; Canadell et al., 2012) provided a first evaluation of global and regional budgets, budget component fluxes and uncertainties. RECCAP produced multiple global syntheses...
of regional CO₂ fluxes between 1990 and 2009 and their driving processes (Andres et al., 2012; Ciais et al., 2014; Chen et al., 2013; Houghton et al., 2012; Khatiwala et al., 2013; Peters et al., 2012; Peylin et al., 2013; Sitch et al., 2013; Wanninkhof et al., 2013) as well as independent budgets for nine land, four ocean and coastal regions.

In this study, we take a closer look at the regional mismatch between the top-down and bottom-up data sets in GCB2018 (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b) to identify the sources of uncertainty in the terrestrial components, specifically to:

(i) identify the regions driving uncertainties in global budget estimates;
(ii) quantify the relative contribution of uncertainty in net CO₂ surface flux estimates from inversions (top-down) and from DGVMs (bottom-up);
(iii) identify those processes contributing the most to uncertainty in global and regional budgets.

We focus on a group of 18 land regions from Tian et al. (2018), which are consistent with the previous nine RECCAP large regions and with ecoclimatic specificities, but follow political borders. Evaluating the agreement between top-down and bottom-up estimates of CO₂ fluxes over such regions allows assessing the confidence and limitations of the state-of-the-art science to provide information on regional budgets, relevant for the global stocktaking process yet with traceable consistency with the GATM.

In this study, we rely on net CO₂ exchange from inversions, considered as uncertain estimates of surface CO₂ fluxes consistent with GATM and compare inversions’ estimates to those from the 16 DGVMs in GCB2018. Since Gaubert et al. (2019) recently analyzed the sources of uncertainty between inversions, here we focus on predictors that can explain the differences between DGVMs and inversions, or differences between individual DGVMs. By statistically modeling the mismatch between inversions and DGVMs, we are able to attribute uncertainty in the global budgets to a few key regions mainly in the tropics, and to identify those processes contributing the most to regional and global uncertainties.

2. Data
2.1. Atmospheric Inversions
Atmospheric inversions use an optimization process by which atmospheric CO₂ mole fraction measurements are used to constrain a priori estimates of the spatial distribution of surface CO₂ fluxes using an atmospheric transport model. In the optimization, information about errors of measurements and of priors, errors in the transport model, and the spatiotemporal structure of fluxes is also included (Peylin et al., 2013).

The four inversion systems used in this study (Table 1) are all based on in situ CO₂ observation measurements but differ in several aspects, such as the observational data assimilated and period covered, the transport model and prior fluxes used, the model grid, the spatiotemporal a priori correlation structure, as well as in the fossil fuel data sets used. A complete description of the different atmospheric inversions can be found in Table A3 in Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b). Here we use two versions of CarboScope (Rödenbeck et al., 2003), both covering a period longer than 30 years but with variable number of assimilated sites: s76 (1976–2017, eight stations) and s85 (1985–2017, 21 stations). Thus, we have used in total five inversion data sets (CAMS, two CarboScope versions, MIROC, and CarbonTracker Europe; Table 1).

Because they are based on CO₂ concentration measurements, atmospheric inversions estimate the net surface-atmosphere CO₂ fluxes including both the natural (fires, storms, pests, and diseases) and anthropogenic disturbance terms, the subsequent recovery, as well as the carbon taken up from the atmosphere over land but then passed on to the oceans through freshwaters, estuaries, and coastal areas (Hartmann et al., 2009; Mayorga et al., 2010; Regnier et al., 2013).

As in Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b) and Peylin et al. (2013), we adjusted the ocean and land fluxes for differences in fossil fuel emission (EF) priors over large latitudinal bands. As reference EF, we chose the data used by CAMS, the Emission Database for Global Atmospheric Research (Olivier et al., 2017) scaled to the Carbon Dioxide Information Analysis Center (Marland et al., 2008) estimates. However, this is only a first-order correction as the biases in EF not only affect the flux estimation of the region in question but also the neighboring regions (Saeki & Patra, 2017). The inversion surface fluxes were then remapped to a regular 1° × 1° latitude/longitude grid and then aggregated to the 18 land regions.
Table 1
Top-Down and Bottom-Up Data Sets Used in This Study, Including References and Period Covered

<table>
<thead>
<tr>
<th>Data set/model</th>
<th>Reference</th>
<th>Period</th>
<th>Fire</th>
<th>N-dep</th>
<th>SC</th>
<th>WH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inversions (top-down)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copernicus Atmosphere Monitoring Service (CAMS)</td>
<td>Chevallier et al. (2005)</td>
<td>1979–2017</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CarboScope s76</td>
<td>Rödenbeck et al. (2003)</td>
<td>1976–2017</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CarboScope s85</td>
<td></td>
<td>1985–2017</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CarbonTracker Europe (CTE)</td>
<td>van der Laan-Luijkx et al. (2017)</td>
<td>2001–2017</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td><strong>DGVMs (bottom-up)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLASS-CTEM</td>
<td>Melton and Arora (2016)</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>CLM5.0</td>
<td>Oleson et al. (2013)</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>DLEM</td>
<td>Tian et al. (2015)</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>ISAM</td>
<td>Meiyappan et al. (2015)</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JSBACH</td>
<td>Mauritsen et al. (2018)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>JULES</td>
<td>Clark et al. (2011)</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LPJ</td>
<td>Poulter et al. (2011)</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>LPJ-GUESS</td>
<td>Smith et al. (2014)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>LPX-Bern</td>
<td>Lienert and Joos (2018)</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>OCN</td>
<td>Zaehle et al. (2010)</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>ORCHIDEE</td>
<td>Krinner et al. (2005)</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>ORCHIDEE-CNP</td>
<td>Goll et al. (2017)</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>SDGVM</td>
<td>Walker et al. (2017)</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>SURFEX</td>
<td>Joetzjer et al. (2015)</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>VISIT</td>
<td>Kato et al. (2013)</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Note. For the DGVMs, we indicate whether they simulate fires, nitrogen deposition (N-dep), shifting cultivation (SC), and wood harvest (WH), and in some models further include irrigation and nitrogen fertilization. A complete description of the processes in each model can be found in Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b).

Inversions determine total CO₂ fluxes between the surface and the atmosphere whereas the GCB approach with land and ocean biogeochemical models determines the anthropogenic budget of CO₂. The difference between total and anthropogenic CO₂ fluxes is that there is a background preindustrial uptake of CO₂ on land that sustains carbon export from soils to rivers and to the ocean, where compensatory outgassing of CO₂ occurs. This requires an adjustment of inversion CO₂ fluxes to transform them into anthropogenic CO₂ fluxes, as performed in Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al. (2018b) over latitudinal bands.

Based on the data-driven estimates of fluvial exports of organic (Mayorga et al., 2010) and inorganic carbon (Hartmann et al., 2009) to the coast, Zscheischler et al. (2017) produced a spatially explicit data set of climatological land-ocean carbon transfers at 1° × 1° latitude/longitude resolution and includes the fluxes from dissolved inorganic carbon from atmospheric origin and from weathering and dissolved and particulate organic carbon (DOC and POC). In this study, the DOC and POC exports of this data set were rescaled per basin to match the estimates of Resplandy et al. (2018). After aggregating these rescaled estimates to the 18 land regions, we subtracted the fluvial carbon exports from the regionally aggregated inversion net surface CO₂ fluxes, to calculate regional net biospheric production (NBP) that can be compared with the DGVM estimates. These data are available in (Bastos, 2019).

2.2. DGVMs

DGVMs simulate water, energy and biogeochemical exchanges between the surface and the atmosphere through ecosystems activity, including growth, turnover and decomposition of vegetation, and soil carbon...
DGVMs simulate the response of ecosystems to changes in environmental conditions (increasing CO₂, changing climate, and altered nitrogen deposition) and land use activity.

Here we use a set of 16 DGVMs from the TRENDY-v7 intercomparison (Sitch et al., 2015) that contributed to GCB2018. In the simulations for GCB2018 (Simulation S3), all 16 models are forced with: (i) observed climate from the Climate Research Unit (Harris et al., 2014) and the Japanese 55-year Reanalysis (Kobayashi et al., 2015) data sets, the CRU-JRA reanalysis, following the methodology in Viovy (2016); (ii) global CO₂ concentration from NOAA/ESRL (Dlugokencky & Tans, 2018); (iii) land use change transitions and land management fields from the Land Use Harmonization (LUH2v2.1h) from Hurtt et al., 2011 (2011, 2017), based on HYDE 3.1 (Klein Goldewijk et al., 2011); and (iv) gridded data of nitrogen (N) deposition when N cycling is simulated by models.

Some models simulate fire emissions \( (E_{\text{Fire}}) \) and nutrient cycling (nitrogen in ten models, and nitrogen and phosphorus in one model), including the effects of N deposition (Table 1). While all DGVMs simulate the fluxes resulting from forest clearing, pasture and crop conversion, abandonment and regrowth and crop harvest, they implement them using different assumptions about the areas being converted (e.g., gross versus net conversion). The way DGVMs simulate management practices and the fate of released carbon also varies between models. Ten models simulate wood harvest and six include shifting cultivation, roughly corresponding to gross transitions (Table 1). Only few DGVMs simulate crop fertilization or irrigation. To calculate \( \Delta F_{\text{LUC}} \), the S3 simulation is compared to the S2 simulation, which is forced with CO₂ and climate changes only, and keeping a fixed land cover map in 1700 (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b).

For \( \Delta F_{\text{LUC}} \) estimated in this way, fluxes resulting from environmental change over managed lands as compared to intact vegetation are included (Pongratz et al., 2014). Due to the effect of CO₂ fertilization, intact vegetation would at present provide a slightly stronger sink, which implies that emissions from, for example, deforestation are now slightly higher than if compared to the preindustrial state. This term is referred to as loss of additional sink capacity (LASC) as proposed by Pongratz et al. (2014). In TRENDY-v7, an additional simulation was performed where preindustrial CO₂ and climate were kept fixed, and only land use changes and management were allowed to change between 1700 and 2017 (S4). \( \Delta F_{\text{LUC}} \) estimated in this way does not include the LASC term and is therefore more comparable in terms of processes to the ones estimated by the bookkeeping methods, except for the values of C densities used.

Outputs of monthly NBP from all three simulations (S2, S3, and S4) were first resampled to a common \( 1^\circ \times 1^\circ \) latitude/longitude grid, and then aggregated for each region. The regional fluxes from S2 and S3 simulations are available from (Bastos, 2019).

### 2.3. Bookkeeping Models

Bookkeeping models track changes in above- and below-ground biomass carbon densities resulting from land use change processes such as deforestation/afforestation, cropland or pasture expansion, wood harvest, shifting cultivation, and forest regrowth after land abandonment. In GCB2018, two bookkeeping models that estimate \( \Delta F_{\text{LUC}} \) at the global scale have been considered: the bookkeeping model of Houghton and Nassikas (2017), referred as HN2017 henceforth, and the “Bookkeeping of Land Use Emissions” model (BLUE) described in Hansis et al. (2015). To estimate \( \Delta F_{\text{LUC}} \), bookkeeping models first calculate the changes in biomass, soil and the atmosphere carbon pools resulting from a given transition following specific response curves. The resulting fluxes for a given year can be calculated as the difference in carbon stocks between two consecutive years. The carbon density values for above- and below-ground biomass used by both models are based on recent history (ca. 1980s) measurements. This means that the C densities used have an implicit transient effect of increasing CO₂ and climate, leading to somewhat higher CO₂ emissions than S4 by DGVMs.

Even though based on similar principles, the two models differ in their mathematical formulation, spatial implementation, assumptions made about LUC transitions, processes included, forcing data used and, consequently, in the sources of uncertainty.

The HN2017 model covers the period of 1700–2015 and is based on regional statistics from the Food and Agricultural Organization (FAO) on changes in the areas of croplands and pastures since 1961 and changes in the areas of forests and other land since 1990 (FAO, 2015; FAOSTAT, 2015), while BLUE used the same data set as the DGVMs, that is, LUH2v2.1h. In the HN2017 model, calculations are performed at the country
scale, and in the recent version, the model now incorporates country specific C densities. In this version, the model does not include shifting cultivation. From 1997 onward, peat fires (Giglio et al., 2016) and emissions from peat drainage (Hooijer et al., 2010) are added in Southeast Asia. The fluxes resulting from fire suppression are considered, but only in the United States. The HN2017 model allocates pasture preferentially to grasslands. This may result in lower CO2 emissions by reducing deforestation rates (Reick et al., 2010).

The BLUE model uses a spatially explicit modeling scheme to calculate $F_{\text{LUC}}$ on a pixel basis between 850 and 2017. It relies on biome-specific C densities and exponential response curves to track the changes in soil and biomass carbon pools following land use conversion or due to management. New cropland and pasture are taken proportionally from natural vegetation types and rangelands clear the natural vegetation for forest areas and degrade other natural land (Hansis et al., 2015; Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b). BLUE further includes shifting cultivation. In the GCB2018 version, BLUE used the same data set as several of the DGVMs, that is, the LUH2v2.1h, to calculate $F_{\text{LUC}}$ at 0.25° × 0.25° latitude/longitude resolution. Because BLUE uses a spatially explicit framework and the same LUC forcing as DGVMs, we use it as a primary basis for comparison with process-based model estimates of $F_{\text{LUC}}$. We then compare BLUE with HN2017 to evaluate the potential contribution of the forcing used to estimate $F_{\text{LUC}}$ by DGVMs and BLUE to the mismatch between DGVMs and inversions.

### 2.4. Additional Data

In order to evaluate possible sources of uncertainty, we considered additional data sets to explain the differences between DGVM and inversion fluxes (see section 2.5): the Oceanic Niño Index (ONI) from NOAA’s Climate Prediction Center (https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v4.shtml) from December to March; annual average total water storage (TWS) over each of the 18 regions from the Gravity Recovery and Climate Experiment (GRACE) reconstruction (Humphrey et al., 2018); satellite-based land cover maps from the European Space Agency’s Climate Change Initiative Land-Cover (LC-CCI) (ESA, 2017) from 1992–2015 aggregated from 250 m × 250 m to 1° × 1° latitude/longitude resolution; carbon emissions data from the Global Fire Emissions Database version 4.1s (GFED4.1s, https://doi.org/10.3334/ORNLDAAC/1293) available from 1997–2017, described in Randerson et al. (2017); and burned area from ESA-CCI Fire-CCI v5.1 (Fire-CCI; Chuvieco et al., 2018), covering the period 2001–2017.

### 2.5. Methods

The goal here is to quantify the mismatch between NBP estimated by inversions and DGVMs. As discussed in section 2.1, inversions differ in several aspects between each other. Likewise, DGVMs show variable skill in simulating fundamental processes of the terrestrial C cycle (see Figure B2 in Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b) and do not all simulate the same processes (Table 1). Some of the differences across data sets may arise from complex interactions between processes within models, and not be traceable to a particular process.

We define $D_{kj}$ as the difference in year $j$ ($j = 1, \ldots, J$) between NBP estimates of the $k$th inversion ($k = 1, \ldots, n_j$, with $n_j \leq 5$, depending on the year $j$) available over that year and the $i$th DGVM ($i = 1, \ldots, 16$).

$$D_{kj} = \text{DGVM}_i - \text{INV}_k$$

For a given year, the NBP values estimated by the inversions ($\text{INV}_k$) can be then considered as a sample from a distribution of surface fluxes compatible with $G_{\text{ATM}}$, with the size of this sample changing over the study period (1979–2017) as more inversions become available. For each year we can generate a set of $n_j \times 16$ values of $D$ for the globe ($D_{\text{Globe}}$), as well as for each of the 18 regions ($D_{\text{reg}}$).

For each (inversion, DGVM) pair, the $D_{\text{Globe}}$ can be decomposed into the 18 region components:

$$D_{\text{Globe}ij} = \sum_{r=1}^{18} D_{\text{reg}rj}$$

We can, therefore, quantify the contribution of each region to the variance ($V$) of $D_{\text{Globe}}$ as

$$V(D_{\text{Globe}ij}) = \sum_{r=1}^{18} V(D_{\text{reg}rj}) + 2 \sum_{r=1}^{18} \sum_{p=1}^{18} \text{Cov}(D_{\text{reg}rj}, D_{\text{reg}pj})$$

BASTOS ET AL. 6 of 21
The contribution of each region \( r \) to the variance of \( D_{\text{Globe}} \) is therefore the quotient between the sum of the regional variance and covariance terms in \( r \) and the variance of \( D_{\text{Globe}} \) for each (inversion, DGVM) pair.

To identify possible drivers of \( D \), we test several statistical models using likely predictors for the mismatch between inversions and DGVMs. Linear mixed effects (LME) models allow modeling \( D \) as a function of fixed effects (e.g., El Niño/La Niña cycles), but also of random effects describing several sources of variability (DGVMs, years, inversions). We start by fitting to \( D_{ijk} \) the following simple LME model without any predictor region by region (the lme function from the R package lme4; Bates et al., 2014) described in equation (4) (LME\(_{\text{FE}}\)):

\[
D_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ikj}
\]  

where \( \mu \) is the average mismatch (i.e., \( D \) averaged over all models, years, and inversions) over the considered region. The random effects \( \alpha_i, \beta_j \) and \( \gamma_{ij} \) (all with zero mean and variance \( V_{\alpha_i}, V_{\beta_j}, \) and \( V_{\gamma_{ij}} \) respectively) describe the variability of \( D \) across the DGVMs with index \( i \), the between-year variability of \( D \) with index \( j \), and the interaction between models and years on \( D \), respectively. The interaction term accounts for the possibility that some models may have larger \( D \) in some years, and smaller \( D \) in other years. The \( \epsilon_{ikj} \) term describes the variability across the five inversions and their interactions with year and DGVMs with index \( k \) and variance \( V_{\epsilon} \). For each region, a set of LME models with the possible combinations of one, two or all three random effects are fit to \( D \), the model that best explains \( D \) for each region is retained, by choosing the one with lowest Akaike Information Criterion and with significant fit (\( p \) value < 0.05). We then expand the LME in equation (4) in order to explain part of the variance of \( D \) using different sets of predictors. We test four hypotheses (equation (5)) that are supported by literature and for which data are available. We define a set of LME models (equation (5)) (LME\(_{\text{FE}}\)) with one or more predictors \((X_a, X_b, X_c, X_d)\) and associated coefficients \((c_a, c_b, c_c, c_d)\) corresponding to the following hypotheses:

a. \( D \) can be explained by a linear time trend, for example, if models under- or overestimate the sensitivity to a process with a strong trend component such as CO\(_2\) increase (Graven et al., 2013; Thomas et al., 2016);

b. Since \( D \) corresponds to the difference between top-down and bottom-up estimates of NBP, and NBP includes \( F_{\text{LUC}} \), which is highly uncertain (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b; Piao et al., 2018), \( D \) might be explained by errors in \( F_{\text{LUC}} \);

c. Differences can be due to too strong response of DGVMs to ENSO (Bastos et al., 2018);

d. In fire-dominated regions, \( D \) may be explained by the fact that some DGVMs do not simulate fires (Table 1), or because that do still show limited ability to correctly represent spatial and interannual variability in burned area and fire emissions (Li et al., 2014; Poulter et al., 2015; Yue et al., 2014, 2015).

The full model includes the four predictors and is expressed as

\[
D_{ijk} = c_a \times X_a + c_b \times X_b + c_c \times X_c + c_d \times X_d + \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ikj}
\]  

Because the predictors have different units, and variability across regions may also vary significantly for \( F_{\text{LUC}} \) and \( E_{\text{fire}} \), we centered and standardized these variables (i.e., the mean was subtracted and the result divided by the standard deviation). As done for the random effects, we define multiple models with one, two, three, or four predictors as fixed effects, and choose the best model fit (lowest Akaike information criterion and significant fit). We fit the LME\(_{\text{FE}}\) separately for 1997–2017 when including fire emissions as a predictor.

When any of these variables is found to provide the best fit, we explore how the predictors may contribute to explain the differences. In the case of \( F_{\text{LUC}} \) we compare the estimates from DGVMs with those of BLUE and HN2017 and evaluate the role of the forcing LUC to explain \( D \) in that region. As ENSO effects on NBP are mainly related to changes in temperature and water availability in the tropics (warm/dry during El Niño and cool/wet during La Niña) we evaluate the sensitivity of fluxes to soil water availability and temperature by performing a multiple linear regression of spatially aggregated fluxes with regionally averaged annual temperature from CRU-JRA and water availability. Water availability was estimated by annual TWS from GRACE for inversions and simulated soil moisture for DGVMs. Where \( E_{\text{fire}} \) is found to contribute to \( D \), we compare simulated emissions from fire as well as burned area with the reference data sets.

We further group DGVMs by “flavors,” corresponding to whether they represent (or not) the processes highlighted in Table 1: fire, N deposition, shifting cultivation and wood harvest. We evaluated whether
the different groups of DGVMs have significantly lower biases (μ) than the set of 16 DGVMs at global and regional scale (Supporting Information Table S1), and additionally add the process representation as a predictor in the LMEFE fit (Figure S7).

### 3. Results

#### 3.1. Global Differences and the “Budget Imbalance”

The values of $G_{ATM}$ reconstructed from the inversions fluxes on average match those from GCB2018 (Figure S1), with differences of $-0.04 \pm 0.41$, $0.11 \pm 0.36$, $0.14 \pm 0.36$, $0.29 \pm 0.42$, and $0.00 \pm 0.48$ Pg C yr$^{-1}$, for CAMS, CarboScope s76, CarboScope s85, MIROC, and CarbonTracker Europe, respectively, over the period covered by each inversion. For comparison, the budget imbalance term in GCB2018 for 2008–2017 was 0.5 Pg C yr$^{-1}$.

The ensemble of $D_{Globe}$ calculated following equation (1) for all (inversion, DGVM) pairs is shown in Figure 1. Over the 1979–2017 period, $D_{Globe}$ averaged across all DGVMs for each inversion (bold lines) show strong multiannual fluctuations. From 1983–1989 (except 1982) and over the late 1990s, $D_{Globe}$ was strongly positive, which indicates stronger NBP from DGVMs compared to inversions; following the year 2000, $D_{Globe}$ showed a decreasing trend. These multiannual tendencies are punctuated by years of strong peaks, for example, 1982, 1991/1992, 2009/2010, and 2015 (negative $D_{Globe}$) and 1984, 1988/1989, 2000, and 2011 (positive). The trough in 1991/1992 suggests that DGVMs miss the large abnormal sink deduced from $G_{ATM}$, which is a known feature from all previous budgets (Le Quéré et al., 2009; Le Quéré, Andrew, Friedlingstein, Sitch, Pongratz, et al., 2018; Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b) and likely linked to the underestimation by DGVMs of the land sink enhancement in response to the eruption of Mt. Pinatubo (Lucht et al., 2002; Mercado et al., 2009). Other years are associated with positive and negative phases of El Niño (Bastos et al., 2018; Bowman et al., 2017), suggesting a possible contribution to $D_{Globe}$ of mismatches in the response of DGVMs and inversions to ENSO.

The “budget imbalance” time series from GCB2018 (black line, Figure 1) shows very similar variability to the group of ensemble means, supporting the hypothesis that errors in the land sink representation explain a large fraction of the GCB2018 imbalance term. The values of $D$ show an offset compared to this term though, which is in part explained by the adjustment of lateral fluxes to Resplandy et al. (2018) values. Even though over the 39-year period no clear trend can be distinguished in the whole ensemble, seven out of the 80 pairs inversion-DGVM show significant trends, evaluated by the Mann-Kendall test.
Figure 2. The 18 study regions: United States (USA), Canada (Canada), Europe (EU), Northern Africa (NAF), central Asia (CAS), Russia (RUS), Korea and Japan (KAJ), China (CHN), Southeast Asia (SEAS), Oceania (OCE), south Asia (SAS), Middle East (MIDE), southern Africa (SAF), equatorial Africa (EQAF), southern South America (SSA), Brazil (BRA), northern South America (NSA), and central America (CAM) are shown in the top panel. In the lower panel, the contribution of each region to the variance of $D$ is shown for each inversion, calculated for the multi-DGVM mean for each of the five inversions. The bar order corresponds to CAMS, CarboScope s76 and s85, MIROC, and CarbonTracker Europe.

The variability across (inversion, DGVM) pairs is much higher than the variability from year-to-year of the ensemble means. The 1σ of $D_{\text{Globe}}$ is 0.9 Pg C yr$^{-1}$ (Figure 1). This value is comparable to the uncertainty reported for the GCB2018 imbalance term ($\pm 1$ Pg C yr$^{-1}$).

3.2. Regional Contributions to the Global Differences

We quantify the contribution of each of the 18 regions to the variance of the spatially averaged differences over the globe ($D_{\text{Globe}}$, Figure 2) for the $n_{\text{inversion}} \times 16$ DGVM pairs. All (inversion, DGVM) pairs single out Brazil (BRA) as the region contributing the most to $D$ at the global scale (16–27%, averaged across DGVMs), followed by Oceania (OCE, 7–17%). Inversions also agree on a moderate contributions of northern South America (NSA, 4–7%), southern South America (SSA, 3–6%) and China (CHN, 3–11%). Four out of five inversions estimate a strong contribution of Southeast Asia (SEAS) to $D_{\text{Globe}}$ (−1% for CAMS, 6–11% for the other four). CAMS assigns a strong negative contribution of EU (−6%) versus a strong positive contribution of RUS (21%) to $D_{\text{Globe}}$, while the other inversions indicate comparable contributions of EU and RUS (0–7%). This suggests that inversions have dipoles in their allocation of the Eurasian CO$_2$ sink with inversions having more uptake in Russia and less in Europe. This is likely due to the relatively sparse observation network in RUS that does not allow to separate easily the flux of nearby regions.
Figure 3. Coefficients of the linear mixed effects model fit only with random effects only (equation (4)) to the time series of $D$ over each region. The intercept $\mu$ corresponds to the error $D$ averaged over all models and all years over each region (top panel). A positive (negative) value of $\mu$ indicates an overestimate (underestimate) of NBP by the DGVMs compared to inversions over the study period (1979–2017). Regions where $\mu$ values are not significantly different from zero are masked out in white. The coefficients for the random effects for each region and for the globe are shown in the bottom panel. A high value indicates a strong contribution of random effects from differences between DGVMs ($\alpha$), between-year variability ($\beta$), model year variability ($\gamma$), and variability between inversions ($\epsilon$).

If the contribution of a given region to the variance of $D_{\text{Globe}}$ changed over time, differences between the three “long” and the two short inversions (Table 1) can be explained partly by their different validity periods. However, large differences are found even between inversions with comparable validity periods, so they are most likely due to differences in the inversion systems. Moreover, we find a large spread of results between (inversion, DGVMs pairs) for several regions, indicating a relevant contribution of the uncertainty in spatial patterns of NBP from DGVMs to $D_{\text{Globe}}$.

As for the globe, the spread of regional $D$ ($D_{\text{reg}}$) for the $5 \times 16$ members ensemble in certain regions can be very high (Figure S2). There is not necessarily a correspondence between the regions with larger $D_{\text{reg}}$ spread and those regions identified above for their contribution to $D_{\text{Globe}}$. The regions showing larger average spread over the 39-year period are, in decreasing order, RUS ($1.3 \, \text{Pg C yr}^{-1}$), USA, and EU ($1.0 \, \text{Pg C yr}^{-1}$), and BRA and CAN ($0.9 \, \text{Pg C yr}^{-1}$). In some years the spread in $D_{\text{reg}}$ can reach 2.6 and 1.8 Pg C yr$^{-1}$ for RUS and BRA, respectively, and 1.6 Pg C yr$^{-1}$ in USA, EU, and CHN, which is close to the magnitude of the global sink (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b). In most regions, it is hard to distinguish between individual inversions, showing that the variability across DGVMs also contributes considerably to the range of $D_{\text{reg}}$. Moreover, some regions show consistent interannual or long-term variability in $D_{\text{reg}}$, and its spread. For example, $D_{\text{BRA}}$ is predominantly positive in the 1980s and 1990s, and becomes mostly negative following 2010. In OCE, a very large spread across data sets is observed for some years (e.g., 1999 and 2011) but not during most of the study period.

3.3. Decomposition of $D$

The intercept ($\mu$) of the LME$_{\text{reg}}$ fit corresponds to the average stationary value of $D_{\text{reg}}$ for the period 1979–2017 (Figure 3 top panel). For the globe, $\mu$ is positive, indicating that DGVMs estimate a stronger global CO$_2$ sink than inversions. Regionally, negative $\mu$ values are found over most of the Northern Hemisphere's
extratropics (except CAN and RUS) and positive values in the tropics and southern extratropics. Negative (positive) values indicate that DGVMs report a weaker (stronger) sink in the northern (tropical and southern) regions. The largest $\mu$ terms are found in South America ($>0.2$ Pg C yr$^{-1}$), followed by EUR, CHN and USA ($<−0.15$ Pg C yr$^{-1}$). The tropics and southern extratropics have average $D$ of 1.0 Pg C yr$^{-1}$ and the northern regions of $−0.5$ Pg C yr$^{-1}$.

The variance of $D_{\text{globe}}$ and $D_{\text{reg}}$ attributed to each random term of LME$_{\text{FE}}$ is shown in Figure 3 (bottom panel). Interannual variability ($\beta$) is the term contributing the most to $D_{\text{globe}}$, followed by the between-DGVM variability ($\alpha$) and model year combined ($\gamma$). Variability between inversions ($\epsilon$) contributes the least to $D_{\text{globe}}$, but at the regional scale it is the most important factor in half of the regions, and is especially high in USA, CAN, EU, RUS, and CHN.

Interannual variability is the second most relevant term regionally, especially in BRA, OCE, NSA, and SSA where the interannual term dominates over the difference between inversions, although for the latter two the variance is rather small. Between-DGVM differences do not contribute much to the regional differences, except in RUS.

If random effects could explain all of the variance of $D$, the residuals of the model fit should be randomly distributed around 0 and not change over time. However residuals of the LME$_{\text{FE}}$ fit show clear trends or multiannual patterns (Figure S3). This misfit of LME$_{\text{FE}}$ suggests that deterministic processes that explain interannual variability in $D$ are missing or have a common bias in either inversions or DGVMs (probably being attributed to $\beta$). The most evident cases are the decadal variations in the residuals for the globe, the sharp decreasing trend in the misfit in BRA since circa 2005, and long-term trends in SSA, EU, and RUS.

### 3.4. Effects of Predictors

For each region and for the globe, we test a set of key predictors (fixed effects model, LME$_{\text{FE}}$, equation (5)) that might help explain $D$. The predictors tested are the (i) a simple linear trend (“year” as predictor); (ii) the Oceanic Nino Index (ONI); (iii) $F_{\text{LUC}}$ from the BLUE bookkeeping model between 1979 and 2017, and (iv) the same predictors plus $F_{\text{fire}}$ for 1997–2017 (the period covered by GFED4.1s). For each region, we tested additional predictors such as TWS, burned area and changes in forest and cropland areas. These additional predictors are correlated with the previous ones (e.g., soil water with ENSO, or $F_{\text{LUC}}$ from cropland/forest area changes) and were not found to provide additional information. The regression coefficients of the LME$_{\text{FE}}$ fit to $D_{\text{reg}}$ and $D_{\text{globe}}$ are shown in Figure 4 for the two periods.

The coefficients indicate the effect of a unit change in the predictor (e.g., $F_{\text{LUC}}$) to a corresponding unit change in $D$. We find that $D_{\text{globe}}$ is dominated by ONI and $F_{\text{LUC}}$ with a small trend component, during 1979–2017, and only by the trend during 1997–2017, as other effects cancel out regionally. All coefficients for the globe have negative sign, indicating that positive anomalies of each predictor contribute to explain an underestimate of the CO$_2$ sink by DGVMs relative to inversions. This means, for example, that a positive anomaly in land use change emissions (above average) would coincide with lower NBP in DGVMs compared to inversions.

For the 39-year period, a strong negative relationship is found between $D_{\text{globe}}$ and ONI, implying that DGVMs tend to estimate a weaker sink or stronger source than inversions in response to El Niño, and the opposite during La Niña events. The global relationship of $D_{\text{globe}}$ and ONI is mainly driven by NSA, BRA, SAF, and SEAS, which also show strong negative effects of ONI, and partly offset by an opposing effect in CAS, MIDE, and SAS.

The LME$_{\text{FE}}$ indicates $D_{\text{globe}}$ is explained by emissions from land use change in those regions where land use changes are more intense. The minus sign indicates that the higher $F_{\text{LUC}}$ (positive for emissions), the more DGVMs underestimate the net sink globally compared to inversions. Such a negative contribution from LUC emissions is found in BRA, EU, EQAF, and SEAS. We tested LME$_{\text{FE}}$ using $F_{\text{LUC}}$ simulated by DGVMs (Experiments S2 and S3), which yielded very similar results. This suggests that too strong emissions from LUC in DGVMs might result in a weaker sink compared to inversions (which implicitly include $F_{\text{fire}}$).

We find a weak negative trend contribution to $D_{\text{globe}}$, revealing a significant, though small, divergence of inversions and DGVMs over time by $−26$ Tg C decade$^{-1}$. In other words, DGVMs underestimate the rate of increase of the global terrestrial sink compared to inversions. Significant trend components of $D$ are found in 8 regions (negative in NSA, BRA, SSA, RUS, and CAS and positive in EU, KAJ, and SEAS), but their values are rather small (from 1–10 Tg C yr$^{-1}$ in absolute magnitude).
Global Biogeochemical Cycles

Figure 4. Coefficients of the LME model fit with fixed effects (equation (5)). $F_{\text{LUC}}$, ONI, and $E_{\text{Fire}}$ correspond to emissions from land use change, the Oceanic Niño Index, and emissions from fire, respectively. The regions abbreviations are described in Figure 2. The symbols indicate statistically significant results: crosses for $p$ value $< 0.05$ and asterisks for $p$ value $< 0.01$. The top panel shows values for the period 1979–2017, and the bottom panel for the period 1997–2017.

Significant effects of $E_{\text{Fire}}$ are found in NSA, CAS, and SEAS (Figure 4), the first two negative (DGVMs estimating lower sink/stronger source associated with fires than inversions) and positive in SEAS (DGVMs estimating stronger sink/weaker source associated with fires than inversions). Including $E_{\text{Fire}}$ reduces the variance of the residuals of the LME fit to $D_{\text{Globe}}$ during 1997–2017 by 80%, compared to a reduction of 58% for the fit without fire in that period (Figure S3). For one DGVM simulating fire using a state-of-the-art fire module, Yue et al. (2015) have shown a tendency of the model to overestimate fire emissions and burned area in regions roughly corresponding to NSA, BRA, and CAS, and to underestimate $E_{\text{Fire}}$ and burned area in SEAS. This is discussed in section 3.5.3.

3.5. Sources of Uncertainty

3.5.1. ENSO, Temperature, and Soil Moisture Variability

The sensitivity of soil moisture from DGVMs to ONI in the regions with significant effects (NSA, BRA, SAF, and SEAS, negative and CAS, MIDE, and SAS, positive) is generally consistent with that of TWS from the GRACE reconstruction (Humphrey et al., 2018), although slightly underestimated for NSA and BRA (Figure S4). In SAF and SEAS the sensitivity of SM from DGVMs to ONI is close to that of TWS, but in the former DGVMs show a large range. The differences between DGVMs and inversions in these regions can be further related with the sensitivity of modeled NBP to the hot/cooler (e.g., by too strong respiration response to warming) or dry/wet anomalies linked with El Niño/La Niña events (e.g., by too strong water stress controls on productivity). Therefore, we evaluate the sensitivity of NBP to temperature and simulated soil moisture.

In NSA and BRA, inversions tend to estimate a positive sensitivity of NBP to temperature (higher uptake for warmer conditions, Figure 5), indicating a positive effect of temperature on productivity, which is mainly limited by radiation rather than water (Nemani et al., 2003). On the contrary, DGVMs tend to estimate negative sensitivity of NBP to temperature (lower NBP with warming). On the other hand, DGVMs estimate higher sensitivity of NBP to water availability than inversions (Figure 5). Higher sensitivity to water availability is also observed in SEAS, but inversions and DGVMs show similar sensitivity to temperature in this region.
3.5.2. Land Use Change

The results of LME_F (Figure 4) point to a negative effect of $F_{LUC}$ in $D_{Globe}$, indicating that DGVMs might overestimate emissions from LUC or underestimate LUC-related recovery sinks. However, the $F_{LUC}$ data used relies on the same LUC forcing as DGVMs (BLUE) as a predictor. Therefore, the agreement between BLUE and DGVMs’ estimates of $F_{LUC}$ is partly explained by their use of a common forcing. In this sense, BLUE cannot be seen as a fully independent estimate of $F_{LUC}$. We find a remarkable disagreement between regional estimates of $F_{LUC}$ from DGVMs and BLUE with those from HN2017, especially in NSA, BRA, and SEAS. This points to a strong influence of the underlying LUC data used to force models, more than to structural differences between bookkeeping models or differences between bookkeeping and process-based models.

We compare $F_{LUC}$ from DGVMs with the two bookkeeping models (BLUE and HN2017, $F_{LUC}$−BLUE and $F_{LUC}$−HN2017, Figure 6). Simulation S3 (climate, CO2 and LUC), although more suitable to compare with NBP from inversions, includes in their $F_{LUC}$ the LASC term ($F_{LUC}$−LASC). The difference between S3 and S2 (climate and CO2 only) is thus not directly comparable to bookkeeping model estimates. Therefore, we further include results from Simulation S4 (LUC only under preindustrial climate and CO2) that does not include the LASC to calculate $F_{LUC}$−noLASC.

Globally, the multimodel ensemble mean (MMEM) of $F_{LUC}$−LASC is above the estimates by the two bookkeeping models during most of the 39-year period (Figure 6), but $F_{LUC}$−noLASC values are mostly between the two bookkeeping models, when they would be expected to be slightly lower. The two bookkeeping models differ on average by 0.5 Pg C yr$^{-1}$. with BLUE systematically above HN2017, possibly because it includes more processes, particularly shifting cultivation (Arneth et al., 2017), or because HN2017 use grasslands preferentially over forest for pasture expansion (Hansis et al., 2015).

In BRA, the LASC has a relatively small effect (Figure 6). The MMEM for $F_{LUC}$−noLASC is generally close to the estimates of BLUE and shows the same dynamics: a decreasing trend from the 1980s until circa 2009, and then a sudden increase in $F_{LUC}$ of about 0.5 Pg C yr$^{-1}$, followed by a period with high emissions and strong interannual variability. This 2009–2017 period coincides with a strong decrease in $D_{reg}$ in BRA (Figure S2),
Figure 6. Comparison of $F_{LUC}$ from different data sets for the (top) globe, (middle) BRA, and (bottom) SEAS (where $F_{LUC}$ is a significant predictor of $D$).

that is, DGVMs simulating a much weaker sink than inversions. HN2017, on the contrary, indicates lower emissions in the 1980s, a slow increasing trend until circa 2005 and a decreasing trend afterward.

In SEAS, peat burning and drainage are a major component of LUC-related fluxes (Moore et al., 2013). Since DGVMs do not simulate peat, the comparison of $F_{LUC}$ with bookkeeping models should be adjusted for peat burning and drainage emissions (also note that part of these fluxes in models could be included in $E_{fire}$, which are discussed in the next section). When these emissions are removed from $F_{LUC}$ estimated by bookkeeping models (Figure 6): $F_{LUC - LASC}$ is close to the higher value estimated by BLUE, and $F_{LUC - noLASC}$
is between the two bookkeeping models, but only following the year 2000. Before that, $F_{\text{LUC+noLASC}}$ is well below both bookkeeping models, which is consistent with the lower equilibrium C densities in this simulation, as compared with bookkeeping values. The year 2000 also marks a period with higher $F_{\text{LUC}}$ in BLUE, indicating a possible effect of the LUC forcing.

3.5.3. Fire

Fire emissions from GFED4.1s are found to be a significant predictor for $D_{\text{reg}}$ in NSA, CAS and SEAS. In the first two regions, $E_{\text{fire}}$ effects on $D_{\text{reg}}$ are negative, indicating that higher fire emissions in GFED4.1s are associated with too weak CO$_2$ uptake by DGVMs compared with inversions. In SEAS, the effect is the opposite, which suggests an underestimation by DGVMs of $E_{\text{fire}}$.

In Figure 7 (left panel) we compare $E_{\text{fire}}$ from those DGVMs that simulate fire (7 out of 16) with GFED4.1s fire emissions (Giglio et al., 2016). In NSA and CAS the ensemble of DGVMs estimates fire emissions about twice as high on average as those reported by GFED4.1s, but also much more variable (Figure S6). On the contrary, in SEAS, DGVMs estimate fire emissions that are, on average lower than the total emissions reported by GFED4.1s (about 60%). These results are consistent with overestimation of fire emissions in NSA and CAS leading to an underestimation of the CO$_2$ sink compared with inversions, and the opposite in SEAS.

Biases in emissions from fire can be due to biases in simulated burned area or in the fuel properties and fire characteristics (e.g., human influence on ignitions). The comparison of simulated burned area with Fire-CCI burned area product (Chuvieco et al., 2016) indicates an overestimation by DGVMs of burned area in CAS and underestimation in SEAS, consistent with $E_{\text{fire}}$ biases. However, this is not the case in NSA, where DGVMs estimate too high fire emissions in spite of underestimating burned area.

4. Discussion

Tropical regions explained most of the variance in $D_{\text{globe}}$, especially Brazil and northern South America, Southeast Asia and Oceania. These are, at the same time, the regions less constrained by observations. However, the regions with higher biases (Figure 3) are generally those with better observational coverage (USA, CAN, EU). These regions are strong $E_{\text{FF}}$ emitters, suggesting that spatial and temporal differences between prescribed $E_{\text{FF}}$ might still affect surface fluxes estimated by inversions, which were only adjusted for latitudinal bands.

At regional scale, negative values of $D$ (DGVMs underestimating the long-term sink) dominate in the Northern Hemisphere (especially Europe, China and USA), and positive values in the south, mostly in South America. We find that regional differences can be mostly attributed to variability across inversions, followed by interannual variability in the fluxes.
We found a strong contribution of interannual variability in NBP to $D$, associated with El Niño–Southern Oscillation (ENSO) and fluxes from land use change ($F_{LUC}$), especially in Brazil and Southeast Asia, the two regions also contributing the most to $D_{Globe}$ variance. Using a similar set of DGVMs and some of the inversions used here, Bastos et al. (2018) have shown that DGVMs estimated a positive source anomaly in response to El Niño in 2015/2016 that was 0.3–0.6 Pg C yr$^{-1}$ stronger than inversions. The coefficient of ONI estimated here ($−0.3$ Pg C yr$^{-1}$/s.d.u., s.d.u. being the ONI values in standard deviation units) would imply a source anomaly difference between the inversions and DGVMs of about 0.5 Pg C yr$^{-1}$ for these years ($ONI_{2015/16} = 1.6$s.d.u.), consistent with their estimate. Several studies have shown that variability in tropical CO$_2$ fluxes (mainly driven by ENSO) is better explained by variations in soil water storage than by temperature alone (Bastos et al., 2013; Humphrey et al., 2018; Poulter et al., 2014). The regions with stronger negative coefficients for ONI (NSA, BRA, SAF, and SEAS, Figure 4) are those where El Niño events impose strong warm and dry conditions, and those with positive coefficients (CAS, MIDE, and SAS) are associated with cool and wet conditions during El Niño (Bastos et al., 2013; Mason & Goddard, 2001). The combination of these two results hints at differences in the sensitivity of NBP from inversions and DGVMs to changes in water availability or temperature related with ENSO (section 3.5.1), or possibly also to changes in atmospheric circulation/mixing patterns linked to ENSO. The opposing sign in the sensitivity to temperature between inversions and DGVMs (Figure 5) may be due to an overestimate of the sensitivity of respiration to temperature (Bastos et al., 2018; van Schaik et al., 2018), or indirectly linked to the higher sensitivity of NBP to water availability in DGVMs compared to inversions. Models differ in prescribed soil depth and root water access for transpiration, but in general do not have deep rooting and ground water access of plants (Fan et al., 2017), which may explain why the sensitivity of regional CO$_2$ fluxes to water availability in DGVMs is higher than that of inversions with TWS. Still, inversions also show high uncertainty in these regions, which hampers the attribution of errors to one or other data source.

Excluding the years corresponding to the Mt. Pinatubo eruption, the strong peaks in $D_{Globe}$ are associated with moderate or strong El Niño (1982, 2009/2010, and 2015) or La Niña events (1984, 1988/1989, 2000, and 2011). In most of these years, the GCB2018 budget imbalance term shows similar changes as $D_{Globe}$. This suggests that improvements in DGVMs modeling of tropical vegetation sensitivity to water availability could help reducing this gap in future GCBs.

The higher values of $F_{LUC-\text{noLASC}}$ are consistent with the effect of CO$_2$ fertilization on natural ecosystems resulting in higher LUC-related emissions if a pristine forest is converted to managed land. However, $F_{LUC-\text{noLASC}}$ values (Figure 6) remain mostly between the two bookkeeping models, when they would be expected to be slightly lower, given that they rely on preindustrial C densities. Since most DGVMs do not include shifting cultivation, which is included in the BLUE model, estimates of $F_{LUC}$ from DGVMs are likely too high. The negative effect of $F_{LUC}$ to $D_{reg}$ in BRA and SEAS, and consequently in $D_{Globe}$, suggests therefore a possible overestimation of LUC-related emissions due to the forcing used, for example, higher deforestation rates in LUH2 than in FAO/FRA.

The strong peak in emissions from LUC following 2009 in DGVMs and in BLUE is mainly explained by sharp increases in cropland area during 2009–2017 coinciding with increased forest loss in the LUH2v2.1h land cover data set (Figure S5). On the contrary, in the FAOSTAT (2015) data used by HN2017, most of the cropland area increase occurs before 2009, and forest loss slows down only after 2005. The LC-CCI also reports a slow down of forest area loss and stabilization of cropland in Brazil after 2005. A slowdown in deforestation rates in Brazil over this period has further been reported by Hansen et al. (2013) using high-resolution satellite imagery. This indicates that the overestimation of DGVMs of $F_{LUC}$ inferred from the MMEM appears to be related with biases in deforestation rates from LUH2v2.1h in this region.

In SEAS, the two forcings show opposing changes in deforestation rates starting from circa 2000 onward (Figure S5), coinciding with higher $F_{LUC}$ in DGVMs and BLUE (Figure 6) LUH2v2.1h indicates a faster decrease of forest area, while (FAO, 2015) shows a tendency for deceleration in deforestation. LUH2v2.1h also reports higher cropland expansion values, with much faster rates following 2005. The study by Hansen et al. (2013) indicated that gross forest loss rates increased in Indonesia between 2000–2012. Li et al. (2018) showed that net forest area changes in Southeast Asia and Indonesia from Houghton et al. (2012) and from LUH2v2h between 1992–2012 are consistent with the changes reported by Hansen et al. (2013), but much
suggests that uncertainty from other terms is probably higher.

fuel dryness in this region (Van Leeuwen et al., 2013). For example, some models show very high

simulated by DGVMs, perhaps leading to an overestimation of fire intensity and combustion completeness

since burned area is still lower than that of Fire-CCI in those years). In order to understand the errors in

simulated by DGVMs, a much more detailed analysis of the different factors controlling

how vegetation response to drought during El Niño and consequent fire emissions (Patra et al., 2005) are

1997/1998 and 2015 in NSA compared to GFED4.1s. These very strong peaks are possibly a consequence of

of warming and decrease in water availability concurrent with increasing CO2 in some regions (Buermann

in the land sink compared to inversions. This might be because DGVMs underestimate the negative effects

positive values for the trend in EU, KAJ, and SEAS indicate that DGVMs show a stronger long-term increase

al., 2016), or the effect of N deposition in Asia (Liu et al., 2013), or forest regrowth (Pugh et al., 2019). The

mate the response of vegetation to the effect of CO2 fertilization (Fernández-Martínez et al., 2019; Thomas et

We showed that the difference between NBP estimates

of the LUC forcing.

We found that biases in burned area explain well biases in $E_{fire}$ and its contribution to $D_{SEAS}$ and $D_{CAS}$, but

not in NSA. A possible explanation for this disagreement might be related with simulated soil moisture and

fuel dryness in this region (Van Leeuwen et al., 2013). For example, some models show very high $E_{fire}$ in

1997/1998 and 2015 in NSA compared to GFED4.1s. These very strong peaks are possibly a consequence of

how vegetation response to drought during El Niño and consequent fire emissions (Patra et al., 2005) are

simulated by DGVMs, perhaps leading to an overestimation of fire intensity and combustion completeness

since burned area is still lower than that of Fire-CCI in those years). In order to understand the errors in

simulated by DGVMs, a much more detailed analysis of the different factors controlling $E_{fire}$, from burned area to fuel characteristics and fire dynamics, is needed, which is beyond the scope of this study.

DGVMs tend to simulate a slower increase in CO$_2$ uptake than reported by inversions in several regions and

globally. The weak negative trend in $D_{globe}$ is consistent with studies showing that DGVMs may underestimate

the response of vegetation to the effect of CO$_2$ fertilization (Fernández-Martínez et al., 2019; Thomas et

al., 2016), or the effect of N deposition in Asia (Liu et al., 2013), or forest regrowth (Pugh et al., 2019). The

positive values for the trend in EU, KAJ, and SEAS indicate that DGVMs show a stronger long-term increase

in the land sink compared to inversions. This might be because DGVMs underestimate the negative effects

of warming and decrease in water availability concurrent with increasing CO$_2$ in some regions (Buermann

et al., 2018). However, inversions might also indirectly overestimate the land sink by underestimating the

ocean sink intensification in recent years (Landschützer et al., 2015).

We found a significant contribution of uncertainty in $F_{LUC}$ and of $E_{fire}$, as well as a small trend component,
to regional and global $D$. It is thus worth evaluating whether those DGVMs including representation of key

LUC processes such as shifting cultivation and wood harvest (Arneth et al., 2017), N deposition or fires

perform better compared to the ensemble of 16 DGVMs.

By fitting the LME to each subset of DGVMs, we find that simulating fire, N deposition or wood harvest
does not lead to major reductions in uncertainty (Table S1). The DGVMs including shifting cultivation,
though, show significant reduction of the bias globally, as well as in SEAS and EQAF, regions where these
practices are relevant (Heinimann et al., 2017). When adding the process representation as a predictor in
the LME$_{EG}$ (Figure S7), including some of these processes as predictors are found to improve the LME fit,
but the coefficients are generally nonsignificant, excepting N deposition globally and wood harvest in some
regions. These results indicate that including gross transitions in $F_{LUC}$ estimates is a key improvement to
DGVMs, in order to reduce uncertainties in global and regional budgets.

The fact that models with fire do not perform much better than those without fire may be because models
do not realistically simulate burned area and $E_{fire}$, as discussed above, but also because those DGVMs not
simulating fire might compensate by having higher sensitivity of decomposition to temperature (as found
for tropical regions in Figure 5). Simulating wood harvest also does not seem to necessarily reduce the bias
in $D$, and is only a weak predictor in NSA, NAF and KAJ. This is possibly because the wood harvest fluxes
are generally small, compared to other terms, for example, shifting cultivation (Wilkenskjeld et al., 2014), or
because trade of wood products is not accounted for, and therefore the location of sinks and sources related
with wood production and consumption is not well captured by DGVMs. Finally, N deposition if found to
significantly contribute to $D_{globe}$ (the negative sign indicating that models simulating N deposition tend
to have lower $D$), but when fitting the LME$_{EG}$ model to that subset of DGVMs, no reduction in $D$ is found. This
suggests that uncertainty from other terms is probably higher.

5. Conclusions

In this study we attempted to identify the sources of uncertainty explaining the budget imbalance term of the
latest GCB (Le Quéré, Andrew, Friedlingstein, Sitch, Hauck, et al., 2018b). We compared DGVM outputs of
the NBP between 1979 and 2017 with results from an ensemble of atmospheric inversions, whose fluxes are
consistent with the growth rate of atmospheric CO$_2$. We showed that the difference between NBP estimates
by DGVMs and inversions matches the main variation patterns of the GCB2018 budget imbalance term. At the global scale, interannual variability and between-DGVM variability contribute the most to $D$, and variability across inversions contributes the least. Based on our results, we could identify key regions and processes that deserve special attention from the Earth System and C cycle communities:

(i) the reduction of uncertainty in atmospheric inversions, for example, through more in situ observations in the tropics, or the use of remote-sensing;

(ii) improvements in $F_{\text{LCC}}$ estimates, by including shifting cultivation DGVMs and also by evaluating thoroughly land use change estimates at country level, especially for the period overlapping high-resolution satellite records;

(iii) the improvement of tropical hydroecological processes and fire representation within DGVMs;

(iv) accurate estimates of lateral carbon fluxes and their variability, to guarantee a fair comparison between top-down and bottom-up approaches.

There are several ongoing efforts to tackle each of the processes highlighted above. Satellite-based inversions might help to reduce uncertainty in the flux partitioning between land, ocean, and different continents, and regional inversions might provide better constraints of regional CO$_2$ fluxes. Above-ground biomass products at annual time steps and moderate resolution are currently available at least for the tropics (Brandt et al., 2018), providing a truly independent estimate of losses and gains of carbon in ecosystems due to both natural and anthropogenic processes. Considerable work is being put into improving tropical phenology, vegetation-water interactions, mortality, and nutrient limitations by several modeling teams.

The predictors tested in this work reduced the variance of residuals of $D$ only by 25–51%, and residuals still show large variability at the global and regional scale. This points to other processes not discussed in this study either on the DGVM side (e.g., response of forests to climate change, forest regrowth or forest aging, or their response to nitrogen deposition to nutrient limitations) or on the inversion side (e.g., errors in transport modeling, inversion set up, atmospheric monitoring network) that might deserve more attention in the future. Such efforts will hopefully allow reducing uncertainty in global and regional CO$_2$ budgets in the near future. This, in turn, is fundamental for a science-based evaluation of the effectiveness of mitigation policies implemented in light of the Paris Agreement.

References


Acknowledgments

This work was partly supported by the European Space Agency Climate Change Initiative ESA-CCI RECCAP2 project (ESRIN/4000123002/18/I-NB). P. K. P. is partly supported by the Environment Research and Technology Development UT-Battelle, LLC, under contract DE-AC05-76OR00001, Oak Ridge National Laboratory is operated by UT-Battelle, LLC, under contract DE-AC05-76OR00001, Oak Ridge National Laboratory, U.S. Department of Energy. A. J.’s effort was supported in part by the Department of Energy (DESC0016323) and NSF (NSF AGS 12-43071). P. C. acknowledges support from the European Research Council Synergy Project SyG-2013-610028 IMBALANCE-P, and P. C. and D. M. the ANR CLAND Convergence Institute. W. P. received funding from the European Research Council (ERC) for the Airborne Stable Isotopes of Carbon from the Amazon (ASICIVA) project, Contract 640987. N. E. S. received funding from the Netherlands Organisation for Scientific Research (NWO) for the Ruisdael Observatory. I. T. L. received funding from NWO under Contract 016.Veni.171.095. CT Europe simulations were performed using a grant for computing time (SH-312, 16666) from NWO. The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/06RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. Oak Ridge National Laboratory is operated by UT-Battelle, LLC, under Contract DE-AC05-00OR22725 with the U.S. Department of Energy. The authors would like to thank Corinne Le Quéré and Johannes Winckler for constructive comments in the preparation of this study and the ESA Soil Moisture CCI and the Fire-CCI teams for producing and maintaining the data sets used. The GRACE reconstructed data set was kindly provided by V. Humphrey.


UNFCCC (2015). Adoption of the Paris Agreement. FCCC/CP/2015/L.9/Rev.1


