Response to comments

Paper #: *bg-2017-186*
Title: *Land-use and land-cover change carbon emissions between 1901 and 2012 constrained by biomass observations*
Journal: *Biogeosciences*

Editor:

General Comments:

Comment:

Response comments and proposed edits are satisfactory overall, and if implemented fully I expect the manuscript will be accepted for publication. There are, however, some remaining issues I would ask the authors to address.

Response:

We thank the editor for the suggestions. We fully implemented the proposed edits in the revised manuscript. Please see the point-by-point responses below.

Comment:

The abstract ends with the claim that constraint from biomass observations reduced the uncertainty of the historical carbon budget, however this does not appear to be consistent with the results of the manuscript. The study reports that the mean and standard deviation on the constrained LUC emissions is almost identical to that from the TRENDY DGVMs if one assumes that they have a normal distribution. It would thus be difficult to argue that uncertainty has been reduced and such statements need to be removed from the paper. You could potentially claim to have identified one or more of the TRENDY members that are inconsistent with the observations. You can also comment, as in the Conclusions, that the wide range in the constrained estimates derives from wide uncertainty in biomass observations which acts as the constraint, and that reductions in uncertainties on observed biomass would improve the constraint that these observations could provide.

Response:

We went through the manuscript and revised the related sentences. All the "reduced uncertainty" description has been deleted. We pointed out that the wide range in the constrained estimates are from the biomass observations in Abstract and also in Results and Discussion.

Comment:

R1 asks a good question about what this study offers given the above result. The so-called "lessons learned" are well communicated in the response comment and should be reflected in the paper's discussion and/or conclusions.

Response:

As suggested, we added these "lessons learned" in Conclusion.

Comment:

Given concerns about non-normality of the distribution of ELUC estimates across the TRENDY DGVM ensemble, you might consider using quantiles for both the original _and_ the constrained estimates when you compare them.

Response:

As suggested, we added the interquantiles of global constrained E^{c}_{LUC} in **Figure 6** and gave the interquantile ranges for each region in **Table S1**. Texts are also revised accordingly.

Land-use and land-cover change carbon emissions between 1901 and 2012 constrained by biomass observations

Wei Li[1], Philippe Ciais[1], Shushi Peng[1,2], Chao Yue[1], Yilong Wang[1], Martin Thurner[3], Sassan S. Saatchi[4], Almut Arneth[5], Valerio Avitabile[6], Nuno Carvalhais[7,8], Anna B. Harper[9], Etsushi Kato[10], Charles Koven[11], Yi Y. Liu[12], Julia E.M.S. Nabel[13], Yude Pan[14], Julia Pongratz[13], Benjamin Poulter[15], Thomas A. M. Pugh[5,16], Maurizio Santoro[17], Stephen Sitch[18], Benjamin D. Stocker[19,20], Nicolas Viovy[1], Andy Wiltshire[21], Rasoul Yousefpour[13,‡], Sönke Zaehle[7]

[1]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
[2]Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
[3]Department of Environmental Science and Analytical Chemistry (ACES) and the Bolin Centre for Climate Research, Stockholm University, SE-106 91 Stockholm, Sweden
[4]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[5]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research - Atmospheric Environmental Research (IMK-IFU), Garmisch-Partenkirchen, Germany
[6]Centre for Geo-Information and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708PB Wageningen, The Netherlands
[7]Department for Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany
[8]CENSE, Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Caparica, Portugal
[9]College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, UK
[10]Institute of Applied Energy, Minato, Tokyo 105-0003, Japan
[11]Climate and Ecosystem Sciences Department, Lawrence Berkeley Lab, Berkeley, CA, USA
[12]ARC Centre of Excellence for Climate Systems Science & Climate Change Research Centre, University of New South Wales, Sydney, New South Wales 2052, Australia
[13]Max Planck Institute for Meteorology, Hamburg, Germany
[14]USDA Forest Service, Durham, New Hampshire, USA
[15]Department of Ecology, Montana State University, Bozeman, MT 59717
[16]School of Geography, Earth & Environmental Science and Birmingham Institute of Forest Research, University of Birmingham, Birmingham, B15 2TT, United Kingdom
[17]GAMMA Remote Sensing, 3073 Gümligen, Switzerland
[18]College of Life and Environmental Sciences, University of Exeter, Exeter, UK
[19]Climate and Environmental Physics, and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[20]Imperial College London, Life Science Department, Silwood Park, Ascot, Berkshire SL5 7PY, UK
[21]Met Office Hadley Centre, Exeter, Devon. EX1 3PB, UK
[‡]Current address: Chair of Forestry Economics and Forest Planning, University of Freiburg, 79106 Freiburg, Germany

Correspondence to: Wei Li (wei.li@lsce.ipsl.fr)

Abstract. The use of dynamic global vegetation models (DGVMs) to estimate CO_2 emissions from land-use and land-cover change (LULCC) offers a new window to account for spatial and temporal details of emissions, and for ecosystem processes affected by LULCC. One drawback of LULCC emissions from DGVMs however is ~~their large uncertainty~~lack of observation constraint. Here, we propose a new method of using satellite- and inventory-based biomass observations to constrain historical cumulative LULCC emissions (E^c_{LUC}) from an ensemble of nine DGVMs based on emerging relationships between simulated vegetation biomass and E^c_{LUC}. This method is applicable at global and regional scale. The original DGVM estimates of E^c_{LUC} range from ~~Compared to the large range of E^c_{LUC} in the original ensemble (~~94 to 273 Pg C~~)~~ during 1901-2012.~~,~~ After constrained by current biomass observations, we ~~allow us to~~ derive a ~~new~~ best estimate of 155 ± 50 (1-σ Gaussian error) Pg C. The constrained LULCC emissions are higher than prior DGVM values in tropical regions, but significantly lower in North America. Our emergent constraint approach independently verifies the median model estimate by biomass observations, giving support to the use of this estimate in carbon budget assessments. The uncertainty in the constrained E^c_{LUC} is still relatively large because of the uncertainty in the biomass observations, and thus more accurate biomass observations in future will help improve the constraint. ~~of constraining cumulative LULCC emissions based on biomass observations reduces the uncertainty of the historical carbon budget, and~~This constraint method can also be applied to evaluate the impact of land-based mitigation activities.

1 Introduction

Carbon emissions from land-use and land-cover change (LULCC) are part of the human perturbation to the global carbon cycle (Houghton et al., 2012; Le Quéré et al., 2015) and started in fact before the Industrial Era when fossil fuel CO_2 emissions appeared. Since 1850, estimated cumulative LULCC emissions, E^c_{LUC}, represent one-third of total cumulative anthropogenic CO_2 emissions (Boden et al., 2013; Houghton et al., 2012; Le Quéré et al., 2015). Annual LULCC emissions have been higher than those from fossil fuel burning until the 1930s (Boden et al., 2013; Houghton et al., 2012; Le Quéré et al., 2015), and today represent a smaller but persistent perturbation in the global carbon cycle. Unlike fossil fuel emissions, relative uncertainties in LULCC emissions are high, due to the difficulty of assessing this flux from measurements. Some progress has been made to better quantify gross tropical deforestation emissions by combining spatial biomass data with satellite-derived maps delineating forest cover loss (Harris et al., 2012). But such spatially resolved data are not available beyond the last decade and provide only gross deforestation emissions, i.e. do not track the regrowth of secondary ecosystems or legacy soil carbon losses that can persist long after deforestation.

Bookkeeping models (Hansis et al., 2015; Houghton, 1999) based on historical LULCC area data and tabulated functions of carbon losses and gains are one approach to estimating E^c_{LUC}, but they do not include the effects of environmental changes on carbon stocks before and after LULCC happens (Gasser and Ciais, 2013; Pongratz et al., 2014). The bookkeeping model of Houghton (1999) used for the annual update of the global carbon budget (Le Quéré et al., 2015) is based on regionally

aggregated data, and does not consider spatial differences of LULCC fluxes within a region. Alternatively, the estimated LULCC fluxes by dynamic global vegetation models (DGVMs) account for spatial and temporal variations of carbon stock densities and land-cover change, as well as for delayed ("legacy") carbon fluxes. In DGVMs, LULCC fluxes are related to environmental conditions through simulated carbon cycle processes, i.e., net primary production (NPP) and respiration resulting in changes of biomass and soil carbon stocks are simulated with variable atmospheric CO_2 concentration and climate. Yet, LULCC emissions from DGVMs differ greatly, even when these models are prescribed with the same inputs of land-cover change data (such as time-variable areas of pasture and crops) (Pitman et al., 2009). Several factors are responsible for differences of E^c_{LUC} among DGVMs, including 1) different representations of processes that determine the carbon densities of vegetation and soils subject to land-use change; 2) using dynamic vegetation or prescribing a fixed vegetation distribution; and 3) use of different rules assigning how natural vegetation types change to agricultural areas (Peng et al., 2017; Pitman et al., 2009; Reick et al., 2013).

Carbon initially stored in forest biomass contributes the predominant portion of the LULCC emissions after deforestation (Hansis et al., 2015). Thus, an accurate representation of the biomass carbon density exposed to LULCC is crucial to reduce uncertainties of DGVM-based E^c_{LUC} estimates. Global biomass datasets based on inventories and satellites recently became available. These datasets (Table 1) provide the spatially distributed biomass carbon density at regional or global scales (Avitabile et al., 2016; Baccini et al., 2012; Carvalhais et al., 2014; Liu et al., 2015; Pan et al., 2011; Saatchi et al., 2011; Santoro et al., 2015; Thurner et al., 2014), but differ in terms of their coverage of aboveground or belowground biomass and whether they provide only forest biomass or biomass for all vegetation types.

In this study, we propose a new method to combine recent satellite- and inventory-based biomass datasets to constrain E^c_{LUC} simulated by DGVMs (Figure 1). We analyzed the outputs from nine DGVMs (Table 2) of the Trends in Net Land-Atmosphere Exchange (TRENDY-v2) project (Sitch et al., 2015) (http://dgvm.ceh.ac.uk/node/9) and developed global and regional regressions between initial biomass in 1901 and present-day biomass (average of 2000-2012), and regressions between E^c_{LUC} during 1901-2012 and initial biomass across the DGVMs. The former set of regressions is used to extrapolate present-day observation-based biomass (Table 1) to initial biomass in the year 1901. The latter set of regressions is applied to provide an emerging constraint on E^c_{LUC} as a function of initial biomass (Figure 1). Using the Gaussian uncertainties associated with the observation-based biomass datasets and the uncertainties in the two regressions, Gaussian errors of E^c_{LUC} can be derived after applying the biomass-constraint. ~~uncertainties of E^c_{LUC} can be reduced by providing Gaussian errors after applying the biomass-constraint, compared to the original large range from TRENDY models.~~

2 Materials and Methods

2.1 LULCC emissions and biomass from the DGVMs

The DGVMs in TRENDY v2 conducted two simulations (labeled S2 and S3) between 1860 (except JSBACH from 1850, Table 2) and 2012, with outputs quantifying LULCC emissions over the period 1901-2012 (Sitch et al., 2015). Both simulations are performed with changing climate and CO_2 concentration, but one (called S3) has variable LULCC maps based on Land Use Harmonization (LUH) dataset (Hurtt et al., 2011) (with an extension until 2012), and the other (called S2) has a time-invariant land-cover map representing the state in 1860. The difference of net biome production (NBP, the net carbon exchange between the biosphere and the atmosphere) between these two simulations (S3 and S2) defines modeled LULCC emissions. This ~~definition~~ calculation of LULCC emissions by DGVMs includes the "lost sink capacity" (called "altered sink capacity" in (Gasser and Ciais, 2013) and "the loss of additional sink capacity" in (Pongratz et al., 2014)), because simulated NBP in the S2 simulation without LULCC ~~represents~~ is a net sink over areas affected by LULCC in S3. For example, forests have larger carbon storage and slower turnover time than croplands and thus are expected to be carbon sinks when atmospheric CO_2 level increases. After deforestation to croplands, this sink capacity due to CO_2 fertilization is lost. Modeled LULCC emissions include the legacy emissions from soil carbon losses, and emissions from wood and other products produced by LULCC, as far as the latter are included in the TRENDY v2 models (Table 2). The DGVMs used in this study are CLM4.5 (Oleson et al., 2013), JSBACH (Reick et al., 2013), JULES3.2 (Best et al., 2011; Clark et al., 2011), LPJ (Sitch et al., 2003), LPJ-GUESS (Smith et al., 2001), LPX-Bern (Stocker et al., 2014), ORCHIDEE (Krinner et al., 2005), VISIT (Ito and Inatomi, 2012; Kato et al., 2013), and OCN (Zaehle and Friend, 2010). Each DGVM is described briefly in Table 2.

LULCC can either reduce or increase the biomass amount over time depending on the LULCC types. For example, forest clearing turns forest biomass into atmospheric CO_2 eventually, while secondary forest regrowth can increase biomass. The overall effects of LULCC on biomass during the historical period is a net loss of carbon (Houghton, 1999) due to converting natural vegetation into cultivated lands by human (Klein Goldewijk et al., 2011). Identifying the LULCC-affected grid cells in each model is thus critical, because only biomass in these grid cells should be used to constrain LULCC emissions. Grid cells affected by LULCC differ among models. Although all models share the same pasture and cropland areas from LUH dataset (Hurtt et al., 2011), the models have different numbers of PFT, use different PFT definitions and have different allocation rules for translating the shared agricultural data into the new vegetation cover (Peng et al., 2017; Pitman et al., 2009; Reick et al., 2013). As a result, there is no unified map to determine the LULCC-affected grid cells in all models. For the same reasons, the forest areas and the LULCC types are also different among models.

In this study, we adopted the "deforestation grid cells" in their corresponding PFT maps as a criterion to locate the LULCC-affected grid cells from DGVM outputs. Thus we used the PFT maps from each model to first calculate the temporal change of forest area (total area of all forest PFTs) during 1901-2012 and then select the grid cells that experienced deforestation by

Formatted: Subscript

Formatted: Subscript

comparing the forest area maps between year 1901 and year 2012 (net deforestation). This procedure produces a good approximation, given the continuously decreasing trend of forest area in LULCC hotspot regions like South and Central America (Figure 2). We also tested an alternative method to determine the LULCC-affected grid cells in TRENDY model outputs, i.e., PFT maps were compared year-by-year during 1901-2012, and grid cells with deforestation were selected (gross deforestation). This method tends to give a greater number of LULCC-affected grid cells, reducing the goodness of fitting in the regression between the biomass in 1901 and E^{c}_{LUC} during 1901-2012 (Figure S1 and Figure S2). Therefore, the method of gross deforestation is not used for further analyses.

We verified that deforestation grid cells are responsible for most of the total net LULCC flux. In fact, the average of the different model simulations of LULCC emissions from deforestation grid cells between 1901 and 2012 is approximate 90% of the total LULCC emissions from all grid cells (Figure S1). The LULCC emissions in this study are thus taken to equal the sum of LULCC emissions from the selected deforestation grid cells using our criterion. It should be noted that, although only deforestation is used as a single criterion to define grid cells affected by LULCC in DGVMs, modeled LULCC emissions also include other types of land-use transitions involving pairs of non-forest PFTs in the selected grid cells.

In each model, only biomass in deforestation grid cells is considered. Biomass in the year 1901 is thereby defined as *initial* biomass, and biomass averaged during 2000-2012 is defined as *present* biomass. An ordinary least squares linear regression is performed with the outputs of all models between initial biomass and E^{c}_{LUC} from 1901 to 2012, and between the initial and the present biomass at both global and regional scales. Our division of nine regions in the world (Figure 2) for estimating LULCC fluxes is the same as Houghton et al. (1999).

2.2 Observation-based biomass datasets

Several biomass datasets (Avitabile et al., 2016; Baccini et al., 2012; Carvalhais et al., 2014; Liu et al., 2015; Pan et al., 2011; Saatchi et al., 2011; Santoro et al., 2015; Thurner et al., 2014) based on inventories and remote sensing can be potentially used to constrain E^{c}_{LUC} through the set of regressions from DGVMs. However, these biomass datasets cover different parts of biomass (aboveground, belowground or total) and different regions (tropics, Northern Hemisphere or the globe) at different spatial resolutions (Table 1). We choose the global grid-based biomass dataset from Carvalhais et al. (2014) to derive an observational constraint which brings a best estimate of E^{c}_{LUC}. This map merges the Northern Hemisphere biomass dataset from Thurner et al. (2014) and the tropical biomass dataset from Saatchi et al. (2011). An advantage of this map is its consistency in biomass terms with the outputs of TRENDY models, because it documents aboveground + belowground, and forest + herbaceous biomass (Tables 1, 2). Three other biomass maps are used as alternative datasets for sensitivity tests: 1) the global biomass map from the GEOCARBON project, a merged product of the biomass datasets in the Northern Hemisphere (Santoro et al., 2015) and tropics (Avitabile et al., 2016); 2) regional biomass estimates from Pan et al. (2011) based on forest inventory data; and 3) the biomass map from Liu et al. (2015) derived from satellite vegetation optical depth. The GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and Liu et al. (2015) datasets that only provide

aboveground biomass were extended to total forest biomass using the conversion factors for the nine regions (Liu et al., 2015). The global biomass maps from GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and Pan et al. (2011) are only for forest (Table 1), and we do not add the herbaceous biomass to these two datasets because the global herbaceous biomass only accounts for about 3% of the global total biomass (Carvalhais et al., 2014). Note that the uncertainties in the corresponding constrained results using these three alternative datasets do not include 1) the uncertainties of converting aboveground biomass to the total of aboveground and belowground biomass for datasets from Liu et al. (2015) and GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and 2) the uncertainties of ignoring non-woody biomass in datasets from GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and Pan et al. (2011). The biomass maps of Carvalhais et al. (2014), GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and Liu et al . (2015) with different spatial resolutions were aggregated to 1°×1° resolution before selecting the deforestation grid cells.

2.3 Methods to identify grid cells subject to past deforestation in biomass datasets

It is not practical to use PFT maps from DGVMs to define deforestation grid cells in the observation-based biomass datasets because PFT maps and forest area change since 1901 differ across DGVMs. Instead, we diagnosed deforestation grid cells in the biomass maps using three harmonized methods (Method-A, Method-B and Method-C). All the methods are based on the reconstructed historical agricultural area from the History Database of the Global Environment (HYDE v3.1) (Klein Goldewijk et al., 2011) but with different hypotheses regarding how agricultural expansion has affected forests. These harmonized methods are representative of the different rules of assigning LULCC data to natural vegetation types in DGVMs. Method-A assumes that the increase of cropland area in a grid cell between 1901 and 2012 is taken from forest; Method-B assumes that the increase of cropland and pasture is taken proportionally from all natural vegetation types; and Method-C (like the “BM3” scenario in (Peng et al., 2017)) assumes that the increase of cropland and pasture is first taken from forest and then from natural grassland if no more forest area is available, and that the regional forest area change is set to match the historical forest reconstruction from (Houghton, 2003). Because the biomass distribution in Pan et al. (2011) is given as regional mean values and not resolved on grid cell basis, it is impossible to select deforestation grid cells directly from this dataset using the above methods. Therefore, for each region, we calculated the ratios of biomass in deforestation grid cells according to Method-A, Method-B and Method-C to the total biomass in all grid cells in each of the other three biomass datasets (Carvalhais et al. (2014), GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and Liu et al. (2015)). For each method (Method-A, -B, and -C), the three ratios corresponding to the three biomass datasets were further averaged in each region. The total biomass amount from Pan et al. (2011) in each region was multiplied by the average ratio to derive the biomass equivalent to using Method-A, Method-B and Method-C for the dataset from Pan et al. (2011).

These three methods applied to the above-listed biomass datasets are also applied as sensitivity tests to select the deforestation grid cells since 1901 in the TRENDY model outputs. Identically, regressions are performed using initial biomass amount and E^{c}_{LUC} from these selected grid cells. Due to the inconsistencies between the three methods and the

historical PFT maps of each DGVM, the biomass amount in 1901 in the selected grid cells using these three methods is higher than using PFT maps, but the E^c_{LUC} are lower, reflecting a lower representativeness of the deforestation grid cells using these three methods for DGVM outputs (Figure S1). As a consequence, a weaker goodness of regression fit was found between E^c_{LUC} and initial biomass (Figure S2).

2.4. Uncertainties of constrained LULCC emissions

The biomass from Method-A, Method-B and Method-C obtained from each dataset is extrapolated into biomass for the year 1901 using the regression between initial biomass and present biomass modeled by the DGVMs. This biomass in 1901 is then applied in the regression between modeled E^c_{LUC} and modeled initial biomass among different DGVMs to calculate constrained E^c_{LUC}. In this emerging constraint approach (Figure 1), the uncertainties of constrained E^c_{LUC} are a function of the uncertainties of the observed biomass datasets, of the linear regression goodness of fit for the two regressions (regressions between E^c_{LUC} and the initial biomass, and between the initial and present biomass), and of the slopes of the regressions. The uncertainty of constrained LULCC emissions is calculated as (Stegehuis et al., 2013):

$$\sigma_{LULCC} = \sqrt{\alpha^2 \sigma^2_{initial\_biomass} + \sigma^2_{res\_LULCC}} \quad (1)$$

$$\sigma_{initial\_biomass} = \sqrt{\beta^2 \sigma^2_{present\_biomass} + \sigma^2_{res\_biomass}} \quad (2)$$

where σ_{LULCC}, $\sigma_{initial\_biomass}$ and $\sigma_{present\_biomass}$ are the uncertainties of constrained E^c_{LUC}, the uncertainty of initial biomass and the uncertainty of present biomass. α and σ_{res_LULCC} represent the slope and the standard deviation of the residuals from the linear regression fit between E^c_{LUC} and initial biomass. β and $\sigma_{res\_biomass}$ represent the slope and standard deviation of the residuals from the linear regression between initial biomass and present biomass.

2.5. Two supplementary methods to constrain E^c_{LUC} using biomass observations

We also tested two supplementary methods to constrain E^c_{LUC}: Method-S1, using the regression between E^c_{LUC} and present-day biomass from TRENDY models rather than extrapolating present biomass to biomass in 1901 first, and Method-S2, using ΔB (biomass difference between present biomass and biomass in 1901 derived from the model simulations) instead of a regression between biomass in 1901 and present-day biomass to extrapolate the observation-based biomass in 1901. In Method-S1, the uncertainties in the biomass observations and in the regression between E^c_{LUC} and present biomass from models are used to calculate the uncertainties in the constrained E^c_{LUC}. In Method-S2, the uncertainties in the biomass observations and the standard deviation of ΔB among models are used.

Formatted: Heading 2

3. Results

3.1. Forest area change and cumulative LULCC emissions in DGVMs

As expected, a general decrease of forest area is found between 1901 and 2012, especially in regions subject to extensive deforestation over the last decades, namely South and Central America, south and southeast Asia, and tropical Africa (Figure 2), which is in support of our methods of defining deforestation grid cells, although the forest area in some regions differs substantially across DGVMs. Differences in forest area are large in tropical Africa, North America and the former Soviet Union while they are smaller in South and Central America, and south and southeast Asia (Figure 2). There are several reasons for these differences of forest area: 1) models have different initial distribution of PFTs (the TRENDY v2 protocol only prescribed the same initial area of natural vegetation, but did not specify the PFTs that compose natural vegetation); 2) some models consider only net LULCC, but others have gross LULCC including some sub-grid transitions (Table 2, e.g., see a comparison using the JSBACH model (Wilkenskjeld et al., 2014)); 3) models have different treatments for changing pasture areas (either proportional from natural vegetation or preferential from natural grasslands). In North America, China region and western Europe, the forest area decreased in the first half of the 20th century and then increased in recent decades. Yet, the magnitude of the increase is smaller than that of the previous decrease in these regions, and the global average is net forest loss between 1901 and 2012 (ranging from 2.3 Mkm^2 to 16.8 Mkm^2 across the nine models).

E^c_{LUC} from the nine DGVMs between 1901 and 2012 ranges from 1.7 (-0.6 to 6.0) Pg C (median and range, a positive number indicating a net cumulative flux to the atmosphere) in north Africa and Middle East to 42.6 (33.5 to 81.4) Pg C in South and Central America, resulting in a global total of 148 (94 to 273) Pg C (Table 3). Tropical Africa and south and southeast Asia have the second large E^c_{LUC} of 21.8 (15.8 to 57.8) and 21.8 (9.6 to 46.6) Pg C, respectively. Although afforestation / reforestation occurred in North America after around 1960, and in China after 2000 (Figure 2), E^c_{LUC} of these two regions are positive since 1901, with median values of 19.9 and 10.7 Pg C, respectively (Table 3).

3.2. Relationship between cumulative LULCC emissions and initial biomass

We found a positive linear relationship between E^c_{LUC} and initial biomass in deforestation grid cells of each model, both at global scale and in the regions considered (Figure 3). The coefficients of determination (r^2) are 0.61, 0.58 and 0.76 in South and Central America, south and southeast Asia and tropical Africa, respectively. Due to stable or slightly increasing forest area (Figure 2), the correlation between initial biomass and E^c_{LUC} is small in western Europe (Figure 3). The slopes of the relationships between E^c_{LUC} and initial biomass shown in Figure 3 range from 0.13 Pg C/Pg C in western Europe to 0.63 Pg C/Pg C in north Africa and Middle-East. In tropical regions with intensive LULCC, the slope is similar between south and southeast Asia (0.36 Pg C/Pg C) and tropical Africa (0.37 Pg C/Pg C), but lower in South and Central America (0.21 Pg C/Pg C). These slopes reflect the sensitivity of cumulative carbon loss to initial biomass carbon stock. They are mainly influenced by the fraction of deforested area relative to the initial forest area in each region, which explains 46% of the

variations of the slopes across regions (Figure S3). Differences in biomass density across regions and in the use of gross or net transitions among DGVMs (Table 2) also contribute to variations of slopes.

3.3. Cumulative LULCC emissions constrained by present-day biomass observations

There is also a strong positive relationship between initial biomass in 1901 and present-day biomass in grid cells experienced deforestation (Figure 4). The r^2 of this regression is higher than 0.92 in most regions, except in North America and China region (0.89 and 0.76, respectively). The regression between present-day and initial biomass was applied to extrapolate current observation-based biomass back to the year 1901. The extrapolated biomass in 1901 is higher than that in the present day, mainly due to a larger forest area — although it is difficult to discriminate other effects such as CO_2 fertilization that might have increased biomass between 1901 and 2012.

Using the chain of emerging constraints between present-day and initial biomass (Figure 4), and between E^c_{LUC} and initial biomass (Figure 3), with all uncertainties being propagated (Eqs. (1) and (2)), we were able to constrain E^c_{LUC} during 1901-2012 by biomass observations (Figure 3, S4, S5, and Table 3). E^c_{LUC} constrained by the biomass dataset of Carvalhais et al. (2014) is 155 ±50 (mean and 1-σ Gaussian error) Pg C and this estimate is robust to the choice of the methods to define deforestation grid cells in biomass datasets (constrained E^c_{LUC} = 152 ±49, 154 ±50 and 159 ±51 Pg C for Method-A, Method-B and Method-C, respectively). The difference between the global constrained E^c_{LUC} and the median value of original E^c_{LUC} (148 Pg C) from TRENDY DGVMs is not significant, suggesting that the median model estimate is independently verified by biomass observations. Still, some models that are inconsistent with the observations can be identified (Figure 3).

The uncertainties reported in our constrained estimate of E^c_{LUC} include uncertainties in the biomass observations, and in the scatter of the two regressions (Figure 3, 4) used to construct the emerging constraint. The uncertainties in the constrained E^c_{LUC} are still relatively large, resulting from the large uncertainties in the biomass observations. ~~Compared to the original global total E^c_{LUC} range (94 to 273 Pg C) in DGVMs, the constrained estimate has a smaller uncertainty and a larger median value.~~ However, ~~I~~it should be noted that we summed the biomass uncertainty in each deforestation grid cell to give the regional biomass uncertainty, which gives a maximum uncertainty with a potential assumption that the uncertainties in all grid cells are fully correlated. In reality, the regional biomass uncertainty should be lower, thus leading to a lower uncertainty of constrained E^c_{LUC}. ~~However,~~But it is difficult to estimate the error correlations of observation-based biomass between different grid cells at this stage.

Although the constrained global E^c_{LUC} value is only 7 Pg C higher than the median of the original DGVM ensemble (Table 3), larger differences can be found at regional scale (Figure 5). Constrained E^c_{LUC} estimates are higher than the original modeled values in south and southeast Asia, tropical Africa and South and Central America (Table 3). For example, the constrained E^c_{LUC} is 37.2 ±14.4 Pg C in south and southeast Asia, compared to the original TRENDY median value of 21.8 Pg C (range of 9.6 to 46.6 Pg C) for that region. The constrained emissions are also higher in China region and the Pacific developed region compared to the prior median value (see Table 3). A significantly large reduction in E^c_{LUC} through the emerging constraint is

found in North America because of the lower biomass amount from observation-based datasets than from DGVMs. The original median E^c_{LUC} of that region is 19.9 Pg C (range of 8.6 to 40.8 Pg C) while the constrained result is 10.8 ±7.1 Pg C. Constrained E^c_{LUC} are also lower than original estimates in western Europe, and north Africa & Middle East, although their contributions to the global total emissions are very small (Table 3).

Alternative estimates of E^c_{LUC} constrained by three other biomass datasets (Liu et al. (2015), GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015), and Pan et al. (2011)) are provided in Figure 6 and Table 3. In general, the constrained E^c_{LUC} using biomass maps from Liu et al. (2015) and GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) are rather consistent with ~~slightly higher (7 Pg C on average) than~~ (on average only 4.5% higher than) those from Carvalhais et al. (2014), implying the robustness of our estimates. The biomass dataset from Pan et al. (2011) leads to lower LULCC emission estimates at global scale, mainly due to a lower estimate in south and southeast Asia (Table 3) compared to the other products. In the Pacific developed region, GEOCARBON-based estimates (Avitabile et al., 2016; Santoro et al., 2015) are much higher than those from Carvalhais et al. (2014), because the latter has a gap in their biomass map in the southern part of Australia (Carvalhais et al., 2014). In Figure 6, we show the original E^c_{LUC} from TRENDY DGVMs as quantiles because we don't know whether they follow a normal distribution, and to be comparable, the interquantiles of the constrained E^c_{LUC} are also shown. The interquantile range of constrained E^c_{LUC} is larger than that of the original E^c_{LUC} (Figure 6). This, however, doesn't mean that our emerging constraint method is not effective, but that the relatively large uncertainty of the constrained E^c_{LUC} is propagated from the biomass observation uncertainty, which is about one-third of the mean biomass at global level (Carvalhais et al., 2014).

The global constrained E^c_{LUC} using the two supplementary methods is almost identical with that using our original method in Figure 1 (see an example in Figure S6). The difference of E^c_{LUC} between supplementary and original methods at the global level is <1% for all biomass observation datasets (Carvalhais et al. (2014), Liu et al. (2015), GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and Pan et al. (2011)) and all methods to select LULCC grid cells (Method-A, -B and -C). This suggests that our constrained results are very robust. The change in the uncertainty in global constrained E^c_{LUC} is also very small (<2%), because most of the uncertainties are from the biomass observations (see Discussion) and the regression between E^c_{LUC} and biomass (see r^2 in Figure 3), rather than converting present-day biomass to biomass in 1901 (see r^2 in Figure 4). The difference in regional E^c_{LUC} between different constraint methods is relatively larger (12% on average), but the difference remains very small in tropical regions (~1%). However, we note that the results from the two supplementary methods (Method-S1 and –S2) should be cautiously treated. First, because E^c_{LUC} is related to the biomass that have been affected since the start of the land use perturbation, only biomass in 1901 (rather than that left out of land use in the 2000s) in LULCC-affected grid cells is logically related to historical E^c_{LUC}. Thus, converting present-day biomass to biomass in 1901 (the original method, Figure 1) is a more direct and process-justified approach compared to regressing present-day biomass versus E^c_{LUC} (Method-S1) which is not justified by a logical mechanism. Second, using ΔB in Method-S2 is not a perfect solution to extrapolate biomass in 1901 from present-day biomass, because the change in biomass is not solely impacted by

Formatted: Space Before: 2.5 pt

land use change. The interactions between biomass and climate conditions, disturbances and nutrient limitation are also very important in DGVMs. For example, historical LULCC may reduce biomass over LULCC-affected regions, e.g. by replacing forests with croplands. On the contrary, the CO_2 fertilization effects may increase biomass over LULCC and non-LULCC regions. Therefore, ΔB reflects a mixed effect of different factors, not a sole response to LULCC. In addition, as the ΔB has a higher relative uncertainty between models (~53% at the global level), using the regression ($r^2 > 0.92$ in seven regions, Figure 4) to calculate biomass in 1901 could include relatively less noisy information than using ΔB.

4. Discussion

Our approach to constraining E^c_{LUC} from an ensemble of DGVMs provides a best estimate that is between those from two bookkeeping models (~130 Pg C from Houghton et al. (2012) and 212 Pg C for the default dataset from Hansis et al. (2015)). Although the bookkeeping model from Hansis et al. (2015) was driven by the same agricultural land use maps as the TRENDY models (the model of Houghton et al. (2012) uses FRA/FAO data), E^c_{LUC} from Hansis et al. (2015) is different from that constrained from the DGVMs. Differences of estimates between DGVMs and bookkeeping models have been attributed to different definitions of LULCC emissions (Pongratz et al., 2014; Stocker and Joos, 2015). Indeed, LULCC emissions from DGVM simulations in TRENDY include the "missed sink capacity in the deforested area" (Gasser and Ciais, 2013; Pongratz et al., 2014), and so, all else being equal, should simulate higher emissions than bookkeeping models which do not include this term. However, bookkeeping models take forest degradation into account, while this process is ignored in DGVMs. Bookkeeping models also represent shifting cultivation (resulting in larger sub-grid scale gross land transitions as opposed to net transitions) and wood harvest; processes which are accounted for in only a subset of the TRENDY models (see Table 2). In addition to different driving LULCC area data, differences between the two bookkeeping models were discussed by Hansis et al. (2015); for example, Houghton et al. (2012) assumed a preferential allocation of pastures on natural grasslands, while Hansis et al. (2015) assumed a proportional allocation of both cropland and pasture on all available natural vegetation types.

We are aware that our truncated diagnostic of a set of deforestation grid cells, instead of grid cells affected by all LULCC types, is an under-estimate of the total area subject to LULCC, because we ignore grid cells that experienced land-use transitions between non-forest vegetation only (e.g. only conversions from grasslands to cropland happening in a grid cell). However, the conversion of forest to croplands and pasture dominates the total net LULCC flux (Houghton, 2003, 2010), while the contribution of transitions between non-forest vegetation and agriculture to E^c_{LUC} is comparatively small (Figure S1). In fact, the annual LULCC emission from deforestation was estimated to be 2.2 Pg C yr^{-1} during 1990s, and the total emissions from other activities (e.g., afforestation, reforestation, non-forest transitions) is nearly neutral (Houghton, 2003).

The lack of direct biomass observations at the initial state forces us to hindcast biomass in 1901 based on present-day observations; an extrapolation that also comes with uncertainties. Some of the observed biomass datasets only cover forests,

and satellite measurements usually quantify aboveground biomass carbon stocks and not total biomass stocks (Table 1). In addition, the regression of modeled biomass between 1901 and 2000-2012 (average) to extrapolate the biomass amount in 1901 is only a statistical approach. This regression cannot be mechanistically explained because its slope and intercept are impacted by multiple factors in the models like land clearing, secondary vegetation regrowth, ~~the~~ CO_2 fertilization, climate, disturbances, and the nutrient limitation on biomass. Despite these uncertainties, the high coefficient of determination in the regression increases our confidence in the biomass extrapolation to 1901. For a given biomass dataset, the choice of a method for defining deforestation grid cells (Method-A, Method-B and Method-C) has very small influence on our results (Table 3). ~~Methods of defining deforestation grid cells (Method-A, Method-B and Method-C) to be associated with a certain biomass dataset had very small influence on our results (Table 3).~~

LULCC carbon emissions are influenced not only by changes in biomass, but also by how these are prescribed in the model to influence posterior changes in detrital and soil organic carbon pools. However, LULCC emissions are dominated by changes in biomass. For example, LULCC results in a net carbon loss of 110 Pg C in biomass during 1850-1990, accounting for 89% of the total E^{c}_{LUC} (Houghton, 1999). The soil carbon changes after LULCC is also indirectly impacted by initial biomass, since the dead roots and remaining aboveground debris turn into soil organic carbon after land clearing, which takes longer time to return into the atmosphere. In addition, it is not necessary to account for all factors when applying an emergent constraint approach (e.g. ~~(~~Cox et al., 2013; Kwiatkowski et al., 2017; Wenzel et al., 2016). The regression between E^{c}_{LUC} and biomass in 1901 in models in our study is satisfying (e.g. r^2 = 0.66 at global scale, Figure 3) to constrain E^{c}_{LUC} through biomass observations.

The required model outputs for carbon stocks and fluxes in the TRENDY project are not PFT-specific; only the mean PFT-mixed variables in each grid cell are required. Such an aggregation prevents a rigorous separation of biomass between forest and other biomes in each grid cell. It was thus impossible for us to calculate individual contributions of different LULCC types to the overall LULCC emissions, which induces uncertainties when matching model results with observed forest biomass distributions (e.g. only forest biomass in datasets from GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015) and Pan et al. (2011)). Therefore, we suggest that the next generation of DGVM comparisons reports PFT-specific carbon stock and fluxes, and other model inter-comparison exercises should follow suit. The approach of using multiple biomass observation datasets to constrain the LULCC emissions could also be applied in other modeling projects, such as Coupled Model Intercomparison Project Phase 5 (CMIP5) and CMIP6.

Currently, the uncertainties in the satellite-based biomass datasets are relatively large (e.g. 38% on average in tropics (Saatchi et al., 2011)) at the pixel level (< 1 km). ~~that may~~This introduces uncertainties ~~of~~ in the constrained cumulative LULCC emissions, depending on the forest types and biomass range. For example, on the average at the global scale, the uncertainty at the ~~modeling~~ resolution of DGVM grid cells (0.5° × 0.5°) is about one-third of the mean biomass (Carvalhais et al., 2014) and the relative uncertainty is~~ly~~ smaller for high biomass areas in tropics (Avitabile et al., 2016; Saatchi et al., 2011).

Formatted: Superscript

The main sources of uncertainties in satellite-based biomass datasets depend on the specific product, spatial resolution of the datasets, and the methodology used to validate the data. For instance, in the case of radar remote sensing used for biomass mapping in Northern Hemisphere boreal and temperate forests, the uncertainty is largely due to the sensitivity of the signal to other properties than vegetation structure (e.g. moisture), the influence of non-forest vegetation on the signal (especially in fragmented landscapes; (Santoro et al., 2015)), and uncertainties in additional datasets (allometric databases, land cover) used for conversion of satellite measurements to biomass estimates (Thurner et al., 2014). At the pixel level and modeling grid cells, uncertainties may also be strongly influenced by the quality and size of inventory data used for validation and the significant mismatch between pixel area and the plot data, and the difference of between the dates of satellite and ground observations (Saatchi et al., 2015, 2011; Thurner et al., 2014).

Moreover, the satellite-derived biomass datasets used in this study represent different dates. The tropical biomass products represent circa 2000 status of forests, whereas the boreal and temperate biomass maps are based on spaceborne radar data from the year 2010. These differences in the date of observations introduce additional uncertainty in the biomass estimates due to changes of forest cover from disturbance and recovery and land use activities (Hurtt et al., 2011) occurring annually and regionally.

However, in boreal/temperate as well as in tropical regions, the estimated relative uncertainties were lowest in high biomass areas (Avitabile et al., 2016; Thurner et al., 2014), which dominates the contribution to our results. Moreover, the relatively high accuracy of biomass datasets when aggregated to modeling grid cells from higher resolution maps (< 1 km) (Saatchi et al., 2011; Thurner et al., 2014) suggest that biomass datasets implemented in our study provide a realistic representation of carbon stocks to constrain the historical cumulative LULCC emissions from vegetation.

5. Conclusions

Uncertainties in LULCC carbon emissions are relatively large, compared to other terms in the global carbon budget. The wide spread is partly due to the differences in model structure but also because of the difficulty in constraining models by observations of LULCC, particularly emissions resulting from deforestation. ~~We use satellite and inventory-based biomass observations to constrain simulated historical LULCC emissions based on relationships between simulated vegetation biomass and cumulative LULCC emissions in nine DGVMs.~~ We propose an observationally constrained global cumulative LULCC emission of 155 ±50 Pg C during 1901 and 2012. Although the constrained cumulative LULCC emissions are close to the unconstrained ones from models, our study offers an evaluation of the modeling results using the observation-based biomass. More importantly, we combine the uncertainties in the regressions from state-of-the-art models with uncertainties in multiple observation-based biomass datasets and give a constrained E_{LUC}^{c} with a 1-σ Gaussian uncertainty. The idea of an emergent constraint approach is to give a more accurate estimate and / or a reduced uncertainty on an unknown variable by combining a heuristic relationship between two modeled variables (an observable and an unknown one) with actual

observations of the observable variable. Thus, our study shows 1) that there is a heuristic relationship between initial biomass and E_{LUC}^{c} among different models, 2) that available biomass observation data confirm independently the median of modeled emission estimates, and 3) that more accurate biomass data in the future would allow to falsify some of the modeled estimates of emissions. Although the uncertainties in current observation-based biomass datasets are relatively high, as more accessible and accurate observation data become available, many data-driven opportunities are being created to improve the accuracy of DGVM predictions.

Acknowledgements

W.L., C.Y., T.A.M.P. and A.A. were supported by the European Commission-funded project LUC4C (grant No. 603542). P.C. and S.P. acknowledge support from the European Research Council through Synergy grant ERC-2013-SyG-610028 "IMBALANCE-P". J.P., J.N., and R.Y. were supported by the German Research Foundation's Emmy Noether Program (PO 1751/1-1). B.D.S. was supported by the Swiss National Science Foundation and FP7 funding through project EMBRACE (282672). A.H. was supported by the UK Natural Environment Research Council Joint Weather and Climate Research Programme. M.T. acknowledges funding from the Vetenskapsrådet grant 621-2014-4266 of the Swedish Research Council. The biomass maps and model outputs can be freely accessed following instructions in the original publications. All the biomass constrained LULCC emission data are freely obtained from WL (email: wei.li@lsce.ipsl.fr).

References

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L. F., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M. J., Ryan, C. M., Ferry, S. J. W., Sunderland, T., Laurin, G. V., Gatti, R. C., Valentini, R., Verbeeck, H., Wijaya, A., Willcock, S., Asthon, P., Banin, L. F., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M. J., Ryan, C. M., Slik, F., Sunderland, T., Vaglio Laurin, G., Valentini, R., Verbeeck, H., Wijaya, A. and Willcock, S.: An integrated pan-tropical biomass map using multiple reference datasets, Global Chang. Biol., 22(4), 1406–1420, doi:10.1111/gcb.13139, 2016.

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P. S. A., Dubayah, R., Friedl, M. A., Samanta, S. and Houghton, R. A.: Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps, Nat. Clim. Chang., 2(3), 182–185, doi:10.1038/nclimate1354, 2012.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. . L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R.

J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4(3), 677–699, doi:10.5194/gmd-4-677-2011, 2011.

Boden, T. A., Marland, G. and Andres, R. J.: Global, regional, and national fossil-fuel CO_2 emissions, Carbon Dioxide Inf. Anal. Center, Oak Ridge Natl. Lab. USA. Oak Ridge, TN Dep. Energy, doi:10.3334/CDIAC/00001, 2013.

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T. and Reichstein, M.: Global covariation of carbon turnover times with climate in terrestrial ecosystems., Nature, 514(7521), 213–7, doi:10.1038/nature13731, 2014.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C. and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4(3), 701–722, doi:10.5194/gmd-4-701-2011, 2011.

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D. and Luke, C. M.: Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability, Nature, 494(7437), 341–344, doi:http://www.nature.com/nature/journal/v494/n7437/abs/nature11882.html#supplementary-information, 2013.

Gasser, T. and Ciais, P.: A theoretical framework for the net land-to-atmosphere CO_2 flux and its implications in the definition of "emissions from land-use change," Earth Syst. Dyn., 4(1), doi:10.5194/esd-4-171-2013, 2013.

Hansis, E., Davis, S. J. and Pongratz, J.: Relevance of methodological choices for accounting of land use change carbon fluxes, Global Biogeochem. Cycles, 29(8), 1230–1246, doi:10.1002/2014GB004997, 2015.

Harris, N. L., Brown, S., Hagen, S. C., Saatchi, S. S., Petrova, S., Salas, W., Hansen, M. C., Potapov, P. V and Lotsch, A.: Baseline map of carbon emissions from deforestation in tropical regions., Science, 336(6088), 1573–6, doi:10.1126/science.1217962, 2012.

Houghton, R. A.: The annual net flux of carbon to the atmosphere from changes in land use 1850-1990, Tellus B, 51(2), 298–313, doi:10.1034/j.1600-0889.1999.00013.x, 1999.

Houghton, R. A.: Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850-2000, Tellus Ser. B-Chemical Phys. Meteorol., 55(2), 378–390, doi:10.1034/j.1600-0889.2003.01450.x, 2003.

Houghton, R. A.: How well do we know the flux of CO_2 from land-use change?, Tellus, Ser. B Chem. Phys. Meteorol., 62(5), 337–351, doi:10.1111/j.1600-0889.2010.00473.x, 2010.

Houghton, R. a., House, J. I., Pongratz, J., Van Der Werf, G. R., Defries, R. S., Hansen, M. C., Le Quéré, C. and Ramankutty, N.: Carbon emissions from land use and land-cover change, Biogeosciences, 9(12), 5125–5142, doi:10.5194/bg-9-5125-2012, 2012.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S.,

Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P. and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Clim. Change, 109(1–2), 117–161, doi:10.1007/s10584-011-0153-2, 2011.

Ito, A. and Inatomi, M.: Use of a process-based model for assessing the methane budgets of global terrestrial ecosystems and evaluation of uncertainty, Biogeosciences, 9(2), 759–773, doi:10.5194/bg-9-759-2012, 2012.

Kato, E., Kinoshita, T., Ito, A., Kawamiya, M. and Yamagata, Y.: Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model, J. Land Use Sci., 8(1), 104–122, doi:10.1080/1747423X.2011.628705, 2013.

Klein Goldewijk, K., Beusen, A., Van Drecht, G. and De Vos, M.: The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years, Global Ecol. Biogeogr., 20(1), 73–86, doi:10.1111/j.1466-8238.2010.00587.x, 2011.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S. and Prentice, C. I.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cycles, 19(1), doi:10.1029/2003GB002199, 2005.

Kwiatkowski, L., Bopp, L., Aumont, O., Ciais, P., Cox, P. M., Laufkötter, C., Li, Y. and Séférian, R.: Emergent constraints on projections of declining primary production in the tropical oceans, Nat. Clim. Chang., 7(5), 355–358, doi:10.1038/nclimate3265, 2017.

Liu, Y. Y., van Dijk, A. I. J. M., de Jeu, R. A. M., Canadell, J. G., McCabe, M. F., Evans, J. P. and Wang, G.: Recent reversal in loss of global terrestrial biomass, Nat. Clim. Chang., 5(5), 470–474, doi:10.1038/nclimate2581, 2015.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R. A., Heald, C. L., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W. J., Sun, Y., Tang, J. and Yang, Z.-L.: Technical Description of version 4.5 of the Community Land Model (CLM)., 2013.

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(6045), 988–993, doi:10.1126/science.1201609, 2011.

Peng, S., Ciais, P., Maignan, F., Li, W., Chang, J., Wang, T. and Yue, C.: Sensitivity of land use change emission estimates to historical land use and land cover mapping, Global Biogeochem. Cycles, 31(4), 626–643, doi:10.1002/2015GB005360, 2017.

Pitman, A. J., de Noblet-Ducoudré, N., Cruz, F. T., Davin, E. L., Bonan, G. B., Brovkin, V., Claussen, M., Delire, C., Ganzeveld, L., Gayler, V., van den Hurk, B. J. J. M., Lawrence, P. J., van der Molen, M. K., Müller, C., Reick, C. H., Seneviratne, S. I., Strengers, B. J. and Voldoire, A.: Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study, Geophys. Res. Lett., 36(14), L14814, doi:10.1029/2009GL039076, 2009.

Pongratz, J., Reick, C. H., Houghton, R. a. and House, J. I.: Terminology as a key uncertainty in net land use and land cover change carbon flux estimates, Earth Syst. Dyn., 5(1), 177–195, doi:10.5194/esd-5-177-2014, 2014.

Le Quéré, C., Moriarty, R., Andrew, R. M., Peters, G. P., Ciais, P., Friedlingstein, P. and Jones, S. D.: Global carbon budget 2014, Earth Syst. Sci. Data, 47–85, doi:10.5194/essd-7-47-2015, 2015.

Reick, C. H., Raddatz, T., Brovkin, V. and Gayler, V.: Representation of natural and anthropogenic land cover change in MPI-ESM, J. Adv. Model. Earth Syst., 5(3), 459–482, doi:10.1002/jame.20022, 2013.

Saatchi, S., Mascaro, J., Xu, L., Keller, M., Yang, Y., Duffy, P., Espírito-Santo, F., Baccini, A., Chambers, J. and Schimel, D.: Seeing the forest beyond the trees, Global Ecol. Biogeogr., 24(5), 606–610, doi:10.1111/geb.12256, 2015.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M. and Morel, A.: Benchmark map of forest carbon stocks in tropical regions across three continents., Proc. Natl. Acad. Sci. USA, 108(24), 9899–904, doi:10.1073/pnas.1019576108, 2011.

Santoro, M., Beaudoin, A., Beer, C., Cartus, O., Fransson, J. E. S., Hall, R. J., Pathe, C., Schmullius, C., Schepaschenko, D., Shvidenko, A., Thurner, M. and Wegmüller, U.: Forest growing stock volume of the northern hemisphere: Spatially explicit estimates for 2010 derived from Envisat ASAR, Remote Sens. Environ., 168, 316–334, doi:10.1016/j.rse.2015.07.005, 2015.

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., Bopp, L., Canadell, J. G., Chevallier, F., Ciais, P., Ellis, R., Gloor, M., Peylin, P., Piao, S. L., Le Quéré, C., Smith, B., Zhu, Z. and Myneni, R.: Recent trends and drivers of regional sources and sinks of carbon dioxide, Biogeosciences, 12(3), 653–679, doi:10.5194/bg-12-653-2015, 2015.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K. and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global Chang. Biol., 9(2), 161–185, doi:10.1046/j.1365-2486.2003.00569.x, 2003.

Smith, B., Prentice, I. C. and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecol. Biogeogr., 10(6), 621–637, 2001.

Stegehuis, A. I., Teuling, A. J., Ciais, P., Vautard, R. and Jung, M.: Future European temperature change uncertainties reduced by using land heat flux observations, Geophys. Res. Lett., 40(10), 2242–2245, doi:10.1002/grl.50404, 2013.

Stocker, B. D., Feissli, F., Strassmann, K. M., Spahni, R. and Joos, F.: Past and future carbon fluxes from land use change, shifting cultivation and wood harvest, Tellus B, 66, doi:10.3402/tellusb.v66.23188, 2014.

Stocker, B. D. and Joos, F.: Quantifying differences in land use emission estimates implied by definition discrepancies, Earth Syst. Dyn., 6(2), 731–744, doi:10.5194/esd-6-731-2015, 2015.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R. and Schmullius, C.: Carbon stock and density of northern boreal and temperate forests, Global Ecol. Biogeogr., 23(3), 297–310, doi:10.1111/geb.12125, 2014.

Wenzel, S., Cox, P. M., Eyring, V. and Friedlingstein, P.: Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO2, Nature, 538(7626), 499–501, doi:10.1038/nature19772, 2016.

Wilkenskjeld, S., Kloster, S., Pongratz, J., Raddatz, T. and Reick, C. H.: Comparing the influence of net and gross anthropogenic land-use and land-cover changes on the carbon cycle in the MPI-ESM, Biogeosciences, 11(17), 4817–4828, doi:10.5194/bg-11-4817-2014, 2014.

Zaehle, S. and Friend, A. D.: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates, Global Biogeochem. Cycles, 24(1), doi:10.1029/2009GB003521, 2010.

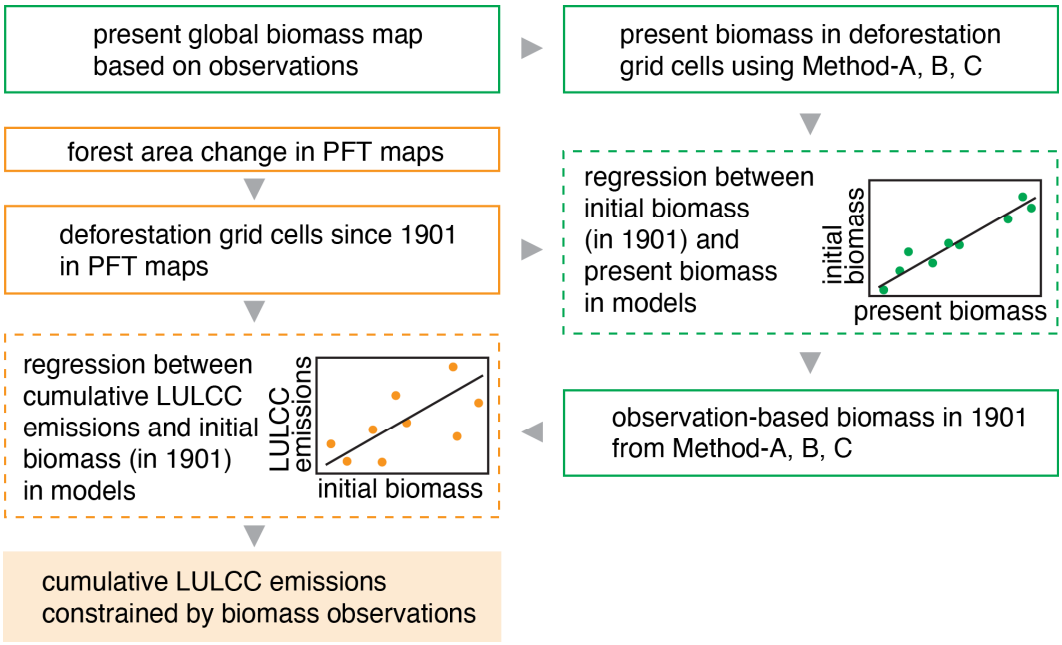

Figure 1. The framework of this study.

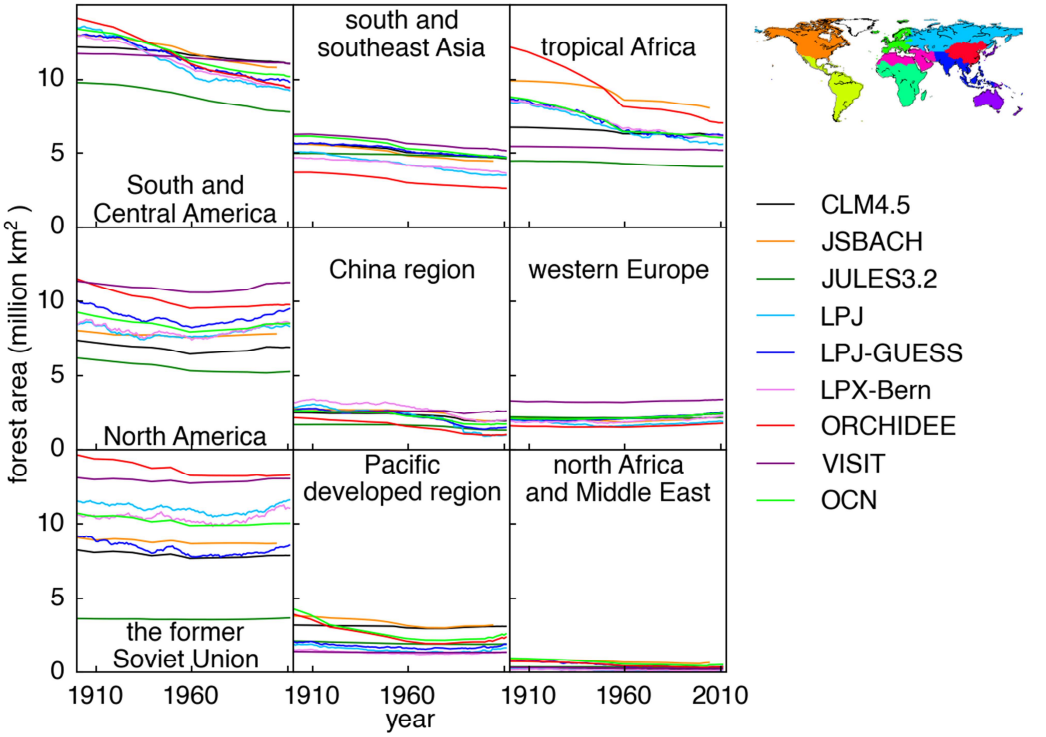

Figure 2. Temporal change of forest area from TRENDY v2 models in each of the nine regions. Differences between models arise from their specific vegetation maps and rules through which natural PFTs are chosen to give land to agriculture.

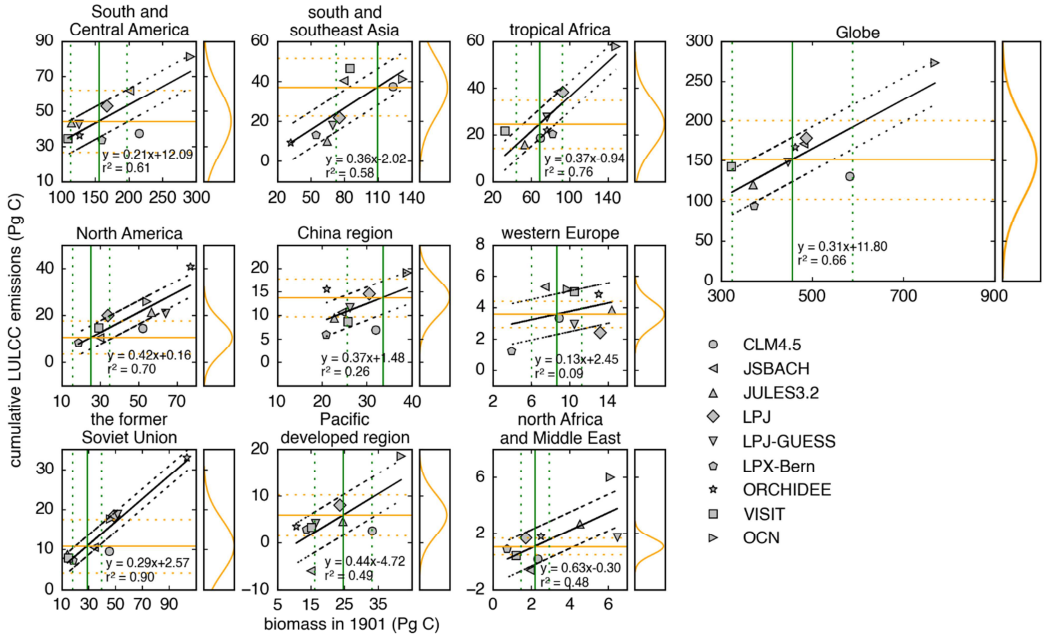

Figure 3. Relationship between biomass in 1901 and cumulative land-use and land-cover change (LULCC) emissions during 1901-2012 across the nine TRENDY v2 models. The black solid line is the linear regression line. The vertical green solid line indicates the reconstructed biomass in 1901 from Carvalhais et al. (2014) by applying Method-A (all the increase of cropland in HYDE v3.1 data from forest, see Figs. S4 and S5 for results of Method-B and Method-C) to define deforestation grid cells. The orange solid horizontal line indicates the cumulative LULCC emissions constrained by reconstructed biomass in 1901. Dashed lines represent 1-σ uncertainties. The probability density function of the constrained cumulative LULCC emissions is shown on the right.

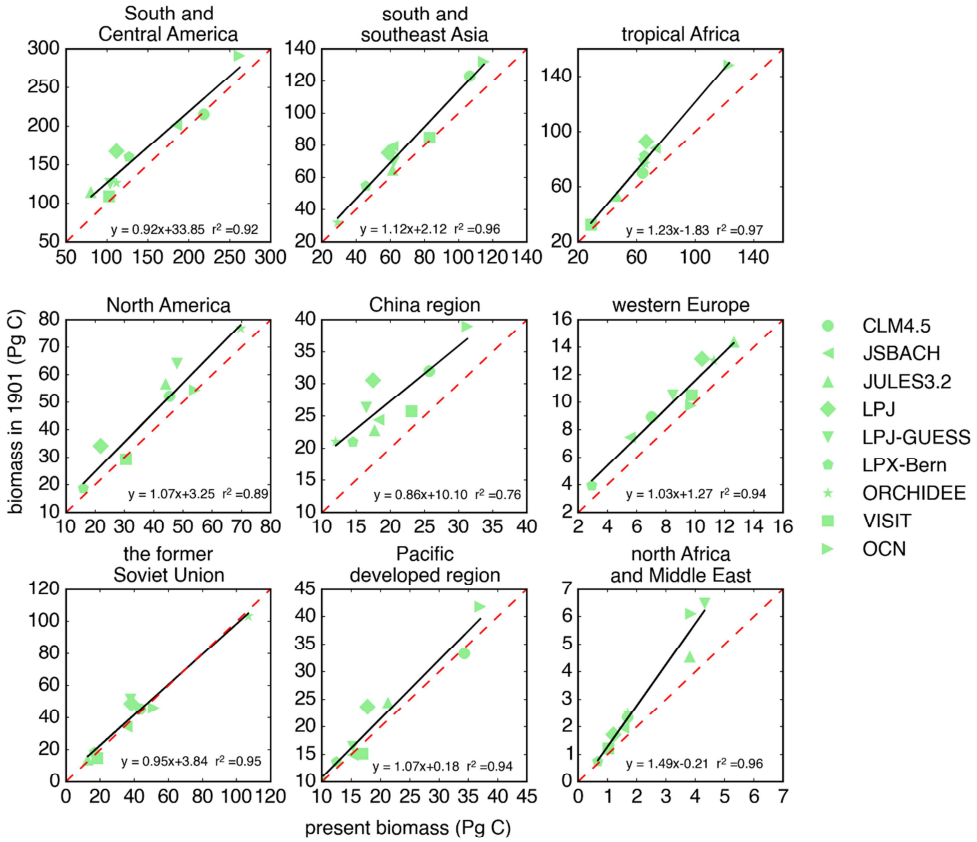

Figure 4. The relationship between initial biomass in 1901 and present biomass (average of biomass from 2000 to 2012) across the TRENDY v2 models for each region. Note that both biomass in 1901 and present biomass are from TRENDY models, not the observations. Dashed line is the 1:1 line.

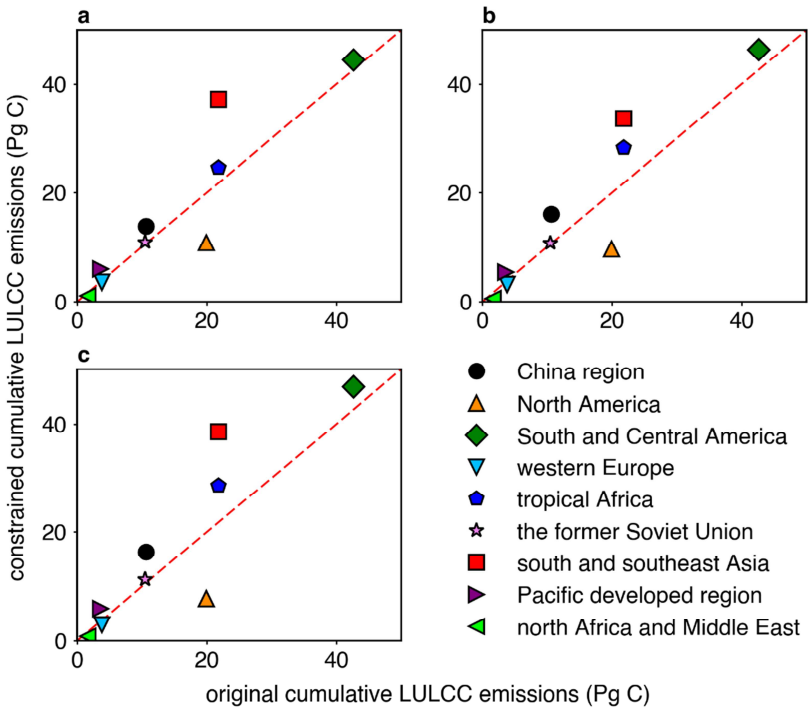

Figure 5. Comparisons between the original TRENDY land-use and land-cover change (LULCC) emissions and the cumulative LULCC emissions constrained by biomass dataset from Carvalhais et al. (2014). (a), (b) and (c) are results from Method-A, Method-B, and Method-C, respectively. The original TRENDY emissions are shown as the median value of all models. ~~Solid and dotted lines represent the 1-σ Gaussian error after biomass constraint and the original modeled range.~~ Dashed line is the 1:1 line.

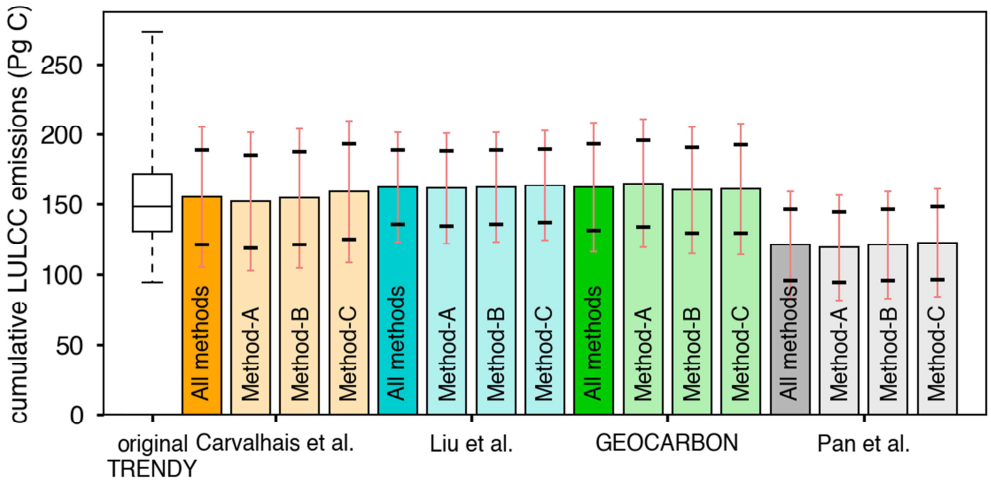

Figure 6. The global cumulative land-use and land-cover change (LULCC) emissions during 1901-2012 from original TRENDY models and from the estimates constrained by different biomass datasets with different methods to define deforestation grid cells. "All methods" represents the ensemble mean and uncertainty of the constrained results from "Method-A", "Method-B" and "Method-C" for each biomass dataset. The whisker-box plot represents the minimum and maximum values, 25th and 75th percentiles and the median value of original TRENDY models. In the bar plot for the constrained estimates, ~~T~~the red error ~~bars in the bar plot~~ represent the 1-σ Gaussian errors; the black ticks represent the 25th and 75th percentiles.

Table 1. The different biomass datasets based on observations. The information of biomass from TRENDY v2 project is also listed for comparison.

dataset	coverage	resolution	biome type	aboveground / belowground	note
Thurner et al. (Thurner et al., 2014)	30°N~80°N	0.01 degree	forest	aboveground + belowground	Growing stock volume from Santoro et al. (Santoro et al., 2015)
Saatchi et al. (Saatchi et al., 2011)	30°N – 40°S	1 km	forest	aboveground	
Carvalhais et al. (Carvalhais et al., 2014)	global (without South Australia)	0.5 degree	forest + herbaceous	aboveground + belowground	a merged map of Thurner et al. (Thurner et al., 2014) and Saatchi et al. (Saatchi et al., 2011)
Baccini et al. (Baccini et al., 2012)	23°N – 23°S	500 m	forest	aboveground	
Liu et al. (Liu et al., 2015)	global	0.25 degree	all	aboveground	calibration based on Saatchi et al. (Saatchi et al., 2011)
Avitabile et al. (Avitabile et al., 2016)	30°N – 40°S	1 km	forest	aboveground	a fusion of Saatchi et al. (Saatchi et al., 2011) and Baccini et al. (Baccini et al., 2012)
Santoro et al. (Santoro et al., 2015)	30°N+	0.01 degree	forest	aboveground	sharing growing stock volume with Thurner et al. (Thurner et al., 2014)
Pan et al. (Pan et al., 2011)	global	regional	forest	aboveground + belowground	based on FAO data
TRENDY v2	global	1 degree	all	aboveground + belowground	

Table 2. Description of TRENDY model set-ups used in this study.

model	PFT number	allocation rules of changes in agriculture area	spatial resolution	dynamic vegetation activated	wood harvest	shifting agriculture	explicit gross LULCC transitions	start of transient simulation	reference
CLM4.5	17	pasture; crop model exists but not used in these simulations	0.9° ×1.25°	no	yes	no	no	1860	(Oleson et al., 2013)
JSBACH	12	proportional allocation of cropland; preferential allocation of pasture on natural grassland	T63[a]	no	yes	yes	yes	1850	(Reick et al., 2013)
JULES3.2	5	crop and pasture added together to create a single agricultural mask, where trees and shrubs are excluded from growing. There is no assumption for which PFTs the agriculture replaces	N96[b]	yes	yes	no	no	1860	(Clark et al., 2011); (Best et al., 2011)
LPJ	9	crop and pasture were added together to create a single managed lands fraction	0.5° × 0.5°	yes	no	no	no	1860	(Sitch et al., 2003)
LPJ-GUESS	11	proportional allocation of cropland and pasture	0.5° × 0.5°	yes	no	no	no	1860	(Smith et al., 2001)
LPX-Bern	9	proportional allocation of cropland and pasture	1.0° × 1.0	yes	no	no	no	1860	(Stocker et al., 2014)
ORCHIDEE	13	proportional allocation of cropland and pasture	2° × 2°	no	no	no	no	??	(Krinner et al., 2005)
VISIT	16	no specific rule applied because only 1 natural PFT exists for primary and secondary land in a grid cell	0.5° × 0.5°	no	yes	yes	yes	1860	(Kato et al., 2013); (Ito and Inatomi, 2012)
OCN	12	proportional allocation of cropland and pasture	2.5° x 3.75°	no	no	no	no	1860	(Zaehle and Friend, 2010)

[a]T63 grid has an approximate resolution of 1.9° × 1.9°.

[b]N96 resolution is equivalent to 1.25° latitude × 1.875° longitude.

Formatted Table

Table 3. The global and regional cumulative land-use and land-cover change (LULCC) emissions (Pg C) during 1901-2012 from original TRENDY models and from the estimates constrained by different biomass datasets with different methods to define deforestation grid cells. The interquantile ranges are shown in Table S1.

	TRENDY			Carvalhais et al. (Carvalhais et al., 2014)			Liu et al. (Liu et al., 2015)			GEOCARBON (Avitabile et al., 2016; Santoro et al., 2015)			Pan et al. (Pan et al., 2011)		
	median	min	max	Method-A	Method-B	Method-C	Method-A	Method-B	Method-C	Method-A	Method-B	Method-C	Method-A	Method-B	Method-C
China region	10.7	6.0	19.1	13.8±4.0	16.0±4.3	16.3±4.3	10.5±3.1	11.1±3.1	11.1±3.1	10.0±4.0	10.5±4.5	10.5±4.6	7.3±2.9	7.6±2.9	7.7±2.9
North America	19.9	8.6	40.8	10.8±7.1	9.6±7.0	7.8±6.7	14.7±6.9	13.6±6.8	9.3±6.6	17.8±8.3	15.4±7.7	13.0±7.6	9.5±6.4	8.5±6.4	6.7±6.4
South and Central America	42.6	33.5	81.4	44.4±17.8	46.4±18.1	46.8±18.2	48.3±17.0	50.1±17.0	50.6±17.0	44.5±16.6	46.4±16.7	46.8±16.8	43.1±17.0	44.8±17.1	45.1±17.2
western Europe	3.8	1.2	5.3	3.6±0.8	3.2±0.8	3.0±0.8	4.1±0.8	3.4±0.8	3.2±0.8	5.0±1.2	3.8±0.9	3.4±0.8	3.6±0.8	3.2±0.8	3.0±0.8
tropical Africa	21.8	15.8	57.8	24.6±10.3	28.2±11.4	28.6±11.5	31.4±8.8	36.3±9.4	36.9±9.4	22.7±12.8	26.2±14.6	26.4±16.1	23.8±10.3	27.5±11.4	27.8±11.5
the former Soviet Union	10.5	7.2	33.0	10.9±6.7	10.7±6.7	11.3±6.8	14.2±6.5	14.3±6.5	14.9±6.5	14.7±6.6	13.0±6.5	13.5±6.5	10.1±6.2	9.8±6.2	10.2±6.2
south and southeast Asia	21.8	9.6	46.6	37.2±14.4	33.6±13.3	38.5±14.8	27.8±8.5	24.1±7.9	27.9±8.3	32.9±10.1	29.0±9.6	33.6±10.4	15.1±9.5	13.3±8.9	15.5±9.6
Pacific developed region	3.6	-6.0	18.6	6.0±4.4	5.4±4.2	6.0±4.3	7.1±3.5	6.0±3.4	5.6±3.3	18.0±3.1	16.4±2.9	14.3±3.0	-1.7±2.6	-1.9±2.6	-2.0±2.6
north Africa and Middle East	1.7	-0.6	6.0	1.1±0.6	0.6±0.6	0.8±0.6	4.3±1.1	3.1±0.8	3.5±0.9	4.5±4.5	3.0±3.0	3.1±3.7	-0.1±0.5	-0.2±0.5	-0.1±0.5
globe	148	94	273	152±49	154±50	159±51	161±40	162±39	163±39	165±46	160±45	161±47	119±37	121±38	122±38

5