Global Rules for Translating Land-use Change (LUH2) To Land-cover Change for CMIP6 using GLM2

Lei Ma[1], George C. Hurtt[1], Louise P. Chini[1], Ritvik Sahajpal[1], Julia Pongratz[2], Steve Frolking[3], Elke Stehfest[4], Kees Klein Goldewijk[4,5], Donal O' Leary[1], Jonathan C. Doelman[4]

[1]Department of Geographical Sciences, University of Maryland, College Park, MD, USA

[2]Department of Geography, Ludwig-Maximilians-Universität, 80333 München, Germany and Max Planck Institute for Meteorology, Bundesstr. 53, 20143 Hamburg, Germany

[3]Institute of the Study of Earth Oceans and Space, University of New Hampshire, Durham, NH, USA

[4]PBL Netherlands Environmental Assessment Agency, The Hague, the Netherlands

[5]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, P.O. Box 80115, the Netherlands

Correspondence to: Lei Ma (lma6@umd.edu)

Abstract.

Anthropogenic land-use and land-cover change activities play a critical role in Earth system dynamics through significant alterations to biogeophysical and biogeochemical properties at local to global scales. To quantify the magnitude of these impacts, climate models need consistent land-cover change time-series at a global scale, based on land-use information from observations or dedicated land-use change models. However, a specific land-use change cannot be unambiguously mapped to a specific land-cover change. Here, various translation rules are evaluated based on assumptions about the way land-use change could potentially impact land-cover. Utilizing the Global Land use Model 2 (GLM2), the model underlying the latest Land Use Harmonization dataset (LUH2), the land-cover dynamics resulting from land-use change were simulated based on multiple alternative translation rules from 850 to 2015 globally. For each rule, the resulting forest cover, carbon density, and carbon emissions were compared with independent estimates from remote sensing observations, U.N. Food and Agricultural Organization reports, and other studies. Examinations at global, country, and grid scales indicate that the recommended translation rule for CMIP6 models is 1) completely clear vegetation in land-use changes from primary and secondary land (including both forested and non-forested) to cropland, urban land, and managed pasture; 2) completely clear vegetation in land-use changes from primary forest and/or secondary forest to rangeland; 3) keep vegetation in land-use changes from primary non-forest and/or secondary non-forest to rangeland. This confirms the translation rules suggested earlier in the HYDE dataset underlying LUH2. According to this rule, contemporary global forest area is estimated to be 37.42 10^6 km^2, and forest area estimates at global and country scales both stay within the range derived from remote sensing products. Likewise, the estimated carbon stock is in close agreement with reference biomass datasets, particularly over regions with 50% forest cover.

This rule also mitigates the anomalously high carbon emissions from land-use change observed in previous studies in the 1950s.

1 Introduction

Historical land-use activities have been significantly affecting the global carbon budget in both direct and indirect ways, and changing Earth's climate through altering land surface properties (e.g. surface albedo, surface aerodynamic roughness, and forest cover) (Betts, 2007; Bonan, 2008; Brovkin et al., 2006; Claussen et al., 2001; Feddema et al., 2005; Guo and Gifford, 2002; Pongratz et al., 2010; Post and Kwon, 2000). It has been estimated that, during the past 300 years, >50% of the land surface has been affected by human land-use activities, >25% of forest has been permanently cleared, and 10-44 10_6 km_2 of land are recovering from previous human land-use disturbances (Hurtt et al., 2006). Impacts on the carbon cycle result from several processes among others: deforestation removes natural forest and its corresponding carbon biomass is used for wood products, burning, or decay by microbial decomposition (DeFries et al., 2002). Afforestation/reforestation, in contrast, recovers forest which accumulates carbon but sequestration potential are constrained by water and nutrient availability (Smith and Torn, 2013). Wood harvesting is one of the largest source contributing gross carbon emission by modifying the litter input into various soil pools, stand age, and biomass of secondary forest (Dewar, 1991; Hurtt et al., 2011; Nave et al., 2010) . Cumulatively, models estimate that land-use and land-use change have contributed to a net flux 190±75 Pg C to the atmosphere during 1870-2017 (Le Quéré et al., 2018). While emissions from land-use and land-use change only account for 10% of current anthropogenic carbon emissions, they were a dominant contributor to increasing the atmospheric CO_2 above pre-industrial levels before 1920 (Ciais et al., 2014).

Quantification of historical Land-Use and Land-Cover Change (LULCC) is important because it serves as the basis for examining the role of human activities in the global carbon budget and the resulting impacts to Earth's climate system. For this purpose, LULCC reconstructions enter Earth System Models (ESMs) (Lawrence et al., 2016), Dynamic Global Vegetation Models (DGVMs) (Le Quéré et al., 2018) and bookkeeping models (Hansis et al., 2015) to quantify biogeochemical and biophysical impacts of historical land-use change as part of historical simulates (DECK and CMIP6 historical simulations), future projections (scenarioMIP), impacts studies (ISIMIP), paleoclimate studies (PMIP), land-use specific simulations (LUMIP), and biodiversity studies (IPBES). Considerable efforts have been devoted to modelling historical land-use states (Goldewijk et al., 2017; Kaplan et al., 2009; Pongratz et al., 2008; Ramankutty and Foley, 1999) and land-use transitions (Houghton, 1999; Hurtt et al., 2006, 2011). In particular, the recent Land-Use Harmonization 2 (LUH2) dataset (Hurtt et al., 2017) has been developed to provide global gridded land-use states and transitions in a consistent format for use in ESMs as part of CMIP6 experiments. However, large uncertainties still exist in the carbon/climate studies based on many of the above LULCC products (Chini et al., 2012; Houghton et al., 2012; Pongratz et al., 2014). For example, the Global Carbon Budget reports the spread of cumulative LULCC carbon emission during 1870-2017 estimated by DGVMs is as large as 75 Pg C

though all models are forced by the LUH2 (Le Quéré et al., 2018). LULCC carbon emissions in CMIP5 have an anomalous spike during the years 1950-1960. These anomalous emission estimates by ESMs (hereinafter referred to as the "pasture anomaly") are caused by an implausible high conversion rate of natural and secondary vegetation to pasture, with the 1950s having double the conversion rate of the 40's or 60's. Because of this, the simulated terrestrial land flux has a two decade delay in the switch from a land carbon source to a land carbon sink compared to observations (Shevliakova et al., 2013).

One reason for the above uncertainties is the lack of a globally consistent rule that translate land-use change estimates into land-cover changes, which is critical for ESM models (Brovkin et al., 2013; Di Vittorio et al., 2018, 2014; de Noblet-Ducoudré et al., 2012). Although land-use changes are generally associated with a change in land-cover and carbon stocks, these two changes are not always equivalent (see Figure 1 in (Pongratz et al., 2018)), and the degree of land-cover alteration varies with the types of land-use changes and the location where a land-use change happens. For example, the conversion from forested land to managed pasture and/or cropland tends to be associated with the full removal of native vegetation due to intensive human management, whereas vegetation may be less disturbed during the land conversion from non-forest (e.g. grassland) to rangeland. To enable the inclusion of such land-cover change processes, the HYDE 3.2 dataset has redefined the former pasture category used in CMIP5 into the two sub-categories of "managed pasture" and "rangeland" (with the total being termed "grazing land"). This redefinition intends to suggest different treatments of vegetation and carbon removal in ESMs and DGVMs for these two types of land-use changes (Goldewijk et al., 2017). However, explicit suggestions for land-cover and carbon stock modifications resulting from these new defined land-use types are not yet provided, but are crucial for the translation of land-use change to land-cover change within ESMs or DGVMs. An inconsistent land-cover translation of these land-use products within an ESM or DGVM will potentially produce very different land-cover dynamics, which will impact the land surface biophysical and biochemical processes. Therefore, a globally consistent rule for translating land-use products to land-cover change could eliminate added uncertainties from translation inconsistency in studying land-use effects through ESMs and DGVMs.

To recommend a global translation rule for translating historical land-use changes for CMIP6 models, this study investigates the impacts of land-use change on land-cover by proposing several alternative sets of translation rules, which ae then integrated into the Global Land use Model 2 (GLM2) model (Hurtt et al., 2017, 2019) to simulate the forest cover and carbon dynamics. These simulations are then evaluated against estimates of contemporary forest cover and carbon density from remote sensing observations, and the resulting cumulative LULCC carbon emissions are compared with a range of independent estimates. This recommended rule combined with LUH2 could improve estimates of forest area and carbon stock at global, country and grid-cell scales when compared to remote sensing data and reduce the 1950s pasture anomaly.

2 Methodology

In this study, two key land-cover properties (i.e. forest cover and vegetation carbon) are simulated by combining historical land-use change with translation rules. The historical land-use change information is specified by the LUH2 dataset (v2h, available at http://doi.org/10.22033/ESGF/input4MIPs.1127) which serves as the forcing data for a new generation of advanced ESMs as part of CMIP6. Section 2.1 describes the details of land-use change characterization, and section 2.2 defines each translation rule. The resulting forest cover and vegetation carbon is tracked at each grid cell (0.25×0.25˚) for the year 850 to 2015 using methods described in section 2.3 and 2.4. The simulated forest cover and vegetation carbon are then compared with multiple published datasets of land-cover , carbon stock, and estimates of land-use change emission (see details in section 2.5).

2.1 Land-use change characterization

The LUH2 dataset was generated with the GLM2 (Hurtt et al., 2017, 2019), which like its predecessors (Hurtt et al., 2006, 2011), estimates annual sub-grid-cell land-use states and transitions by including multiple constraints such as gridded patterns of historical land-use from the HYDE database (Goldewijk et al., 2017), historical national wood harvest reconstructions, potential biomass and recovery rates, and others. Building upon previous work from CMIP5, for which the original LUH1 dataset was used, LUH2 has extended the timespan to 850-2100 and increased spatial resolution to 0.25×0.25˚. In addition, LUH2 includes 12 different land-use types (i.e. forested and non-forested primary and secondary land, cropland of C3 annual, C3 perennial, C4 annual, C4 perennial and C3 nitrogen-fixing, urban, managed pasture and rangeland) and includes transitions between all combinations of these categories.

In LUH2, “primary” refers to land previously undisturbed by any human activities since 850AD, while “secondary” refers to land undergoing a transition or recovering from previous human activities. Global secondary land area was specified as zero in 850. Note that primary and secondary lands are further sub-divided into forested and non-forested grids using a definition based on the potential aboveground biomass density (forested land requiring an aboveground biomass density ≥2 kg C/m$_2$).

2.2 Translation rules

Nine translation rules are proposed (Table 1) to analyse the effects of land-use change on land-cover dynamics, whereby each rule differs in treatment of vegetation cover and vegetation carbon stock during land-use changes. Rules 1-4 all assume complete clearance of vegetation for cropland and vary on vegetation clearance for managed pasture and rangeland. The rules 5-9 are added for analytical purposes, rather than as realistic possibilities. For example, Rule 3 presumes all land-use changes alter land-cover and reduce carbon stock, and this rule would produce the least global forest cover and carbon stock. Rule 1 and 3 differ in treatment of vegetation in non-forested land when converted to rangeland, and the resulting difference between their carbon stocks indicate the impact of rangeland expansion on non-forests, and also tests whether the disaggregation of

grazing land into managed pasture and rangeland will address the pasture anomaly issue in 1950-1960. Rule 1 (clearance of all vegetation for cropland and managed pasture, and only forest clearance for rangeland) is in fact the rule suggested in the underlying HYDE dataset and its distinction between pasture and rangeland (Goldewijk et al., 2017). For simplicity, we do not consider partial removal of vegetation in this study; vegetation is either fully removed or fully remains as these land-cover transitions represent the maximum and minimum bounds for land-cover alteration. In this study, the translation rules are applied to all regions and are constant across the whole simulation period. Although the impacts of land-use change on land-cover may vary in different regions, the discussion of region-varied and time-varied translation rules is beyond the scope of this study.

It is important to note that these nine rules are not equally realistic, and the purpose of including Rules 5-9 is to investigate individual or joint contributions of cropland, managed pasture and rangeland expansion on forest and carbon. For example, forest and carbon dynamic resulting from Rule 6 could suggest individual impact of cropland expansion.

2.3 Simulation of land-cover change

In this study, land-cover change is simulated by performing a modified GLM2 simulation in which the computed land-use transition rates (using the same methodology as LUH2) are supplemented with a set of translation rules (Table 1) to track forest cover change and carbon dynamics at 0.25º spatial resolution. GLM2 uses a statistical model to estimate ecosystem stocks and fluxes with temperature and precipitation as inputs (see (Hurtt et al., 2002) for details). Climatological temperature and precipitation during 1901-2000 were produced from the MSTMIP (Wei et al., 2014) and used to spin up the GLM2 globally at 0.25×0.25˚ resolution for 500 years. The climatology stays as constant over the spin up period, and other environmental factors were not taken into consideration such as CO_2 fertilization, nitrogen limitation and climate variability.

When land is converted to cropland, managed pasture, and/or rangeland, each translation rule indicates that vegetation in primary and secondary may be cleared or remain intact as the result of land-use changes. For example, for a given land-use transition rate from forest to pasture, if the applied translation rule indicates to clear the vegetation completely, then the resulting grid cell vegetation fraction in forest land-use type is reduced equal to the amount of pasture gained. If the rule indicates not to clear vegetation, then only the land-use type will be changed to pasture and the vegetation area will be unchanged, but the vegetation will be influenced by the management in terms of stand age/biomass, which are assumed to cease growing due to pressure from subsequent human management. If this pasture land is further converted to other non-primary and non-secondary land (e.g. cropland, rangeland or urban), the vegetation remaining from previous forest-pasture conversion then will be totally cleared. Therefore, the vegetation fraction existing within the cropland, managed pasture, rangeland and urban of each grid-cell can be tracked via the following equation:

$$f(i,t+1) = f(i,t) + f^{gained}(i,t) - f^{lost}(i,t), (i = 5,6,7,8)\ , \quad (1)$$

Where $f(i,t)$ is the fraction of grid-cell that is vegetated in land-use type *i* (i.e. classes 5-8: cropland, managed pasture, rangeland, urban) at time *t*, $f^{gained}(i,t)$ and $f^{lost}(i,t)$ are gained and lost vegetation fractions respectively. The vegetation fraction could only be gained in land-use change from primary and secondary land (both forested and non-forested), and be lost in land-use change to any other land use types except forested and non-forested primary land.

$$f^{gained}(i,t) = \sum_{j=1}^{4} a_{ij}\gamma_{ij}, (i = 5,6,7,8; j = 1,2,3,4), \quad (2)$$

$$f^{lost}(i,t) = \frac{f(i,t)}{l(i,t)} \sum_{k=1,k\neq i}^{8} a_{ki}, (i = 5,6,7,8; k = 3,4,\cdots,8), \quad (3)$$

The possible values of *i*, *j* and *k* are 1, 2, … , 8 representing primary forested land, primary non-forested land, secondary forested land, secondary non-forested land, cropland, managed pasture, rangeland and urban respectively. a_{ij} is the land-use transition fraction estimate by LUH2 from land-use type *j* (i.e. primary forested land, primary non-forested land, secondary forested land, secondary non-forested land) to land-use type *i*, γ_{ij} represents the translator factor to convert land-use change to land-cover change, it equals to 1 if the translation rule in Table 1 indicates an 'X' or 'F' for this land-use change. For example, γ_{ij} is 1 for land-use change from primary land (forested, non-forested grids) to cropland in Rules 1 and 2, but 0 for the same type of change in Rules 8 and 9. This translator factor is 1 for all types of land-use change in Rule 3 since all vegetation is cleared during all land-use changes. $l(i,t)$ is the land-use fraction estimate by LUH2 for type *i* at time *t*, and this fraction is larger than or equal to its vegetation fraction $f(i,t)$.

Vegetation in primary and secondary land can remain or be lost in land-use changes to cropland, pasture or rangeland depending on translation rules. According to the definition of primary land in the LUH2, its transition to other land-use types is unidirectional, thus primary land could not gain vegetation from any land-use changes. Wood harvest on primary land will result in vegetation loss and a change of land-use type to secondary land, but harvest on secondary land will not change the land-use type. Furthermore, vegetation in secondary land could be gained from harvest on primary land and may be gained through the process of abandonment of cropland, pasture or rangeland depending on translation rules. Note that reforestation but not afforestation is also considered in this study. The former is to re-establish forest on the land which has been forested before, while the latter is an anthropogenic activity to establish forests on land which has never been forested. Thus, the vegetation of primary and secondary land is tracked by the following equation:

$$f(i,t+1) = f(i,t) - f^{lost}(i,t) + f^{gained}(i,t), (i = 1,2,3,4), \quad (4)$$

$$f^{lost}(i,t) = \begin{cases} \sum_{j=5}^{8} a_{ji}\gamma_{ji} + b_i, (i = 1,2; j = 5,6,7,8) \\ \sum_{j=5}^{8} a_{ji}\gamma_{ji}, (i = 3,4; j = 5,6,7,8) \end{cases} \quad (5)$$

$$f^{gained}(i,t) = \sum_{k=5}^{8} \frac{f(k,t)}{l(k,t)} a_{ik} + b_j, (i = 3,4; j = 1,2;\ k = 5,6,7,8) \quad (6)$$

Where $f(i,t)$ is fraction of vegetation at land-use category i (primary forested land, primary non-forested land, secondary forested land, secondary non-forested land) at time t. a_{ji} is land-use transition fraction from primary and secondary land to cropland, managed pasture, rangeland and urban in LUH2, γ_{ji} is the translator factor, as is γ_{ij} in Eq.2; both indicate whether to clear the vegetation during land-use changes. b_i or b_j is wood harvest fraction from primary or secondary (forested or non-forested) land. $f(k,t)$ and $l(k,t)$ are vegetation fraction and land-use fraction in land-use type k (i.e. cropland, managed pasture, rangeland, urban), and a_{ik} is land-use transition due to land-use abandonment.

2.4 Simulation of vegetation carbon dynamics

Vegetation carbon stocks fluctuate through releasing and accumulating carbon in response to natural growing conditions, disturbances, and anthropogenic land-use changes, which can vary widely in terms of their carbon impacts. For land-use changes associated with clearing or harvesting vegetation, the forest biomass is either released immediately (e.g. burning) or stored in soil pools or as timber products (both of which eventually decay over decades). However, when managed land is abandoned and allowed to recover, the vegetation takes up CO_2 from the atmosphere through photosynthesis, resulting in increasing carbon stocks in vegetation and possibly soils. The magnitude of each of these bi-directional carbon flows ultimately determine if the land is a net carbon sink or carbon source. In this study, the temporal dynamics of carbon fluxes after land-use change are simplified, with all biomass (above- and below-ground) being released instantaneously to the atmosphere. Note that the biomass stock change is a rough proxy of actual net land-use change fluxes, for which delayed emissions from litter and soil carbon and product pools needed to be accounted for as well as instantaneous emissions from burning biomass. Changes in soil carbon associated with loss of vegetation biomass are usually associated with carbon losses, but are likely less important than biomass changes, as are net fluxes from product pool changes (Erb et al., 2018).

Similar to land-cover change simulation in section 2.3, if translation rules indicate vegetation clearing at expansion of cropland, managed pasture, rangeland or urban land, vegetation biomass is totally released as a carbon emission, and its age is set as zero. If vegetation is not cleared based on translation rules, the biomass remains but ceases to increase, and the age of this vegetation also remains unaffected, because the age is used in this model only for the calculation of biomass density. Keeping age fixed corresponds to keeping biomass from further growing, which represents the influences of management. If the land is abandoned and converted back to secondary land, a mean age are calculated over all vegetation with different ages, then the mean age increases year by year and biomass regrows towards equilibrium. Thus, the biomass density in secondary vegetation at time t is calculated for each grid cell using its mean age, potential biomass, and potential NPP:

$$B(t) = B_0\left(1 - e^{-NPP_0 \times G(t)/B_0}\right), \quad (7)$$

Where $B(t)$ is the aboveground biomass density of vegetation at secondary land at time t, and B_0 is the potential aboveground biomass density from the GLM2 model and varied by grid location, and NPP_0 is the potential NPP of the wood fraction that is allocated to cumulate stem and branch biomass annually, and $G(t)$ is the mean age of secondary vegetation. Note that B_0

and NPP_0 is constant over simulation period from 850 to 2015. Above- to below-ground biomass ratio is assumed as 3:1 when converting aboveground biomass to total biomass (above- and belowground), and biomass density is converted to carbon by a ratio of 0.5.

Plants cultivated by human management (e.g. crops and orchards) are not tracked in this study; zero biomass is assigned to cropland, managed pasture, rangeland and urban use types. However, carbon is tracked for vegetation remaining from primary or secondary due to the translation rules, as well as lands that convert from human management back to natural lands. Thus, the total carbon stocks in this study are expected to be lower than other estimates (Houghton, 2003; Saatchi et al., 2011), especially in the grids with a higher fraction of non-primary and non-secondary land-use.

2.5 Diagnostics for evaluating translation rules

To evaluate which translation rules best translate land-use changes to land-cover changes, the simulation results were compared with contemporary forest cover and carbon density maps from remote sensing observations and other estimates, as well as LULCC carbon emissions from other studies using different models. Contemporary values of forest cover and carbon density are used for two reasons. First is the lack of multiple diagnostics of forest cover and carbon density across the whole simulation period (i.e. 850 to 2015). Second is that contemporary values could potentially reflect cumulative error in converting land-use change to land-cover change since 850. We assume that if a translation rule produces a best match with the diagnostic maps of forest cover and carbon density, then it would also produce the best estimate for the historical period.

Diagnostics of contemporary forest cover consist of six widely used satellite-based land-cover and tree coverage datasets (Bartholomé and Belward, 2005; Bicheron et al., 2008; DeFries et al., 2000; Friedl et al., 2010; Hansen et al., 2010; Loveland et al., 2000) (see Table 2) and the Global Forest Resources Assessment (FRA) 2015 (FAO, 2015). In Table 2, GLC, GLC2000, GlobCover and MODIS LC are land-cover datasets rather than tree cover and were produced based on different classification schemes resulting in different land-cover legends. Prior to being used as diagnostics in this study, they needed further reclassification of their land-cover legends into a common representation of forest canopy cover at the same spatial resolution (0.25˚) by the following procedures: First, the GLCC, GLC2000, GlobCover and MODIS LC were converted to tree cover fraction based on Table S1 at their native resolutions (Song et al., 2014). Then, all six datasets were resampled to 1 km resolution and translated to a binary (forest versus non-forest) map by applying a 30% tree-cover threshold (Sexton et al., 2016). Through counting the percentage of pixels marked as forest within each 0.25x0.25˚ grid cell, six global gridded forest cover maps at 0.25º spatial resolution were generated, and resulting global forest area of each dataset are shown in Table 2. As these satellite-based datasets were developed from different sensors (e.g. AVHRR, SPOT-4, MERIS, MODIS, Landsat) and models (regression trees, decision tree, clustering labels and random forests), an averaged map (hereinafter referred to as 'Averaged satellite-based forest cover') was generated in accompany with the six forest cover maps to examine spatial pattern

of contemporary forest cover simulated by each translation rule. In addition, since FAO only reports national forest cover (not spatially explicit), these data were only used for comparison at the country level.

Carbon density maps are employed as the second metric to evaluate the translation rules. Two datasets were employed: the IPCC Tier-1 biomass carbon map for the year 2000 (Ruesch and Gibbs, 2008) and a pantropical biomass map (hereinafter referred to as the Baccini's product (Baccini et al., 2012). The former, a global above- and below-ground carbon density map, is created by dividing the globe into 124 carbon zones by land-cover, continental regions, eco-floristic zones, and forest age and assigning each zone a unique carbon stock value. The latter is estimated by combining ground plots, GLAS LiDAR observations and optical reflectance of MODIS. This dataset employs the empirical relationship between aboveground biomass and tree diameter at breast height and estimates aboveground biomass density for pantropical regions (40˚S-30˚N). Both carbon density maps were resampled to 0.25˚ before evaluation.

In addition, the ability of the translation rules to reproduce LULCC carbon emissions is also assessed. The estimates of LULCC carbon emissions were compiled from published papers (Table 3) (Houghton, 2010; Houghton and Nassikas, 2017; Le Quéré et al., 2018; Pongratz et al., 2009; Reick et al., 2010; Shevliakova et al., 2009; Stocker et al., 2011). These studies have significant discrepancy in emissions estimates as they employed various methods (e.g. book-keeping methods and different process-based models), LULCC datasets, and considered different types of land-use change activities. They also differ in treatment of environmental change, for example, (Pongratz et al., 2009; Reick et al., 2010; Shevliakova et al., 2009; Stocker et al., 2011) include effects of evolving climate or atmospheric CO_2 concentration on LULCC emissions, which is not accounted for in bookkeeping mode based studies (Houghton, 2010; Houghton and Nassikas, 2017). In this study, only the range of these estimates during the pre-industrial and industrial periods are chosen to evaluate the translation rules. We posit that the recommended translation rule should not produce anomalous carbon emissions that are outside the compiled range.

In summary, the GLM2-based estimates of forest cover and carbon density in the year 2000 and LULCC carbon emissions during the periods 850-1850 and 1850-2000, based on nine different translation rules are compared with the above three types of diagnostics (i.e. contemporary forest cover/area and carbon density maps, LULCC emissions). The final recommended translation rules should produce: 1) the forest cover with the smallest difference with diagnostic maps at global, country and grid scale, the total forest cover at global and country level should be comparable to the range of diagnostics, and spatial pattern should also be closed to diagnostics; 2) the closest carbon density map compared to diagnostics with the smallest difference, comparable spatial pattern and total carbon stock as well; and 3) reasonable LULCC carbon emissions within the range from other diagnostic estimates and minimizing the anomalous emissions during 1950-1960. Finally, if several rules have a reasonably good fit to these three diagnostics, other criteria, such as the definition characteristics for managed pasture and rangeland has handled in HYDE (Goldewijk et al., 2017) will also be taken into account in identifying the recommended rule.

3 Results

3.1 Potential forest cover and biomass carbon

The GLM2 estimates global vegetation carbon stock (including above- and belowground) in 850 as 718 Pg C, and the resulting potential biomass map is shown in Figure 1a. For comparison, global potential vegetation carbon stock was estimated as 557 Pg C in (Kucharik et al., 2000), 772 Pg C in (Pan et al., 2013) and 923 Pg C in (Sitch et al., 2003). Forested land in GLM2 is defined as land which has aboveground potential biomass of at least 2 kg C/m^2 (Hurtt et al., 2006, 2011). With this definition, global potential forest area was estimated as 47.82 million km^2, and the resulting potential forest cover map is shown in Figure 1b. For comparison, global potential forest area was estimated as 48.68 million km^2 in (Pongratz et al., 2008), and potential forests and woodlands area was 55.3 million km^2 in (Ramankutty and Foley, 1999).

3.2 Forest cover evaluation

The global gridded forest cover maps resulting from Rules 1-4 in 2000 are generally consistent in forest extent with satellite-based observations (shown in Figure 2). For example, they all estimate high forest cover in tropical rainforests and northern boreal forests but low cover in Western USA, Eastern Europe and Central Asia. As Rules 1, 2, and 3 only differ in whether to clear vegetation and carbon in the conversion from non-forest to pasture or rangeland, the forest cover resulting from Rules 1, 2, and 3 are the same. All rules of 1-4 consistently estimate higher forest cover than the averaged satellite-based forest cover in West Siberia and South China, and lower forest cover in African savannas and East Siberia, Western Mexico and Argentina. Separately, Rule 4 shows larger forest cover than Rules 1-3 in South and Southeast of Brazil and Tiber in China. The spatial pattern of negative bias in estimated forest cover of Rules 1-4 well corresponds to where the GLM2 model and Satellite-based datasets disagree the presence of forest.

The total area of global forest in 850 amounts to 47.82 million km^2 according to the GLM2 model (Figure 1b and Figure 3a) when all forested lands were in a primary state by definition and decreased thereafter (Figure 3a). Forest loss has accelerated since the beginning of the Industrial Revolution and shows relatively high annual change rates (shown in Figure 3c). The translation rules produce a wide range of global forest cover in 2000 from 37.42 to 45.89 million km^2. In Rules 1, 2, and 3, the global forest is lost at the highest rate due to all land-use change activities on forested land resulting in the clearing of forest, and only 37.42 million km^2 of global forest is left in 2000 under these three rules. In contrast, under Rule 4 forest remains during rangeland expansion, and this would result in greater forest cover (e.g. 41.80 million km^2 in 2000, Table 4). The forest losses in Rules 6, 8, and 9 indicate the individual contribution of cropland, manged pasture and rangeland expansion. For example, rangeland and cropland expansion results in the most and second most of forest loss with an area of 4.34 million km^2 and 4.06 million km^2 respectively during 850-2000.

Six satellite-based forest cover datasets and FAO data report the global forest area around the year 2000 ranging from 35.66 to 42.74 million km2. One of major reasons underlying the discrepancy in global forest area is the difference in defining 'forest', particularly in the regions with intermediate tree cover (Sexton et al., 2016). The global forest area in the year 2000 resulting from the translation rules are compared to the range of seven diagnostic estimates (Figure 3b). The forest cover based on Rules 6, 8 and 9 is beyond the range of the diagnostics, indicating that these rules underestimate the impacts of land-use change on land-cover and overestimate the global forest existing in the present day. The excessive remaining forest cover in these three rules also rejects these rules' assumptions that only a particular type of land-use change would alter the land-cover. In contrast, Rules 1-4, 5 and 7 produced estimates of global forest area within the range of diagnostics.

The forest cover estimation from translation rules are further compared with diagnostic datasets at the country level. In the diagnostic forest cover datasets, three-fourths of global forest cover lies within eight countries: the Russian Federation, Brazil, Canada, United of States of America, China, Democratic Republic of the Congo, Indonesia and Peru. The forest cover estimates from Rules 1-4 are generally well within the range of diagnostics for most of the eight countries (e.g. Brazil, Indonesia, and United States of America) in terms of forest area and slightly overestimated in the Russian Federation and Canada, where the estimates of Rules 1-3 are closer to the upper bound of the diagnostics than Rule 4. China and Brazil are the two countries where Rules 1-4 shows relatively larger difference in forest area and the difference are 1.17 million and 1.08 million respectively. Rule 5 and 7 overestimated forest area of China, Russian Federation and Canada though their global forest areas are within the range of diagnostic.

These comparisons evaluate the resulting gross forest cover of the translation rules at global and country level. Further examination at the grid level is also needed. Since the FAO report only provides national forest cover, the averaged satellite-based forest cover map and each of the six satellite-based forest cover maps were used to calculate the average of absolute difference across global grids (Figure 4) respectively. Rule 1, 2, and 3 consistently produce the smallest overall difference than Rule 4 and other rules regardless of which satellite-based forest cover is chosen as the reference. The average absolute difference (AAD) of Rule 1, 2, 3 is under 90 km2 comparing to the averaged satellite-based forest cover map, and even smaller comparing to the GFC. Regional comparison of average of absolute difference (Figure S1) suggests Rule 1, 2, 3 give better estimate of forest cover at the north and south temperate zones (i.e. 60°N ~ 23°N and 23°S ~ 60°S) than tropical zone (23°N ~ 23°S). All rules have similar AAD at 60°N ~ 90°N zone.

3.3 Evaluation of carbon dynamics

The net carbon emissions of the nine translation rules were calculated over two periods (850 to 1850 and 1850 to 2000) and compared to other studies (Table 5). Rules 1-4 produced similar patterns to other studies, specifically that global carbon emissions of 1850-2000 are twice as large as that of 850-1850. However, the emissions estimates of each period varied among Rules 1-4, from 55 to 77 Pg C during 850-1850 and from 142 to 185 Pg C during 1850-2000, due to the assumptions for

clearing vegetation during land-use change. For example, Rule 3 produced the largest emissions as the carbon in both forested and non-forested land is released for all land-use changes, and Rule 1 produces fewer emissions since the vegetation is not cleared and carbon is not released when non-forested land is converted to rangeland. In general, Rule 1, 2, 3 and 4 estimated comparable emissions with other studies, while the emissions of the Rules 6-9 are out of range (Table 5).

Carbon emissions from pasture expansion were calculated for LUH1 (Hurtt et al., 2011) and this is used as a baseline to assess the improvement of translation rules on the pasture anomaly. Rules 1-4 estimate fewer emissions during this decade and decrease the anomaly between 4 to 10 Pg C. In LUH1, the anomalous emissions spike during 1950-1960 mainly arises from overestimating the emissions from pasture expansion, especially in three regions (i.e. Africa, East, South and Central Asia, and North America). The carbon flux from expansion of managed pasture and rangeland in LUH2 was reduced at global (Figure 5) and regional (Figure 6) scales in simulations based on Rules 1, 2, and 3. Note that the pasture land in LUH1 corresponds to rangeland and managed pasture together in LUH2. Rule 2 reduces more anomalous emissions than Rule 1 (reduced 6 Pg C in Rule 1 and 7 Pg C in Rule 2), because Rule 1 completely clears vegetation when transitioning to managed pasture, whereas Rule 2 only removes vegetation if the preceding land cover is primary or secondary forest.

Rules 1-4 generally capture the spatial pattern that carbon density in tropical rainforest regions is much higher than northern boreal forests (Figure 7). These four rules overestimate carbon density at high latitudes of the Northern Hemisphere, in South China and in the Amazon rainforests but underestimate density across much of Sub-Saharan Africa, Mexico and the Southwestern part of the United States (Figure S2 and Figure S3). To further examine the spatial pattern of estimated carbon density, the estimates from all rules were compared to the carbon density maps of IPCC Tier-1 (above- and belowground) globally and the Bacchini's dataset (only aboveground) at the pantropical scale by calculating averaged absolute difference (Figure 8). According to this comparison, Rules 1-3 best capture the carbon density heterogeneity (Figure 8). Regional comparison of the IPCC Tier-1 biomass map and rule estimates indicate Rules 1-4 have comparable AAD of carbon density at the zone of 90°N ~ 60° N, the AAD difference between four rules is largest at 23°S ~ 60°S, followed by 23°N ~ 23°S and 23°N ~ 60°N (Figure S4). Carbon density estimates of Rules 1-3 were further examined at regions where their estimates have difference (shown in Figure S5a). The spatial pattern (Figure S5c-S5f) and histogram (Figure S5b) of carbon density difference between rules and IPCC Tier-1 biomass estimates shows that all of these three rules underestimate carbon density and more grids are less underestimated in Rules 1-2 than Rule 3. The underestimation is expected because biomass of human cultivated vegetation is not tracked, and nor is growth of natural vegetation on cropland and pasture and rangeland. However, uncertainty level of the IPCC Tier-1 biomass should be taken into account when determining rule performance. Three bias levels of IPCC Tier-1 biomass map (i.e. ±10%, ±20% and ±30%) were considered (Figure S5b). At these levels of uncertainty in the reference, Rules 1-3 could not be distinguished in performance. Finally, the carbon stock comparison between Rules 1-3 (Figure 9) shows these three rules underestimate carbon stock at low forest fraction, but give better agreement with diagnostics as forest fraction increases.

4 Discussion and Conclusions

This study quantified the result of multiple alternative translation rules for estimating the potential effects of land-use change on landcover utilizing the LUH2 dataset, and the underlying land model embedded in it (GLM2). The evaluations of forest cover and carbon at both global, country and grid level jointly indicate that Rules 1-3 outperform other rules and are able produce the closest estimates of contemporary forest cover an carbon to global diagnostics. Differentiation between Rules 1-3 depends largely on estimates of forest carbon because these rules produce equivalent estimates of forest cover. Comparisons of carbon stock and gridded difference in carbon density have shown that Rules 1 and 2 produce better estimates of carbon density than Rule 3 relative to references. Factoring in the uncertainty in the carbon density reference map, and prior recommendation from HYDE 3.2 (Goldewijk et al., 2017), Rule 1 is recommended for model implementation, namely removes all vegetation when establishing cropland, urban land, or managed pasture, and leaves all vegetation when establishing rangeland only if the land is previously non-forested.

A key feature of this study is to explicitly link land-use change and land-cover change and to suggest a suitable method to incorporate the LUH2 land-use transition dataset into ESMs and DGVMs. With recommended translation rule from this study and the LUH2, historical land-cover change could be reconstructed over the period of 850-2015, and the resulting contemporary forest cover and carbon density are comparable to independent estimates. Furthermore, this study also provides insights into the uncertainty attribution of LULCC emissions estimates from ESMs and DGVMs. The CMIP5 models estimate vegetation carbon changes during 1850-2005 in a wide range of -151 to +27 Pg C, which is contributed by different implementations of land-use change as well as inter-model uncertainty of strength of the CO_2 fertilization (Jones et al., 2013). Table 4 in this study indicates differences in choice of Rules 1-4 result in an uncertainty of 43 Pg C in LULCC emissions during 1850-2000 and it is about 24% of the uncertainties in estimates of vegetation carbon changes in CMIP5. Therefore, in addition to uncertainties stemming from inter-model difference in potential vegetation cover and biomass as well as effects of CO_2 fertilization, an added uncertainty of 43 Pg C could be expected in CMIP6 solely from inconsistent choices in Rules 1-4 and it could be larger when other rules are implemented.

When considering translation rules, it is important to note that the results of this study may depend on the land-use change dataset being used, the land-cover properties being evaluated, reference datasets, and the models. This study used the LUH2 dataset because of its required used in CMIP6 and widespread used in other studies. The land cover properties addressed here include two critical variables (i.e. forest cover and carbon stock) due to their biophysical and biogeochemical significance. Multiple datasets based on remote sensing and other sources were selected for evaluation with the intention to provide a robust reference. The use of GLM2 model was selected to provide the most internally consistent treatment of these issues given its role in producing the LUH2 dataset. Given these considerations, it is possible that different results could be obtained for different systems. Although multiple of satellite-based land-cover datasets were included, they disagree the presence or

absence of forest over low forest cover regions such as shrublands and semi-arid savannahs, and the discrepancies due to technical challenges and disagreement of forest definition. In addition, global vegetation carbon mapping is still challenging and uncertain mainly because of indirect proxies of biomass and paucity of in situ measurements and observations from space. Combined uncertainties from forest cover and vegetation carbon diagnostics may limit the evaluation of translation rules, especially in locations where forest cover or vegetation carbon is low. Furthermore, dynamics of forest cover and vegetation carbon from past to present interact with climate change and increasing atmospheric CO_2, which are not considered in this study. Finally, the carbon emission estimates using the same translation rules and land-use change dataset may be different using other carbon models/DGVMs.

Future research is needed to investigate both the robustness of these findings, and potentially identify even better implementations. The CMIP6 LUMIP study is designed to quantify some of these effects (Lawrence et al., 2016) through model inter-comparison. Additional work on translation rules should include possible spatial/temporal varying rules, partial land clearing, and more land cover variables (e.g. forest age, height, soil carbon, energy balance).

Code and data availability. The source code of the modified GLM2, source and citation of inputs, results of all translation rules and scripts for producing figures and tables are archived at https://doi.org/10.5281/zenodo.3533792 , LUH2 dataset is available at http://doi.org/10.22033/ESGF/input4MIPs.1127. IPCC Tier-1 biomass is available at https://cdiac.ess-dive.lbl.gov/epubs/ndp/global_carbon/carbon_documentation.html, Baccini's aboveground biomass is available at https://doi.org/10.3334/ORNLDAAC/1337. TCCF, MODIS LC, GLCC, GFC, GLC2000 and GlobCover can be obtained from http://www.landcover.org/data/treecover/, https://doi.org/10.5067/MODIS/MCD12Q1.006, https://doi.org/10.5066/F7GB230D, https://earthenginepartners.appspot.com/science-2013-global-forest, https://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php, http://due.esrin.esa.int/page_globcover.php respectively.

Author contributions. LM, GH, LC and RS designed this study. LM conducted the simulations and wrote the main body of the paper. All authors discussed the results and commented on the paper at all stages.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We gratefully acknowledge the support to DOE-SCIDAC DESC0012972, and NASA-TE NNX13AK84A and NASA-IDS 80NSSC17K0348.

References

Baccini, A., Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., and Friedl, M.: Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps, Nature climate change, 2, 182-185, 2012.

Bartholomé, E., and Belward, A.: GLC2000: a new approach to global land cover mapping from Earth observation data, International Journal of Remote Sensing, 26, 1959-1977, 2005.

Betts, R. A.: Biogeophysical impacts of land use on present-day climate: Near-surface temperature change and radiative forcing, Atmospheric Science Letters, 2, 39-51, 2001.

Bicheron, P., Amberg, V., Bourg, L., Petit, D., Huc, M., Miras, B., Brockmann, C., Delwart, S., Ranéra, F., and Hagolle, O.: Geolocation assessment of 300 m resolution MERIS Globcover ortho-rectified products, Proceedings of the' 2nd MERIS/(A) ATSR User Workshop', Frascati, Italy, 22– 26 September 2008 (ESA SP-666, November 2008), 2008,

Bonan, G. B.: Forests and climate change: forcings, feedbacks, and the climate benefits of forests, science, 320, 1444-1449, 2008.

Brovkin, V., Claussen, M., Driesschaert, E., Fichefet, T., Kicklighter, D., Loutre, M.-F., Matthews, H. D., Ramankutty, N., Schaeffer, M., and Sokolov, A.: Biogeophysical effects of historical land cover changes simulated by six Earth system models of intermediate complexity, Climate Dynamics, 26, 587-600, 2006.

Brovkin, V., Boysen, L., Arora, V., Boisier, J., Cadule, P., Chini, L., Claussen, M., Friedlingstein, P., Gayler, V., and Van Den Hurk, B.: Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century, Journal of Climate, 26, 6859-6881, 2013.

Chini, L. P., Hurtt, G., Klein Goldewijk, K., Frolking, S., Shevliakova, E., Thornton, P., and Fisk, J.: Addressing the pasture anomaly: how uncertainty in historical pasture data leads to divergence of atmospheric CO2 in Earth System Models, AGU Fall Meeting Abstracts, 2012,

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., and Heimann, M.: Carbon and other biogeochemical cycles, in: Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 465-570, 2014.

Claussen, M., Brovkin, V., and Ganopolski, A.: Biogeophysical versus biogeochemical feedbacks of large-scale land cover change, Geophysical research letters, 28, 1011-1014, 2001.

de Noblet-Ducoudré, N., Boisier, J.-P., Pitman, A., Bonan, G., Brovkin, V., Cruz, F., Delire, C., Gayler, V., Van den Hurk, B., and Lawrence, P.: Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: Results from the first set of LUCID experiments, Journal of Climate, 25, 3261-3281, 2012.

DeFries, R., Hansen, M., Townshend, J., Janetos, A., and Loveland, T.: A new global 1-km dataset of percentage tree cover derived from remote sensing, Global Change Biology, 6, 247-254, 2000.

DeFries, R. S., Houghton, R. A., Hansen, M. C., Field, C. B., Skole, D., and Townshend, J.: Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s, Proceedings of the National Academy of Sciences, 99, 14256-14261, 2002.

Dewar, R. C.: Analytical model of carbon storage in the trees, soils, and wood products of managed forests, Tree Physiology, 8, 239-258, 1991.

Di Vittorio, A., Mao, J., Shi, X., Chini, L., Hurtt, G., and Collins, W.: Quantifying the effects of historical land cover conversion uncertainty on global carbon and climate estimates, Geophysical Research Letters, 45, 974-982, 2018.

Di Vittorio, A. V., Chini, L., Bond-Lamberty, B., Mao, J., Shi, X., Truesdale, J., Craig, A., Calvin, K., Jones, A., and Collins, W. D.: From land use to land cover: restoring the afforestation signal in a coupled integrated assessment–earth system model and the implications for CMIP5 RCP simulations, Biogeosciences, 11, 6435-6450, 2014.

Erb, K.-H., Kastner, T., Plutzar, C., Bais, A. L. S., Carvalhais, N., Fetzel, T., Gingrich, S., Haberl, H., Lauk, C., and Niedertscheider, M.: Unexpectedly large impact of forest management and grazing on global vegetation biomass, Nature, 553, 73, 2018.

FAO: Global Forest Resources Assessment 2015, 2015.

Feddema, J. J., Oleson, K. W., Bonan, G. B., Mearns, L. O., Buja, L. E., Meehl, G. A., and Washington, W. M.: The importance of land-cover change in simulating future climates, Science, 310, 1674-1678, 2005.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote sensing of Environment, 114, 168-182, 2010.

Goldewijk, K. K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land use estimates for the Holocene–HYDE 3.2, Earth System Science Data, 9, 927-953, 2017.

Guo, L. B., and Gifford, R.: Soil carbon stocks and land use change: a meta analysis, Global change biology, 8, 345-360, 2002.

Hansen, M. C., Stehman, S. V., and Potapov, P. V.: Quantification of global gross forest cover loss, Proceedings of the National Academy of Sciences, 107, 8650-8655, 2010.

Hansis, E., Davis, S. J., and Pongratz, J.: Relevance of methodological choices for accounting of land use change carbon fluxes, Global Biogeochemical Cycles, 29, 1230-1246, 2015.

Houghton, R.: The annual net flux of carbon to the atmosphere from changes in land use 1850–1990, Tellus B, 51, 298-313, 1999.

Houghton, R., and Nassikas, A. A.: Global and regional fluxes of carbon from land use and land cover change 1850–2015, Global Biogeochemical Cycles, 31, 456-472, 2017.

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 B, 55, 378-390, 10.1034/j.1600-0889.2003.01450.x, 2003.

Houghton, R. A.: How well do we know the flux of CO2 from land-use change?, Tellus B, 62, 337-351, 2010.

Houghton, R. A., House, J., Pongratz, J., Van der Werf, G., DeFries, R., Hansen, M., Quéré, C. L., and Ramankutty, N.: Carbon emissions from land use and land-cover change, Biogeosciences, 9, 5125-5142, 2012.

Hurtt, G., Pacala, S., Moorcroft, P. R., Caspersen, J., Shevliakova, E., Houghton, R., and Moore, B.: Projecting the future of the US carbon sink, Proceedings of the National Academy of Sciences, 99, 1389-1394, 2002.

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J., Krisztin, T., Lawrence, D., Lawrence, P., Mertz, O., Pongratz, J., Popp, A., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., van Vuuren, D., and Zhang, X.: Harmonization of global land use scenarios (LUH2): SSP585 v2.1f 2015 - 2100. Earth System Grid Federation, 2017a.

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J., Lawrence, D., Lawrence, P., Mertz, O., Popp, A., Stehfest, E., Thornton, P., van Vuuren, D., and Zhang, X.: Harmonization of global land use scenarios (LUH2): Historical v2.1h. Earth System Grid Federation, 2017b.

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J., Lawrence, D., Lawrence, P., Mertz, O., Popp, A., Stehfest, E., Thornton, P., van Vuuren, D., and Zhang, X.: Harmonization of global land-use change and management for the period 850–2100, Geoscientific Model Development (In prep), 2019.

Hurtt, G. C., Frolking, S., Fearon, M., Moore, B., Shevliakova, E., Malyshev, S., Pacala, S., and Houghton, R.: The underpinnings of land-use history: Three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands, Global Change Biology, 12, 1208-1229, 2006a.

Hurtt, G. C., Frolking, S., Fearon, M. G., Moore, B., Shevliakova, E., Malyshev, S., Pacala, S. W., and Houghton, R. A.: The underpinnings of land-use history: Three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands, Global Change Biology, 12, 1208-1229, 2006b.

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, Climatic Change, 109, 117, 10.1007/s10584-011-0153-2, 2011.

Jones, C., Robertson, E., Arora, V., Friedlingstein, P., Shevliakova, E., Bopp, L., Brovkin, V., Hajima, T., Kato, E., Kawamiya, M., Liddicoat, S., Lindsay, K., Reick, C. H., Roelandt, C., Segschneider, J., and Tjiputra, J.: Twenty-First-Century Compatible CO2 Emissions and Airborne Fraction Simulated by CMIP5 Earth System Models under Four Representative Concentration Pathways, Journal of Climate, 26, 4398-4413, 10.1175/JCLI-D-12-00554.1, 2013.

Kaplan, J. O., Krumhardt, K. M., and Zimmermann, N.: The prehistoric and preindustrial deforestation of Europe, Quaternary Science Reviews, 28, 3016-3034, 2009.

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land-use estimates for the Holocene - HYDE 3.2, Earth Syst. Sci. Data, 9, 927 - 953, https://doi.org/10.5194/essd-9-1-2017, 2017.

Kucharik, C. J., Foley, J. A., Delire, C., Fisher, V. A., Coe, M. T., Lenters, J. D., Young-Molling, C., Ramankutty, N., Norman, J. M., and Gower, S. T.: Testing the performance of a dynamic global ecosystem model: water balance, carbon balance, and vegetation structure, Global Biogeochemical Cycles, 14, 795-825, 2000.

Lawrence, D. M., Hurtt, G. C., Calvin, K. V., Jones, A. D., Jones, C. D., Lawrence, P. J., and Seneviratne, S. I.: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: rationale and experimental design, Geoscientific Model Development, 9, 2973, 2016.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S. I., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, Earth Syst. Sci. Data, 10, 2141-2194, 10.5194/essd-10-2141-2018, 2018.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, International Journal of Remote Sensing, 21, 1303-1330, 2000.

Nave, L. E., Vance, E. D., Swanston, C. W., and Curtis, P. S.: Harvest impacts on soil carbon storage in temperate forests, Forest Ecology and Management, 259, 857-866, 2010.

Pan, Y., Birdsey, R. A., Phillips, O. L., and Jackson, R. B.: The structure, distribution, and biomass of the world's forests, Annual Review of Ecology, Evolution, and Systematics, 44, 593-622, 2013.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, Global Biogeochemical Cycles, 22, 2008.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: Effects of anthropogenic land cover change on the carbon cycle of the last millennium, Global Biogeochemical Cycles, 23, 2009.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: Biogeophysical versus biogeochemical climate response to historical anthropogenic land cover change, Geophysical Research Letters, 37, 2010.

Pongratz, J., Reick, C. H., Houghton, R., and House, J.: Terminology as a key uncertainty in net land use and land cover change carbon flux estimates, Earth System Dynamics, 5, 177-195, 2014.

Pongratz, J., Dolman, H., Don, A., Erb, K. H., Fuchs, R., Herold, M., Jones, C., Kuemmerle, T., Luyssaert, S., and Meyfroidt, P.: Models meet data: Challenges and opportunities in implementing land management in Earth system models, Global change biology, 24, 1470-1487, 2018.

Post, W. M., and Kwon, K. C.: Soil carbon sequestration and land-use change: processes and potential, Global change biology, 6, 317-327, 2000.

Ramankutty, N., and Foley, J. A.: Estimating historical changes in global land cover: Croplands from 1700 to 1992, Global biogeochemical cycles, 13, 997-1027, 1999.

Reick, C. H., Raddatz, T., Pongratz, J., and Claussen, M.: Contribution of anthropogenic land cover change emissions to pre-industrial atmospheric CO2, Tellus B, 62, 329-336, 2010.

Ruesch, A., and Gibbs, H. K.: New IPCC Tier-1 global biomass carbon map for the year 2000, 2008.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., and Hagen, S.: Benchmark map of forest carbon stocks in tropical regions across three continents, Proceedings of the National Academy of Sciences, 108, 9899-9904, 2011.

Sexton, J. O., Noojipady, P., Song, X.-P., Feng, M., Song, D.-X., Kim, D.-H., Anand, A., Huang, C., Channan, S., and Pimm, S. L.: Conservation policy and the measurement of forests, Nature Climate Change, 6, 192-196, 2016.

Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P., Caspersen, J. P., Sentman, L. T., Fisk, J. P., Wirth, C., and Crevoisier, C.: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, Global Biogeochemical Cycles, 23, 2009.

Shevliakova, E., Stouffer, R. J., Malyshev, S., Krasting, J. P., Hurtt, G. C., and Pacala, S. W.: Historical warming reduced due to enhanced land carbon uptake, Proceedings of the National Academy of Sciences, 110, 16730-16735, 2013.

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 Change Biology, 9, 161-185, 10.1046/j.1365-2486.2003.00569.x, 2003.

Smith, L. J., and Torn, M. S.: Ecological limits to terrestrial biological carbon dioxide removal, Climatic Change, 118, 89-103, 2013.

Song, X.-P., Huang, C., Feng, M., Sexton, J. O., Channan, S., and Townshend, J. R.: Integrating global land cover products for improved forest cover characterization: an application in North America, International Journal of Digital Earth, 7, 709-724, 2014.

Stocker, B., Strassmann, K., and Joos, F.: Sensitivity of Holocene atmospheric CO2 and the modern carbon budget to early human land use: analyses with a process-based model, Biogeosciences, 8, 69, 2011.

Wei, Y., Liu, S., Huntzinger, D. N., Michalak, A. M., Viovy, N., Post, W. M., Schwalm, C. R., Schaefer, K., Jacobson, A. R., Lu, C., Tian, H., Ricciuto, D. M., Cook, R. B., Mao, J., and Shi, X.: NACP MsTMIP: Global and North American Driver Data for Multi-Model Intercomparison. ORNL Distributed Active Archive Center, 2014.

Figures & Tables

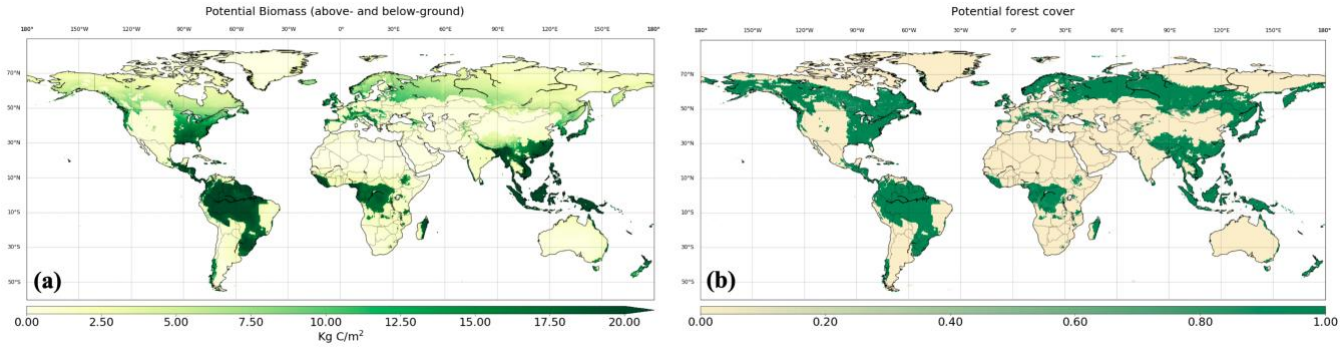

Figure 1. Potential biomass density (a) and potential forest cover (b) in 850 estimated by GLM2 model.

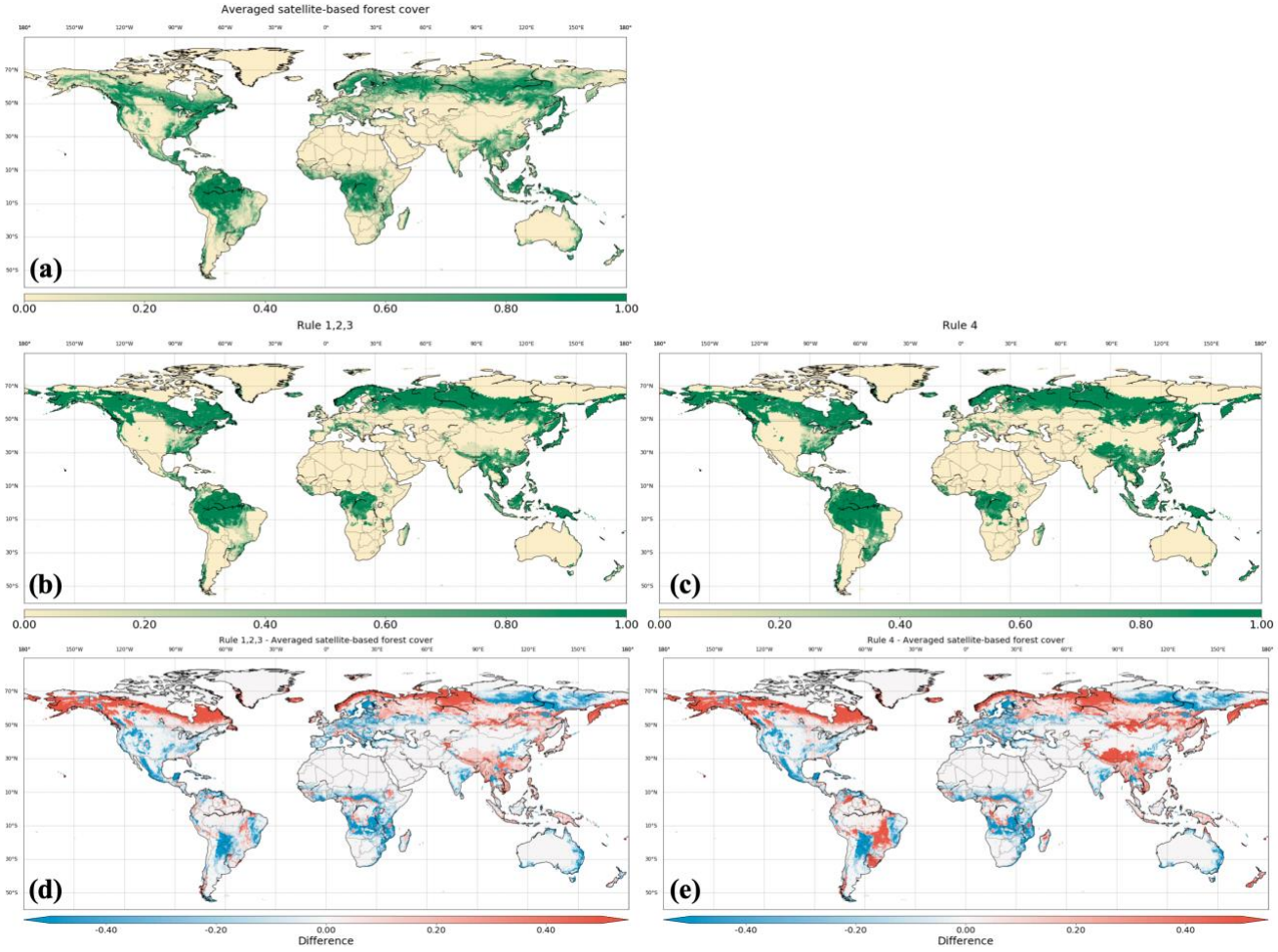

Figure 2. Forest cover in 2000 from the Averaged satellite-based forest cover in (a), Rule 1, 2, 3 in (b) and Rule 4 in (c). (d) and (e) are maps of forest cover difference between (b) and (a), and (c) and (a) respectively.

Table 1. Rules for vegetation clearance during cropland, pasture and rangeland expansion. 'X' indicates complete removal of vegetation if the primary and secondary land state is altered. 'O' indicates no vegetation removal when land-use change occurs. 'F' indicates that vegetation is only removed if the preceding land cover is forested primary or forested secondary land.

Transition Rule	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8	Rule 9
->Crop	X	X	X	X	X	X	O	O	O
->Managed pasture	X	F	X	X	O	O	X	X	O
->Rangeland	F	F	X	O	X	O	X	O	X

Table 2. Summary of land cover products used in this study including six satellite-based datasets and FAO FRA report.

Product	Global Forest Area (10_6 km_2)	Time	Publication	Data Type/Classification Scheme
GLCC	40.89	1992-1993	Loveland et al. 2000	Land Cover (IGBP)
GLC2000	38.22	1999-2000	Bartholome et al. 2005	Land Cover (GLC 2000)
GlobCover	35.66	2004-2006	Bicheron et al. 2008	Land Cover (GlobCover)
MODIS LC	41.05	2001	Friedl et al. 2010	Land Cover (IGBP)
1 Kilometer Tree Cover Continuous Fields (TCCF)	42.74	1992-1993	DeFries et al. 2000	Tree Percentage
Global Forest Change (GFC)	41.71	2000	Hansen et al. 2010	Tree Percentage
FAO	40.55	2000	FRA 2015	National Censuses

Table 3. Summary of carbon emissions due to LULCC from available studies at pre-industrial and industrial period.

Reference	Time span	Carbon Emissions (Pg C)	LULCC types
Pre-industrial Period			
Reick et al., 2010 (bookkeeping model)	1100-1850	80	Cropland/Pasture Change
Reick et al., 2010 (DGVM)	1100-1850	47	
Pongratz et al., 2009	850-1850	53	Cropland/Pasture Change
Stocker et al., 2011	until 1850	69	Cropland/Pasture Change, Urban
Industrial Period			
Houghton 2010	1850-2005	156	Cropland/Pasture Change, shifting cultivation in tropics, and wood harvest
Houghton and Nassikas, 2017	1850-2015	145	Cropland/Pasture Change, shifting cultivation in tropics, and wood harvest
Shevliakova et al.,2009	1850-2000	164 - 188	Cropland/Pasture Change, shifting cultivation in tropics, and wood harvest
Pongratz et al.,2009	1850-2000	108	Cropland/Pasture Change
Reick et al.,2010 (bookkeeping model)	1850-1990	153	Cropland/Pasture Change
Reick et al.,2010 (DGVM)	1850-1990	110	Cropland/Pasture Change
Stocker et al., 2011	1850-2004	164	Cropland/Pasture Change, Urban
Le Quéré et., 2018	1850-2014	195	Cropland/Pasture Change, shifting cultivation in tropics, and wood harvest

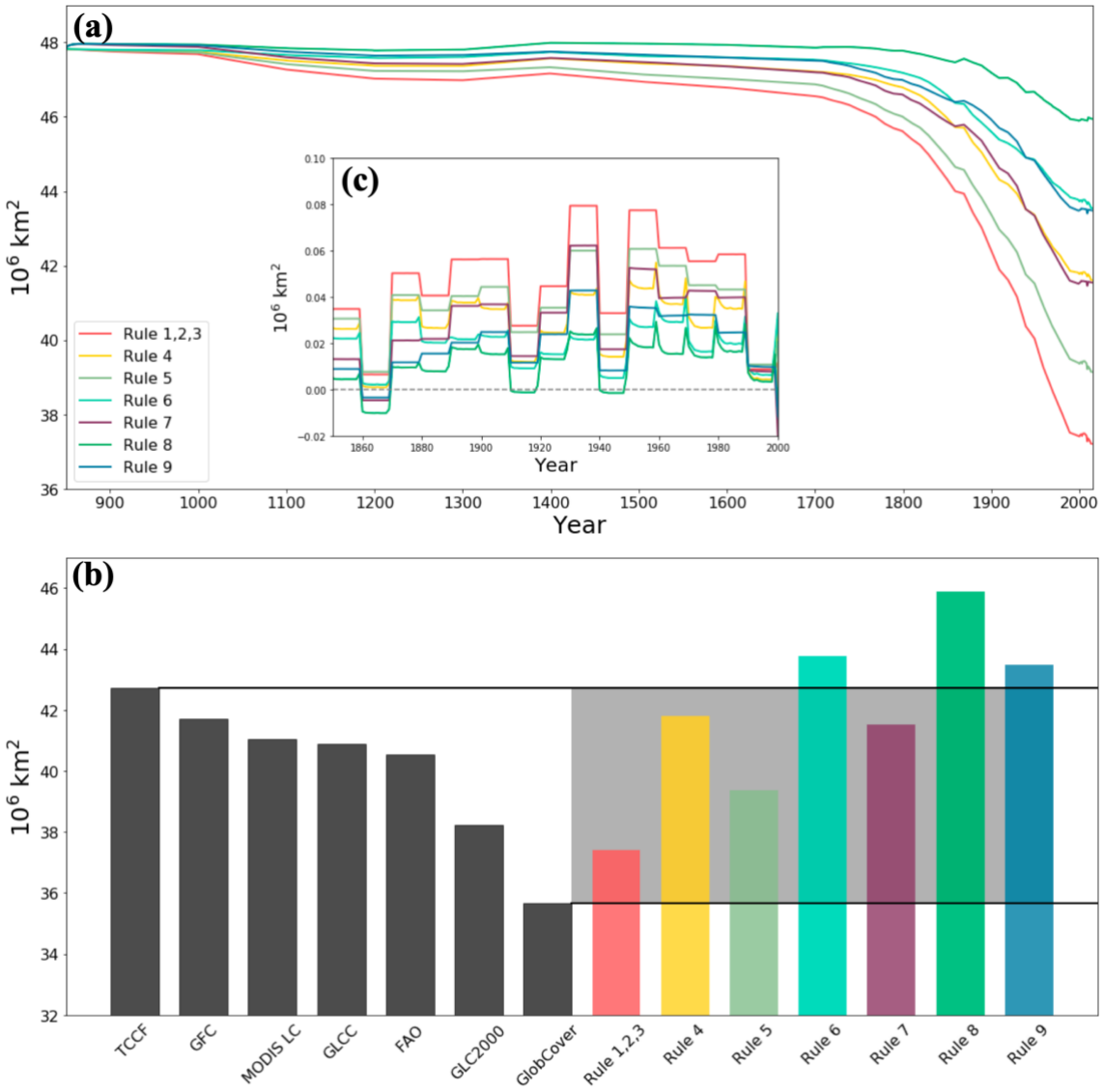

Figure 3. (a) Global forest area resulting from translation rules from 850 to 2015; (b) Comparison of global forest area in 2000 between remote sensing and FAO (shown as black bars) and results of translation rules (colored bars); (c) Annual change rate from 1850 to 2000. Positive value indicates the forest loss.

Table 4. Forest area (10^6 km^2) in 2000 of eight countries with the largest forest area, and all other countries combined ('Others'), estimated by the 9 translation rules, range compiled from satellite-based datasets and FAO report.

Country	Forest Area (10^6 km^2)							Range from satellite-based products and FAO
	Rule 1, 2, 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8	Rule 9	
Russian Federation	8.72	9.15	8.80	9.23	9.01	9.44	9.10	6.65-8.62
Brazil	4.61	5.69	4.89	5.96	5.05	6.12	5.33	4.19-5.92
Canada	5.59	5.63	5.59	5.64	5.76	5.81	5.77	3.27-4.36
United States of America	2.81	2.94	3.06	3.19	3.62	3.76	3.87	2.65-3.36
China	2.04	3.22	2.44	3.61	2.45	3.63	2.85	1.34-2.14
Democratic Republic of the Congo	1.57	1.61	1.60	1.64	1.63	1.67	1.66	1.57-2.11
Indonesia	1.30	1.33	1.36	1.38	1.58	1.60	1.64	0.99-1.64
Peru	0.76	0.78	0.78	0.80	0.77	0.79	0.79	0.69-0.79
Others	10.02	11.47	10.86	12.31	11.63	13.08	12.48	12.21-17.08
World	**37.42**	**41.80**	**39.38**	**43.76**	**41.52**	**45.89**	**43.48**	**35.66-42.74**

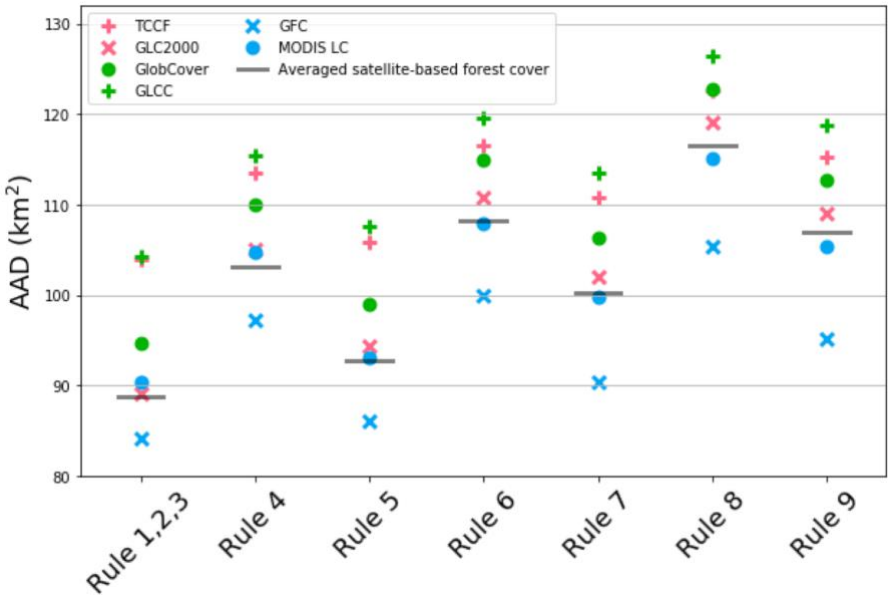

Figure 4. Global average of absolute difference in forest area between maps estimated by translation rules, and each of the six satellite-based forest cover maps as well as the averaged satellite-based forest cover map.

Table 5. Summary of LULCC carbon emissions estimated by the 9 translation rules and those from other studies in Table 3

Translation Rule	Carbon Emissions Estimation (Pg C)			Emission Range from Table 3		Estimation using LUH1
	850-1850	1850-2000	1950-1960	850-1850	1850-2015	1950-1960
Rule 1	72	175	20			
Rule 2	70	170	19			
Rule 3	77	185	22			
Rule 4	55	142	16			
Rule 5	63	146	17	47-80	108-195	26
Rue 6	41	104	11			
Rule 7	28	107	13			
Rule 8	5	65	7			
Rule 9	13	67	7			

Figure 5. Carbon emission due to vegetation (forests and non-forests) removal in expansion of managed pasture and rangeland. Black line represents emissions from pasture expansion in LUH1. Orange and green lines represent emissions from expansion of managed pasture and rangeland and from expansion of just managed pasture respectively in LUH2. Note that the pasture category in LUH1 corresponds to managed pasture and rangeland together in LUH2.

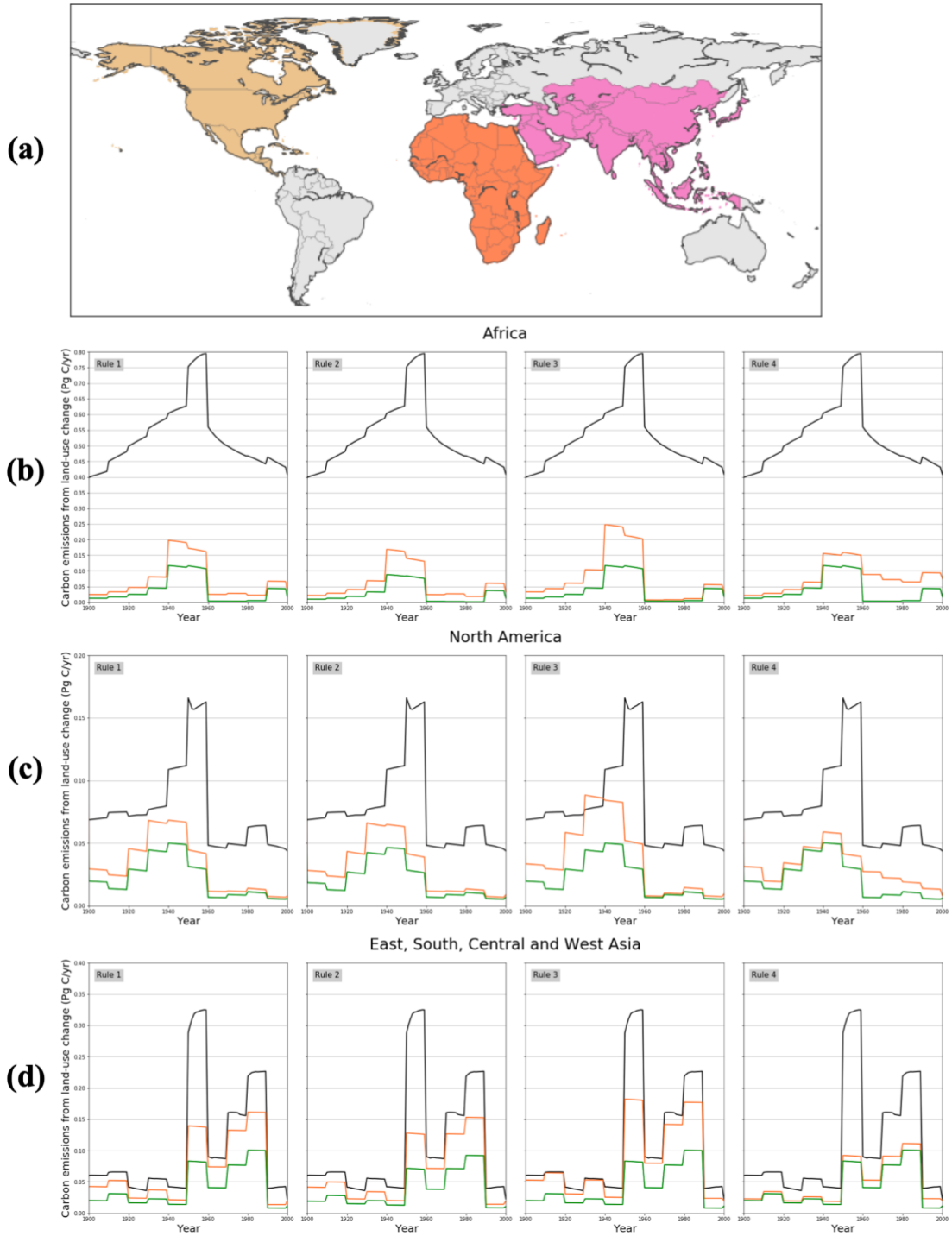

Figure 6. As in Figure 5 but three regions: (b) Africa; (c) East, South, Central and West Asia; (d) North America. (a) illustrates the defined boundaries of (b) - (d).

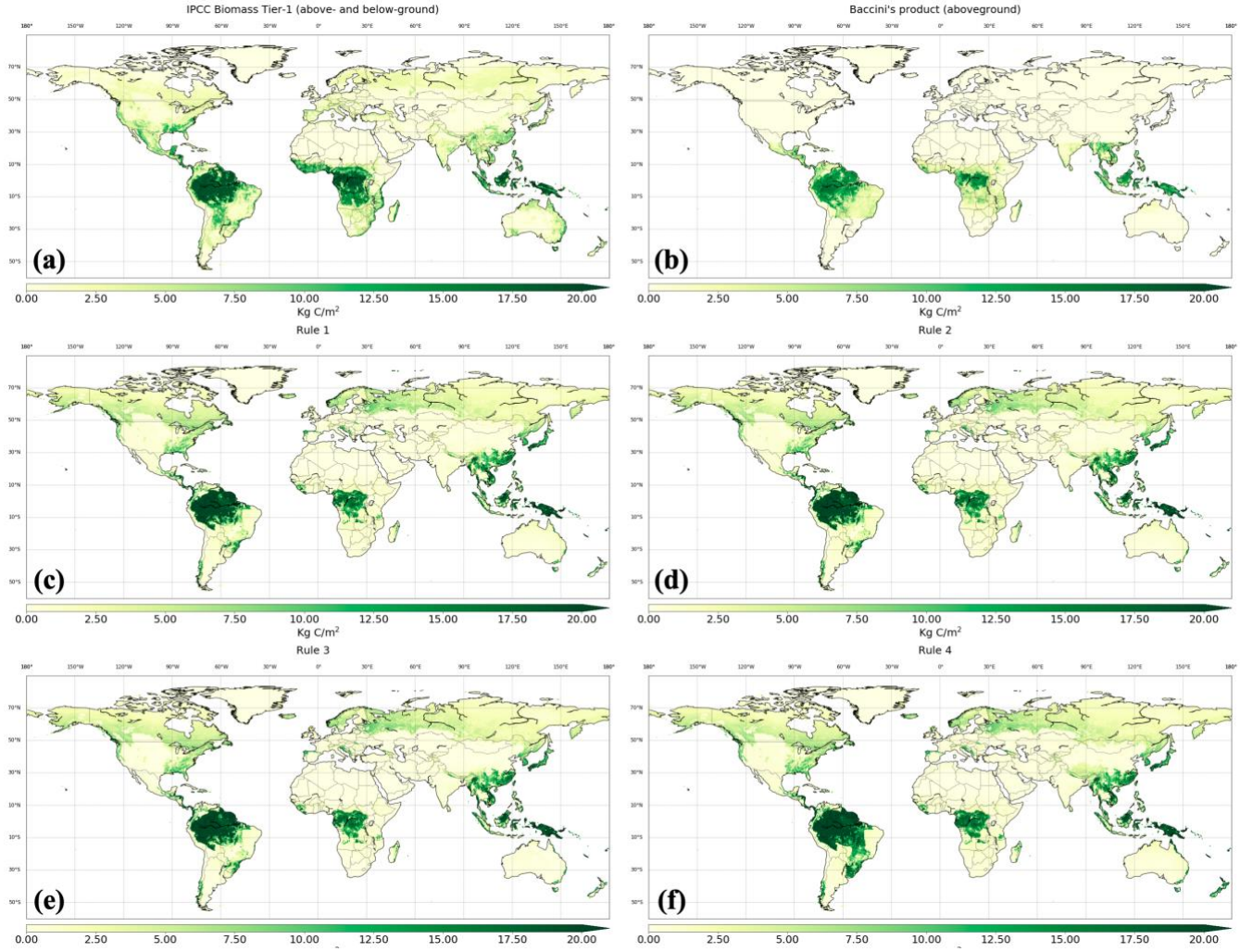

Figure 7. (a) IPCC Biomass Tier-1 density; (b) Baccini's product (only aboveground) at pantropical; global carbon density (above- and below-ground) maps estimated by Rules 1-4 from (c) to (f).

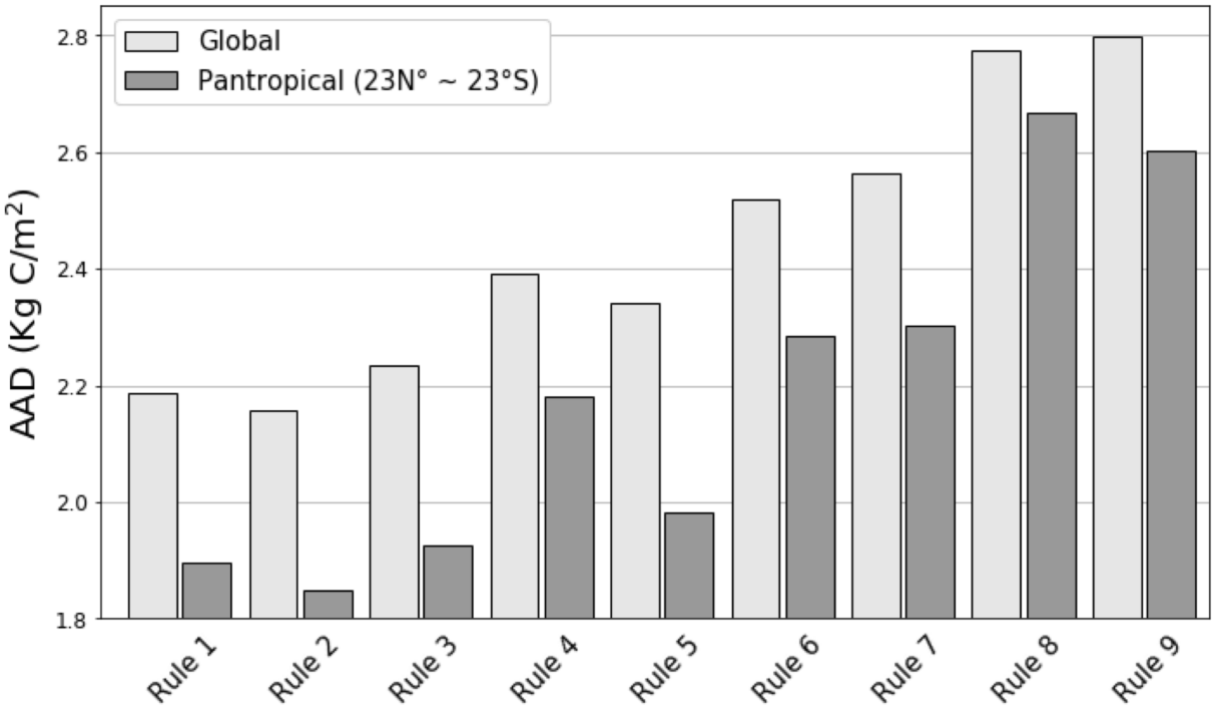

Figure 8. Average of absolute difference in carbon density between estimations of the 9 translation rules and two diagnostic maps: global comparison with IPCC Tier-1 biomass density map (incl. above- and below-ground); tropical comparison with Baccini's carbon density map (only aboveground).

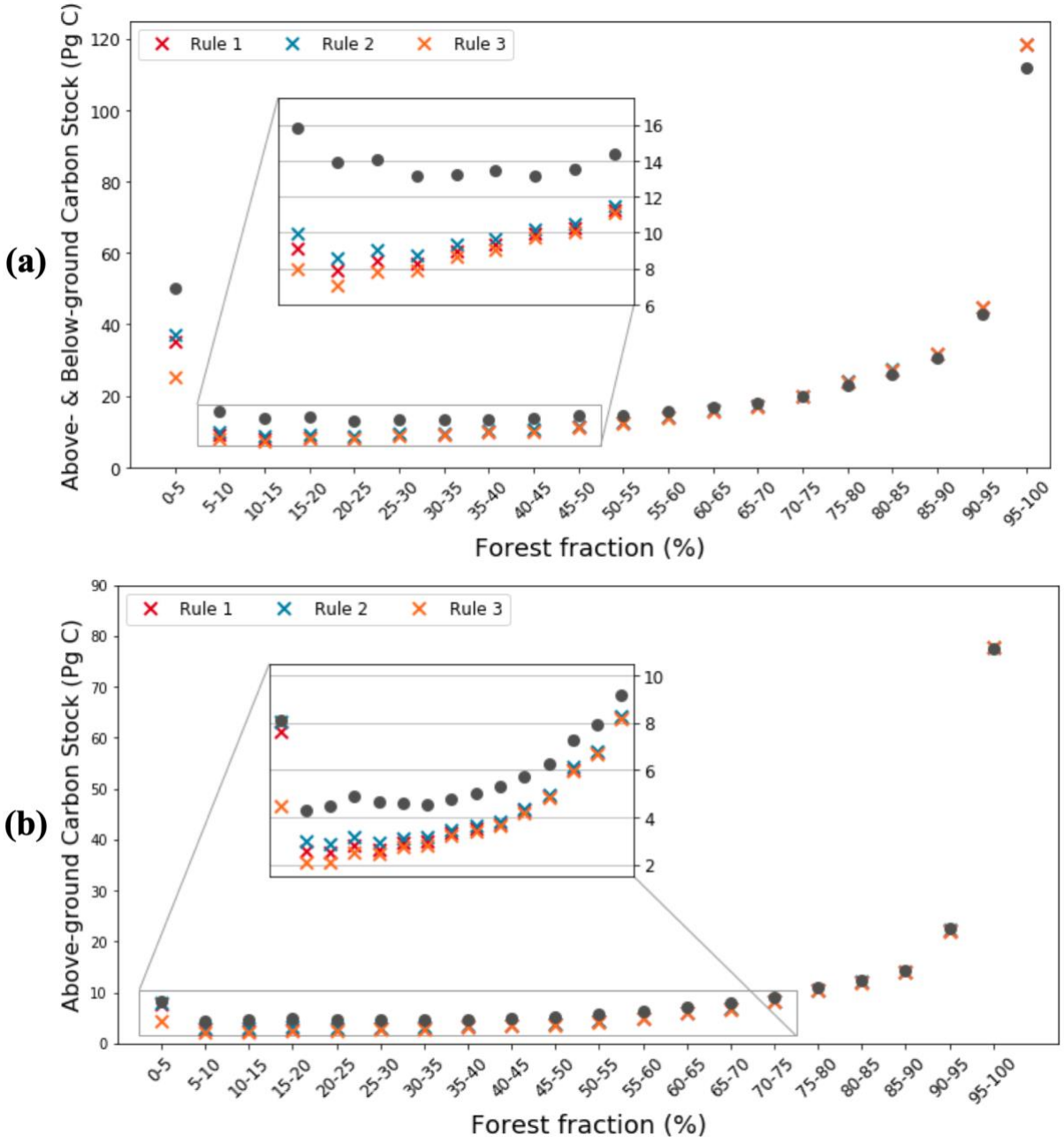

Figure 9. **Total carbon stock grouped by forest fraction from the averaged satellite-based forest cover map. (a) global (above- and below-ground); (b) pantropical (aboveground).**