RESEARCH ARTICLE

Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years

Prasanth MEIYAPPAN, Atul K. JAIN (✉)

Department of Atmospheric Sciences, University of Illinois, Urbana, IL 61801, USA

Abstract Earth's land cover has been extensively transformed over time due to both human activities and natural causes. Previous global studies have focused on developing spatial and temporal patterns of dominant human land-use activities (e.g., cropland, pastureland, urban land, wood harvest). Process-based modeling studies adopt different strategies to estimate the changes in land cover by using these land-use data sets in combination with a potential vegetation map, and subsequently use this information for impact assessments. However, due to unaccounted changes in land cover (resulting from both indirect anthropogenic and natural causes), heterogeneity in land-use/cover (LUC) conversions among grid cells, even for the same land use activity, and uncertainty associated with potential vegetation mapping and historical estimates of human land use result in land cover estimates that are substantially different compared to results acquired from remote sensing observations. Here, we present a method to implicitly account for the differences arising from these uncertainties in order to provide historical estimates of land cover that are consistent with satellite estimates for recent years. Due to uncertainty in historical agricultural land use, we use three widely accepted global estimates of cropland and pastureland in combination with common wood harvest and urban land data sets to generate three distinct estimates of historical land-cover change and underlying LUC conversions. Hence, these distinct historical reconstructions offer a wide range of plausible regional estimates of uncertainty and the extent to which different ecosystems have undergone changes. The annual land cover maps and LUC conversion maps are reported at $0.5° \times 0.5°$ resolution and describe the area of 28 land-cover types and respective underlying land-use transitions. The reconstructed data sets are relevant for studies addressing the impact of land-cover change on biogeophysics, biogeochemistry, water cycle, and global climate.
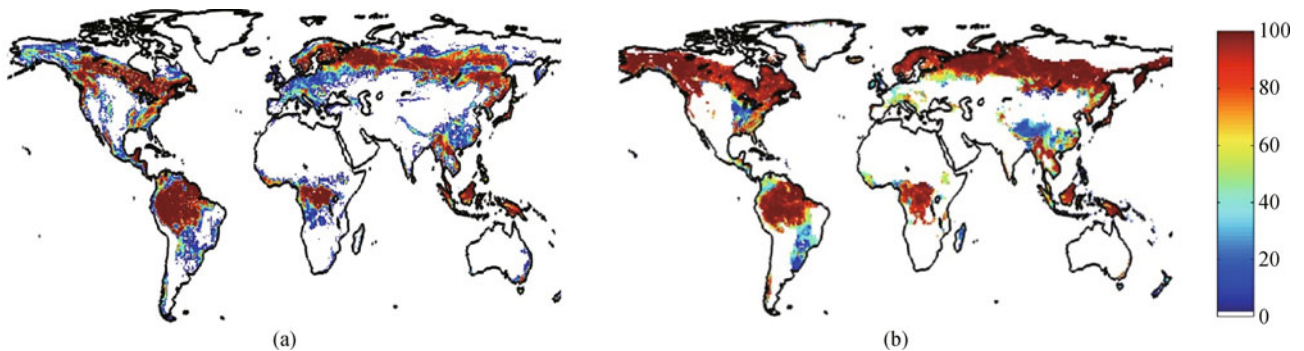
1 Introduction

Human activities have transformed natural ecosystems into managed areas in almost every part of the world (Foley et al., 2005; 2011). At present, nearly 40% of the Earth's ice-free land surface is being used for agricultural activities, all of which had previously been covered by natural vegetation (Ramankutty et al., 2008; Ellis et al., 2010). Such large-scale changes in land cover affect regional and global climate through biogeophysical (Bonan et al., 1992; Pielke et al., 2002; 2011; Feddema et al., 2005; Brovkin et al., 2006; Bala et al., 2007; Pitman et al., 2009; 2011; Findell et al., 2009) and biogeochemical (Jain and Yang, 2005; Canadell et al., 2007; Bonan, 2008; Jain et al., 2009; Pongratz et al., 2009; Shevliakova et al., 2009; Houghton et al., 2012) pathways.

Assessing the historical impacts of land-use/cover change (LUCC) at global scale (e.g. biogeophysical, biogeochemical, and climate effects) requires spatially and temporally explicit data sets on land cover and land-use/cover (LUC) conversions (replacement of one land-cover type by another) spanning several hundred years. Though remote sensing data provides a globally consistent picture of land cover, these data are only available for the past four decades (Houghton et al., 2012). Hence, several studies (e.g., Ramankutty and Foley, 1999; Klein Goldewijk, 2001; Hurtt et al., 2006; 2011; Klein Goldewijk et al., 2007; 2010; 2011; Olofsson and Hickler, 2008; Pongratz et al., 2008) have adopted different approaches in order to reconstruct spatially explicit data sets of dominant land-use activities (e.g., cropland, pastureland, urban land, wood harvest) covering several centuries. Typically, process-based modeling studies combine one or more of these land-use data sets with a map of potential

vegetation (representing primary land cover in the absence of human activities) to estimate the changes in land cover. The method adopted to replace potential vegetation varies from simple proportional clearing (e.g., Jain and Yang, 2005; Pitman et al., 2009) to a rule-based approach based on several logical assumptions and prioritizations that best describe the trends associated with historical LUCC (e.g., Hurtt et al., 2006; 2011).

Hurtt et al. (2006) developed a Global Land-use Model (GLM) to provide historical estimates of LUCC and LUC conversions due to expansion of cropland and pastureland, shifting cultivation and wood harvest at 1° spatial resolution. An updated version of GLM framework has recently been used in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) to provide estimates of LUCC and LUC conversions among five simple classes (cropland, pastureland, urban land, primary land, and secondary land) at 0.5°×0.5° resolution annually from 1500 to 2100 (AD) (Hurtt et al., 2011). This includes historical input data covering the period 1500–2005 and data for the four Representative Concentration Pathways (RCP) scenarios (Moss et al., 2010) for the future (2005–2100). The LUCC and LUC conversion estimates are usually translated to the specific land-cover classes suitable for use in a process-based model and subsequently used for impact assessments (e.g., Lawrence et al., 2012). Jain and Yang (2005) used a much simpler technique of superimposing the historical cropland data (based on Ramankutty and Foley, 1999) on a 0.5°×0.5° potential vegetation map (with each grid cell occupied by one potential vegetation) to estimate the changes in land cover. Similar but varying methods for superimposing a common cropland and pastureland were adopted by each of the seven climate models that participated in an inter-comparison study aimed at understanding the historical impact of land-cover change (Pitman et al., 2009). These estimates have been used as inputs to terrestrial carbon models, dynamic vegetation models, and earth system models to assess the impacts of LUCC (e.g.,

Shevliakova et al., 2009; Yang et al., 2010; Lawrence et al., 2012) on biogeophysics and/or biogeochemistry. However, most of these previous studies have not considered land-cover change arising due to indirect anthropogenic (e.g., climate driven land-cover change) or natural disturbances like fires, blowdowns, and insect outbreaks. Several local- to national-scale studies have demonstrated their importance and ecological significance (e.g., Giglio et al., 2010; van der Werf et al., 2010; also see Foley et al., 2003 and Lambin et al., 2003). For example, according to Forest Resources Assessment (FAO, 2006), 104 million hectares of forest on average were reported to be significantly affected each year by forest fire, pests (insects and disease), or climatic events such as drought, wind, snow, ice, and floods, with many countries missing this crucial information. In addition to differences arising from unaccounted land-cover change (indirect anthropogenic and natural causes), significant uncertainties could also arise due to heterogeneity associated with LUCC at temporal and spatial scales which cannot be captured using a rule-based approach of converting vegetation generalized at a regional or global scale. As a result, the global land cover estimated by most of the previous studies does not match estimates based on remote sensing data, a valuable tool in detecting several types of land-cover changes and land-cover modifications (subtle changes in land cover) that are difficult to map using other methods. For example, a comparison of forest area in 2005 from Hurtt et al. (2011) (estimated by combining information on primary and secondary land with a basemap which classifies each grid cell as either forest or non-forest based on potential vegetation biomass, as provided by Hurtt et al. (2006)) and 500 m resolution Moderate Resolution Imaging Spectro-radiometer (MODIS) Collection 5 land-cover data (Friedl et al., 2010) following International Geosphere-Biosphere Programme (IGBP) classification scheme (Loveland and Belward, 1997) aggregated to 0.5° × 0.5° resolution (hereafter referred to as MODIS-IGBP data) indicates pronounced differences in magnitude and spatial distribu-

Fig. 1 Global distribution of forest area during 2005 based on (a) MODIS-IGBP data, and (b) estimates by Hurtt et al. (2011) (Unit: % per grid cell area)

tion (Figs. 1(a) and (b)). Globally, Hurtt et al. (2011) estimated forest area was about 8×10^6 km^2 higher than the MODIS-IGBP estimated value of 31.5×10^6 km^2 in 2005. Similarly, other studies also overestimated the global forest extent for the recent past at similar magnitudes (refer to Sect. 4). It is essential to reconcile such differences in estimates, especially in the context of studies addressing the biogeophysical impacts of land-cover change.

The objective of this study is to build upon and extend the approaches of previous studies in order to provide estimates of historical land-cover change (and underlying LUC conversions) that are consistent with satellite observations. We use a rule-based approach to assign priorities for converting land cover due to various human land-use activities. Multiple years of satellite data sets are used to quantify the differences in estimates that may be arising due to unaccounted land-cover change and heterogeneity associated with LUCC that cannot be captured using simple rules for clearing vegetation. These differences are used to constrain and accordingly adjust the priorities for changing land cover, thereby producing land cover maps consistent with satellite observations for recent years. The work presented here takes into account land-cover change due to four major land-use activities: 1) cropland expansion and abandonment, 2) pastureland expansion and abandonment, 3) urbanization, and 4) regrowth due to wood harvest. Due to uncertainties associated with historical agricultural land-use, we have used three global historical data sets of cropland and pastureland (refer Sect. 2.1) in combination with a common data set for historical wood harvest and urban land, to produce three distinct estimates. The core products we generated were annual maps (at 0.5° × 0.5° resolution) of land cover and LUC conversions starting from the pre-industrial year of 1765 until 2010 or before (based on the ending time of the three cropland and pastureland data sets). The annual land cover data sets are reported as area fractions of 28 land-cover types (Table 1) for each 0.5° × 0.5° grid cell and the annual LUC conversion maps are reported as the area converted for each of the 92 unique conversions possible (refer supplementary Table S1) among the 28 land-cover types. The results are compared with other recently published model results and data-based studies. Finally, the sources of uncertainties in the present study are discussed.

2   Methods

The method used to characterize historical land-cover change can be described in five steps: 1) Historical land-use change data sets are processed to suit this study; 2) Land cover map for the year 1765 are generated by combining potential vegetation map, cropland, pasture-land, and urban land map for that year; 3) Land-cover change and LUC conversions starting with the 1765 land

cover map are estimated using a rule-based approach for prioritizing LUC conversion for each of the four land-use activities; 4) Estimates from the previous step are compared with satellite data. Priorities are accordingly adjusted to correct for the differences; 5) Grassland, pastureland, and cropland estimates are separated into C$_3$/ C$_4$ photosynthetic pathways.

2.1   Step 1: Processing of historical land-use change data sets

The three different data sets on cropland and pastureland are based on: 1) HYDE 3.1 (Historical Database of the Global Environment) (Klein Goldewijk et al., 2011), 2) new pastureland estimates and updated cropland estimates based on Ramankutty and Foley (1999) (N. Ramankutty, personal communication, 2011), and 3) regional estimates based on Houghton (2008). These three agricultural land use data sets are henceforth referred to as HYDE, RF, and HH data, respectively. The urban land data set is from Klein Goldewijk et al. (2010). Historical wood harvest data are based on annual wood harvesting rates from Hurtt et al. (2011). RF and HH data are at an annual time scale. The decadal time resolution HYDE data was linearly inter-polated to yield annual maps. All these data sets except HH data are gridded data sets at 0.5° × 0.5° or finer resolution. Finer resolution data were aggregated to 0.5° × 0.5° resolution. The HH data set provides the annual rate of deforestation/reforestation due to cropland, pastureland, wood harvest and shifting cultivation for ten regions (defined in Houghton et al., 1983) covering the entire globe, rather than by geographic details. HH regional data sets for cropland and pastureland resulting from deforesta-tion were converted to gridded estimates using the LUC conversion estimates derived based on RF data. Additional details on the method used to spatialize HH data, details and processing of other data sets are available in supplementary text. The three land-cover change and LUC conversion estimates generated from this study (henceforth referred to as ISAM-HYDE, ISAM-RF and ISAM-HH) based on three agricultural data sets (HYDE, RF, and HH) utilized the same wood harvest and urban land data. ISAM-HYDE, ISAM-RF, and ISAM-HH estimates extend to the year 2010, 2007 and 2005, respectively.

2.2   Step 2: Land cover map of 1765

A land cover map for the year 1765 was generated as a reference map to track land-cover change and LUC conversions. We started with the global map of potential vegetation derived at 5 min spatial resolution by Raman-kutty and Foley (1999). Fourteen of the 15 vegetation classes present in the potential vegetation map directly correspond to the potential land-cover types used in this study (Table 1). The land-cover classification used in this

Table 1   Land-cover classifications used in this study

No.	Land-cover type	Symbol	No.	Land-cover type	Symbol
1*	Tropical evergreen broadleaf forest	TrpEBF	15*	Polar desert/rock/ice	PdRI
2*	Tropical deciduous broadleaf forest	TrpDBF	16	Secondary tropical evergreen broadleaf forest	SecTrpEBF
3*	Temperate evergreen broadleaf forest	TmpEBF	17	Secondary tropical deciduous broadleaf forest	SecTrpDBF
4*	Temperate evergreen needleleaf forest	TmpENF	18	Secondary temperate evergreen broadleaf forest	SecTmpEBF
5*	Temperate deciduous broadleaf forest	TmpDBF	19	Secondary temperate evergreen needleleaf forest	SecTmpENF
6*	Boreal evergreen needleleaf forest	BorENF	20	Secondary temperate deciduous broadleaf forest	SecTmpDBF
7*	Boreal deciduous needleleaf forest	BorDNF	21	Secondary boreal evergreen needleleaf forest	SecBorENF
8*	Savanna	Savanna	22	Secondary boreal deciduous needleleaf forest	SecBorDNF
9*	C_3 grassland/steppe	C_3grass	23*	Water/Rivers	Water
10*	C_4 grassland/steppe	C_4grass	24	C_3 cropland	C_3crop
11*	Dense shrubland	Denseshrub	25	C_4 cropland	C_4crop
12*	Open shrubland	Openshrub	26	C_3 pastureland	C_3past
13*	Tundra	Tundra	27	C_4 pastureland	C_4past
14*	Desert	Desert	28	Urban land	Urban

Note: * Natural land-cover classes used in this study. Except for water/rivers (No. 23), all other natural land-cover classes were directly derived from the potential vegetation map of Ramankutty and Foley (1999). Please note that C_3 and C_4 grassland (Nos. 9 and 10) are considered to be a single land-cover class in the potential vegetation map and during the initial stages of calculation. Partitioning to C_3 and C_4 types is carried out in the last step (Sect. 2.5)

study is chosen to be consistent with the land-cover types required for the Integrated Science Assessment Model (ISAM) (Jain and Yang, 2005; Yang et al., 2009; 2010) for which we originally produced these data sets. Mixed forest (which is not part of ISAM land-cover classification) from the potential vegetation map was reclassified into any one of the seven forest types by searching for dominant (greater than 70% of the area considered) forest type within a $4° \times 4°$ resolution window around the grid cell. The window size was increased until the requirements for dominant forest type were satisfied. Savanna (usually defined as tropical grasslands) present outside tropical regions was reclassified to other herbaceous types, using the method adopted for reclassifying mixed forest. Ramankutty and Foley (1999) assigned single potential vegetation to each 5 min grid cell from 1 km DISCover satellite-based global land cover data (Loveland and Belward, 1997) even in grid cells where anthropogenic land cover was absent. In such grid cells, we used MODIS-IGBP data for the year 2005 to reassign the grid cell area (currently occupied by either 100% forest or non-forest) to fractional area of forest and non-forest. The forest and non-forest types were determined using a combination of MODIS-IGBP data and the method adopted to reclassify mixed forest. This reduced the total area of forest in the potential vegetation map from 55.2×10^6 km² to about 48.6×10^6 km². An additional land-cover class (water-covered areas) map was derived at 5 min resolution using MODIS-IGBP data, and was included in the potential vegetation map by proportional adjustment of potential vegetation areas.

Next, we aggregate the 5 min resolution potential vegetation map to $0.5° \times 0.5°$ resolution to yield the fractional areas of 15 land-cover types within each grid cell. Hence, each grid cell in our potential vegetation map can be occupied by more than one type of natural vegetation. We assume water-covered areas to be constant for every year.

Finally, we derive the land cover map for the year 1765 by including the 1765 cropland and pastureland maps from RF and the urban land map (Klein Goldewijk et al., 2010); the $0.5° \times 0.5°$ resolution potential vegetation map is generated by simple proportional adjustments to the potential vegetation areas within each grid cell. The map was used as a starting point to produce all three estimates of land-cover change and LUC conversions. We also assume all forest in the 1765 land cover map as primary forest. At this stage, we do not distinguish between C_3/C_4 types for grassland, pastureland, and cropland. Classification to C_3/C_4 pathways is accomplished in the final step (Sect. 2.5).

2.3   Step 3: Estimating historical LUCC and LUC conversions
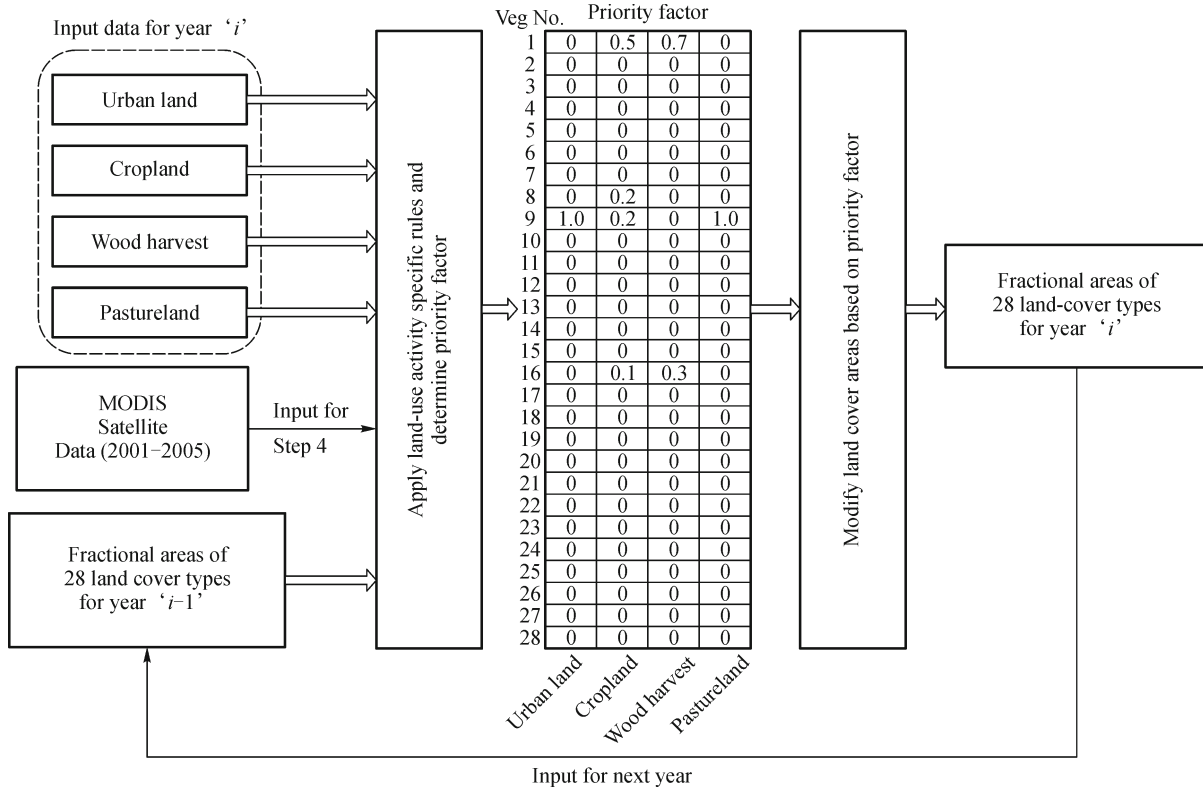
To derive the LUCC and LUC conversion estimates, we define a set of rules to characterize each of the four land-use activities. These rules impose a logical sequence and priority order in which land cover is modified. Based on these rules, a priority factor is assigned to each land-cover type within each grid cell, corresponding to each of the four land-use activities (Fig. 2). The priority factor for a land-cover type indicates the probability of that vegetation being altered due to that particular land-use activity. The priority factor for an individual land-cover type within

each grid cell varies from 0 to 1.0, and the sum of priority factors for all land-cover types corresponding to each land-use activity sums up to 1.0.

The rules that determine the priority factors for a land-use activity depend on the magnitude of that land-use activity for that year, the land cover map from the previous year, and the potential vegetation map. For example, for an increase in cropland area between two consecutive years in a grid cell, a priority factor is assigned to each land-cover type (except for water, pastureland and urban land for which priority factor is assumed to be 0), which is proportional to the total area of natural vegetation in that grid cell. The increase in cropland area is accounted by converting each land-cover type to cropland based on its designated priority factor. In the case of cropland abandonment (characterized by decrease in cropland area with time), the abandoned land reverts back to the potential vegetation level present in that grid cell. In such cases, the potential vegetation map was used to determine the priority factors. Usually, grasses and other herbaceous land-cover types are faster colonizers than forests (Arora and Boer, 2006). They invade the abandoned land initially, while woody vegetation grows later. However, the rationale here is that a one-year time gap is sufficient for woody vegetation to reappear. This method provides a simple

representation of successions. LUCC treatment due to urbanization is similar to that described for cropland, with the exception that in case of decrease in urban land area with time, the decreased area is reverted to grasses (i.e., priority factor for grasses was assigned as 1.0), irrespective of the potential vegetations present within that grid cell. For wood harvest, preference is given to primary forests over secondary forests. Priority factors were assigned proportional to the area of each of the seven primary forests within that grid cell. In cases where total primary forest was insufficient to account for wood harvest, clearing was done from secondary forests following a similar approach. For an expansion of pastureland, clearing of grassland is preferred (Houghton, 1999). In cases where grassland is insufficient, we followed the method adopted for increase in cropland area. In case of decrease in pastureland area, the abandoned area was reverted back to grassland.

There are a few exceptions to these rules. In cases where cropland is abandoned and pastureland/urban land con-currently increases with time, a part of the abandoned area was considered a source for pastureland/urban land. The fraction of abandoned cropland area used as a source of pastureland/urban land is determined by the likelihood that the other vegetations present in the grid cell are sources for the growth in pastureland/urban land. For example, a grid

Fig. 2 Schematic diagram showing the process involved in Step 3 to estimate LUCC and LUC conversions. Step 4 involves modification of priority factors estimated from Step 3 using forest area estimated from MODIS-IGBP data. 'i' denotes year, which increases from 1765 to 2005/2007/2010 (ISAM-HH/ISAM-RF/ISAM-HYDE) in annual time steps. The priority factors shown here are just an example, and they vary for each land-cover type from year to year between each grid cell

cell dominated by forest is more likely to have a higher fraction of abandoned cropland area to be used as a source of pastureland than a grid cell dominated by grassland. Similar treatment exists for decrease in pastureland area accompanied by increase in cropland/urban land, in which a part of cleared pastureland area is considered a source for cropland/urban land. It should be noted that in case of succession, forest returns as secondary forest (vegetation numbers 16 to 22 in Table 1), whereas we have not differentiated herbaceous land-cover types as primary/ secondary. Because the data sets for four land-use activities considered in this study come from more than one source, certain cases exist where the desired conversations, as determined by the assigned prioritization factor, could not be carried out for all four land-use activities. In such cases, we assign the following order of preference to modify land cover: urban land, cropland, wood harvest, and pasture-land. This order of preference was chosen considering the uncertainties in magnitude, spatial distribution, and definitions associated with each land-use activity. Hence, the cropland and pastureland areas in ISAM-RF, ISAM-HYDE, and ISAM-HH will be slightly less compared to the original RF, HYDE, and HH data sets in certain grid cells.

The land cover map of 1765 derived from Step 2 (Sect. 2.2) is used as the initial condition from which we move forward in time, modifying land cover by superimposing the year-to-year land-use activities following the method described above.
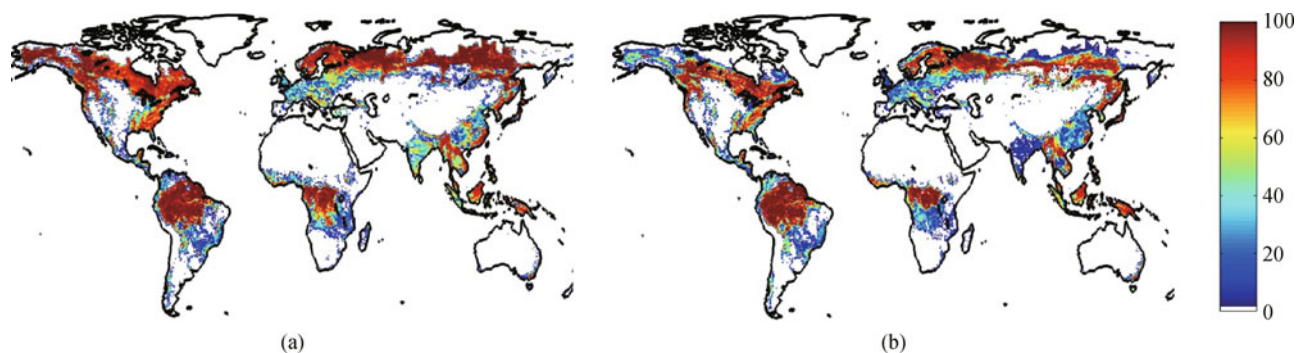
2.4   Step 4: Calibration using satellite data

Historically, substantial land cover-changes have occurred due to climate feedbacks (Parry et al., 2007) and through natural disturbances like forest fires (Giglio et al., 2010; van der Werf et al., 2010), blowdowns, and insect outbreaks (Foley et al., 2003; Lambin et al., 2003). Due to the unavailability of information on the magnitude and spatial extent to which these effects have altered land cover historically at a multi-centennial time scale, their impacts on land cover have been excluded from the rule-based

approach for estimating historical land-cover change. Additionally, the rule-based approach is a simplified representation of general trends associated with historical land-cover change due to human land-use activities, which is subject to variations at the regional and grid cell levels. Due to the factors discussed above, there exist differences between satellite observations and estimates from the rule-based approach (Step 3; Sect. 2.3). For example, the total forest area estimated using rule-based approach (Fig. 3(a)) differs from satellite estimates (Fig. 1(a)) for certain grid cells. Our estimated forest area varies from 36.7 to 39.4 million km^2 among the three estimates, compared to 30.7 million km^2 (after changing to the land mask used in this study) estimated using MODIS-IGBP data. We implicitly account for these differences by calibrating with satellite data.

We first classify the 28 land-cover classes into two broad categories: forest and non-forest. Medium resolution satellite data captures forest extent/type with high accuracy compared to other herbaceous types (Jung et al., 2006; Friedl et al., 2010). The basic aim is to reconcile these in a way that will make the magnitude and spatial patterns of present-day forest estimates as close as possible to satellite estimates.

We compare estimated forest area for the year 2005 with estimates from 500 m resolution MODIS-IGBP data for the year 2005. An overestimation of forest area in a grid cell indicates that higher priority factor should be assigned to forest land-cover types for clearing than previously assumed. Similarly, an underestimation of forest area in a grid cell indicates a lesser priority factor should be assigned to forest land-cover types for clearing. To modify the priority factor for each land cover type in a grid cell for a particular year, we determine a "correction factor" using a combination of information from the potential vegetation map, the land cover map for the year 2005, the land cover map of the historical year for which the priority factor is to be adjusted, and the magnitude of underestimation/over-estimation of forest area estimated in comparison to satellite data. The correction factor for each land-cover type is chosen such that the estimated area of forest

Fig. 3   Estimated global forest area for the year 2005 based on ISAM-RF, (a) Without calibration (b) after calibration using MODIS-IGBP data (Unit: % per grid cell area)

matches with satellite data when the correction factor is multiplied by the priority factor estimated in Step 3. The value of the correction factor is > 1 for land-cover types with increased priority and < 1 for land-cover types with decreased priority. An additional constraint is imposed so that the sum of the correction factor multiplied by priority factor for all land-cover types, corresponding to each land-use activity, add up to 1.0, a basic criteria described in Step 3 (Sect. 2.3). For the grid cells where land-use data indicates the absence of anthropogenic land-cover types, a simple linear interpolation is used to adjust the area of natural vegetations between the starting and ending reference years, in order to make the present-day estimates consistent with satellite data. A similar approach was applied to grid cells where the magnitude of historical land-use was small and correction factor alone is insufficient for effecting the changes needed to match satellite estimates. The changes effected through linear interpolation are reflected in annual land cover maps, but are not recorded as LUC conversions. Hence, our estimates of LUC conversions are only attributable to the four direct human land-use activities. To avoid underestimation of forest area from satellite data, which may result due to the exclusion of regrowing forest, we also use four additional years of MODIS-IGBP data covering the period 2001–2004 to estimate the 'correction factor'. This method results in a close match between MODIS-IGBP forest distribution (Fig. 1(a)) and our estimated forest distribution (Fig. 3(b)).

2.5 Step 5: Separation of grassland, pastureland, and cropland to C_3/C_4 types

We only classify grassland, pastureland, and cropland to C_3/C_4 types in annual land cover maps, not annual LUC conversion maps. To separate the grassland and pastureland area fractions into C_3 and C_4 types, we followed the modified approach of Still et al. (2003). If there is at least one month in a year when temperature is above crossover temperature (the temperature at which the C_3 quantum yield equals C_4 quantum yield) and rainfall is concurrently above 25 mm, it is assumed that the C_4 grass fraction is equal to the number of months where C_4 photosynthesis is favored relative to the number of growing season months with a temperature greater than 5°C. Mathematically,

C_4 fraction = (number of months with T_{air} > crossover temperature and rain > 25mm) / (number of months with T_{air} > 5°C)

We use the monthly air temperature (T_{air}) and precipitation data at 0.5°×0.5° resolution based on CRU TS 3.0 (updated based on Mitchell and Jones, 2005), covering the period 1901–2006; a 10-year moving average was calculated for both variables, to avoid sudden fluctuations. For the years 1765 to 1900, average monthly precipitation and temperature values from 1901 to 1910 were used. For the period 2007–2010, the same values were assigned as

for the year 2006. For each year, we calculated the crossover temperature following Collatz et al. (1998), based on global CO_2 concentration values from 1765 to 2010 (Meinshausen et al., 2011). The calculated crossover temperature varies from 18.2°C in 1765 to 24.1°C in 2010. The C_4 fraction generated for the period 1765–2010 was combined with annual pastureland and grassland estimates from Step 4 (Sect. 2.4), to separate them into C_3 and C_4 fractions.

To separate the annual cropland area into C_3 and C_4 fractions, we use the estimates of harvested areas of 175 different crops across the world at 5 min by 5 min spatial resolution for the year 2000 (Monfreda et al., 2008). C_3 and C_4 designations were assigned to each crop type based on known pathway characterizations. A map indicating the fractional coverage of C_4 cropland was generated at 0.5° × 0.5° resolution. In grid cells where there were no crop present, 100% of the grid cell was assigned to C_3 cropland. This map was used to separate annual historical cropland estimates into C_3 and C_4 types.

3 Results

3.1 Comparison of cropland and pastureland estimates among 3 data sets

Comparison of global cropland statistics of RF and HYDE data averaged over the period 2001–2005 shows similar levels of cropland area, varying from 14.3 to 15.3 million km², with HYDE estimates being 8% higher than RF estimates (Table 2). However, this global picture varies regionally. The most pronounced differences are found in Pacific Developed region and China, where the cropland areas estimated by HYDE data are 70% and 23% higher than RF data, respectively. The major differences between the two data sets result from the fact that these data sets adopted different methods (Refer supplementary text) and agricultural inventory data sets. While HYDE inventory data was based on FAO (2008), RF estimates relied more on national-level census statistics, along with FAO estimates for recent years (Ramankutty et al., 2008). HH estimates of both global and regional cropland are lower than RF and HYDE estimates. This is because HH data includes only cropland that were created or abandoned on land originally covered by forest.

While global cropland statistics estimated based on RF and HYDE data match reasonably well with each other, pastureland statistics globally show substantial disagreement, with even more regional disagreement. This is because the global pastureland area estimated by the census report used in RF itself is significantly lower than FAO (2008) estimates of pastureland used in HYDE data. Globally, HYDE data estimates of pastureland are 26% higher than the RF estimated value of 26.3 million km² average over the period 2001–2005. Major disagreement is

Table 2 Regional areas of cropland and pastureland averaged for the period 2001–2005 estimated directly from RF (Updated estimates based on Ramankutty and Foley, 1999), HYDE (Klein Goldewijk et al., 2011) and HH (Houghton, 2008) data sets across nine regions covering the world. The nine regions are based on Houghton et al. (1983). Units are in million km². All values are rounded to one decimal place

Regions	Cropland				Pastureland			
	RF	HYDE	HH	Range	RF	HYDE	HH	Range
North America	2.1	2.3	1.9	1.9 – 2.3	2.4	2.5	0.0	0.0 – 2.5
Latin America	1.6	1.5	1.4	1.4 – 1.6	4.8	5.4	2.8	2.8 – 5.4
Europe	1.2	1.2	0.1	0.1 – 1.2	0.6	0.7	0.0	0.0 – 0.7
North Africa and Middle East	0.8	0.9	0.3	0.3 – 0.9	1.8	3.0	0.0	0.0 – 3.0
Tropical Africa	2.0	2.0	1.9	1.9 – 2.0	7.0	8.0	0.0	0.0 – 8.0
Former USSR	2.0	2.2	0.4	0.4 – 2.2	3.3	3.6	0.0	0.0 – 3.6
China	1.3	1.6	0.7	0.7 – 1.6	3.5	5.2	0.0	0.0 – 5.2
South & South-East Asia	3.0	2.9	1.5	1.5 – 3.0	0.3	0.4	0.0	0.0 – 0.4
Pacific Developed Region	0.4	0.6	0.2	0.2 – 0.6	2.6	4.1	0.0	0.0 – 4.1
World	14.3	15.3	7.6	7.6 – 15.3	26.3	33.0	2.8	2.8 – 33.0

found over 'North Africa and the Middle East' where pasture area estimates for HYDE are 83% higher than RF for 2005. While the percentage difference is highest for 'North Africa and the Middle East', a large difference in pastureland area is found in the Pacific Developed region and China, where the estimated pastureland area averages from 2001 to 2005 for HYDE are 1.5 million km² (~53%) and 1.7 million km² (~43%) higher than RF data, respectively. HH data estimates of pastureland are zero for all regions except Latin America, because Houghton (2008) assumes that all pastures are derived from grasslands, with the exception of Latin America, where significant clearance of forest area for pastureland has taken place due to extensive cattle ranching (Lambin and Geist, 2003).

3.2 Land-cover change estimates during 1765–2005

The 28 land-cover classes have been combined into a broader category for the purpose of analysis (Table 3), and the values are presented in the text as range among three estimates (ISAM-HYDE, ISAM-RF and ISAM-HH). Globally, the total area of forest has decreased from 45.5 million km² (~36% of the total land area) to about 29–30 million km² during this period, a one-third decrease. Of this, human land-use activities have contributed to a net decrease in forest area of about 6.5–8.4 million km² (Table 4), while the rest is attributed to indirect anthropogenic and natural causes. Total deforestation amounts to 14.5–14.7 million km², and forest regrowth ranges between 6 and 8 million km². Forest area in North America shrunk by 3–3.5 million km² (~35%–40%) and Tropical Africa shrank by 2.3–2.6 million km² (~43%–49%) (Table 4). Total forest area in Europe decreased by 44%–52% from its initial value of 2.5 million km². Estimates of forest area in

China and South and South-East Asia (SSEA) regions show the largest difference among the three estimates. Forest area in China and SSEA decreased by 40%–52% and 47%–66%, respectively. Such large differences in these regions are mainly due to uncertainty in estimates of cropland (see Ramankutty et al., 2008; Liu and Tian, 2010). North America, the former USSR and Tropical Africa show a large amount of net forest loss attributed to indirect anthropogenic and natural causes. Total forest regrowth due to human land-use activities is about 6–8 million km². During 2005, roughly 24%–28% of the total forests present are secondary forests (Fig. 4 and Table 3). North America contains about 26% of global secondary forest whereas the former USSR contains 17–23% of global secondary forest (Fig. 5). ISAM-RF estimates show higher secondary forest in all regions due to more abandonment of cropland present in RF data compared to HYDE data.

Global area of savanna shrank by 5.4–7.1 million km² (i.e., 38%–50%) and shrubland decreased by 6.8–8.9 million km² (i.e., 40%–53%) (Table 3). The area of grassland and pastureland combined increased by about 19.7–24 million km² (i.e., 83%–101%). However, regional comparisons show more disagreement than global estimates of change (Fig. 5). For a single time snap during 2005, ISAM-HYDE estimates show 57% less shrubland area in the Pacific Developed region compared to ISAM-RF estimates. Except for North America, ISAM-RF shows more cropland expansion in regions that were originally shrubland, compared to ISAM-HYDE. The area of grassland in ISAM-RF is higher than ISAM-HYDE for all regions because of lower pastureland estimates by RF data compared to HYDE data. As we have considered only deforestation and reforestation statistics due to agricultural activities from HH data, they have been excluded in the

Table 3 Global area of various land cover types for 4 time slices based on ISAM-RF, ISAM-HYDE, and ISAM-HH estimates. 'Primary forest' includes TrpEBF, TrpDBF, TmpEBF, TmpENF, TmpDBF, BorENF, and BorDNF. 'Secondary forest' includes SecTrpEBF, SecTrpDBF, SecTmpEBF, SecTmpENF, SecTmpDBF, SecBorENF, and SecBorDNF. Shrubland is a combination of Denseshrub and Openshrub. 'Others' category includes Tundra, Desert, and PdRI. The estimates of cropland and pastureland area are slightly lower than the original estimates (Table 2) due to a difference in land mask used and other minor adjustments made in Step 3 (Sect. 2.3) for consistency purposes (Unit: million km^2)

Land-cover type	1765	1900			2000			2005		
	ISAM-RF/ISAM-HYDE/ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH
Primary forest	45.4	34.9	34.8	33.5	22.1	22.5	20.8	21.7	22.2	20.3
Secondary forest	0.0	2.9	2.9	3.1	7.9	7.0	7.5	8.3	7.2	7.8
C_3 cropland	2.9	5.9	6.2	4.2	10.0	11.4	5.5	10.0	11.6	5.6
C_4 cropland	0.6	1.7	1.8	1.2	2.9	3.4	1.5	2.9	3.4	1.5
C_3 pastureland	3.0	9.1	9.1	3.3	18.0	24.4	4.2	18.0	24.6	4.3
C_4 pastureland	1.2	3.0	3.6	1.7	5.9	7.7	2.6	5.5	7.3	2.6
C_3 grassland	14.6	15.6	15.4	20.2	16.5	13.8	26.0	17.2	14.1	26.4
C_4 grassland	4.9	4.1	3.7	5.8	2.7	1.8	4.5	2.7	1.7	4.2
Savanna	14.2	13.0	12.5	14.2	9.1	7.2	14.2	8.8	7.1	14.2
Shrubland	16.9	14.1	14.6	16.8	10.1	8.0	16.8	10.1	8.0	16.8
Others	26.1	25.7	25.4	26.1	24.4	22.5	26.1	24.4	22.5	26.1
Urban land	0.0	< 0.1	< 0.1	< 0.1	0.4	0.4	0.4	0.5	0.5	0.5

Table 4 Area of forest cleared and forest regrown during the period 1765–2005 across nine regions covering the world, based on ISAM-RF, ISAM-HYDE, and ISAM-HH estimates. Total deforested and forest regrowth estimates are based on four land-use activities only. However, changes in forest area effected due to calibration with satellite data (Step 4; Sect. 2.4) are reflected in year 2005 forest estimates (Unit: million km^2)
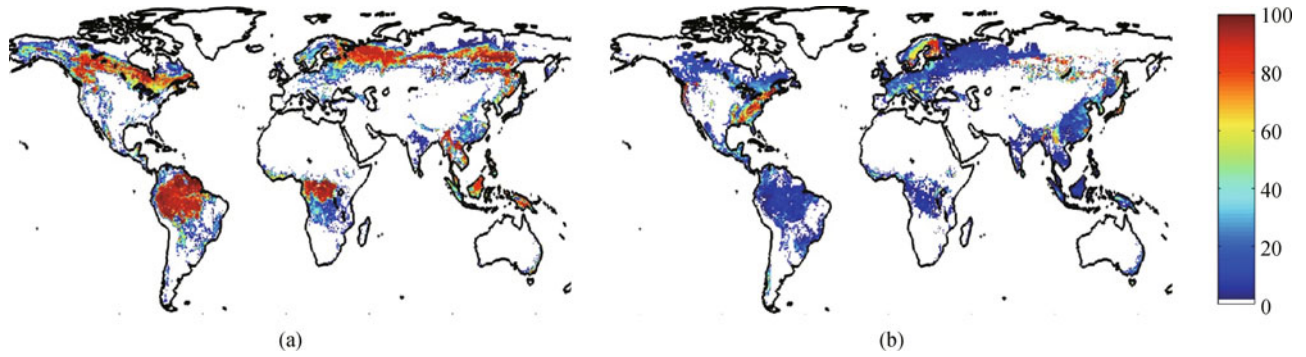
Regions	Forest area in 1765	Total deforested area			Total forest regrowth			Estimated forest area in 2005
		ISAM-RF	ISAM-HYDE	ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH	
North America	9.6	3.3	3.5	3.3	2.4	2.0	2.2	5.8–6.2
Latin America	10.5	3.1	2.4	4.5	1.0	0.6	1.2	8.4–8.8
Europe	2.5	2.0	1.6	1.3	1.5	1.0	1. 1	1.2–1.4
North Africa and Middle East	0.2	0.1	0.1	0.1	< 0.1	< 0.1	< 0.1	~0.1
Tropical Africa	5.3	1.2	1.2	0.9	0.4	0.3	0.5	2.7–3.0
Former USSR	8.1	1.4	1.8	0.9	0.8	1.1	0.7	5.9–6.0
China	2.3	1.1	1.1	0.7	0.8	0.3	0.7	1.1–1.4
South & South-East Asia	5.8	2.0	2.1	2.4	0.7	0.4	1.2	2.0–3.1
Pacific Developed Region	1.2	0.4	0.4	0.4	0.3	0.2	0.3	~1.1
World	45.5	14.7	14.4	14.5	8.0	6.0	8.0	28.3–30.0

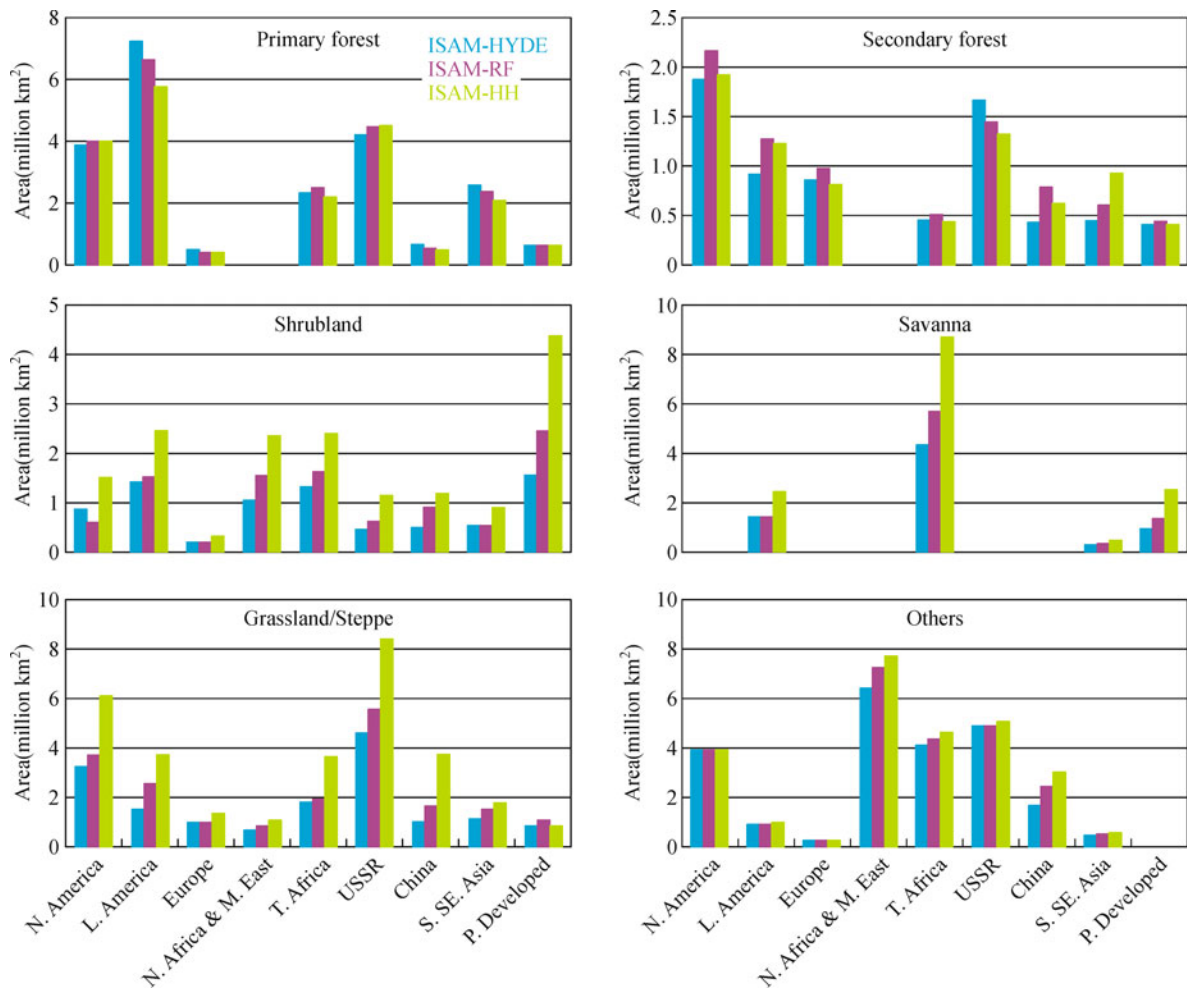discussion relating to comparison of herbaceous land-cover types.

3.3   LUC conversions during 1765–2005

Globally 6.6–6.8 million km^2 of forest loss (~45% of human-caused forest loss) has occurred due to cropland expansion, whereas only 2.7–2.9 million km^2 was due to pastureland expansion (Supplementary Table S2). SSEA contributes to 25%–30% (1.6–2.1 million km^2) of forest loss occurring due to conversion to cropland, followed by

North America (1.1–1.5 million km^2; 16%–23%). Although the cropland estimates for Latin America by RF, HYDE, and HH are in close range of 1.4–1.6 million km^2 for the early 2000s (Table 2), their pathways of expansion are very different. ISAM-HYDE estimates only 0.65 million km^2 of forest loss in Latin America due to cropland expansion, whereas ISAM-RF shows almost double the forest loss estimated by ISAM-HYDE (Supplementary Table S2). Because HH data was spatialized using ISAM-RF estimates, the trend exhibited by ISAM-HH cannot be considered independent from ISAM-RF

Fig. 4 Estimated (a) primary and (b) secondary forest area for the year 2005 based on ISAM-RF (Unit: % per grid cell area)

Fig. 5 Regional comparisons of various natural land-cover types during 2005 based on ISAM-RF, ISAM-HYDE, and ISAM-HH. 'Primary forest' includes TrpEBF, TrpDBF, TmpEBF, TmpENF, TmpDBF, BorENF, and BorDNF. 'Secondary forest' includes SecTrpEBF, SecTrpDBF, SecTmpEBF, SecTmpENF, SecTmpDBF, SecBorENF, and SecBorDNF. Shrubland is a combination of Denseshrub and Openshrub. Grassland is a combination of C_3 grass and C_4 grass. 'Others' category includes Tundra, Desert and PdRI

estimates. Roughly 47%–58% (~7.4–9.6 million km²) of cropland expansion has occurred due to conversion of non-forested land (Supplementary Table S2).

About 49%–62% of forest loss due to human land use in

Latin America occurred due to conversion to pastureland, compared to 29%–36% caused due to cropland expansion (from Supplementary Table S2 and Table 4). Globally, 28.5–31.8 million km² of non-forested land was used for

Table 5 Comparison of regional forest area estimated in this study with other published studies for the year 1990. The results from this study are provided as a range of forest area estimated from ISAM-RF, ISAM-HYDE, and ISAM-HH. An additional 'test case' was performed (following UMD land classification scheme) to facilitate direct comparisons with FAO estimates (Unit: million km^2)

Regions	Yang et al. (2010)	Klein Goldewijk (2001)	Hurtt et al. (2006)	IPCC AR5[a]	This study	Test case	
						FAO[b]	This study (UMD scheme)
North America	9.5	8.7	9.3	9.3	5.8–6.0	5.1	4.1–4.5
Latin America	9.0	9.2	9.0	8.6	7.4–8.3	10.2	9.8–10.1
Europe	2.1	2.2	1.6	1.5	1.3–1.4	1.7	1.5
North Africa and Middle East	0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.1	0.4
Tropical Africa	4.3	3.3	4.4	4.0	2.8–3.15	6.9	7.0–9.8
Former USSR	11.0	11.9	9.7	10.0	5.9–6.0	8.1	6.3–6.5
China	1.0	1.3	2.5	2.0	1.2–1.35	1.7	1.8–2.0
South & South-EastAsia	4.1	3.3	3.3	3.4	3.1–3.2	3.6	3.3–3.4
Pacific Developed Region	1.2	1.4	1.1	1.1	1.1	2.2	2.4–3.7
World	42.3	41.5	40.9	39.9	29.0–30.1	39.6	37.2–41.3

Notes: [a] Based on Hurtt et al. (2011), [b] from Global Forest Resource Assessment (FRA) 2010

pastureland expansion, the majority of which consisted of grasslands. It is interesting to note that though the areas of cropland and pastureland estimated by RF are about 1 million km^2 and 6.7 million km^2 lower than HYDE estimates, respectively, for the period 2001–2005 (Table 2), ISAM-RF estimates show substantially more clearing (and regrowth) of forested and non-forested land compared to ISAM-HYDE (Supplementary Table S2 and Table 4). This is because HYDE data show a consistently increasing trend in cropland and pastureland area over time, compared to RF data which show substantially more abandonment (and thus more regrowth of natural vegetation), leading to more gross conversions by ISAM-RF. ISAM-RF and ISAM-HH estimates show ~42% contribution of total secondary forest regrowth due to cropland abandonment, whereas ISAM-HYDE show a contribution of only about 23%.

3.4   Implications of partitioning to C_3/C_4 type

A significant amount of land in North America and Europe estimated as C_4 grassland in 1765 was classified as C_3 grassland for present-day conditions when changes in CO_2 concentration were taken into account in the simplified method of Still et al., (2003) (Refer to Supplementary Fig. S1). About 10% of the grassland and 22.4% of pastureland from ISAM-HYDE was classified as C_4-type for 2010 (Table 3). Combining the same grassland and pastureland estimates for 2010 from ISAM-HYDE with the C_4 fraction map for the year 1765 resulted in 18.4% and 32% classified as C_4-type, respectively. Both ISAM-HYDE and ISAM-RF estimates show about 23% of the total cropland area as C_4-type throughout the historical period.

4   Comparison with other studies

We compared our estimates of forest for the year 1990 with other studies (Table 5). The year 1990 was chosen for comparison because it is the farthest year from present for which many gridded estimates were available that would facilitate regional comparisons. All the previous modeling (Klein Goldewijk, 2001; Hurtt et al., 2006; Yang et al., 2010; Hurtt et al., 2011) studies show good agreement with one another, even regionally. However, global total forest area estimates from ISAM-HYDE, ISAM-RF, and ISAM-HH are about 10 million km^2 less than previous studies. Major disagreements occur in North America and the former USSR, where our estimates of forest area are reduced by 3.3 and 4.7 million km^2, respectively. Because our forest estimates are a reflection of estimates from satellite-based land cover data, the differences in estimates arising in these regions can be attributed to unaccounted land-cover change, assuming the rule-based approach accurately captures land-cover change occurring due to all major land-use activities.

We compared our estimates with FAO forest statistics for 1990 (FAO, 2010). Due to difference between the definition of forest used in FAO (see FAO, 2001; 2006; 2010) and this study, we performed a 'test case' wherein we repeated the entire calculations using a potential vegetation map derived from MODIS land-cover data from the year 2005 (Friedl et al., 2010) classified under the University of Maryland (UMD) classification scheme (Hansen et al., 2000). Land-cover classification in the UMD scheme is favorable for making direct comparisons with FAO estimates. In this case, the estimates seem to agree reasonably well with FAO statistics, with ISAM-

HYDE estimates being at the high end for Tropical Africa and Pacific Developed regions. Similar trends were observed when MODIS estimated forest area (UMD classification scheme) for the year 2005 was directly compared with FAO forest estimates for the same year (FAO, 2010).

5 Discussion and conclusions

This study focused on characterizing historical land-cover change and LUC conversions using annual maps of cropland, pastureland, wood harvest, and urban land as inputs. Due to uncertainties associated with estimates of historical land-use activities, three different data sets on agricultural extent were used to derive three different estimates, consistently using the same rule-based method of prioritizing and converting vegetation. Information from remote sensing data was used to constrain and modify the rule-based method to implicitly account for land-cover changes due to indirect anthropogenic or natural causes. The differences among the three estimates produced in this study can be largely explained by the spatial and temporal differences in estimates of cropland and pastureland areas among the three data sets. Therefore these data sets offer a wide range of plausible regional estimates of uncertainty and the extent to which different ecosystems have undergone changes historically.

The data sets produced in this study have several associated limitations. Since the annual cropland and pastureland maps reveal only the net changes in area, we could not calculate the effect of shifting cultivation in this study. Hurtt et al. (2006) performed a sensitivity test by assuming a standard land abandonment rate of 6.7% y^{-1} due to shifting cultivation in the tropics, and showed that excluding shifting cultivation could lead to underestimation of secondary land created by agriculture. However, we chose not to include shifting cultivation in our study due to high uncertainty in the magnitude and spatial patterns (Hurtt et al., 2006) historically associated with shifting cultivation. In addition, our assumption that all forest on the land cover map for 1765 (starting year of analysis) was primary forest potentially underestimates the secondary forest area created due to wood harvest and cropland abandonment before 1765. The validity of this assumption is well established due to the fact that the aim of this study is to characterize land-cover change after the pre-industrial era.

There are three major sources of uncertainty. First, the potential vegetation map produced from satellite data is assumed to accurately represent the land cover that would have existed at present if human activities have been non-existent. Hence, the usage of potential vegetation map to represent pre-industrial land-cover assumes that changes in environmental conditions have not changed the land cover. The second source of uncertainty arises from the rule-based approach to prioritize land-cover change used in this method; this is a simple representation of historical trends associated with various land-use activities that are not fully understood, and difficult to generalize at a global or regional scale. As shown, the rule-based approach leads to a land cover map that differs substantially compared to satellite estimates for recent years. However, we attribute the differences to unaccounted land-cover change and grid-cell level variations in land-use trends assumed in our rule-based approach. This difference is subsequently used to revise the rules at the grid-cell level to produce estimates close to satellite observations. Hence, the estimates provided here are largely dependent on the simplified representation of converting land cover assumed in this study. However, we have not performed a systematic sensitivity analysis of the different assumptions made to modify land cover. The third source of uncertainty arises due to land-use data sets used as inputs. Estimates of historical gridded wood harvest data were based on several assumptions, which are subject to uncertainty (Hurtt et al., 2006). As shown in this study and in other previous studies (Klein Goldewijk and Ramankutty, 2004; Jain and Yang, 2005), spatial and temporal patterns of historical cropland and pastureland have significant uncertainties. This is reflected in the distribution of non-forested land-cover types as estimated using three agricultural data sets (Fig. 5). As a result, only the total non-forested land as a single broad category matches with satellite estimates. The individual forest area, however, does seem to agree reasonably well between the three estimates, primarily due to the calibration carried out in Step 4 (Sect. 2.4). Constraining each land-cover type (especially herbaceous types) to be close to satellite estimates is impossible, as the cropland and pastureland estimates prescribed based on input data sets need to remain unaltered. In addition, medium/coarse-resolution satellite data have less accuracy in classifying herbaceous land-cover types than trees or barren land (Friedl et al., 2010).

Several regional and national-level reconstructions using finer resolution census data have revealed significant differences in estimates of cropland and pastureland compared to older versions of RF and HYDE global data sets. For example, Li et al. (2010) found that RF data overestimated cropland area in China by a factor of 21 for the year 1700 and 1.6 for 1990 when compared with the cropland data of Northeast China (Ye and Fang, 2011) reconstructed based on combining calibrated historical data from multiple sources. Similarly, they found significant differences in the spatial distribution of cropland in HYDE data for the 18th and 19th century. Historical reconstructions over Amazonia (Leite et al., 2011) using municipal-level census data with higher level of details also show considerable difference in spatial patterns and magnitude compared to RF data. The range of uncertainties in regional estimates is expected to have narrowed in most recent RF and HYDE data used in this study, but

significant differences still exist. Because the three estimates produced in this study are directly dependent on these global land-use data sets, our global data sets should also be used with caution while drawing inferences from regional-level analysis. Since no single agricultural land-use data set used here can be pointed out as better or worse than another, it is recommended to use all three estimates alongside one another to gain a better picture of the range of uncertainties.

The biggest source of uncertainty in the global C budget remains emissions due to LUCC (Canadell, 2002), and these are estimated to be in the order of ± 0.5 GtC y^{-1} (Houghton, 2005; Houghton et al., 2012). Several multi-model comparison experiments have been performed to determine the uncertainty of LUCC in the global carbon budget (e.g., McGuire et al., 2001; Pitman et al., 2009; Reick et al., 2010). The LUCC uncertainty experiments involve using a common land-use data set (e.g., HYDE or RF) in each of the models and comparing the land-use fluxes. However, due to differences in the structure of each model, the method adopted to implement the common land-use data differs significantly between each model (e.g., see Pitman et al., 2009). As a result, it is impossible to attribute the estimated uncertainty to model-related uncertainty and uncertainties arising due to differences in the method of implementing land-use data between different models. However, driving the same model with multiple LUCC data sets derived consistently using same method, as discussed here, opens a new avenue for studying LUCC data-related uncertainty by eliminating the model-related uncertainty.

Certainly, indirect anthropogenic and natural effects have been dominant factors in historical land-cover change and have been poorly documented at a global scale (Lambin et al., 2003). Additionally, land-cover modifications like agricultural intensification have been thought to have a widespread impact on climate through altered surface attributes and changes in biogeochemical cycles. Recent advances in remote sensing observations have provided a more accurate and globally consistent picture of more subtle changes in land-cover (e.g., changes in tree height, vegetation biomass, and vegetation structure), in addition to capturing land-cover changes. Because globally consistent remote sensing observations are available only for the past four decades, we need to rely on other methods of reconstructing large time-scale changes in land cover. Monitoring all forms of land-cover change extensively and consistently at a global scale for the pre-satellite era, even at medium/coarse spatial and temporal resolution, was impractical. Hence, several assumptions need to be made to account for its impact on LUCC. Future research is required on monitoring long-term changes in all forms of land-cover change and land-cover modifications at higher spatial and temporal resolutions through remote sensing observations. Further, tracking LUC conversions rather than net changes in land cover can help facilitate better understanding of trends and fate of LUCC and its implications.

As pointed out by Pitman et al. (2009), implementing a common LUCC data set among different models is challenging. As a result, implementing the land cover maps and LUC conversion estimates presented here in different models may be subject to different approximations depending on the complexity and parameters associated with each model. However, we have chosen land-cover classifications such that the data can be implemented in models without introducing much uncertainty. Preliminary results of regional and global carbon emissions for the last three decades, estimated by implementing these three sets of data in the ISAM, have already been used in the IPCC AR5. A detailed assessment of the range of biogeophysical and biogeochemical impacts produced by these three estimates is in progress using a coupled ISAM-CESM framework. We believe that the data sets presented here will be useful to modelers interested in studying the effects of historical LUCC on biogeophysics, biogeochemistry and hydrological cycle, as well as in general to the global change community interested in studying the impacts of historical LUCC. Digital versions of these data sets can be downloaded from the webpage (http://www.atmos.illinois.edu/~meiyapp2/datasets.htm).

References

Arora V K, Boer G J (2006). Simulating competition and coexistence between plant functional types in a dynamic vegetation model. Earth Interactions, 10 (10): 1–30

Bala G, Caldeira K, Wickett M, Phillips T J, Lobell D B, Delire C, Mirin A (2007). Combined climate and carbon-cycle effects of large-scale deforestation. Proc Natl Acad Sci USA, 104(16): 6550–6555

Bonan G B (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science, 320(5882): 1444–1449

Bonan G B, Pollard D, Thompson S L (1992). Effects of boreal forest vegetation on global climate. Nature, 359(6397): 716–718

Brovkin V, Claussen M, Driesschaert E, Fichefet T, Kicklighter D, Loutre M F, Matthews H D, Ramankutty N, Schaeffer M, Sokolov A (2006). Biogeophysical effects of historical land cover changes simulated by six Earth system models of intermediate complexity. Clim Dyn, 26(6): 587–600

Canadell J G (2002). Land use effects on terrestrial carbon sources and sinks. Science in China (Series C), 45(Sl): 1–9

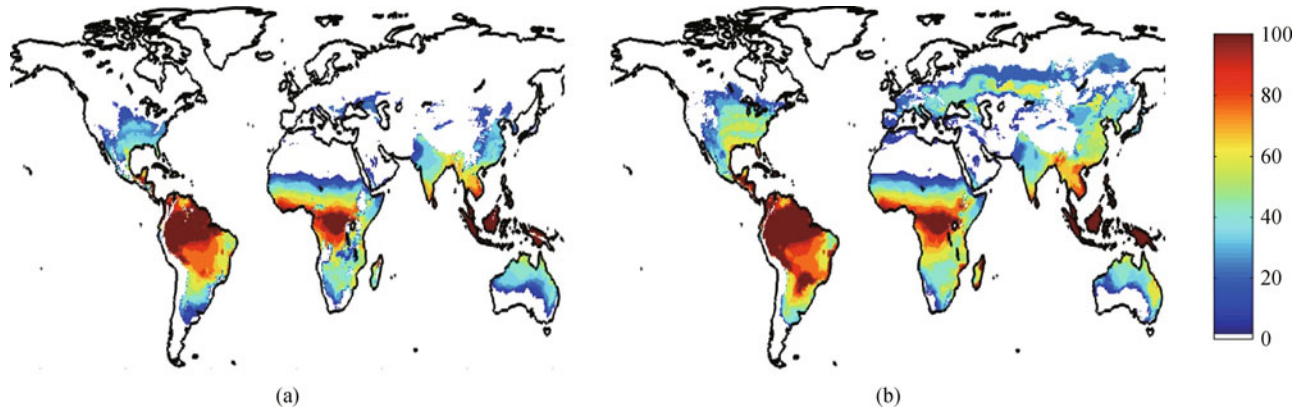Canadell J G, Le Quéré C, Raupach M R, Field C B, Buitenhuis E T,

Ciais P, Conway T J, Gillett N P, Houghton R A, Marland G (2007). Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. Proc Natl Acad Sci USA, 104(47): 18866–18870

Collatz G J, Berry J A, Clark J S (1998). Effects of climate and atmospheric CO_2 partial pressure on the global distribution of C_4 grasses: present, past, and future. Oecologia, 114(4): 441–454

Ellis E C, Klein Goldewijk K, Siebert S, Lightman D, Ramankutty N (2010). Anthropogenic transformation of the biomes, 1700 to 2000. Glob Ecol Biogeogr, 19: 589–606

FAO (2001). Global Forest Resources Assessment 2000. Main Report, FAO Forestry Paper 140, Rome, Italy. Available at http://www.fao.org/forestry/fra/2000/report/en/

FAO (2006). Global Forest Resources Assessment 2005. Main Report, FAO Forestry Paper 147, Rome, Italy. Available at http://www.fao.org/forestry/fra/fra2005/en/

FAO (2008). FAOSTAT. Food and Agriculture Organization of the United Nations, Rome, Italy. Available at: http://www.fao.org

FAO (2010). Global Forest Resources Assessment 2010, Main Report, FAO Forestry Paper 163, Rome, Italy. Available at http://www.fao.org/forestry/fra/fra2010/en/

Feddema J J, Oleson K W, Bonan G B, Mearns L O, Buja L E, Meehl G A, Washington W M (2005). The importance of land cover change in simulating future climates. Science, 310(5754): 1674–1678

Findell K L, Pitman A J, England M H, Pegion P J (2009). Regional and global impacts of land cover change and sea surface temperature anomalies. J Clim, 22(12): 3248–3269

Foley J A, Costa M H, Delire C, Ramankutty N, Snyder P (2003). Green surprise? How terrestrial ecosystems could affect Earth's climate. Front Ecol Environ, 1(1): 38–44

Foley J A, Defries R, Asner G P, Barford C, Bonan G, Carpenter S R, Chapin F S, Coe M T, Daily G C, Gibbs H K, Helkowski J H, Holloway T, Howard E A, Kucharik C J, Monfreda C, Patz J A, Prentice I C, Ramankutty N, Snyder P K (2005). Global consequences of land use. Science, 309(5734): 570–574

Foley J A, Ramankutty N, Brauman K A, Cassidy E S, Gerber J S, Johnston M, Mueller N D, O'Connell C, Ray D K, West P C, Balzer C, Bennett E M, Carpenter S R, Hill J, Monfreda C, Polasky S, Rockström J, Sheehan J, Siebert S, Tilman D, Zaks D P (2011). Solutions for a cultivated planet. Nature, 478(7369): 337–342

Friedl M A, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010). MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens Environ, 114(1): 168–182

Giglio L, Randerson J T, van der Werf G R, Kasibhatla P S, Collatz G J, Morton D C, DeFries R S (2010). Assessing variability and long-term trends in burned area by merging multiple satellite fire products. Biogeosciences, 7(3): 1171–1186

Hansen M C, DeFries R S, Townshend J G R, Sohlberg R (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. Int J Remote Sens, 21(6–7): 1331–1364

Houghton R A (1999). The annual net flux of carbon to the atmosphere from changes in land use 1850–1990. Tellus B Chem Phys Meterol, 51(2): 298–313

Houghton R A (2005). Aboveground forest biomass and the global carbon balance. Glob Change Biol, 11(6): 945–958

Houghton R A (2008). Carbon flux to the atmosphere from land use changes: 1850–2005. In: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN, USA

Houghton R A, Hobbie J E, Melillo J M, Moore B, Peterson B J, Shaver G R, Woodwell G M (1983). Changes in the carbon content of terrestrial biota and soils between 1860 and 1980: a net release of CO_2 to the atmosphere. Ecol Monogr, 53(3): 235–262

Houghton R A, van der Werf G R, DeFries R S, Hansen M C, House J I, Le Quéré C, Pongratz J, Ramankutty N (2012). Chapter G2 carbon emissions from land use and land cover change. Biogeosciences Discuss, 9(1): 835–878

Hurtt G C, Chini L P, Frolking S, Betts R A, Feddema J, Fischer G, Fisk J P, Hibbard K, Houghton R, Janetos A, Jones C D, Kindermann G, Kinoshita T, Klein Goldewijk K, Riahi K, Shevliakova E, Smith S, Stehfest E, Thomson A, Thorton P, Vuuren D P, Wang Y (2011). 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

Hurtt G C, Frolking S, Fearon M G, Moore B III, Shevliakova E, Malyshev S, Pacala S W, Houghton R A (2006). The underpinnings of land use history: three centuries of global gridded land use transitions, wood harvest activity, and resulting secondary lands. Glob Change Biol, 12(7): 1208–1229

Jain A K, Yang X (2005). Modeling the effects of two different land cover change data sets on the carbon stocks of plants and soils in concert with CO_2 and climate change. Global Biogeochem Cycles, 19(2): GB2015

Jain A K, Yang X, Kheshgi H, McGuire A D, Post W P, Kicklighter D (2009). Nitrogen attenuation of terrestrial carbon cycle response to global environmental factors. Global Biogeochem Cycles, 23(4): GB4028

Jung M, Henkel K, Herold M, Churkina G (2006). Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sens Environ, 101(4): 534–553

Klein Goldewijk K (2001). Estimating global land use change over the past 300 years: the HYDE database. Global Biogeochem Cycles, 15(2): 417–433

Klein Goldewijk K, Beusen A, van Drecht G, De Vos M (2011). The HYDE 3.1 spatially explicit database of human-induced land-use change over the past 12,000 years. Glob Ecol Biogeogr, 20(1): 73–86

Klein Goldewijk K, Beusen A, Janssen P (2010). Long term dynamic modeling of global population and built-up area in a spatially explicit way, HYDE 3.1. Holocene, 20(4): 565–573

Klein Goldewijk K, Ramankutty N (2004). Land cover change over the last three centuries due to human activities: the availability of new global data sets. GeoJournal, 61(4): 335–344

Klein Goldewijk K, van Drecht G, Bouwman A F (2007). Mapping contemporary global cropland and grassland distributions on a 5 × 5 minute resolution. Journal of Land Use Science, 2(3): 167–190

Lambin E F, Geist H J (2003). Regional differences in tropical deforestation. Environment, 45(6): 22–36

Lambin E F, Geist H J, Lepers E (2003). Dynamics of land use and land cover change in tropical regions. Annu Rev Environ Resour, 28(1): 205–241

Lawrence P J, Feddema J J, Bonan G B, Meehl G A, O'Neill B C, Oleson

K W, Levis S, Lawrence D M, Kluzek E, Lindsay K, Thornton P E (2012). Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. J Clim: 120208141321006

Leite C C, Costa M H, de Lima C A, Ribeiro C A A S, Sediyama G C (2011). Historical reconstruction of land use in the Brazilian Amazon (1940–1995). Journal of Land Use Science, 6(1): 33–52

Li B B, Fang X Q, Ye Y, Zhang X (2010). Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data—a case study in Northeast China. Science in China (D): Earth Sciences, 53(11): 1689–1699

Liu M, Tian H (2010). China's land cover and land use change from 1700 to 2005: estimations from high-resolution satellite data and historical archives. Global Biogeochem Cycles, 24(3): GB3003

Loveland T R, Belward A S (1997). The IGBP-DIS global 1 km land cover data set, DISCover first results. Int J Remote Sens, 18(15): 3289–3295

McGuire A D, Sitch S, Clein J S, Dargaville R, Esser G, Foley J, Heimann M, Joos F, Kaplan J, Kicklighter D W, Meier R A, Melillo J M, Moore B III, Prentice I C, Ramankutty N, Reichenau T, Schloss A, Tian H, Williams L J, Wittenberg U (2001). Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO_2, climate and land use effects with four process-based ecosystem models. Global Biogeochem Cycles, 15(1): 183–206

Meinshausen M, Smith S J, Calvin K V, Daniel J S, Kainuma M L T, Lamarque J F, Matsumoto K, Montzka S A, Raper S C B, Riahi K, Thomson A, Velders G J M, Vuuren D P P (2011). The RCP greenhouse gas concentrations and their extension from 1765 to 2300. Clim Change, 109(1–2): 213–241

Mitchell T D, Jones P D (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol, 25(6): 693–712

Monfreda C, Ramankutty N, Foley J A (2008). Farming the Planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary prodution in the year 2000. Global Biogeochem Cycles, 22: GB1022

Moss R H, Edmonds J A, Hibbard K A, Manning M R, Rose S K, van Vuuren D P, Carter T R, Emori S, Kainuma M, Kram T, Meehl G A, Mitchell J F B, Nakicenovic N, Riahi K, Smith S J, Stouffer R J, Thomson A M, Weyant J P, Wilbanks T J (2010). The next generation of scenarios for climate change research and assessment. Nature, 463(7282): 747–756

Olofsson J, Hickler T (2008). Effects of human land use on the global carbon cycle during the last 6000 years. Vegetation History and Archaeobotany, 17(5): 605–615

Parry M L, Canziani O F, Palutikof J P, van der Linden P J, Hanson C E (2007). Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 211–272

Pielke R A, Marland G, Betts R A, Chase T N, Eastman J L, Niles J O, Niyogi D D S, Running S W (2002). The influence of land use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. Philos Trans R Soc Lond A, 360(1797): 1705–1719

Pielke R A Sr, Pitman A, Niyogi D, Mahmood R, McAlpine C, Hossain F, Klein Goldewijk K, Nair U, Betts R, Fall S, Reichstein M, Kabat P, de Noblet N (2011). Land use/land cover changes and climate: modeling analysis and observational evidence. Wiley Interdisciplinary Reviews: Clim Change, 2(6): 828–850

Pitman A J, Avila F B, Abramowitz G, Wang Y P, Phipps S J, de Noblet-Ducoudré N (2011). Importance of background climate in determining impact of land cover change on regional climate. Nature Climate Change, 1(9): 472–475

Pitman A J, de Noblet-Ducoudré N, Cruz F T, Davin E, Bonan G B, Brovkin V, Claussen M, Delire C, 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, Voldoire A (2009). Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study. Geophysical Research Letters, 36; L14814

Pongratz J, Reick C, Raddatz T, Claussen M (2008). A reconstruction of global agricultural areas and land cover for the Last Millennium. Global Biogeochem Cycles, 22(3): GB3018

Pongratz J, Reick C H, Raddatz T, Claussen M (2009). Effects of anthropogenic land cover change on the carbon cycle of the last millennium. Global Biogeochem Cycles, 23(4): GB4001

Ramankutty N, Evan A, Monfreda C, Foley J A (2008). Farming the Planet: 1. The geographic distribution of global agricultural lands in the year 2000. Global Biogeochem Cycles, 22(1): GB1003

Ramankutty N, Foley J (1999). Estimating historical changes in global land cover: croplands from 1700 to 1992. Global Biogeochem Cycles, 13(4): 997–1028

Reick C, Raddatz T, Pongratz J, Claussen M (2010). Contribution of anthropogenic land cover change emissions to pre-industrial atmospheric CO_2. Tellus B Chem Phys Meterol, 62(5): 329–336

Shevliakova E, Pacala S W, Malyshev S, Hurtt G C, Milly P C D, Caspersen J P, Sentman L T, Fisk J P, Wirth C, Crevoisier C (2009). Carbon cycling under 300 years of land use change: importance of the secondary vegetation sink. Global Biogeochem Cycles, 23(2): GB2022

Still C J, Berry J A, Collatz G J, DeFries R S (2003). Global distribution of C_3 and C_4 vegetation: carbon cycle implications. Global Biogeochem Cycles, 17(1): 1006–1020

van der Werf G R, Randerson J T, Giglio L, Collatz G J, Mu M, Kasibhatla P S, Morton D C, DeFries R S, Jin Y, van Leeuwen T T (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). Atmos Chem Phys, 10(23): 11707–11735

Yang X, Richardson T K, Jain A K (2010). Contributions of secondary forest and nitrogen dynamics to terrestrial carbon uptake. Biogeosciences Discuss, 7(2): 2739–2765

Yang X, Wittig V, Jain A K, Post W (2009). Integration of nitrogen cycle dynamics into the Integrated Science Assessment Model for the study of terrestrial ecosystem responses to global change. Global Biogeochem Cycles, 23(4): GB4029

Ye Y, Fang X Q (2011). Spatial pattern of land cover changes across Northeast China over the past 300 years. J Hist Geogr, 37(4): 408–417

Supplementary

Land-use change data sets

We apply our method of reconstruction to 3 different sets of cropland and pastureland data along with a common data set for wood harvest and urban land. This results in 3

Fig. S1   C$_4$ fraction map generated using modified method of Still et al. (2003) (a) during 2010, and (b) during 1765 (Unit:% per grid cell area)

Table S1   The 92 types of LUC conversions. Grassland/pastureland and cropland are not distinguished as C$_3$/C$_4$ types while recording LUC conversions

No.	Land-use conversions	No.	Land-use conversions	No.	Land-use conversions	No.	Land-use conversions
1	TrpEBF → SecTrpEBF	24	BorENF → Urban	47	PdRI → Crop	70	SecBorDNF → Urban
2*	TrpEBF → Crop	25	BorDNF → SecBorDNF	48	PdRI → Past	71	Crop → Savanna
3	TrpEBF → Past	26*	BorDNF → Crop	49	PdRI → Urban	72	Crop → Grass
4	TrpEBF → Urban	27	BorDNF → Past	50*	SecTrpEBF → Crop	73	Crop → Denseshrub
5	TrpDBF → SecTrpDBF	28	BorDNF → Urban	51	SecTrpEBF → Past	74	Crop → Openshrub
6*	TrpDBF → Crop	29	Savanna → Crop	52	SecTrpEBF → Urban	75	Crop → Tundra
7	TrpDBF → Past	30	Savanna → Past	53*	SecTrpDBF →Crop	76	Crop → Desert
8	TrpDBF → Urban	31	Savanna → Urban	54	SecTrpDBF → Past	77	Crop → PdRI
9	TmpEBF → SecTmpEBF	32	Grass → Crop	55	SecTrpDBF → Urban	78#	Crop → SecTrpEBF
10*	TmpEBF → Crop	33	Grass → Past	56*	SecTmpEBF → Crop	79#	Crop → SecTrpDBF
11	TmpEBF → Past	34	Grass → Urban	57	SecTmpEBF → Past	80#	Crop → SecTmpEBF
12	TmpEBF → Urban	35	Denseshrub → Crop	58	SecTmpEBF → Urban	81#	Crop → SecTmpENF
13	TmpENF → SecTmpENF	36	Denseshrub → Past	59*	SecTmpENF → Crop	82#	Crop → SecTmpDBF
14*	TmpENF → Crop	37	Denseshrub → Urban	60	SecTmpENF → Past	83#	Crop → SecBorENF
15	TmpENF → Past	38	Openshrub → Crop	61	SecTmpENF → Urban	84#	Crop → SecBorDNF
16	TmpENF → Urban	39	Openshrub → Past	62*	SecTmpDBF → Crop	85	Crop → Past
17	TmpDBF → SecTmpDBF	40	Openshrub → Urban	63	SecTmpDBF → Past	86	Crop → Urban
18*	TmpDBF → Crop	41	Tundra → Crop	64	SecTmpDBF → Urban	87	Past → Grass
19	TmpDBF → Past	42	Tundra → Past	65*	SecBorENF → Crop	88	Past → Crop
20	TmpDBF → Urban	43	Tundra → Urban	66	SecBorENF →Past	89	Past → Urban
21	BorENF → SecBorENF	44	Desert → Crop	67	SecBorENF → Urban	90	Urban → Grass
22*	BorENF → Crop	45	Desert → Past	68*	SecBorDNF → Crop	91	Urban → Crop
23	BorENF → Past	46	Desert → Urban	69	SecBorDNF → Past	92	Urban → Past

Notes: * Forest to cropland conversions; # Cropland to forest conversions

Table S2 Estimates of LUC conversions (grouped into broader category) during the period 1765–2005 based on ISAM-RF, ISAM-HYDE and ISAM-HH estimates (Unit: million km^2)

Regions	Forest → Cropland			Non-forest → Cropland			Forest → Pastureland			Non-forest → Pastureland		
	ISAM-RF	ISAM-HYDE	ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH
North America	1.07	1.54	1.50	1.84	1.35	0	0.32	0.29	0	4.05	2.51	0
Latin America	1.11	0.65	1.31	1.76	0.94	0	1.51	1.19	2.75	5.43	4.17	0
Europe	1.02	0.65	0.46	0.40	0.23	0	0.19	0.13	0	0.86	0.34	0
North Africa and Middle East	0.09	0.07	0.07	1.03	0.72	0	< 0.01	< 0.01	0	2.80	2.81	0
Tropical Africa	0.45	0.40	0.60	1.50	1.39	0	0.51	0.54	0	7.05	6.50	0
Former USSR	0.61	0.95	0.24	1.34	1.44	0	0.12	0.19	0	2.77	3.28	0
China	0.70	0.56	0.44	0.52	0.30	0	0.14	0.25	0	2.53	4.39	0
South & South-East Asia	1.65	1.65	2.06	0.66	0.53	0	0.05	0.09	0	0.28	0.29	0
Pacific Developed Region	0.12	0.12	0.16	0.53	0.52	0	0.06	0.06	0	6.01	4.34	0
World	6.80	6.58	6.85	9.57	7.43	0	2.91	2.75	2.75	31.78	28.63	0

distinct estimates of land-cover change and LUC conversions. In the following sub-sections we briefly describe the land-use change data sets used in this study.

RF cropland and pastureland data

We use the updated version of annual cropland data described in Ramankutty and Foley (1999) and new pastureland data reconstructed using methods similar to those described in Ramankutty et al. (2008) and Ramankutty and Foley (1999). The updated version is a complete revision of the original cropland data set (Ramankutty and Foley, 1999), reconstructed by combining satellite imagery with a revised collection of national and subnational-level statistical database at a higher level of detail. The updated data sets are currently available at 0.5°×0.5° resolution covering the period 1700–2007 (N. Ramankutty, personal communication, 2011).

HYDE cropland and pastureland data

We use the HYDE 3.1 'baseline' cropland and pastureland data set (Klein Goldewijk et al., 2011), reconstructed by combining historical population, cropland and pastureland statistics from FAO (2008) with satellite information (described in Klein Goldewijk et al., 2007). Both the data sets are reconstructed at 5 min spatial resolution at 10-year intervals starting from 10000 BC until 2000 AD and

recently updated until 2010 AD. We linearly interpolate the data between each interval and aggregate the data to yield the annual fractional cropland and pastureland area from 1765 to 2010 at 0.5°×0.5° resolution.

HH cropland and pastureland data

We use the revised Houghton (HH) data set used to estimate the regional and global net carbon fluxes due to land-use change from 1800 to 2005 (Houghton, 2008). These estimates are based on the Forest Resource Assessment (FAO, 2006), which has significantly narrowed down the uncertainty by incorporating satellite data, as compared to the previous estimates based on FAO (2001). Houghton's latest estimate provides the annual rate of deforestation/reforestation due to cropland, pastureland, wood harvest and shifting cultivation, in addition to afforestation rates for ten regions (as defined in Houghton et al., 1983) covering the entire globe, rather than by geographic details, as in the case of RF and HYDE data sets. For this study, we use only cropland and pastureland estimates from Houghton data set.

To estimate the cropland area at 0.5°×0.5° resolution, Jain and Yang (2005) used a simple technique to convert the regional estimates of deforestation due to cropland into gridded estimates. They distribute the regional rate of change of cropland, by proportionally weighing the area of forest present in each grid cell within that region. However,

it is not necessary that each grid cell that contains forest is used for cropland. This method leads to a more homogenous distribution of cropland area. Moreover, a regional increase in cropland does not indicate an increase in cropland area everywhere within that region. Consequently, the area-weighted method of allocating cropland described in Jain and Yang (2005) could not account for sub-regional abandonment of cropland, whenever there is a net regional increase in cropland area and vice-versa.

Here, we overcome both these shortcomings by capitalizing on forest-to-cropland and cropland-to-forest (refer Supplementary Table S1) annual conversion rates from LUC conversion estimates derived at $0.5°\times0.5°$ resolution based on ISAM-RF derived using the method described in Sect. 2. We calculate the forest-to-cropland conversion rate $H_{\mathrm{gridded}}(i,j,k)$ using Houghton's regional forest-to-cropland conversion rate $H(i,j)$, and RF rate of change from forest to cropland $SG_{\mathrm{C}}(i,j,k)$ using the following relationship

$$H_{\mathrm{gridded}}(i,j,k) = H(i,j) \times SG_{\mathrm{fraction}}(i,j,k),$$

where k represents a grid-cell within region i during the year j.

We calculate $SG_{\mathrm{fraction}}(i,j,k)$, from the following relation

$$SG_{\mathrm{fraction}}(i,j,k) = \left(\frac{SG_{\mathrm{C}}(i,j,k)}{\sum\limits_{k=1}^{n} SG_{\mathrm{C}}(i,j,k)} \right),$$

where n represents the last grid cell for a given region, so that $\sum\limits_{k=1}^{n} SG_{\mathrm{fraction}}(i,j,k) = 1$. $SG_{\mathrm{fraction}}(i,j,k)$, gives a value between 0 and 1 for each grid-cell, which is indicative of the relative amount of forest converted to cropland within each region. Similarly, cropland-to-forest conversion rates estimated using ISAM-RF were used to spatialize HH reforestation estimates. The same approach was adopted to spatialize pastureland data. Since Houghton's estimate starts from 1800, we assume cropland and pastureland areas from 1765 to 1799 to be the same as those in the land cover map of 1765 generated from Step 2 (Sect. 2.2).

Urban land and wood harvest data

We use the HYDE 3.1 spatially explicit urban land area data set described in Klein Goldewijk et al. (2010). We process the data using the method described for HYDE cropland data (Sect. 2.1) to derive the fractional area of urban land at $0.5°\times0.5°$ resolution.

Hurtt et al. (2011) estimated the annual rate of wood harvested areas at $0.5°\times0.5°$ resolution for the period 1500–2005. Wood harvesting rates and urban land area after 2005 are assumed to be same as in 2005. The urban land and wood harvest for the period 1765 to 2010 is used in common with all 3 sets of cropland and pastureland data.