

Supplementary Information

S1 Overview of historical land use reconstructions

Several approaches have been published within the last two decades to reconstruct the history of human utilization of land to meet their needs of food, fiber and space for settlement on a global scale. Depending on the objective of the particular study they cover different time periods, spatial resolutions and methods of reconstruction (Table S1). In the following paragraphs we summarize the methodologies of four spatially explicit historical reconstructions. For details, please see the original publications.

S1.1 HYDE

The History Database of the Global Environment (HYDE) was originally developed by Klein Goldewijk (2001), covering spatially explicit historical population estimates and land-use patterns for the past 300 years at 0.5° resolution. Several updates and extensions led to version HYDE 3.1, which was used for the *LUH* data in CMIP5 (Klein Goldewijk et al. (2011); this is the version we refer to here and in the article). Recently there has been a update to version HYDE 3.2, which now covers a time period from 10 000 BC to 2015 AD at 5 arcminute spatial resolution and includes further agricultural management layers, such as irrigation (Klein Goldewijk, 2016).

The underlying principle of the HYDE reconstruction is the relationship between human population and agricultural activity expressed in a per capita use of cropland and pasture area, leading to a spatial dependency of land-use activities to human settlements. Klein Goldewijk et al. (2010) first derived time series of population numbers from a vast number of sources on a subnational or national scale (depending on data availability, e.g., McEvedy and Jones (1978), Livi Bacci (2007) and Maddison (2001); see Klein Goldewijk (2001) and Klein Goldewijk et al. (2011) for details) and translated them to population density maps using patterns from Landsat (2006) for recent time and a combination of suitability maps for historic time. For the period 1961-2000, the per capita use of cropland and pasture was calculated from FAO statistics on country or subnational level. Prior to 1961, the per capita land-use numbers were dynamically estimated country by country following Ruddiman and Ellis (2009) and adjusted, accounting for low population numbers (= higher per capita land use), but also limitations in technology, and a maximum area of land that can be cultivated by a subsistence farmer (= lower per capita land use). Using the per capita usage of cropland and pasture to estimate cropland and pasture total areas on a (sub-) national level for every time step, spatial allocation of the total areas to the 5 arcminute grid was implemented using two sets of weighing maps: On the one side, present distribution of cropland and pasture was derived by integrating FAO statistics and additional subnational statistics for the USA and China with two satellite derived land cover products representative for recent time (DISCover version 2, Loveland et al. (2000); GLC2000, Bartholome and Belward (2005)). The

weighing map for historical time, on the other side, was constructed by combining the earlier described population density maps and different biophysical suitability parameters, namely soil quality, distance to rivers, steepness of terrain, and thresholds for annual mean temperature. Both maps were subsequently used to allocate (sub-) national totals of agricultural areas to specific grid cells, while the influence of the historic map gradually increases when going further into the past.

5 S1.2 Ramankutty and Foley (1999)

Ramankutty and Foley (1999) apply a hindcast modeling technique to derive global scale spatial patterns of cropland for the period 1700-1992. The original reconstruction did not include pasture areas. A revised and updated version covers the years up to 2007, both for cropland and pasture at 5 arcminute spatial resolution. The starting point for the reconstruction is represented by the integration of satellite-derived land-cover products (DISCover in original data set (Loveland and
10 Belward; 1997); BU-MODIS (Friedl et al., 2002) and GLC2000 (Bartholome and Belward, 2005) in the updated version) and FAO statistics. The national and subnational totals of cropland and pasture were calibrated to the spatial distribution of cropland and pasture areas in the earth observation product applying a linear fitting approach. This resulted in a global, 5 arcminute resolution cropland and pasture map for the year 2000, representing the spatial distribution of cropland and pasture areas (Ramankutty et al., 2008). In a second step, a comprehensive data base of historical agricultural areas on (sub-)
15 national level was compiled from different sources. FAO statistics were used for the time period from 1961 to the end point. Prior to 1961, the data base first accounts for census data. Whenever census data were not available, cropland conversion rates of Houghton and Hackler (1995) were applied to the cropland map of Richards (1990) for 1980 with some regional adjustments to avoid unrealistic agricultural areas in particular regions. The spatial allocation of the cropland areas is implemented by applying a simple hindcast model, which preserves the cropland pattern of the start map within each unit of
20 the inventory data base for the whole time period to 1700. For that a change factor between two subsequent years is calculated from the inventory database, dividing the cropland area in the target year by the cropland area in the starting year, which is thereafter applied to each grid cell within a unit.

S1.3 Pongratz et al. (2008)

Pongratz et al. (2008) extended the reconstruction of Ramankutty and Foley (1999) back to 800 AD and presented the first
25 consistent and spatially explicit cropland and pasture reconstruction for pre-industrial times at the date of publication. For the period 1700-1992, the cropland time series is, apart from smaller regional adjustments and updates, the same than the Ramankutty and Foley (1999) data. Since they further had not published their pasture time series at that point, Pongratz et al. (2008) combined the pasture map for 1992 with change rates taken from the HYDE data base to extend it back to 1700. Unlike the pattern maintaining approach applied by Ramankutty and Foley (1999), pasture was spatially distributed around
30 existing cropland while maintaining the pattern of total agricultural area rather than the individual shares of cropland and pasture to allow also for cropland expansion into pasture areas.

Based on these two time series covering the years 1700-1992, an extrapolation to 800 AD was applied on (sub-) national level, while using population data from McEvedy and Jones (1978) as a proxy for land-use change. Similar to HYDE, the simple measure of per capita usage of crop and pasture area was assumed to be the best approximation. However, in this case, per capita use was calculated from the 1700 maps and held constant for the whole period prior to 1700. Spatial distribution of agricultural areas was assumed to represent the patterns of 1700 for the period 800 to 1700. Besides, changes in agricultural patterns, e.g., following the European colonization in North and South America, were especially accounted for by altering the patterns in particular regions. Both time series were aggregated to a 0.5° resolution.

S1.4 KK10

Kaplan et al. (2010) introduce a non-linear relationship between population numbers and area of forest clearance to calculate total areas affected by human land-use change. The basic assumption of this approach is a decreasing per capita land use over time due to intensification of already converted areas rather than the expansion of land use into new areas when population densities increase. With the objective to build an empirical, non-linear model, population time series for the period 6050 BC to AD 1850 were compiled first. Data from McEvedy and Jones (1978) were utilized for the period 1000 BC to AD 1850 with some regional adjustments and subsequently extended back to 6050 BC by a modelling approach (Global Land USE and Technological Evolution Simulator (GLUES, see Lemmen (2009) and Wirtz and Lemmen (2003) for details). Population density was normalized to cultivatable land to prevent the model extending cropland areas into unsuitable land. A sigmoidal log-linear model was fitted to a set of empirical data from various European countries to derive a relationship between forest cover and population density, accounting also for different stages of technological development over time (Kaplan et al., 2009). Concurrently, Kaplan et al. (2010) integrated different climatic and biophysical variables to indices of suitability for cropland and pasture on a 5 arcminute grid following a method of Ramankutty et al. (2002). Combining the regional level estimates of historical forest cover with the suitability datasets led to a spatially explicit representation of area affected by land-use change over time. The integration was done by allocating cropland to high quality and suitable areas first, followed by pasture. As the forest cover – population relationship originally was derived for Europe, it has been adjusted for tropical and boreal regions in the global approach by including a threshold of net primary production, where productivity of agricultural lands is higher and therefore demand for new land lower.

Table S1: Summary of historical LULCC reconstructions.

Reference	Spatial resolution	Temporal coverage and resolution	Input data	Allocation
KK10, Kaplan et al. (2010)	5 x 5 arcminute	6050 BC to AD 1850, annual	population estimates, land suitability maps	based on non-linear population density – forest clearance relationship, high quality land cleared first
HYDE 3.1, Klein Goldewijk et al. (2011)	5 x 5 arcminute	10 000 BC to AD 2005, variable resolution	population estimates, FAO statistics, satellite derived products	dynamic per capita use of cropland and pasture; combination of weighing maps derived from satellite products, population and environmental parameters
Pongratz et al. (2008)	0.5 x 0.5 degree	AD 800 – AD 1992	adjusted Ramankutty and Foley (1999), HYDE 2.0, population data	constant per capita use of cropland & pasture prior to 1700, constant spatial pattern of agriculture prior to 1700
Ramankutty and Foley (1999)	5 x 5 arcminute	AD 1700 – AD 1992; update AD 1700 – 2007	census data and estimates of agricultural area, FAO statistics, satellite derived products	hindcast model, preserving agricultural pattern of 1992 within aggregated units

5 S2 Data and Methods

S2.1 Attribution of uncertainty in land use change projections

Multiple linear regression analysis followed by an ANOVA was used to decompose the variability of 43 projections of regional pasture areas for the year 2030 simulated by 11 global scale IAMs and LUCMs (Alexander et al., 2016; Prestele et al., 2016). Every scenario has been parameterized according to 9 variables (Table S2) that characterize the model structure (model type classification, model resolution), the scenario (socioeconomic and climate scenario variables) and the initial condition (deviation from value reported by FAOSTAT (2015) in the year 2010) prior to the regression analysis. The

modeled pasture area in 2030 was assumed to be a function of these 9 variables. To balance performance and complexity of the resulting regression model, variables were rejected using the Akaike information criterion. Subsequently an ANOVA was conducted on the regression results to identify relative contribution of the variables to the total variation in the 2030 pasture areas. The residual term thus covers all variation that could not be explained by these 9 variables.

5

Table S2: Overview of variables used in the regression analysis and ANOVA (table adopted and modified according to Prestele et al. (2016)).

Variable	Data type	Association
Initial condition delta	Continuous (deviation of model areas from FAO areas in 2010 (FAOSTAT, 2015))	Initial
Model type	Categorical (CGE, PE, Rule-based, Hybrid)	Model
Number of model cells (log)	Continuous	Model
CO ₂ concentration 2100	Continuous	Scenario
Population 2100	Continuous	Scenario
GDP growth rate to 2100	Continuous	Scenario
Inequality ratio 2100	Continuous	Scenario
Technology change	Discrete (0=None, 1=Slow, 2=Medium, 3=Rapid)	Scenario
International trade	Discrete (1=Constrained, 2=Moderate, 3=High)	Scenario

S2.2 Analysis of remote sensing products

10 To derive dominant sources of cropland expansion from remote sensing products, we analyzed high resolution LULCC data from Europe (CORINE, 100 m spatial resolution) and North America (NLCD, 30 m spatial resolution) (Table S3). We downloaded CORINE data from <http://land.copernicus.eu/pan-european/corine-land-cover>. NLCD data were obtained through <http://www.mrlc.gov/>.

S2.2.1 CORINE

15 CORINE was produced by computer assisted visual interpretation of satellite images, processed on a country by country basis, and subsequently merged to a comprehensive European database (EEA, 2007). It covers the years 1990, 2000, 2006 and most recently 2012 with different number of participating countries leading to different overlapping areas between the years. The land-cover classification was derived from different sensors dependent on the final year of the product (1990: Landsat-4/5 TM single date, 2000: Landsat-7 ETM single date; 2006: SPOT-4 and/or IRS P6 LISS III dual date; 2012: IRS
20 P6 LISS III and RapidEye dual date). CORINE is provided at a spatial resolution of 100 m and 250 m in raster data format as well as in vector format. The minimum mapping unit is 25 ha. Besides the products for the years mentioned above, special LULCC products have been produced and are currently available for the periods 1990 to 2000 and 2000 to 2006. For the

change products an enhanced minimum mapping unit of 5 ha was applied. The change products have been used for derivation of agricultural transitions in our analysis, thus covering all changes to agricultural areas larger than 5 ha between start and end year. All CORINE products are accompanied by a three level land-use and land-cover nomenclature varying in detail across the levels (Table S4). The first level only provides very general classes (e.g., artificial surfaces, agricultural areas, forests, etc.). The second level distinguishes 15 different categories and the highest detail is given by the 44 classes at level 3. For our analysis we used a merger of the different levels, as e.g., forests and shrubland could be differentiated at level 2, while low productivity grasslands could be only identified at level 3 (Table S4). See Bossard et al. (2000) for a detailed description of the legend and distinction of individual classes. Thematic accuracy of both products is indicated with larger than 85% (<http://land.copernicus.eu/pan-european/corine-land-cover>).

10 S2.2.2 NLCD

The National Land Cover Database (NLCD) is a high resolution (30 m) land-cover product for the USA. This Landsat-derived product has been provided for the years 1992, 2001, 2006 and 2011 at the latest. For our analysis the 2001, 2006 and 2011 products have been considered, as they are provided in a harmonized collection with special LULCC products. The NLCD dataset is classified according to a 16-class land-cover classification for the United States, developed in the 1970s by Anderson et al. (1976). The classification system distinguishes two agricultural classes, (81) *Pasture/Hay* and (82) *Cultivated Crops* (Table S5). Stehman et al. (2003) report an accuracy level of 55.7 % for the 1992 dataset. Accuracy assessment is not yet available for the 2011 data, but as 2001 and 2006 data showed significantly improved accuracy levels (78.7 % and 78.0 %, Wickham et al. (2010) and Wickham et al. (2013)) a similar (or even better) quality can be assumed for the 2011 data.

20 **Table S3: Summary of land-cover products used for our analysis.**

Product	Temporal coverage	Spatial resolution / Coverage	Legend	Sensor	Classification
CORINE	1990, 2000, 2006, (2012)	100m / Europe	44 classes, hierarchical levels	3 Landsat-4/5 TM, Landsat-7 ETM, SPOT-4, IRS P6 LISS III, RapidEye	change product, supervised, expert knowledge
NLCD	(1992), 2001, 2006, 2011	30m / USA	16 classes	Landsat	change product, spectral and knowledge based change detection

S2.2.3 Change detection

We used the dedicated change products for our analysis, which hold information about source and target classes upon land-use change. Areas of agricultural expansion were identified by every pixel that has an agricultural label (based on the inherent legend) at time t_2 , but not at time t_1 . We calculated the total expansion of agricultural areas by the difference of pixels which were assigned an agricultural label at time t_2 and time t_1 . Subsequently, combining the areas of cropland expansion with the map of time t_1 resulted in a map of sources of agricultural area. The source maps were classified and summarized considering the underlying original legend into grassland, forest, mixed grassland/forest and unvegetated land origin (Table S4, Table S5).

Table S4: CORINE land-cover legend (Bossard et al., 2000) and aggregation applied in our analysis.

Level 1	Level 2	Level 3	Aggregation
(1) Artificial surfaces	(11) Urban fabric; (12) Industrial, commercial and transport units; (13) Mine, dump and construction sites; (14) Artificial, non-agricultural vegetated areas	(111) Continuous urban fabric; (112) Discontinuous urban fabric; (121) Industrial and commercial units; (122) Road and rail networks and associated land; (123) Port areas; (124) Airports; (131) Mineral extraction sites; (132) Dump sites; (133) Construction sites; (141) Green urban areas; (142) Sport and leisure facilities	Other
(2) Agricultural areas	(21) Arable land; (22) Permanent crops; (23) Pastures; (24) Heterogeneous agricultural areas	(211) Non-irrigated arable land; (212) Permanently irrigated land; (213) Rice fields; (221) Vineyards; (222) Fruit trees and berry plantations; (223) Olive groves; (231) Pastures; (241) Annual crops associated with permanent crops; (242) Complex cultivation patterns; (243) Land principally occupied by agriculture, with significant areas of natural vegetation; (244) Agro-forestry areas	Agricultural areas
(3) Forest and semi natural areas	(31) Forests; (32) Scrub and/or herbaceous vegetation associations; (33) Open spaces with little or no vegetation	(311) Broad-leaved forest; (312) Coniferous forest; (313) Mixed forest; (321) Natural grasslands; (322) Moors and heathland; (323) Sclerophyllous vegetation; (324) Transitional woodland-shrub; (331) Beaches, dunes, sands; (332) Bare rocks; (333) Sparsely vegetated areas; (334) Burnt areas; (335) Glaciers and perpetual snow	(311)-(313) Forest (321) Grassland (322)-(324) Shrubland (331)-(335) Other
(4) Wetlands	(41) Inland wetlands; (42) Maritime wetlands	(411) Inland marshes; (412) Peat bogs; (421) Salt marshes; (422) Salines; (423) Intertidal flats	Other
(5) Water bodies	(51) Inland waters; (52) Marine waters	(511) Water courses; (512) Water bodies; (521) Coastal lagoons; (522) Estuaries; (523) Sea and ocean	Other

Table S5: National Land Cover Database (NLCD) classification system according to Anderson et al. (1976) and aggregation applied in our analysis.

Value	Label		Description	Aggregation
11	Open Water		All areas of open water, generally with less than 25 % cover or vegetation or soil	Other
12	Perennial Ice/Snow		All areas characterized by a perennial cover of ice and/or snow, generally greater than 25 % of total cover	Other
21	Developed, Space	Open	Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 % of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes	Other
22	Developed, Intensity	Low	Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 % of total cover. These areas most commonly include single-family housing units.	Other
23	Developed, Intensity	Medium	Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50-79 % of the total cover. These areas most commonly include single-family housing units.	Other
24	Developed, Intensity	High	Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80-100 % of the total cover.	Other
31	Barren (Rock/Sand/Clay)	Land	Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material. Generally, vegetation accounts for less than 15 % of total cover.	Other
41	Deciduous Forest		Areas dominated by trees generally greater than 5 meters tall, and greater than 20 % of total vegetation cover. More than 75 % of the tree species shed foliage simultaneously in response to seasonal change.	Forest
42	Evergreen Forest		Areas dominated by trees generally greater than 5 meters tall, and greater than 20 % of total vegetation cover. More than 75 % of the tree species maintain their leaves all year. Canopy is never without green foliage.	Forest
43	Mixed Forest		Areas dominated by trees generally greater than 5 meters tall, and greater than 20 % of total vegetation cover. Neither deciduous nor evergreen species are greater than 75 % of total tree cover.	Forest
52	Shrub/Scrub		Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20 % of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.	Shrubland
71	Grassland/Herbaceous		Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80 % of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.	Grassland
81	Pasture/Hay		Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 % of total vegetation.	Pasture
82	Cultivated Crops		Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as	Cropland

90	Woody Wetlands	orchards and vineyards. Crop vegetation accounts for greater than 20 % of total vegetation. This class also includes all land being actively tilled. Areas where forest or shrub land vegetation accounts for greater than 20 % of vegetative cover and the soil or substrate is periodically saturated with or covered with water.	Other
95	Emergent Herbaceous Wetlands	Areas where perennial herbaceous vegetation accounts for greater than 80 % of vegetative cover and the soil or substrate is periodically saturated with or covered with water.	Other

S2.3 Derivation of gross vs. net changes due to re-gridding from a CLUMondo simulation

To identify the difference between net and gross changes due to re-gridding of high-resolution modeled land-use change information, we utilized data from a simulation of the CLUMondo model (Van Asselen and Verburg, 2013) based on the

5 FAO 3 demand scenario (Eitelberg et al., 2016). These data are available at a 9.25 x 9.25 km regular grid (~5 arcminute) in an equal area projection and are based on the land system classification described in van Asselen and Verburg (2012). Land systems are characterized by land-cover composition, livestock numbers and land-use intensity. Each grid cell can thus be expressed as a mosaic of five LULC types (cropland, grassland, forest, urban, and bare), whose exact fractions vary with the world region. Upon a change from one land system to another, these characteristics also change.

10 We used the fractions of these five LULC types to track areal changes per grid cell at the original 9.25 x 9.25 km resolution over the whole simulation period (2000-2040). The total area changed at this resolution (sum of gains and losses for each LULC type) was assumed to be the gross changes in our analysis. In a second step, we aggregated the maps to ca. 0.5 x 0.5 degree and calculated the changes between two time steps. Due to bi-directional changes at the higher resolution (which offset each other) the total area affected by change at 0.5 x 0.5 degree resolution is usually smaller. The areal changes at 0.5

15 x 0.5 degree resolution were assumed to be the net changes in our analysis. By adding up the net changes and gross changes across all five LULC types and over the whole simulation period, we identified the amount of actually changed area that would be missed in a net change representation at 0.5 x 0.5 degree for this simulation (Figure S1) .

S2.4 CLUMondo land-use change priority analysis

The CLUMondo data originate from a simulation based on the FAO 3 demand scenario (Eitelberg et al., 2016) and cover the

20 time period from 2000 to 2040 with annual temporal resolution. Data are available at a 9.25 x 9.25 km regular grid (~5 arcminute) in an equal area projection and are based on the land system classification system described in van Asselen and Verburg (2012) (Table S6). In order to detect a particular algorithm, which is valid within a ca. 0.5 x 0.5 degree grid cell, the model output required several steps of preprocessing (Figure S2):

- Aggregation of the CLUMondo land systems legend and reclassification of each map following the PFT scheme of
- 25 DGVMs to cropland, grassland, forest, and mosaics of them. We also kept the bare and artificial classes, since they would have confused the other classes otherwise (Table S6).

- Identification of grid cells with cropland expansion by overlaying maps of two subsequent time steps. Cropland expansion was identified as changes from any other class to the reclassified cropland class or changes from any other classes except than the reclassified cropland class to the reclassified mosaic cropland classes.
 - Tracking of change trajectories, i.e., identification of classes that contributed to cropland expansion. The cropland expansion from the last step was used as a mask to keep only grid cells where cropland actually expanded between two time steps. This step yielded the information, which LULC type was converted to cropland (= ‘contributing source’).
 - Aggregation to ca. 0.5 x 0.5 degree grid. This step yielded the proportion of new cropland that originates in a particular LULC type within each ca. 0.5 x 0.5 degree grid cell.
 - Tracking how much of the original LULC type at t_1 within a ca. 0.5 x 0.5 degree grid cell was converted to cropland in t_2 (= ‘available source’).
 - Division of ‘contribution source’ by ‘available source’. By applying this step we could distinguish grid cells which did not contain a particular LULC type at t_1 (division not defined) from grid cells where a particular LULC type was available, but not converted to cropland (division result equals 0).
- As a result of the preprocessing we obtained maps, where each grid cell contained the fraction of the original LULC type at t_1 that was converted to cropland in t_2 . Subsequently we searched across these maps for priority algorithms of LULCC within ca. 0.5 x 0.5 degree grid cells for decadal time steps following a set of rules (Figure S3). A grid cell was classified as
- UNDEFINED, if either forest or grassland were not available at t_1 . For these cells a classification was not possible, since it is not clear which source class was converted with higher priority. For example, if the grid cell only contains grassland at time t_1 , grassland is logically converted to cropland. However, a forest first algorithm would be also true for this grid cell (and just not executed, because there was no forest to convert). The mosaic class was excluded here, since even it is not available, all algorithms could be detected with the following rules.
 - UNVEGETATED FIRST, if urban or bare classes in a grid cell were converted completely, while at the same time all other sources were available, but not or only partially converted. Additionally, grid cells where urban or bare classes were partially converted, while at the same time all other sources were available, but not converted.
 - FOREST FIRST, if more than 90% of the available forest in a grid cell was converted to cropland, while at the same time grassland was available, but less than 90% of it was converted. Additionally, grid cells where less than 90% of the available forest was converted, while at the same time grassland or mosaic classes were available, but not converted.
 - GRASSLAND FIRST, if more than 90% of the available grassland in a grid cell was converted to cropland, while at the same time forest was available, but less than 90% of it was converted. Additionally, grid cells where less than 90% of the available grassland was converted, while at the same time forest or mosaic classes were available, but not converted.

- PROPORTIONAL, if the mosaic class was converted, while at the same time grassland and forest were available, but not converted. Additionally, grid cells where the ratio of converted grassland and forest was between 0.5 and 1.5 were considered as an indicator for proportional reduction.
- COMPLEX, if at least forest and grassland were available as a source, but neither a preferential conversion nor a proportional conversion could be detected.

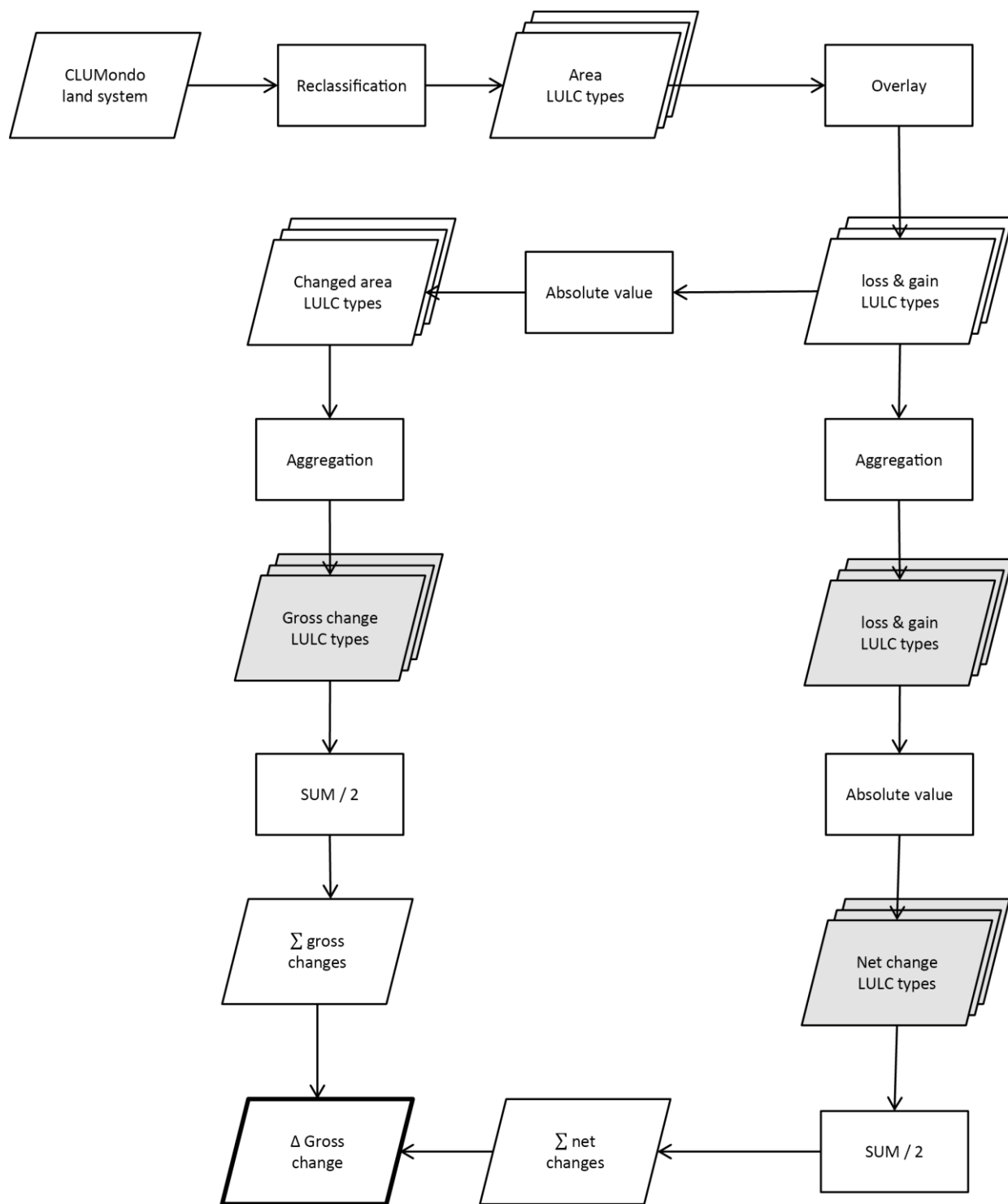


Figure S1: Preprocessing workflow of CLUMondo output for gross change analysis. Rectangles represent processing steps, parallelograms represent data. Grey shaded items emphasize aggregated data at ca. 0.5 x 0.5 degree resolution.

Table S6: CLUMondo land system classification and reclassification to broader LULC types.

LS code	Land system name	Reclassification
0	Cropland; extensive with few livestock	Cropland
1	Cropland; extensive with bovines, goats & sheep	Cropland
2	Cropland; medium intensive with few livestock	Cropland
3	Cropland; medium intensive with bovines, goats & sheep	Cropland
4	Cropland; intensive with few livestock	Cropland
5	Cropland; intensive with bovines, goats & sheep	Cropland
6	Mosaic cropland and grassland with bovines, goats & sheep	Mosaic cropland/grassland
7	Mosaic cropland (extensive) and grassland with few livestock	Mosaic cropland/grassland
8	Mosaic cropland (medium intensive) and grassland with few livestock	Mosaic cropland/grassland
9	Mosaic cropland (intensive) and grassland with few livestock	Mosaic cropland/grassland
10	Mosaic cropland (extensive) and forest with few livestock	Mosaic cropland/forest
11	Mosaic cropland (medium intensive) and forest with few livestock	Mosaic cropland/forest
12	Mosaic cropland (intensive) and forest with few livestock	Mosaic cropland/forest
13	Dense forest	Forest
14	Open forest with few livestock	Forest
15	Mosaic grassland and forest	Mosaic grassland/forest
16	Mosaic grassland and bare	Grassland
17	Natural grassland	Grassland
18	Grassland with few livestock	Grassland
19	Grassland with bovines, goats and sheep	Grassland
20	Bare	Bare
21	Bare with few livestock	Bare
22	Peri-urban & villages	Urban
23	Urban	Urban

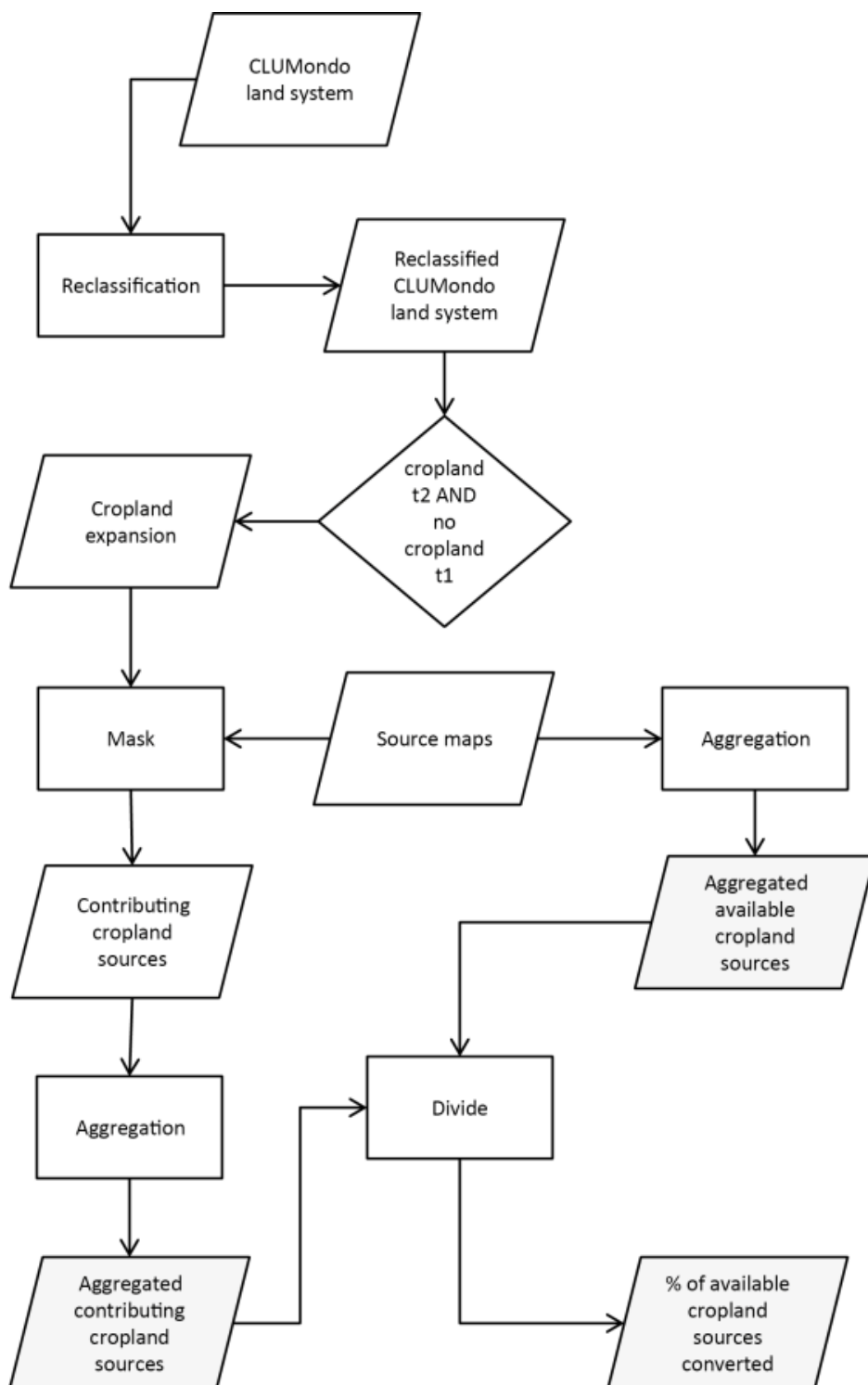


Figure S2: Preprocessing workflow of CLUMondo output for land-use change priority analysis. Rectangles represent processing steps, parallelograms represent data. Grey shaded items emphasize aggregated data at ca. 0.5 x 0.5 degree resolution.

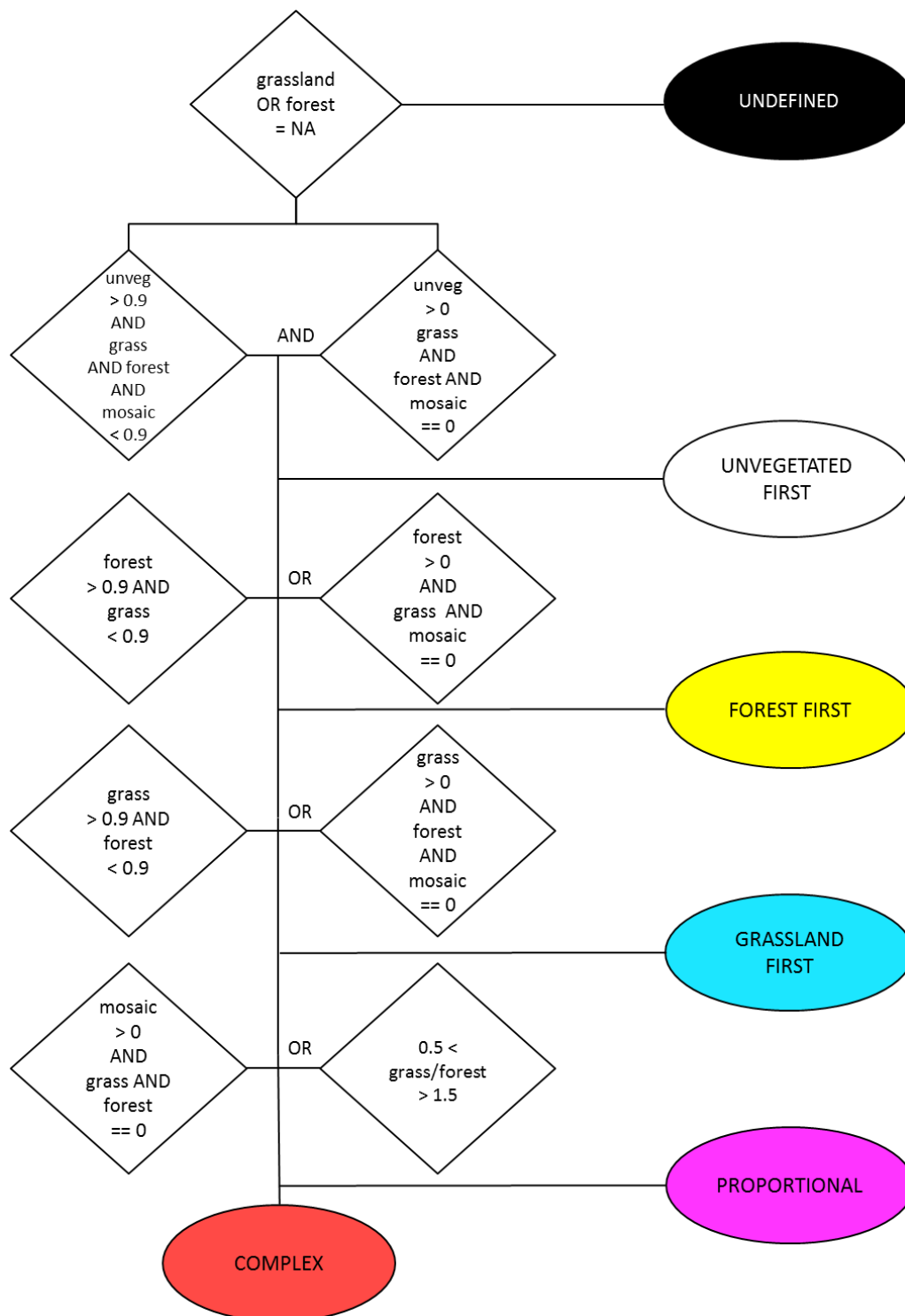
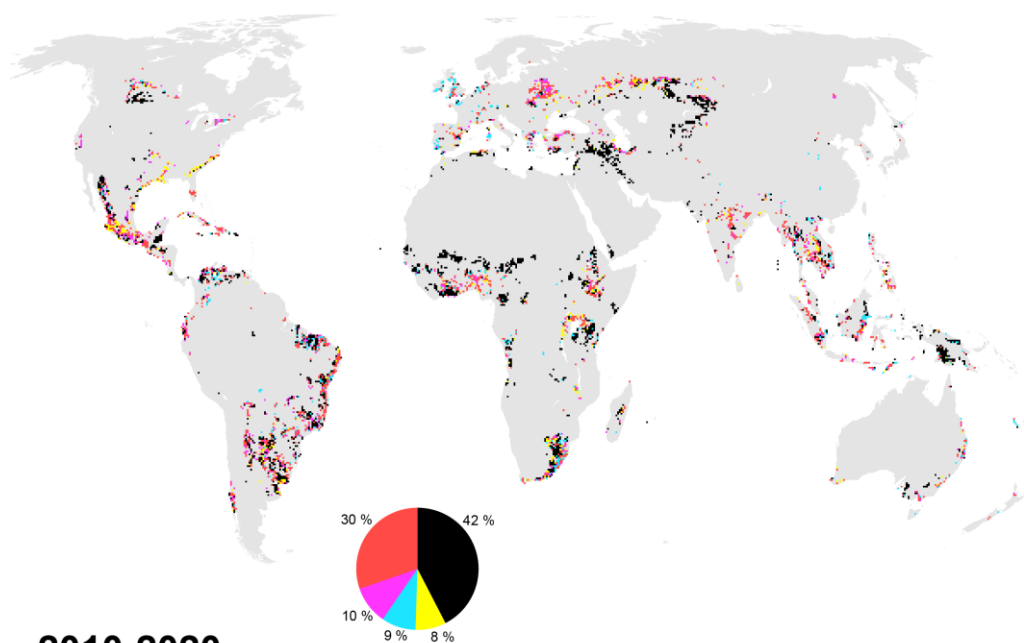
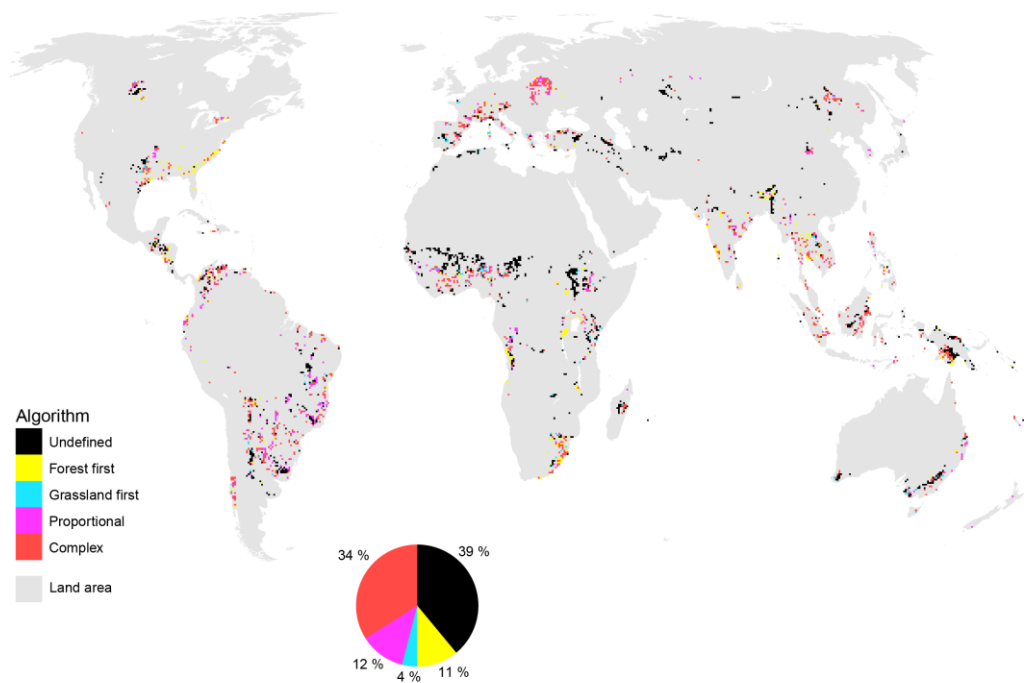


Figure S3: Classification rules applied to each ca. 0.5 x 0.5 degree grid cell to identify a predominant reduction of a particular source LULC type.

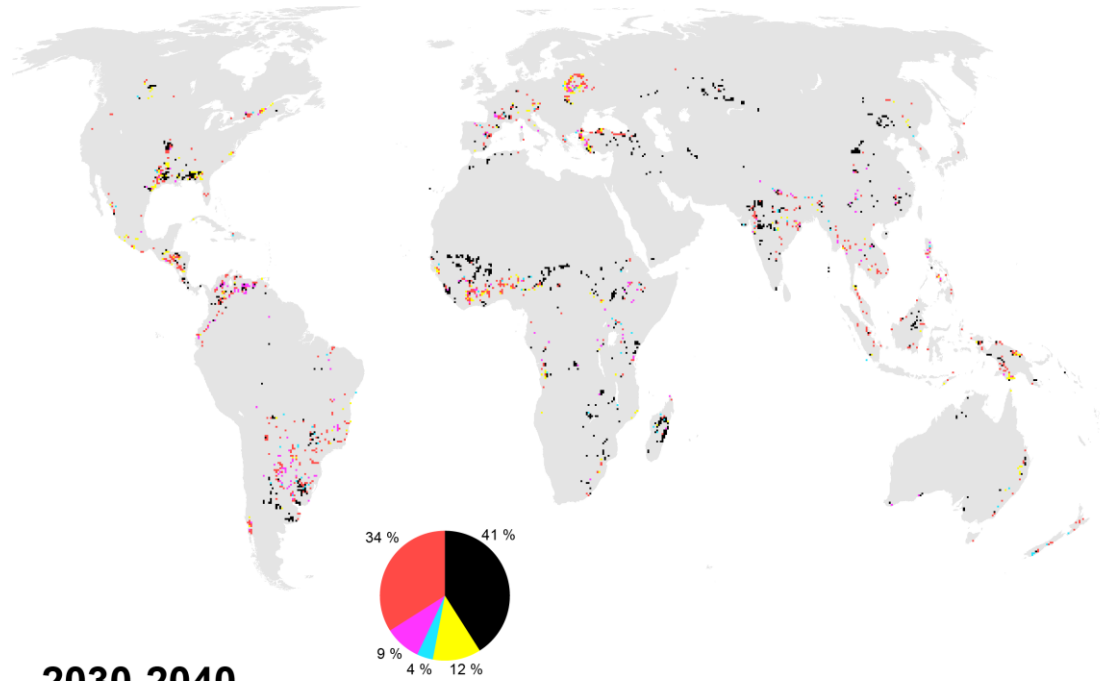
2000-2010



2010-2020



2020-2030



2030-2040

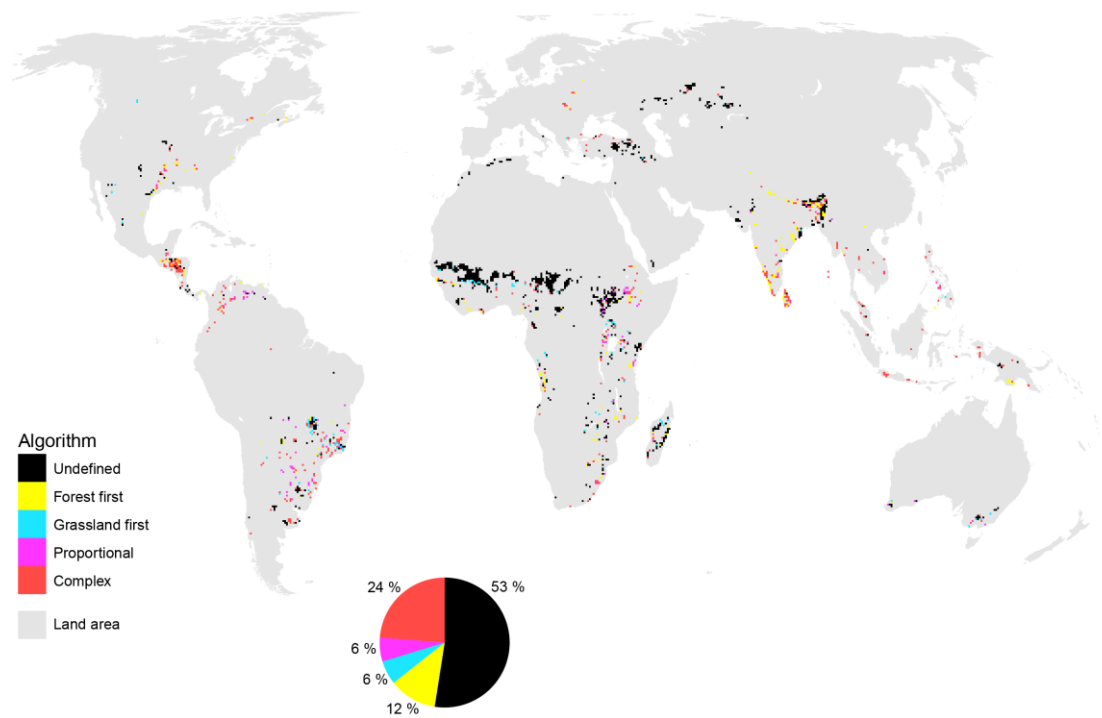


Figure S4: Transitions from natural vegetation to cropland as shown by the CLUMondo model (FAO 3 demand scenario) from 2000 to 2040 in decadal time steps. Colored grid cells represent areas with at least 10 % of cropland expansion within a ca. 0.5 x 0.5 degree grid cell. Grid cells are classified to forest first (yellow), grassland first (cyan), proportional (magenta) and complex reduction (red) algorithm as described in the text (for details see SI). Black grid cells denote areas where the validity of none algorithm could be detected.

References

- Alexander, P., Prestele, R., Verburg, P. H., Arneth, A., Baranzelli, C., Batista e Silva, F., Brown, C., Butler, A., Calvin, K., Dendoncker, N., Doelman, J. C., Dunford, R., Engström, K., Eitelberg, D., Fujimori, S., Harrison, P. A., Hasegawa, T., Havlik, P., Holzhauer, S., Humpenöder, F., Jacobs-Crisioni, C., Jain, A. K., Krisztin, T., Kyle, P., Laval, C., Lenton, T., Liu, J., Meiyappan, P., Popp, A., Powell, T., Sands, R. D., Schaldach, R., Stehfest, E., Steinbuks, J., Tabeau, A., van Meijl, H., Wise, M. A., and Rounsevell, M. D. A.: Assessing uncertainties in land cover projections, *Global Change Biology*, doi: 10.1111/gcb.13447, 2016.
- Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E.: A land use and land cover classification system for use with remote sensor data, USGS, Washington, D.C., 1976.
- Bartholome, E. and Belward, A. S.: GLC2000: a new approach to global land cover mapping from Earth observation data, *International Journal of Remote Sensing*, 26, 1959-1977, 2005.
- Bossard, M., Feranec, J., and Otahel, J.: CORINE land cover technical guide - Addendum 2000, Technical Report, 40, Available from: <http://www.eea.europa.eu/publications/tech40add> 2000.
- EEA: CLC2006 technical guidelines, EEA Technical report, 17/2007, Available from: http://www.eea.europa.eu/publications/technical_report_2007_17, 2007.
- Eitelberg, D. A., van Vliet, J., Doelman, J. C., Stehfest, E., and Verburg, P. H.: Demand for biodiversity protection and carbon storage as drivers of global land change scenarios, *Global Environmental Change*, 40, 101-111, 2016.
- FAOSTAT: Resources/Land. Rome, 2015.
- Kaplan, J. O., Krumhardt, K. M., Ellis, E. C., Ruddiman, W. F., Lemmen, C., and Goldewijk, K. K.: Holocene carbon emissions as a result of anthropogenic land cover change, *The Holocene*, 21, 775-791, 2010.
- 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.: Estimating global land use change over the past 300 years: The HYDE Database, *Global Biogeochemical Cycles*, 15, 417-433, 2001.
- Klein Goldewijk, K.: A historical land use data set for the Holocene; HYDE 3.2, doi: [10.17026/dans-znk-cfy3](https://doi.org/10.17026/dans-znk-cfy3), 2016.
- Klein Goldewijk, K., Beusen, A., and Janssen, P.: Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1, *The Holocene*, 20, 565-573, 2010.
- Klein Goldewijk, K., Beusen, A., van Dreht, G., and de Vos, M.: The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years, *Global Ecol Biogeogr*, 20, 73-86, 2011.
- Landscan: Landscan global population database, the 2004 revision. 2006.
- Lemmen, C.: World distribution of land cover changes during Pre- and Protohistoric Times and estimation of induced carbon releases, *Géomorphologie: relief, processus, environnement*, 15, 2009.

- Livi Bacci, M.: A concise history of world population, Blackwell, Malde, MA, 2007.
- 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.
- Maddison, A.: The world economy: A millenial perspective, OECD Publications, Paris, 2001.
- 5 McEvedy, C. and Jones, R.: Atlas of world population history, Penguin, Harmondsworth, New York, 1978.
- 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.
- 10 Prestele, R., Alexander, P., Rounsevell, M., Arneth, A., Calvin, K., Doelman, J., Eitelberg, D., Engström, K., Fujimori, S., Hasegawa, T., Havlik, P., Humpenöder, F., Jain, A. K., Krisztin, T., Kyle, P., Meiyappan, P., Popp, A., Sands, R. D., Schaldach, R., Schüngel, J., Stehfest, E., Tabeau, A., van Meijl, H., van Vliet, J., and Verburg, P. H.: Hotspots of uncertainty in land use and land cover change projections: a global scale model comparison, *Global Change Biology*, doi: 10.1111/gcb.13337, 2016.
- 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.
- 15 Ramankutty, N., Foley, J. A., and Olejniczak, N. J.: People on the Land: Changes in Global Population and Croplands during the 20th Century, *AMBIO: A Journal of the Human Environment*, 31, 251-257, 2002.
- Ruddiman, W. F. and Ellis, E. C.: Effect of per-capita land use changes on Holocene forest clearance and CO2 emissions, *Quaternary Science Reviews*, 28, 3011-3015, 2009.
- Stehman, S. V., Wickham, J. D., Smith, J. H., and Yang, L.: Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results, *Remote Sensing of Environment*, 86, 500-516, 2003.
- 20 Van Asselen, S. and Verburg, P. H.: Land cover change or land-use intensification: simulating land system change with a global-scale land change model, *Global Change Biology*, 19, 3648-3667, 2013.
- van Asselen, S. and Verburg, P. H.: A Land System representation for global assessments and land-use modeling, *Global Change Biology*, 18, 3125-3148, 2012.
- 25 Wickham, J. D., Stehman, S. V., Fry, J. A., Smith, J. H., and Homer, C. G.: Thematic accuracy of the NLCD 2001 land cover for the conterminous United States, *Remote Sensing of Environment*, 114, 1286-1296, 2010.
- Wickham, J. D., Stehman, S. V., Gass, L., Dewitz, J., Fry, J. A., and Wade, T. G.: Accuracy assessment of NLCD 2006 land cover and impervious surface, *Remote Sensing of Environment*, 130, 294-304, 2013.
- Wirtz, K. W. and Lemmen, C.: A Global Dynamic Model for the Neolithic Transition, *Climatic Change*, 59, 333-367, 2003.