S1 Supplementary text

S1.1 General preprocessing

All spatially explicit datasets are rasterized and regridded to a spatial resolution of $0.0083°$ or $0.25°$, depending on their purpose (see Fig. 1). Environmental predictors are regridded to a spatial resolution of $0.0083°$ using the Climate Data Operators software (CDO; Schulzweida, 2023). A land mask derived from (Buchhorn et al., 2020) is applied to exclude areas with permanent water and snow cover. Datasets required for the reclassification of the global predictions to plant functional types (PFTs) are regridded to a spatial resolution of $0.25°$ likewise using CDO (Schulzweida, 2023), with two exceptions: annual global land cover maps produced by the European Space Agency Climate Change Initiative – Land Cover (referred to as ESA CCI LC) (Copernicus Climate Change Service, 2019) are processed and regridded using the CCI-LC User Tool (European Space Agency, 2019), while modeled C4 grass fractions from Luo et al. (2024) are regridded via nearest-neighbor interpolation using the Python package xarray (Hoyer and Hamman, 2017). In line with the other input datasets used in the BLUE model (Hansis et al., 2015), the LUH2 land area is adopted as the target grid for all datasets regridded to $0.25°$ resolution.

Annual maps of the global PFT dataset by Harper et al. (2023a) based on ESA CCI data (referred to as CCI PFTs; time period: 1992–2020) and the modeled C4 grass fractions by Luo et al. (2024) (time period: 2001–2019) are aggregated by averaging across the respective time steps to condense their information into single maps that approximate the long-term surface cover, factoring out patterns caused by climate variability and disturbances. To reduce computational demands, only the CCI PFT maps for the years 1992, 1995, 2000, 2005, 2010, 2015, and 2020 are regridded and included in the temporal averaging. The resulting CCI PFT dataset encompasses spatially explicit maps of the percentage cover of 14 different PFTs: three abiotic PFTs [(1) permanent inland water bodies, (2) permanent snow and ice cover, (3) bare soil], two anthropogenic PFTs [(4) managed grasses, (5) built-up area], and nine non-anthropogenic vegetation PFTs [(6) natural grass, (7) broadleaf deciduous shrubs, (8) broadleaf evergreen shrubs, (9) needleleaf deciduous shrubs, (10) needleleaf evergreen shrubs, (11) broadleaf deciduous trees, (12) broadleaf evergreen trees, (13) needleleaf deciduous trees, (14) needleleaf evergreen trees].

The ESA CCI LC time series (Copernicus Climate Change Service, 2019) is used to exclude grid cells that do not support vegetation under current environmental conditions from the 25 environmental predictors and global PNV predictions. All grid cells consistently classified as non-vegetated land cover (bare areas, consolidated bare areas, unconsolidated bare areas, water bodies, or permanent snow and ice) over the period 1992–2020 are removed and assigned to the final PFT maps as 100% abiotic.

Three external PNV classification schemes are incorporated into the workflow: the Holdridge life zones (HLZs; Holdridge, 1967), the 2010 update of the global ecological zones for FAO forest reporting (GEZ2010; FAO, 2012), and the terrestrial biomes (TB2017; Dinerstein et al., 2017). Spatially explicit HLZ maps for the period 1979–2013 are obtained from (Elsen et al., 2022). The HLZs are further aggregated across the precipitation and potential evapotranspiration ratio gradients into six life zones: tropical, subtropical/warm temperate, cool temperate, boreal, subpolar, and polar. The ecological zones of the GEZ2010 map are aggregated according to their Level 1 domain (FAO, 2012) into five categories: tropical, subtropical, temperate, boreal, and polar.

S1.2 Biome datasets: Removal of faulty tundra data points

As pollen from intensively managed agricultural land may be misclassified as tundra (Binney et al., 2017), we discard all "Tundra" data points during the processing and harmonization of the biome datasets unless their location meets at least one of the following criteria: (1) HLZ = polar or subpolar, (2) GEZ2010 = polar, or (3) terrestrial biome of TB2017 = tundra or montane grasslands and shrublands.

S1.3 Reclassification scheme: Preparation of the spatially explicit datasets

The preparation of spatially explicit data used to reclassify the 30 global PNV maps into PFT maps compatible with BLUE involves multiple processing steps (see Table S7 for the reclassification scheme).

Global grid cell fractions of plant form, as well as leaf type and phenology of tree fractions, are derived from the averaged and regridded CCI PFTs (Harper et al., 2023a). For each grid cell, plant form fractions are determined as relative proportions of trees (sum of the four tree CCI PFTs), shrubs (sum of the four shrub CCI PFTs), and grasses (values for the CCI PFT "natural grasses") relative to the sum of all nine non-anthropogenic vegetation CCI PFTs. Tree cover is further divided into broadleaf evergreen (BE), broadleaf deciduous (BD), needleleaf evergreen (NE), and needleleaf deciduous (ND) trees to obtain the fractional coverage of leaf type and phenology in each grid cell. Approximately 3% of the grid cells with mixed biome classes lack the CCI PFT values required for their allocation. In these cases, values are estimated through inverse distance weighted interpolation from neighboring grid cells.

The extent and proportion of tropical vegetation (Fig. S1a) are primarily derived from the aggregated HLZs (Elsen et al., 2022). Tropical HLZs are classified as tropical vegetation. Subtropical HLZs are split according to the Level 1 domain of the GEZ2010 map (FAO, 2012): the area of tropical ecological zones is classified as tropical vegetation, while the area of all other ecological zones is classified as non-tropical, in accordance with the equivalent climate regions of the Intergovernmental Panel on Climate Change (IPCC) (FAO, 2012). The remaining HLZs are classified as non-tropical vegetation.

The two shrub PFTs in BLUE are based on the ecosystems "Tropical woodland and shrubland" ("Raingreen shrubs" in BLUE) and "Temperate woodland and shrubland" ("Summergreen shrubs" in BLUE) from Houghton et al. (1983). Accordingly, depending on the tropical vegetation fraction shown in Fig. S1a, the shrub fraction of each grid cell is classified as "Raingreen shrubs" or "Summergreen shrubs".

All grass fractions are allocated to the BLUE PFTs "C3 natural grasses" and "C4 natural grasses" according to the grid cell fractions of C4 grasses shown in Fig. S1b. The map in Fig. S1b is primarily derived from the temporally averaged C4 grass fractions by Luo et al. (2024). Data gaps caused by the restriction of the modeling extent to observed natural grasses are filled with C4 grass fractions obtained through the cross-over temperature approach proposed by Griffith et al. (2015), which calculates the fraction of C4 grass as the fraction of months in a calendar year with a maximum monthly temperature (wclim_tmax) $\geq 27\,^{\circ}\mathrm{C}$ and monthly precipitation (wclim_prec) $\geq 25\,\mathrm{mm}$ (see Table S3 for the datasets). A spatial 3×3 median filter is applied to the resulting composite map to minimize the influence of outliers in individual grid cells.

S2 Supplementary figures

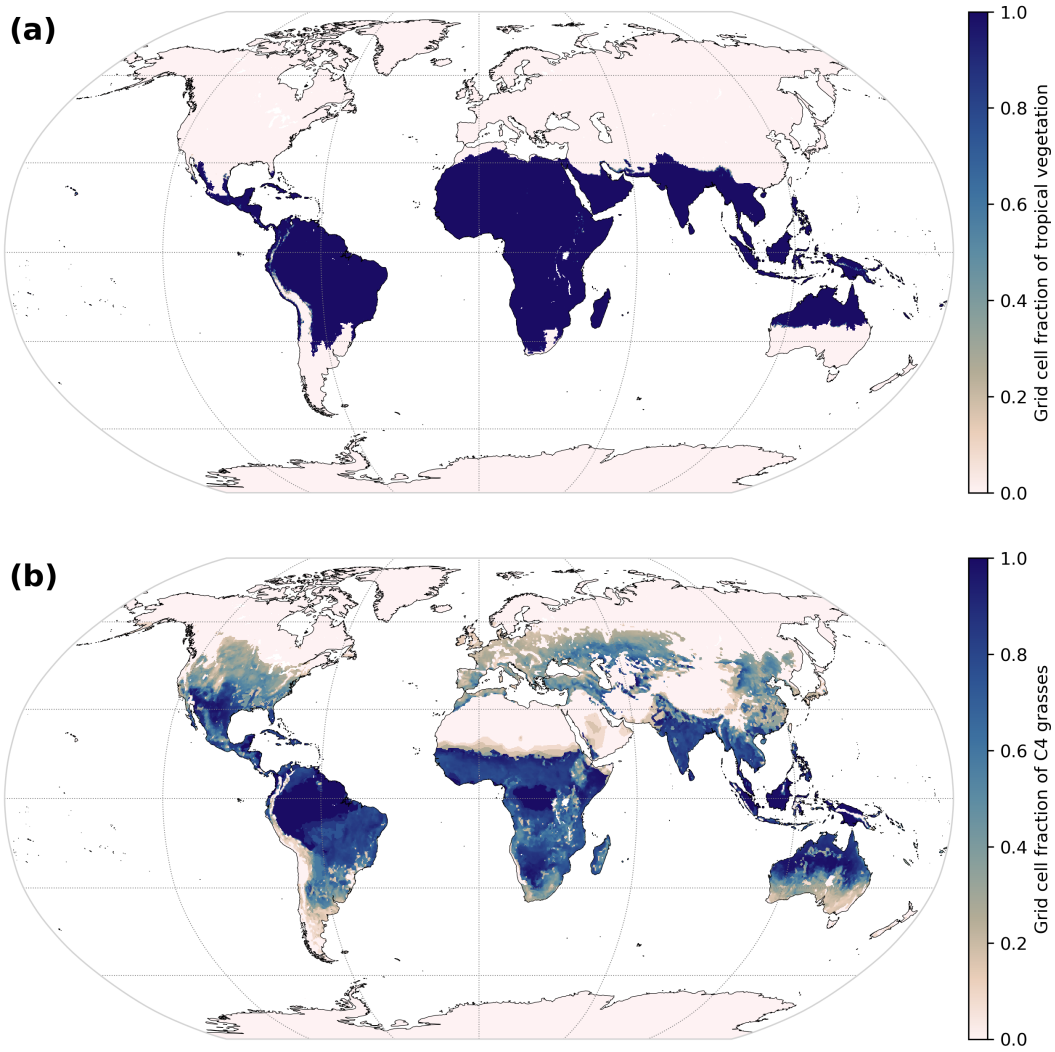

Figure S1. Global grid cell fractions of (a) tropical vegetation used to define the climate region of trees and shrubs during the reclassification of the PNV biomes to BLUE PFTs, and (b) C4 grasses used to define the photosynthetic pathway of grasses during the reclassification of the PNV biomes to BLUE PFTs.

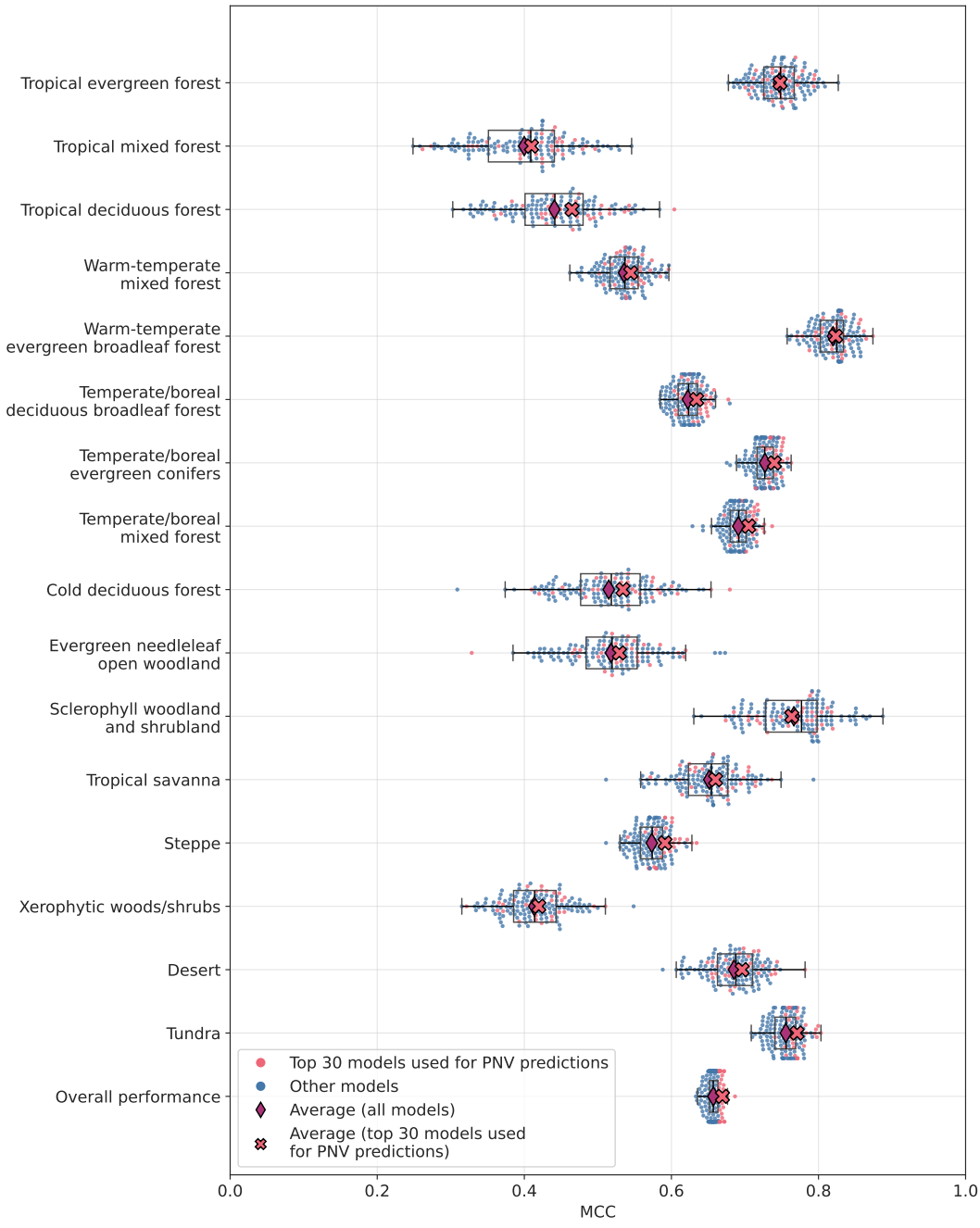

Figure S2. Individual predictive performance for all 150 trained random forest classifiers of each class and overall, quantified through Matthews correlation coefficient (MCC). The 30 models selected for the PNV predictions (best overall MCC values) are highlighted in light coral.

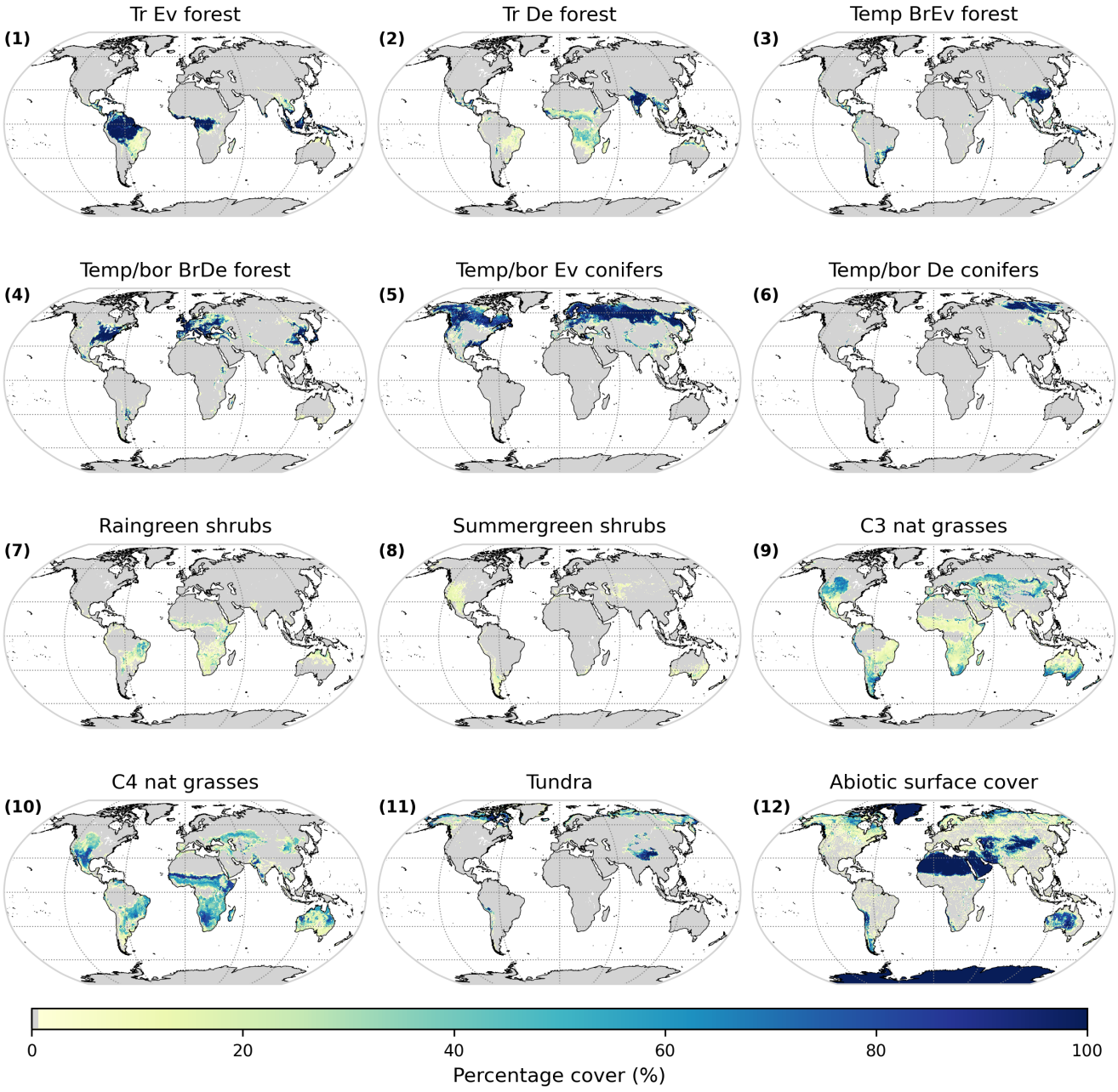

Figure S3. Spatial distribution of the percentage cover of each plant functional type (PFT) of the best-guess PFT map (PFT_{ML}). Grid cells with zero percentage cover are colored gray. Subplot numbers correspond to the IDs in Table S8, which includes the full names of PFTs. Refer to Fig. 4 in the main text for the spatial distribution of the percentage cover of aggregated categories of the best-guess PFT map.

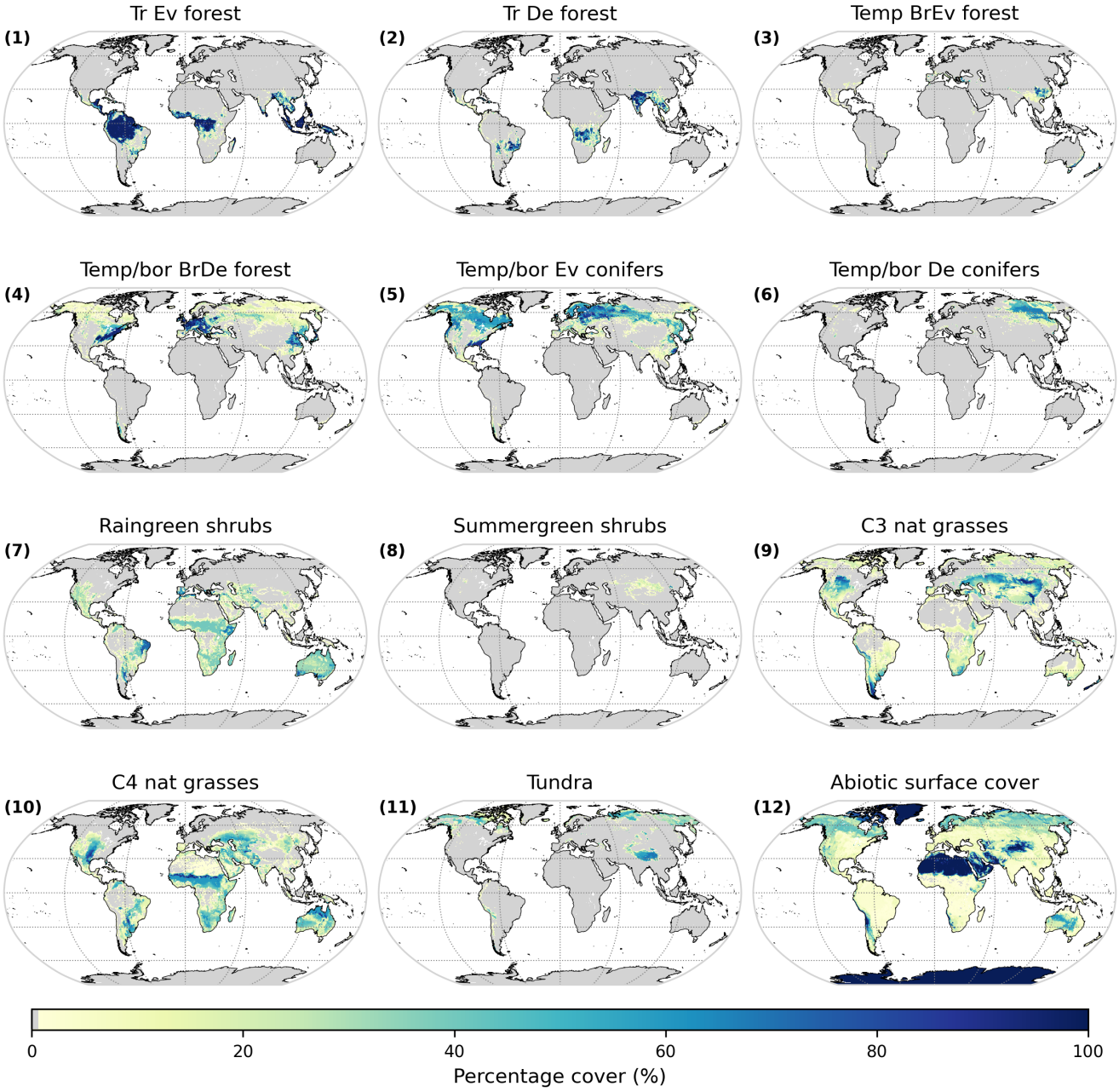

Figure S4. Spatial distribution of the percentage cover of each plant functional type (PFT) of the default PFT map (PFT_{Hansis}). Grid cells with zero percentage cover are colored gray. Subplot numbers correspond to the IDs in Table S8, which includes the full names of PFTs.

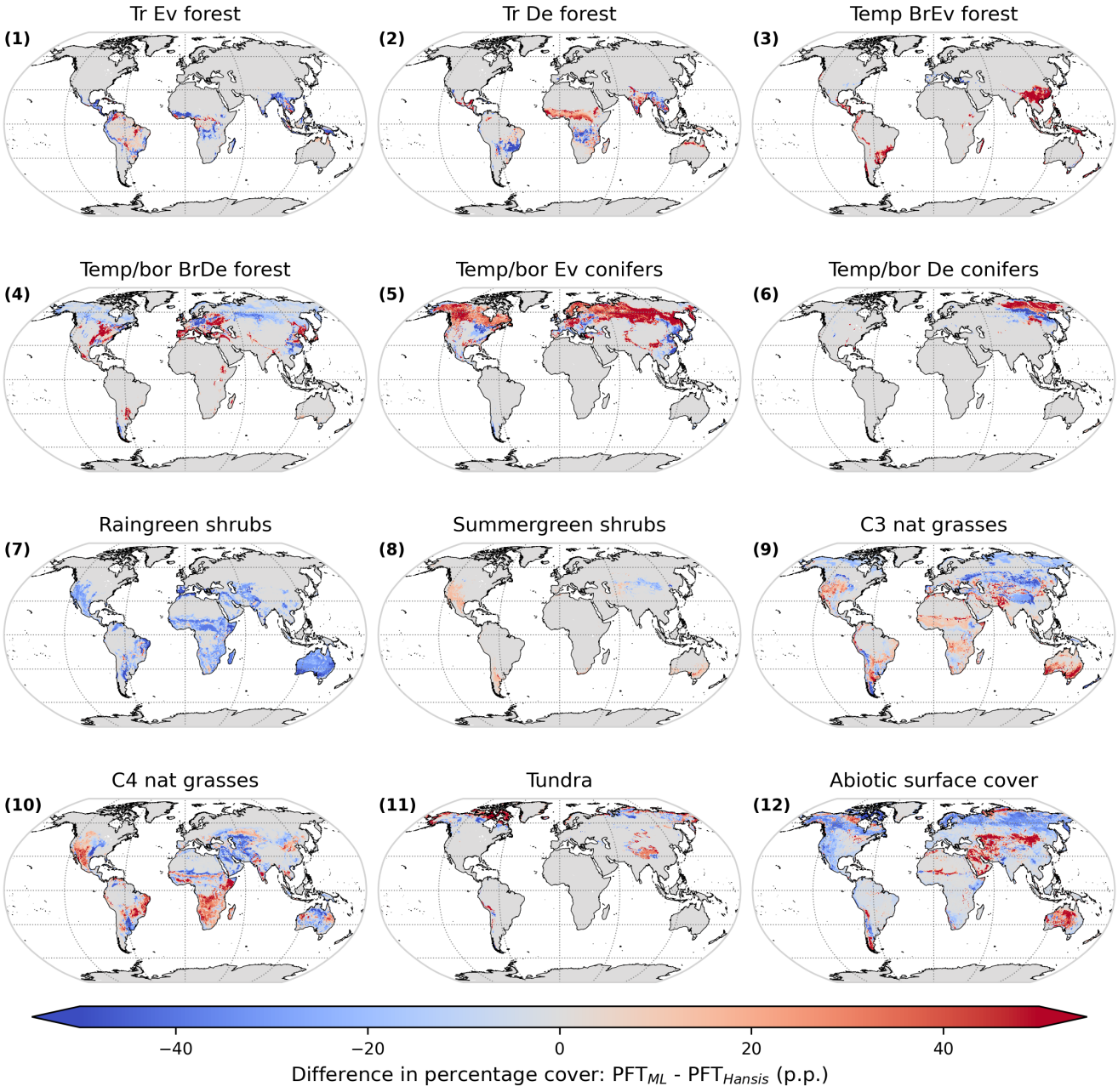

Figure S5. Spatial distribution of the differences in percentage cover of each plant functional type (PFT) between the best-guess PFT map (PFT_{ML}) (see Fig. S3) and the default PFT map of BLUE (PFT_{Hansis}) (see Fig. S4). Subplot numbers correspond to the IDs in Table S8, which includes the full names of PFTs.

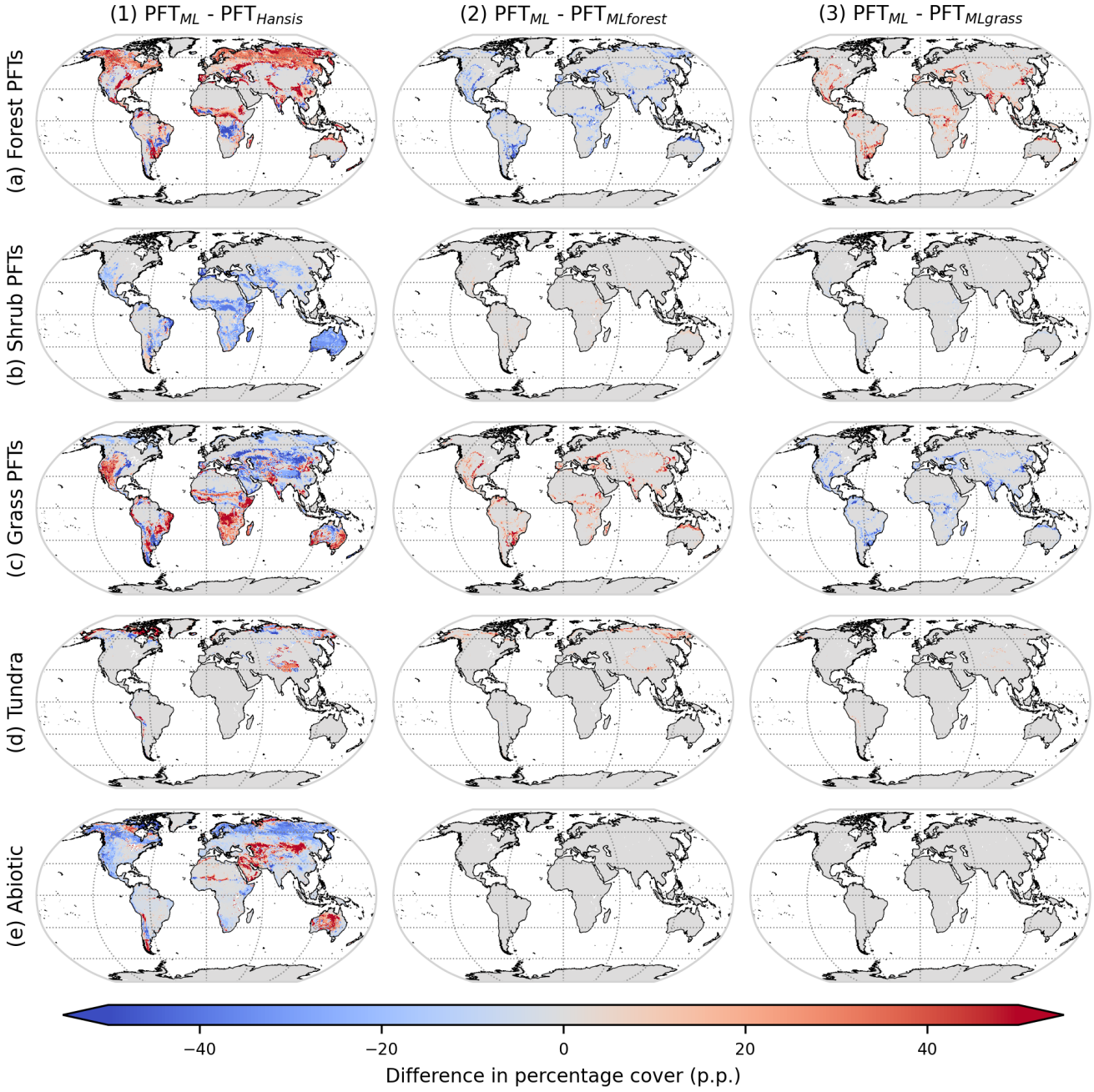

Figure S6. Spatial distribution of the differences in percentage cover of plant functional type (PFT) categories between the best-guess PFT map (PFT_{ML}) and: (1) the default PFT map (PFT_{Hansis}), (2) the sensitivity PFT map with maximum forest extent ($PFT_{MLforest}$), and (3) the sensitivity PFT map with maximum grass extent ($PFT_{MLgrass}$). The PFT categories are: (a) forest PFTs (sum of IDs 1–6), (b) shrub PFTs (sum of IDs 7–8), (c) grass PFTs (sum of IDs 9–10), (d) tundra (ID 11), and (e) abiotic surface cover. Grid cells with zero percentage cover are colored grey.

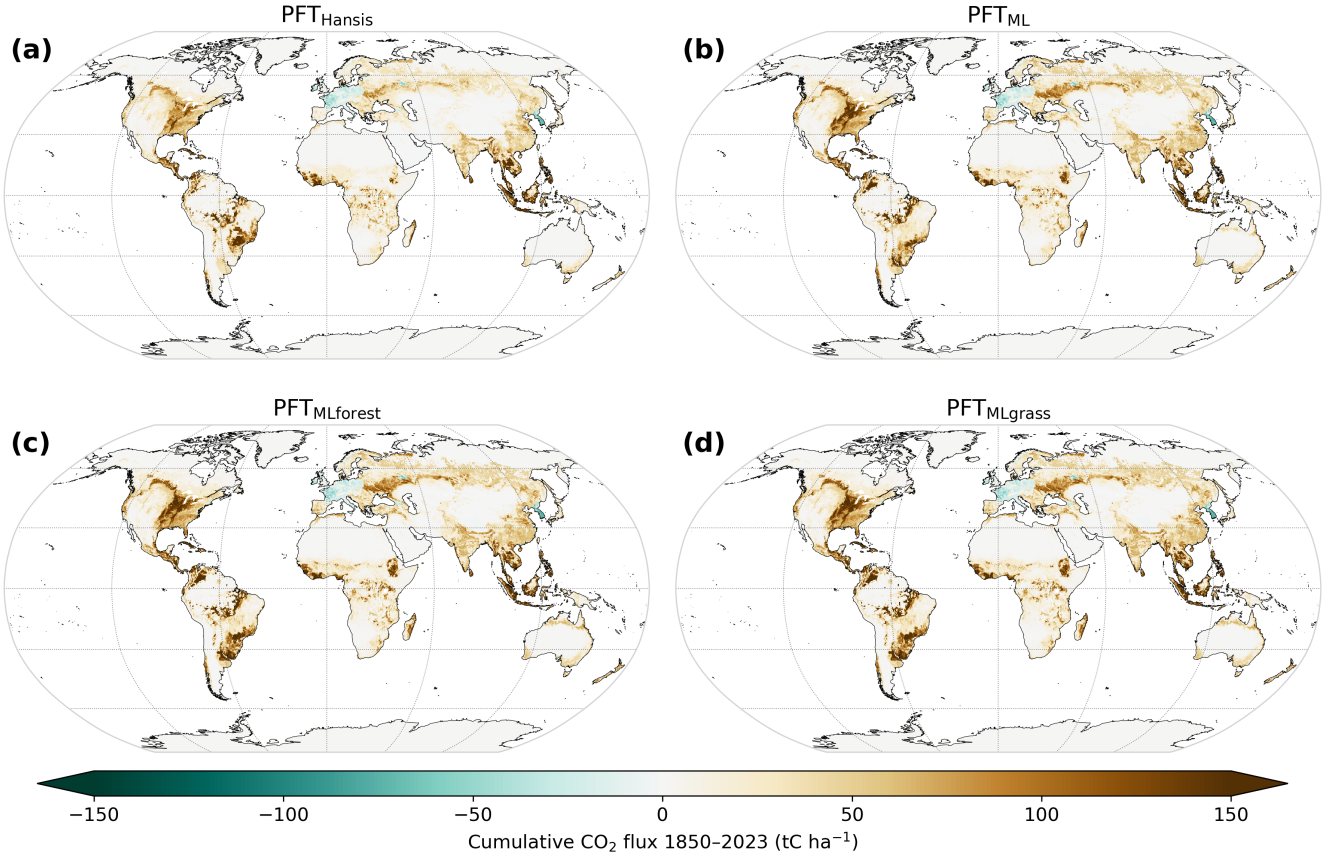

Figure S7. Spatial distribution of cumulative net land use and land-use change flux (F_{LUC}) estimates ($tC\ ha^{-1}$) derived from the bookkeeping model BLUE (1850–2023). Simulation names refer to the underlying plant functional type (PFT) maps: (a) PFT_{Hansis} (default PFT map of BLUE), (b) PFT_{ML} (best-guess PFT map from this study), and the sensitivity maps from this study (c) $PFT_{MLforest}$ and (d) $PFT_{MLgrass}$, representing maximum forest and grassland extent, respectively. Positive values denote carbon emissions, whereas negative values denote carbon removals.

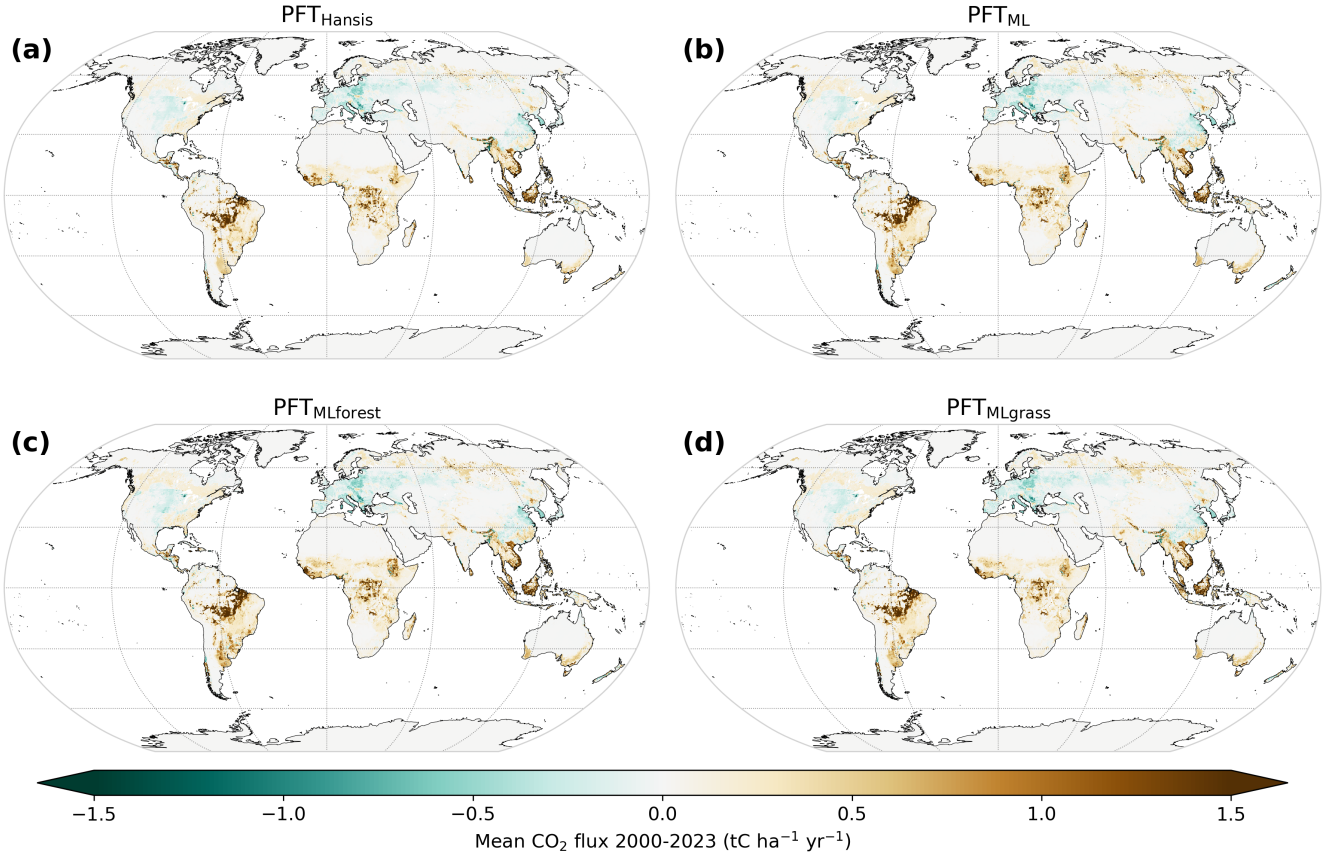

Figure S8. Spatial distribution of mean annual net land use and land-use change flux (F_{LUC}) estimates ($tC\ ha^{-1}$) derived from the bookkeeping model BLUE (2000–2023). Simulation names refer to the underlying plant functional type (PFT) maps: (a) PFT_{Hansis} (default PFT map of BLUE), (b) PFT_{ML} (best-guess PFT map from this study), and the sensitivity maps from this study (c) $PFT_{MLforest}$ and (d) $PFT_{MLgrass}$, representing maximum forest and grassland extent, respectively. Positive values denote carbon emissions, whereas negative values denote carbon removals.

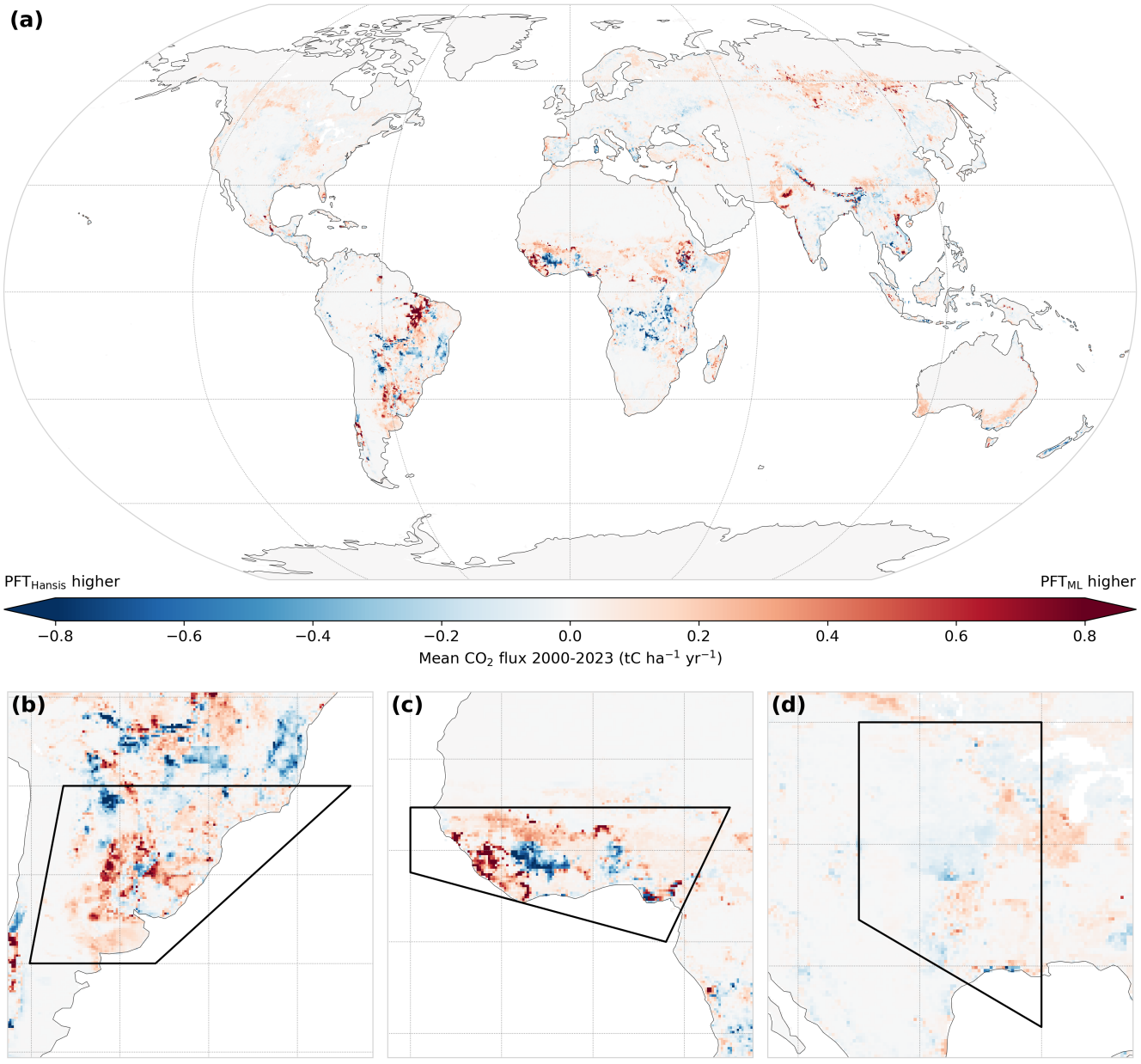

Figure S9. Differences between mean annual net land use and land-use change flux (F_{LUC}) estimates ($tC\,ha^{-1}$) derived from the bookkeeping model BLUE using the plant functional type (PFT) maps PFT_{ML} and PFT_{Hansis} (2000–2023). PFT_{ML} refers to the best-guess PFT map from this study, whereas PFT_{Hansis} refers to the default PFT map of BLUE. Panel (a) shows the global distribution of mean annual F_{LUC} differences, whereas panels (b)–(d) provide insets for IPCC regions discussed in the main text: (b) Southeastern South America (SES), (c) Western Africa (WAF), and (d) Central North America (CNA).

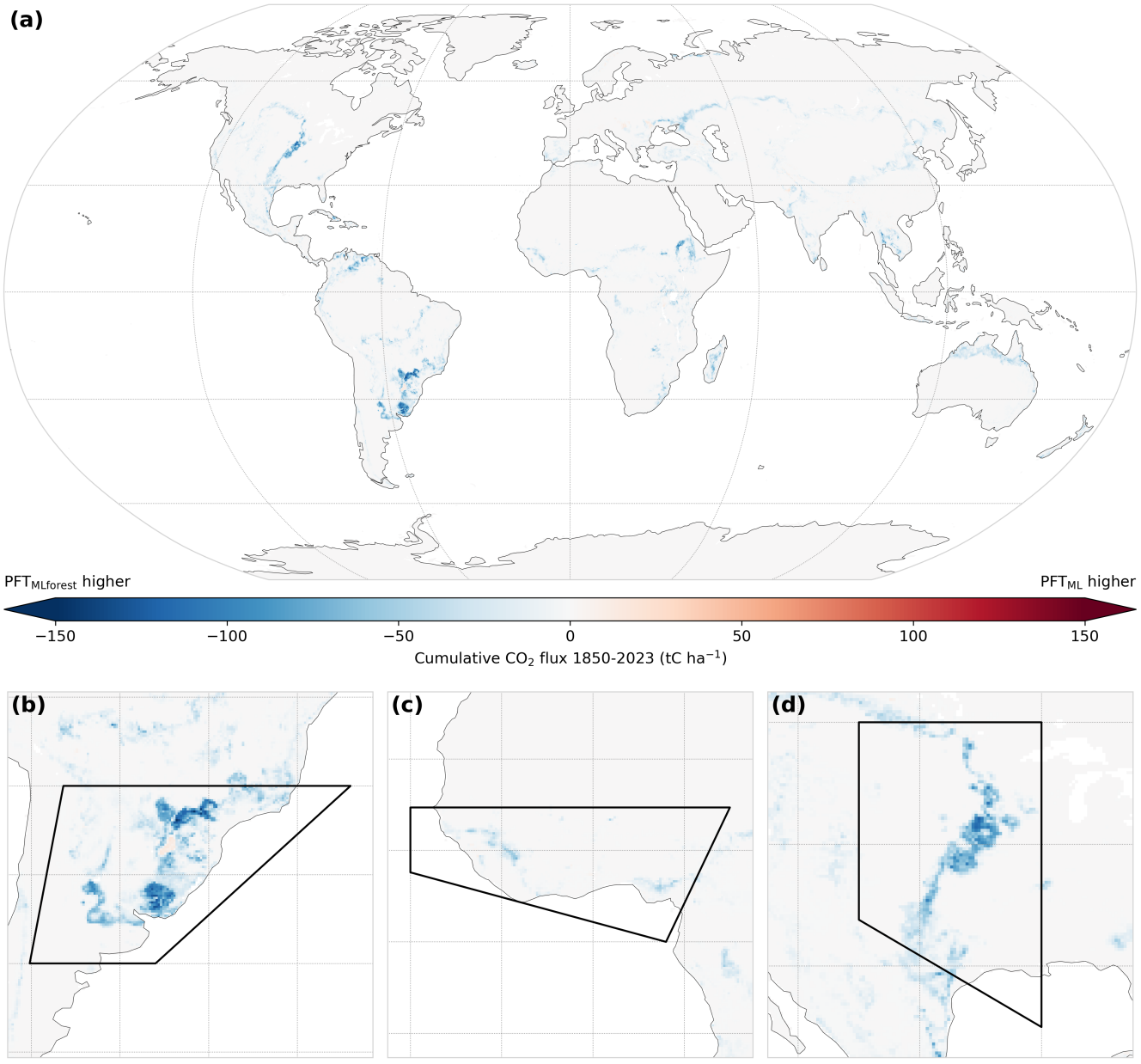

Figure S10. Differences between cumulative net land use and land-use change flux (F_{LUC}) estimates ($tC\,ha^{-1}$) derived from the bookkeeping model BLUE using the plant functional type (PFT) maps PFT_{ML} and $PFT_{MLforest}$ (1850–2023). PFT_{ML} refers to the best-guess PFT map from this study, whereas $PFT_{MLforest}$ refers to the sensitivity map with maximum forest extent. Panel (a) shows the global distribution of cumulative F_{LUC} differences, whereas panels (b)–(d) provide insets for IPCC regions discussed in the main text: (b) Southeastern South America (SES), (c) Western Africa (WAF), and (d) Central North America (CNA).

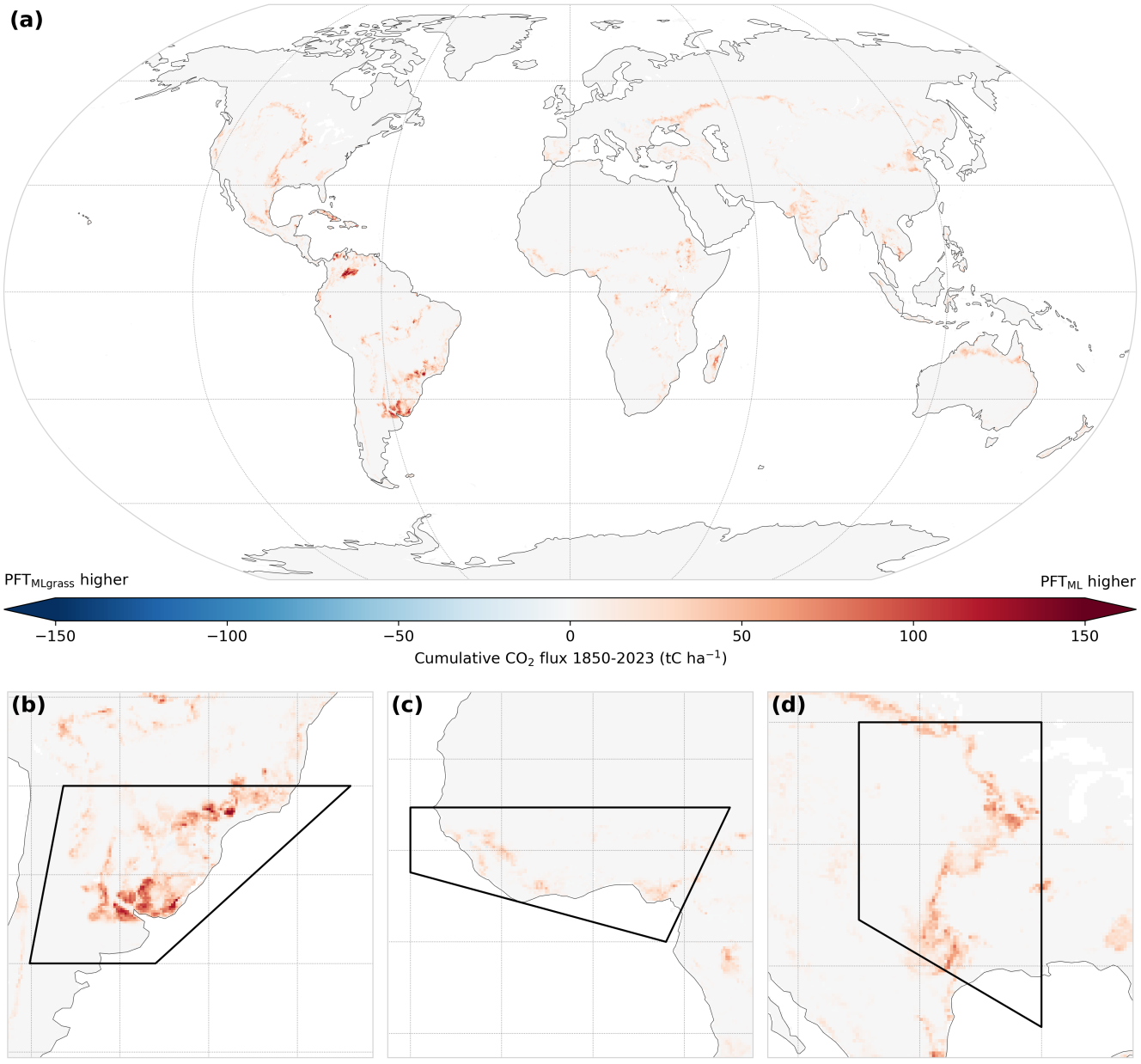

Figure S11. Differences between cumulative net land use and land-use change flux (F_{LUC}) estimates ($tC\,ha^{-1}$) derived from the bookkeeping model BLUE using the plant functional type (PFT) maps PFT_{ML} and $PFT_{MLgrass}$ (1850–2023). PFT_{ML} refers to the best-guess PFT map from this study, whereas $PFT_{MLgrass}$ refers to the sensitivity map with maximum grassland extent. Panel (a) shows the global distribution of cumulative F_{LUC} differences, whereas panels (b)–(d) provide insets for IPCC regions discussed in the main text: (b) Southeastern South America (SES), (c) Western Africa (WAF), and (d) Central North America (CNA).

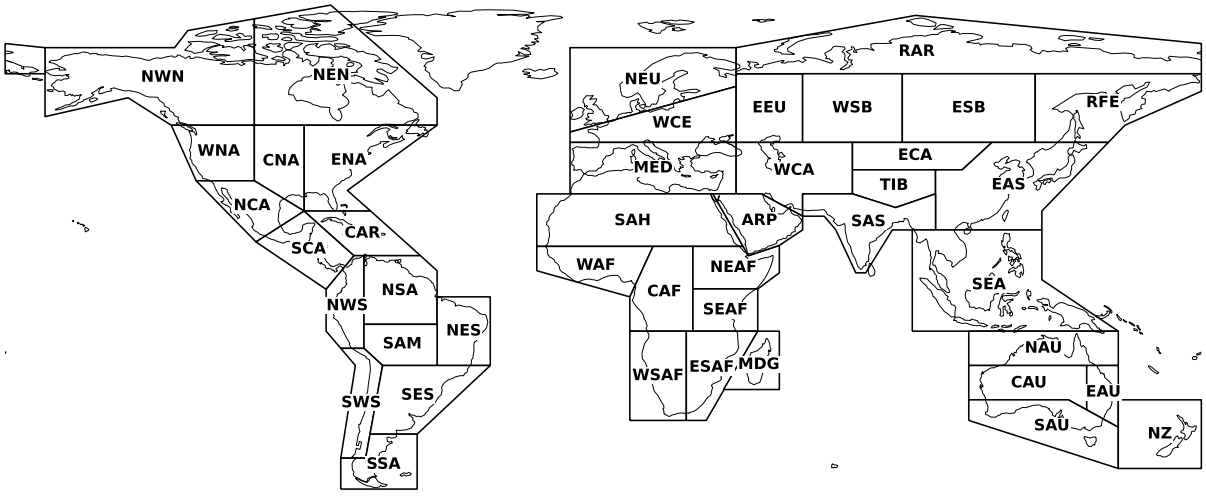

Figure S12. Updated IPCC reference land regions (referred to as IPCC regions in this study; Iturbide et al., 2020) used for regional analysis of net land use and land-use change flux (F_{LUC}) estimates.

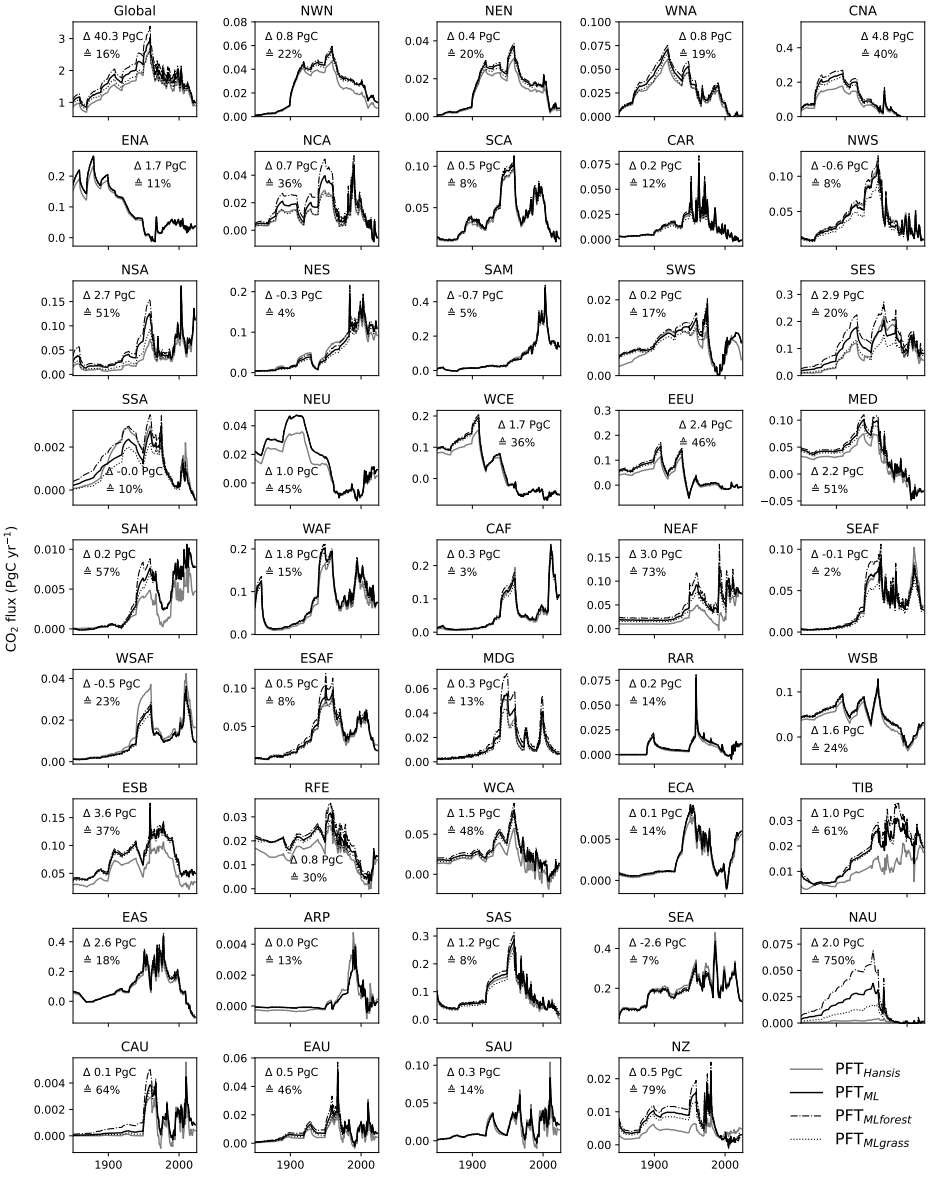

Figure S13. Annual net land use and land-use change flux (F_{LUC}) estimates (PgC yr^{-1}) derived from the bookkeeping model BLUE for the IPCC regions (1850–2023). Simulation names refer to the underlying plant functional type (PFT) maps: (a) PFT_{Hansis} (default PFT map of BLUE), (b) PFT_{ML} (best-guess PFT map from this study), and the sensitivity maps from this study (c) $PFT_{MLforest}$ and (d) $PFT_{MLgrass}$, representing maximum forest and grassland extent, respectively. The upper value (Δ) in each subplot reports the difference in cumulative F_{LUC} estimates between the simulations using PFT_{ML} and PFT_{Hansis}. The lower value ($\widehat{=}$) reports the corresponding relative change, defined as absolute value of the difference in cumulative F_{LUC} relative to the cumulative F_{LUC} estimate based on PFT_{Hansis}.

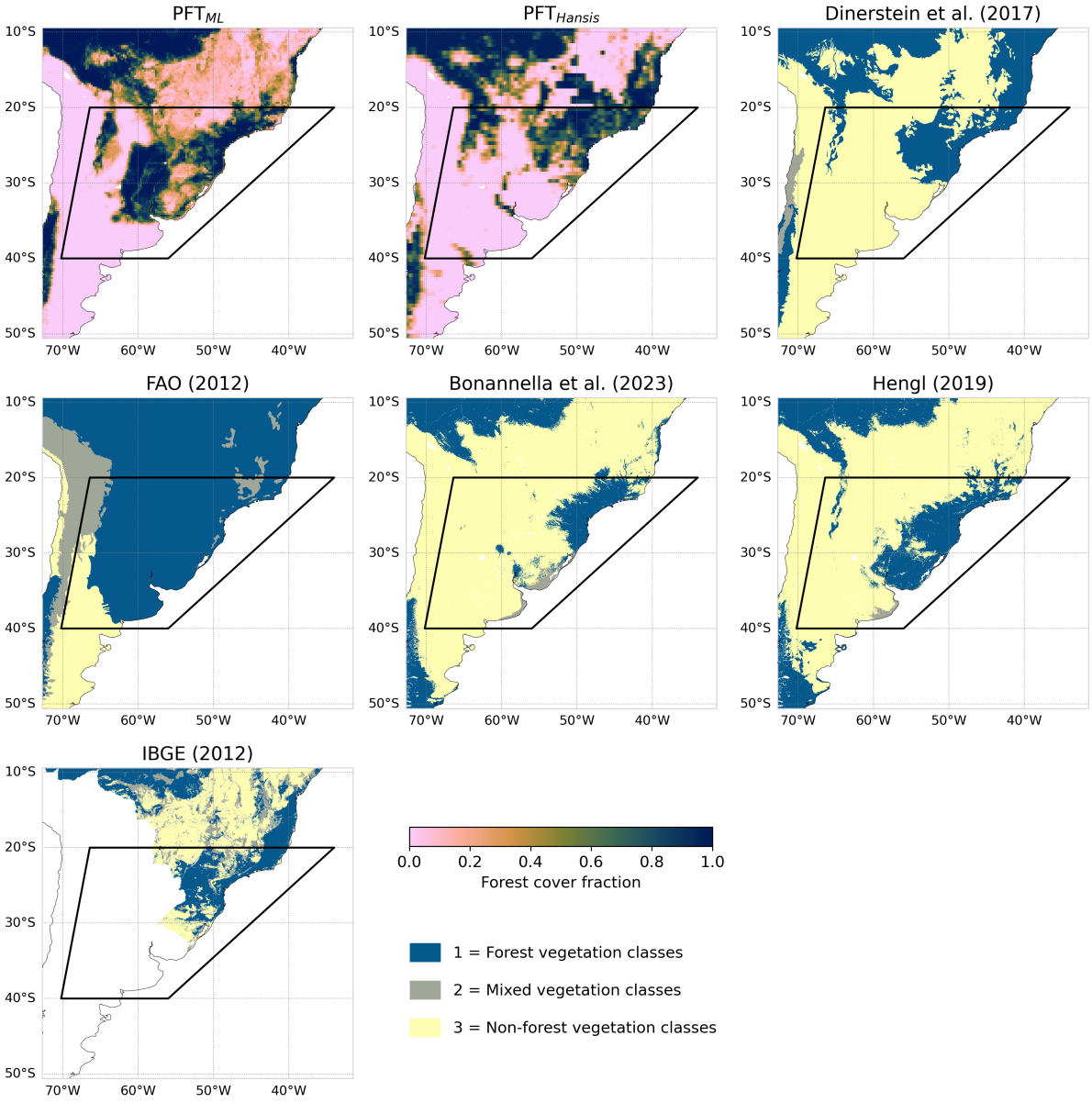

Figure S14. Comparison of forest cover under potential natural vegetation (PNV) for the IPCC region Southeastern South America (SES). Sources for the delineated forest cover include the best-guess plant functional type (PFT) map (PFT_{ML}), the default PFT map (PFT_{Hansis}), the terrestrial biomes (Dinerstein et al., 2017), the 2010 update of the global ecological zones for FAO forest reporting (FAO, 2012), the current global distribution of biomes under potential natural vegetation (Bonannella et al., 2023, version 2), the potential distribution of biomes (Potential Natural Vegetation) at 250 m spatial resolution (Hengl, 2019, version v0.2), and the natural vegetation of Brazil derived from the dominant phytoecological region (IBGE, 2012, version 2023). Forest cover fractions for PFT_{ML} and PFT_{Hansis} are obtained as the sum of the six forest PFTs (IDs 1–6; see Table S8). For the other sources, classes of the respective vegetation maps are summarized into three broad categories: (1) Forest vegetation classes, (2) Mixed vegetation classes (classes defined as a mixture of forests and non-forest vegetation units), and (3) Non-forest vegetation classes (savannas, shrublands, steppes, deserts, tundra, and other distinct non-forest classes).

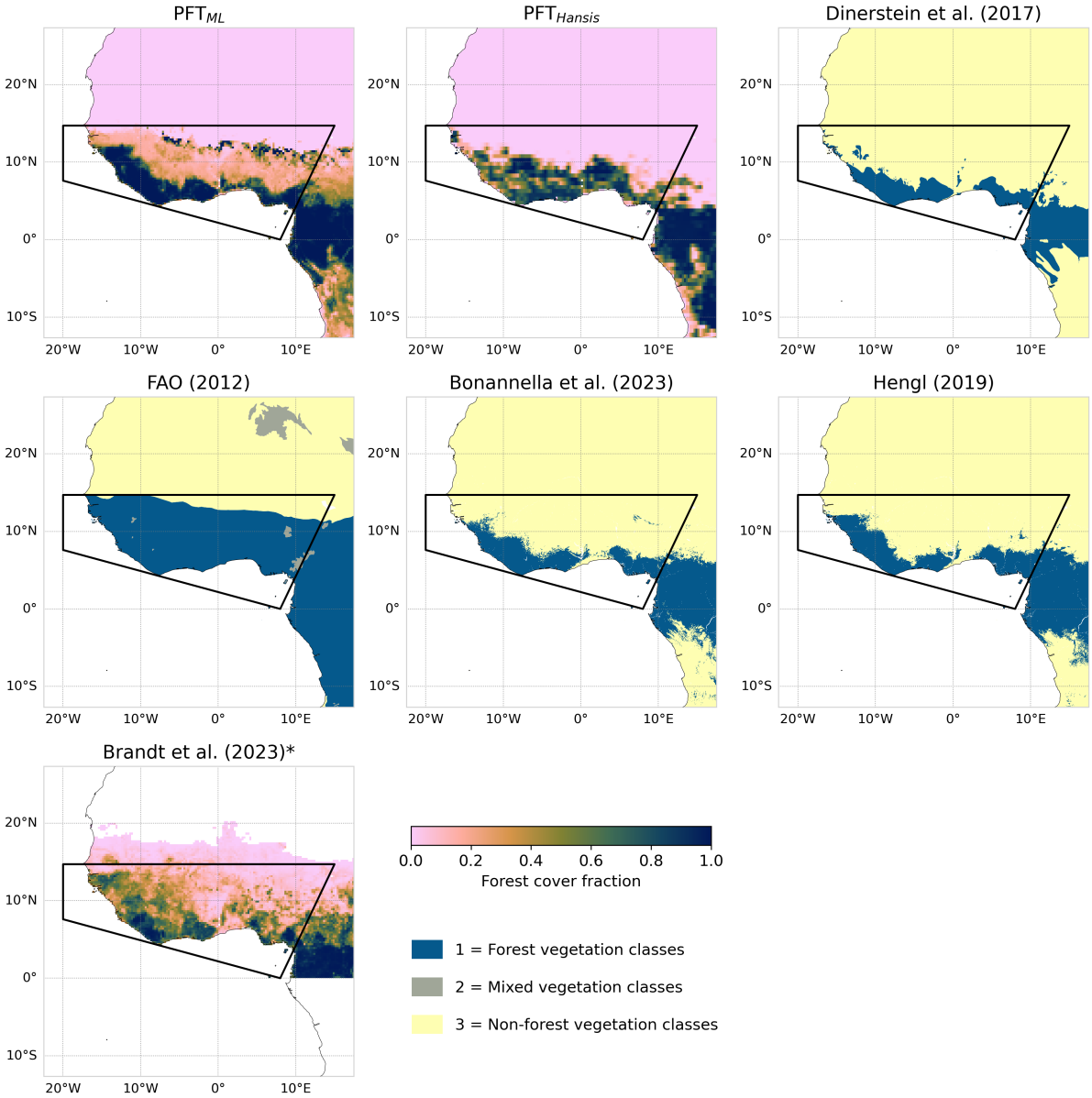

Figure S15. Comparison of forest cover under potential natural vegetation (PNV) for the IPCC region Western Africa (WAF). Sources for the delineated forest cover include the best-guess plant functional type (PFT) map (PFT_{ML}), the default PFT map (PFT_{Hansis}), the terrestrial biomes (Dinerstein et al., 2017), the 2010 update of the global ecological zones for FAO forest reporting (FAO, 2012), the current global distribution of biomes under potential natural vegetation (Bonannella et al., 2023, version 2), the potential distribution of biomes (Potential Natural Vegetation) at 250 m spatial resolution (Hengl, 2019, version v0.2), and tropical tree cover (Brandt et al., 2023). Forest cover fractions for PFT_{ML} and PFT_{Hansis} are obtained as the sum of the six forest PFTs (IDs 1–6; see Table S8). * Brandt et al. (2023) do not represent conditions under PNV, but rather actual tropical tree cover derived from satellite observations, thereby including trees from plantations and agroforestry systems. The source is included due to the lack of suitable regional PNV maps and should be interpreted with this limitation in mind. For the other sources, classes of the respective vegetation maps are summarized into three broad categories: (1) Forest vegetation classes, (2) Mixed vegetation classes (classes defined as a mixture of forests and non-forest vegetation units), and (3) Non-forest vegetation classes (savannas, shrublands, steppes, deserts, tundra, and other distinct non-forest classes).

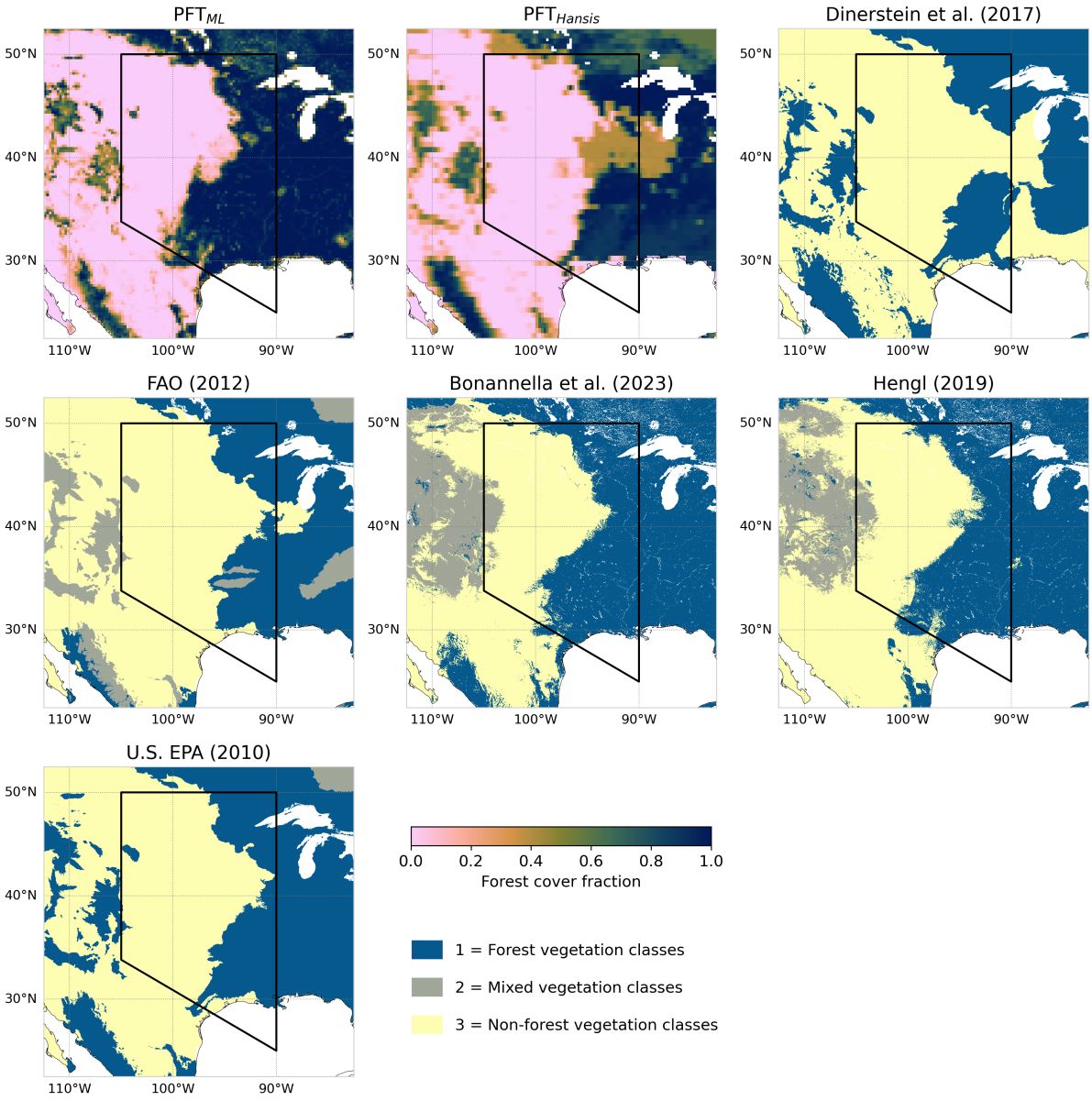

Figure S16. Comparison of forest cover under potential natural vegetation (PNV) for the IPCC region Central North America (CNA). Sources for the delineated forest cover include the best-guess plant functional type (PFT) map (PFT_{ML}), the default PFT map (PFT_{Hansis}), the terrestrial biomes (Dinerstein et al., 2017), the 2010 update of the global ecological zones for FAO forest reporting (FAO, 2012), the current global distribution of biomes under potential natural vegetation (Bonannella et al., 2023, version 2), the potential distribution of biomes (Potential Natural Vegetation; PNV) at 250 m spatial resolution (Hengl, 2019, version v0.2), and the level I ecoregions of North America (U.S. EPA, 2010). Forest cover fractions for PFT_{ML} and PFT_{Hansis} are obtained as the sum of the six forest PFTs (IDs 1–6; see Table S8). For the other sources, classes of the respective vegetation maps are summarized into three broad categories: (1) Forest vegetation classes, (2) Mixed vegetation classes (classes defined as a mixture of forests and non-forest vegetation units), and (3) Non-forest vegetation classes (savannas, shrublands, steppes, deserts, tundra, and other distinct non-forest classes).

S3 Supplementary tables

Table S1. Sources for the biome data used to construct the response dataset for the global PNV predictions. The column "Number of data points" refers to the number of data points included from the dataset after removal of duplicate data points and the removal of tundra data points outside of ecologically plausible locations. This table does not account for BIOME 6000 data points discarded in favor of data points from regional datasets at overlapping sites, which reduces the total number of data points in the response dataset from 14318 to 12785.

Reference	**Spatial extent**	**Approach to determine the PNV**	**Number of data points**
BIOME 6000 database			
Harrison (2017)	global	pollen-based biome reconstructions through biomization	6714
Dataset expansions adopted from Hengl et al. (2018)			
Marinova et al. (2018)	Eastern Mediterranean, Black Sea, and Caspian Sea regions	pollen-based biome reconstructions through biomization	1136
Veloso et al. (1992)	Brazil	random selection of data points from an expert-based regional vegetation map	388
Dataset expansions in this study			
Binney et al. (2017)	Eurasia	pollen-based biome reconstructions through biomization	3473
Sun et al. (2020)	China	pollen-based biome reconstructions through biomization	952
Qin et al. (2022)	Tibetan Plateau	pollen-based biome reconstructions through biomization	1247
Flantua et al. (2015)	South America	expert-based biome assignment based on field observations, regional vegetation maps, and global PNV classifications	408

Table S2. Environmental predictors used for the machine learning procedure and the prediction of the global distribution of potential natural vegetation (PNV).

ID	Description	Spatial res.	Source
Bioclimatic variables[a]			
wclim_bio1	Annual mean temperature	1 km	Fick and Hijmans (2017)
wclim_bio3	Isothermality (day-to-night temperature oscillations relative to summer-to-winter oscillations) ($\times 100$)	1 km	Fick and Hijmans (2017)
wclim_bio4	Temperature seasonality (standard deviation of monthly temperature averages) ($\times 100$)	1 km	Fick and Hijmans (2017)
wclim_bio5	Maximum temperature of the warmest month	1 km	Fick and Hijmans (2017)
wclim_bio6	Minimum temperature of the coldest month	1 km	Fick and Hijmans (2017)
wclim_bio12	Annual precipitation	1 km	Fick and Hijmans (2017)
wclim_bio13	Precipitation of the wettest month	1 km	Fick and Hijmans (2017)
wclim_bio14	Precipitation of the driest month	1 km	Fick and Hijmans (2017)
wclim_bio15	Precipitation seasonality (coefficient of variation of monthly precipitation averages) ($\times 100$)	1 km	Fick and Hijmans (2017)
Soil variables[b]			
sgrids1_bdricm	Depth to bedrock (R horizon) up to 200 cm	250 m	Hengl et al. (2017)
sgrids1_bdrlog	Probability of occurrence of R horizon	250 m	Hengl et al. (2017)
sgrids1_bdticm	Absolute depth to bedrock	250 m	Hengl et al. (2017)
sgrids2_cec	Cation exchange capacity of the soil	250 m	Poggio et al. (2021)
sgrids2_cfvo	Volumetric fraction of coarse fragments (> 2 mm)	250 m	Poggio et al. (2021)
sgrids2_phh2o	Soil pH	250 m	Poggio et al. (2021)
sgrids2_clay	Proportion of clay particles (< 0.002 mm) in the fine earth fraction	250 m	Poggio et al. (2021)
sgrids2_silt	Proportion of silt particles (0.002–0.05 mm) in the fine earth fraction	250 m	Poggio et al. (2021)
sgrids2_sand	Proportion of sand particles (> 0.05 mm) in the fine earth fraction	250 m	Poggio et al. (2021)
Geomorphometric variables			
wclim_elev	Elevation	1 km	Fick and Hijmans (2017)
geom90m_slope	Slope (percentage of elevation change over 100 metres)	250 m	Amatulli et al. (2020)
geom90m_convergence	Convergence index (agreement of the angular direction of the slope in the surrounding cells with the theoretical matrix direction)	250 m	Amatulli et al. (2020)
geom90m_cti	Compound topographic index (logarithm of the cumulative upstream catchment area divided by the tangent of the local slope angle)	250 m	Amatulli et al. (2020)
geom90m_eastness	Eastness (sine of slope multiplied by cosine of the angular direction of the slope)	250 m	Amatulli et al. (2020)
geom90m_northness	Northness (sine of slope multiplied by sine of the angular direction of the slope)	250 m	Amatulli et al. (2020)
geom90m_spi	Stream power index (product between the upstream catchment area and the tangent of the local slope angle)	250 m	Amatulli et al. (2020)

[a] Calculated from the monthly average for the years 1970–2000.

[b] sgrids2_** variables by Poggio et al. (2021) are calculated as weighted average of the three upper soil layers: 0–5 cm, 5–15 cm, and 15–30 cm.

Table S3. Additional datasets used to harmonize the input data for the machine learning procedure, develop the reclassification scheme, and determine the processing extent of the global PNV predictions. The "Usage" column specifies the processing steps that rely on the respective dataset: [1] = Harmonization of the response data, [2] = Reclassification of the biome classes to BLUE PFTs, [3] = Exclusion of non-vegetated grid cells from the environmental predictors and global PNV predictions.

ID	Description	Usage	Spatial res.	Data type	Source
HLZs	Holdridge life zones (Holdridge, 1967)	[1], [2]	1 km	raster	Elsen et al. (2022)
GEZ2010	2010 update of the global ecological zones for FAO forest reporting	[1], [2]	-	vector	FAO (2012)
TB2017	Terrestrial biomes	[1]	-	vector	Dinerstein et al. (2017)
CCI PFTs	Annual fractional PFT maps for the period 1992–2020; 14 PFTs	[2]	300 m	raster	Harper et al. (2023b)
wclim_prec	Average monthly precipitation[a]	[2]	1 km	raster	Fick and Hijmans (2017)
wclim_tmax	Average monthly maximum temperature[a]	[2]	1 km	raster	Fick and Hijmans (2017)
Modeled C4 grass fraction	Annual maps with the C4 grass coverage (share of grassland covered by C4 grasses) for the period 2001–2019	[2]	0.5°	raster	Luo (2024)
ESA CCI LC	Annual land cover maps for the period 1992–2020; 22 land cover classes	[3]	300 m	raster	Copernicus Climate Change Service (2019)

[a] WorldClim 2.1 monthly averages calculated from 1970–2000.

Table S4. Rules for the reassignment of the pollen-based biome reconstructions from Sun et al. (2020) and Qin et al. (2022) into the standardized biome classification system of the BIOME 6000 database (Harrison, 2017). The original biome classes are assigned to the BIOME 6000 database classes based on similarity in plant form, leaf type, phenology, and climate region derived from the class names. In most cases the reassignment is straightforward. Exceptions are detailed in the footnotes.

Original ID	Original biome class	BIOME 6000 class
TRFO	Tropical rain forest	Tropical evergreen broadleaf forest
TSFO	Tropical seasonal forest	Tropical semi-evergreen broadleaf forest
TDFO	Tropical dry forest	Tropical deciduous broadleaf forest and woodland
WAMF	North subtropical mixed forest	Warm-temperate evergreen broadleaf and mixed forest[a,b]
MTFO	Middle subtropical broadleaf evergreen forest	Warm-temperate evergreen broadleaf forest[a,b]
STFO	South subtropical broadleaf evergreen forest	Warm-temperate evergreen broadleaf forest[a,b]
TEDE	Warm-temperate mixed forest	Cool mixed forest[a,c]
COMX	Cool-temperate mixed forest	Cool mixed forest
CLDC	Cold-temperate summergreen conifer forest	Cold deciduous forest
CLEC	Cold-temperate evergreen conifer forest	Cold evergreen needleleaf forest
XERO	Scarse woods/scrubs	Xerophytic woods/scrubs
DESE	Desert	Desert
TEDS	Cool-temperate desert steppe	Steppe
STEP	Cool-temperate steppe	Steppe
TEME	Cool-temperate meadow steppe	Steppe
TEFS	Cool-temperate forest steppe	Steppe[d]
ALME	Alpine meadow	Graminoid and forb tundra[e]
ALST	Alpine steppe	Graminoid and forb tundra[e]

[a] The biome scheme in Sun et al. (2020) and Qin et al. (2022) includes tropical, subtropical, warm-temperate, cool-temperate, and cold-temperate climate region identifiers in the biome names. In contrast, the biomes of Harrison (2017) encompass tropical, warm-temperate, temperate, cool-temperate, and cold-temperate climate regions. The subtropical biomes of these studies are therefore assigned to the warm-temperate biomes in Harrison (2017), whereas the warm-temperate biome of these studies is assigned to the temperate biomes in Harrison (2017).

[b] Additional arguments for the assignment of this class to a warm-temperate biome in the BIOME 6000 classification system include: (1) The taxa used to define the PFTs that make up the original biome overlap best with the taxa used to define the warm-temperate PFTs in the publications incorporated into the BIOME 6000 database. (2) Most sampling sites are located outside the extent of the tropical areas delineated in the HLZ, TB2017, and GEZ2010 maps, making the assignment to tropical BIOME 6000 biomes not sensible. (3) Previous studies, including the biome reconstructions by Ni et al. (2010) integrated into the BIOME 6000 database, define these subtropical biomes as warm-temperate.

[c] As the biome scheme of Harrison (2017) lacks a mixed temperate forest biome, TEDE is assigned to the "cool mixed forest" biome of the BIOME 6000 classification system. Although not ideal, this reassignment is preferred over the reassignment to the BIOME 6000 class "warm-temperate mixed forest": (1) The PFTs of TEDE do not include the subtropical PFTs used to characterize the biomes WAMF, MTFO, and STFO. Instead, TEDE is distinctly different from the subtropical biomes and exists in the climate space between WAMF (North subtropical mixed forest) and COMX (cool-temperate mixed forest). (2) The majority of the sampling sites are located in regions defined as temperate in the TB2017 and GEZ2010 maps and cool-temperate in the HLZs.

[d] Assigned to the BIOME 6000 class "steppe" as the original biome includes only grass and shrub PFTs (Sun et al., 2020).

[e] Assigned to the BIOME 6000 class "graminoid and forb tundra" as the original biome class includes taxa associated with alpine and arctic-alpine PFTs that match well with the taxa used to derive "graminoid and forb tundra". (All BIOME 6000 tundra classes are further aggregated into one tundra class in the subsequent processing step.)

Table S5. Rule set for aggregating the 32 BIOME 6000 classes into the 16 biome classes used for the prediction of the global distribution of potential natural vegetation (PNV). The aggregation scheme relies on four vegetation traits to group the BIOME 6000 classes into the 16 biome classes according to similar physiognomic and climatic properties. The vegetation traits describe the characteristic plant form (grasses, shrubs, trees), leaf type (broadleaf = B, needleleaf = N), phenology (evergreen = E, deciduous = D), and climate region (tropical, warm-temperate, temperate/boreal) of each biome. Traits are generally obtained from the biome name. If the name does not provide sufficient information, vegetation traits are derived from the plant functional types (PFTs) and taxa used during the biomization process. Footnotes provide explanations for exceptions in the aggregation scheme and for cases where the definition of vegetation traits is ambiguous. The plant form "grasses" encompasses not only grass species but all non-woody herbaceous vegetation (e.g., graminoids, forbs, ferns). Leaf type and phenology are only defined for biomes that include trees. If possible, the climate regions are already translated to match the climate regions of the PFTs used in BLUE. Vegetation traits defined in this table are also used in the reclassification scheme in Table S7.

ID	BIOME 6000 class	plant form	leaf type	pheno-logy	climate region	Biome class
1	tropical evergreen broadleaf forest	trees	B	E	tropical	Tropical evergreen forest
2	tropical semi-evergreen broadleaf forest	trees	B	E + D	tropical	Tropical mixed forest
3	tropical deciduous broadleaf forest and woodland	trees	B	D	tropical	Tropical deciduous forest
4	warm-temperate evergreen broadleaf and mixed forest	trees	B + N	E + D	warm-temperate	Warm-temperate mixed forest
5	warm-temperate evergreen broadleaf forest	trees	B	E	warm-temperate	Warm-temperate evergreen broadleaf forest
6	warm-temperate rainforest	trees	B + N	E + D	warm-temperate	Warm-temperate evergreen broadleaf forest[a]
7	wet sclerophyll forest	trees	B + N	E	~	Temperate/boreal mixed forest[b]
8	temperate evergreen needleleaf forest	trees	N	E	temperate / boreal	Temperate/boreal evergreen conifers
9	temperate deciduous broadleaf forest	trees	B	D	temperate / boreal	Temperate/boreal deciduous broadleaf forest
10	cool evergreen needleleaf forest	trees	N	E	temperate / boreal	Temperate/boreal evergreen conifers
11	cool mixed forest	trees	B + N	E + D	temperate / boreal	Temperate/boreal mixed forest
12	cool-temperate rainforest	trees	B + N	E + D	temperate / boreal	Temperate/boreal mixed forest
13	cool-temperate evergreen needleleaf and mixed forest	trees	B + N	E + D	temperate / boreal	Temperate/boreal mixed forest
14	cold mixed forest	trees	B + N	E + D	temperate / boreal	Temperate/boreal mixed forest
15	cold deciduous forest	trees	B + N	D	temperate / boreal	Cold deciduous forest
16	cold evergreen needleleaf forest	trees	N	E	temperate / boreal	Temperate/boreal evergreen conifers
17	temperate evergreen needleleaf open woodland[c]	grasses, shrubs, trees	N	E	temperate / boreal	Evergreen needleleaf open woodland

Continued on next page

Table S5. continued from previous page

ID	BIOME 6000 class	plant form	leaf type	phenology	climate region	Biome class
18	temperate sclerophyll woodland and shrubland[c]	grasses, shrubs, trees	B + N	E + D	temperate / boreal	Sclerophyll woodland and shrubland
19	tropical savanna	grasses, shrubs, trees	B + N	E + D	tropical	Tropical savanna
20	temperate or tropical grassland and xerophytic shrubland	grasses, shrubs	~	~	~	Steppe[d]
21	temperate xerophytic woods/scrubs	grasses, shrubs, trees[e]	B + N	E + D	temperate / boreal	Xerophytic woods/shrubs
22	temperate grassland and xerophytic shrubland	grasses, shrubs	~	~	temperate / boreal	Steppe[f]
23	xerophytic woods/scrubs	grasses, shrubs, trees[e]	B + N	E + D	~	Xerophytic woods/shrubs
24	desert	grasses, shrubs	~	~	~	Desert[g]
25	steppe	grasses, shrubs	~	~	~	Steppe[h]
26	cool grassland	grasses	~	~	temperate / boreal	Tundra[i]
27	low and high shrub tundra	grasses, shrubs	~	~	~	Tundra
28	erect dwarf-shrub tundra	grasses, shrubs	~	~	~	Tundra
29	prostrate dwarf-shrub tundra	grasses[j]	~	~	~	Tundra
30	graminoid and forb tundra	grasses	~	~	~	Tundra
31	cushion-forb tundra (cushion forb, lichen and moss tundra)	grasses	~	~	~	Tundra
32	undifferentiated tundra	grasses	~	~	~	Tundra

[a] The BIOME 6000 class "warm-temperate rainforest" is reconstructed by Marchant et al. (2009) for South America and Pickett et al. (2004) for Australasia (Australia, Southeast Asia, and the Pacific). The descriptions of the climate space and biomization procedure are found in these papers. The biome is assigned to the biome class "Warm-temperate evergreen broadleaf forest" for the following reasons: (1) The biome is described as evergreen closed forest. While the biome type is characterized by broadleaf and needleleaf PFTs, the underlying taxa groups are primarily broadleaf trees and closely match those found in the BIOME 6000 class "warm-temperate evergreen broadleaf forest". (2) The biome occurs in a climate space with the same moisture availability as the BIOME 6000 class "warm-temperate evergreen broadleaf forest" but with higher temperatures. The climate space of the BIOME 6000 class "warm-temperate broadleaf and mixed forests" differs and encompasses drier and colder conditions. (3) The global CCI PFT dataset by Harper et al. (2023a) does not observe needleleaf trees in the regions, where the biome data points are located.

[b] The BIOME 6000 class "wet sclerophyll forest" is reconstructed by Pickett et al. (2004) for Australasia (Australia, Southeast Asia, and the Pacific) and includes seven data points. It is assigned to the biome class "Temperate/boreal mixed forest". The climate region of this biome is treated as temperate/boreal, since 5 out of 7 observations are located in cool temperate HLZs and all data points are located in the cool temperate climate domain of the GEZ2010 map. It is treated as mixed forest due to the inclusion of broadleaf and needleleaf taxa.

[c] While there is no universal definition for woodland, it is generally described as a diverse landscape characterized by open forests, shrubs, and grass patches that transition into shrublands with decreasing temperature or water availability (Audebert et al., 2024; FAO, 2012; Mucina, 2019; Scheffer et al., 2012). The PFTs and taxa used during the biomization procedure reflect this definition and include trees, shrubs, and grasses.

[d] The BIOME 6000 class "temperate or tropical grassland and xerophytic shrubland" is reconstructed by Pickett et al. (2004) for Australasia (Australia, Southeast Asia, and the Pacific) and includes three data points. It is assigned to the biome class "Steppe" for the following reasons: (1) The description of the

biome and the characteristic PFTs and taxa used in the biomization procedure match the definition of steppe biomes in other publications with pollen-based biome reconstructions. (2) Due to a lack of tree PFTs, the assignment to the biome classes "Tropical savanna" or "Xerophytic woods/scrubs" is not feasible, while "Desert" does not match the description.

[e] The PFTs and taxa used during the biomization procedure include trees, shrubs, and grasses.

[f] The BIOME 6000 class "temperate grassland and xerophytic shrubland" encompasses 78 data points, most of which are located in South America. Marchant et al. (2009) describe the biome as mixture of grasses and shrubs that occur above the tree line in the Andes but occupy a warmer climate space than the "cool grasslands". None of the classes in the new classification scheme closely align with the described biome and its characteristic PFTs and underlying taxa. However, the biome needs to be aggregated due to its small class size. Following Hengl et al. (2018) and Harrison (2017) it is treated as a type of steppe.

[g] The PFTs and taxa used during the biomization procedure include shrubs and grasses. The BIOME 6000 database and the expansions encompass 15 publications that describe the biomization procedure for the BIOME 6000 class "desert". Ten of these include shrub and grass PFTs and taxa, three publications include only grass PFTs, and two publications include only shrub PFTs.

[h] The PFTs and taxa used during the biomization procedure include shrubs and grasses. The BIOME 6000 database and the expansions encompass 18 publications that describe the biomization procedure for the BIOME 6000 class "steppe". Twelve of these include shrub and grass PFTs and taxa, while six publications include only grass PFTs and taxa.

[i] The BIOME 6000 class "cool grassland" is reconstructed by Marchant et al. (2009) for South America. It describes the alpine vegetation above the tree line in the highest altitudes of the Andes. Although the climate region derived from the name of the biome equates to a temperate/boreal climate it is assigned to "tundra" under the new classification scheme, as the characteristic PFTs and taxa of the biomization procedure overlap with those used by other publications to reconstruct the BIOME 6000 "tundra" biomes.

[j] While the BIOME 6000 class "prostrate dwarf-shrub tundra" includes the term shrub in its name, it is characterized by very low-stature vegetation (< 15 cm; Raynolds et al. 2019). The CCI PFTs define shrubs as woody vegetation with a height between 3 and 5 m, and therefore cannot distinguish this tundra vegetation from grasses in terms of plant form.

Table S6. Hyperparameter space and values for the random forest classifiers. The value ranges are based on the recommendations by Probst et al. (2019). For all other hyperparameters, the default setting from scikit-learn is used (Pedregosa et al., 2011).

hyperparameter	**selected value or value range for tuning**
n_estimators	{1000}
max_features	{5, 9, 13, 17, 21, 25}
min_samples_leaf	{1, 2, 4}
max_samples	{0.6, 0.7, 0.8, 0.9}

Table S7. Rule set for the reclassification of the 16 biome classes used for the prediction of the global distribution of potential natural vegetation (PNV) into the 11 PFTs of BLUE. As all spatial datasets used for reclassification contain grid cell fractions, the BLUE PFTs in each grid cell are determined by: (1) directly allocating the fraction of matching biome classes, and (2) proportionally allocating the mixed biome classes to the permitted PFT based on the grid cell fractions of the plant traits. The grid cell fractions of plant form, leaf type, and phenology are obtained from the CCI PFTs (see Supplementary text S1.3 for details). They are abbreviated with the following expressions: $\sum$trees = sum of the four tree PFTs, $\sum$shrubs = sum of the four shrub PFTs, BE = broadleaf evergreen trees, BD = broadleaf deciduous trees, NE = needleleaf evergreen trees, and ND = needleleaf deciduous trees. Tropical refers to the fraction of tropical vegetation (Fig S1a), whereas C4 refers to the fraction of C4 grasses (Fig S1b).

Biome class		**Fractional grid cell coverage of the vegetation traits required for allocation**		**BLUE PFT**
1	Tropical evergreen forest			= Tropical evergreen forest
2	Tropical mixed forest	$\times\ (BE+NE)/\sum$trees		= Tropical evergreen forest
		$\times\ (BD+ND)/\sum$trees		= Tropical deciduous forest
3	Tropical deciduous forest			= Tropical deciduous forest
4	Warm-temperate mixed forest	$\times\ BE/\sum$trees		= Temperate evergreen broadleaf forest
		$\times\ BD/\sum$trees		= Temperate/boreal deciduous broadleaf forest
		$\times\ NE/\sum$trees		= Temperate/boreal evergreen conifers
		$\times\ ND/\sum$trees		= Temperate/boreal deciduous conifers
5	Warm-temperate evergreen broadleaf forest			= Temperate evergreen broadleaf forest
6	Temperate/boreal deciduous broadleaf forest			= Temperate/boreal deciduous broadleaf forest
7	Temperate/boreal evergreen conifers			= Temperate/boreal evergreen conifers
8	Temperate/boreal mixed forest	$\times\ BE/\sum$trees		= Temperate evergreen broadleaf forest
		$\times\ BD/\sum$trees		= Temperate/boreal deciduous broadleaf forest
		$\times\ NE/\sum$trees		= Temperate/boreal evergreen conifers
		$\times\ ND/\sum$trees		= Temperate/boreal deciduous conifers
9	Cold deciduous forest	$\times\ BD/(ND+BD)$		= Temperate/boreal deciduous broadleaf forest
		$\times\ ND/(ND+BD)$		= Temperate/boreal deciduous conifers
10	Evergreen needleleaf open woodland	$\times\ (\sum$trees$)/(\sum$trees$+\sum$shrubs$+$natural grasses$)$		= Temperate/boreal evergreen conifers
		$\times\ (\sum$shrubs$)/(\sum$trees$+\sum$shrubs$+$natural grasses$)$		= Summergreen shrubs
		$\times$ natural grasses$/(\sum$trees$+\sum$shrubs$+$natural grasses$)$	$\times$ C4	= C4 natural grasses
		$\times$ natural grasses$/(\sum$trees$+\sum$shrubs$+$natural grasses$)$	$\times\ (\text{C4}-1)$	= C3 natural grasses
11	Sclerophyll woodland and shrubland	$\times\ (BE/(\sum$trees$+\sum$shrubs$+$natural grasses$))$		= Temperate evergreen broadleaf forest
		$\times\ (BD/(\sum$trees$+\sum$shrubs$+$natural grasses$))$		= Temperate/boreal deciduous broadleaf forest[a]
		$\times\ (NE/(\sum$trees$+\sum$shrubs$+$natural grasses$))$		= Temperate/boreal evergreen conifers[a]
		$\times\ (ND/(\sum$trees$+\sum$shrubs$+$natural grasses$))$		= Temperate/boreal deciduous conifers[a]
		$\times\ (\sum$shrubs$)/(\sum$trees$+\sum$shrubs$+$natural grasses$)$		= Summergreen shrubs[a]
		$\times$ (natural grasses$/(\sum$trees$+\sum$shrubs$+$natural grasses$))$	$\times$ C4	= C4 natural grasses
		$\times$ (natural grasses$/(\sum$trees$+\sum$shrubs$+$natural grasses$))$	$\times\ (\text{C4}-1)$	= C3 natural grasses

Continued on next page

Table S7: continued from previous page

Biome class	Fractional grid cell coverage of the vegetation traits required for allocation		BLUE PFT
12 Tropical savanna	$\times\ (BE+NE)/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses})$		= Tropical evergreen forest
	$\times\ (BD+ND)/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses})$		= Tropical deciduous forest
	$\times\ (\sum \text{shrubs})/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses})$		= Raingreen shrubs
	$\times\ (\text{natural grasses}/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses})$	$\times$ C4	= C4 natural grasses
	$\times\ (\text{natural grasses}/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses})$	$\times\ (\text{C4}-1)$	= C3 natural grasses
13 Steppe	$\times\ (\sum \text{shrubs})/(\sum \text{shrubs}+\text{natural grasses})$	$\times$ tropical	= Raingreen shrubs
	$\times\ (\sum \text{shrubs})/(\sum \text{shrubs}+\text{natural grasses})$	$\times\ (\text{tropical}-1)$	= Summergreen shrubs
	$\times\ (\text{natural grasses})/(\sum \text{shrubs}+\text{natural grasses}))$	$\times$ C4	= C4 natural grasses
	$\times\ (\text{natural grasses})/(\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{C4}-1)$	= C3 natural grasses
14 Xerophytic woods/shrubs	$\times\ (BE+NE)/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses})$	$\times$ tropical	= Tropical evergreen forest
	$\times\ (BD+ND)/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses})$	$\times$ tropical	= Tropical deciduous forest
	$\times\ (BE/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{tropical}-1)$	= Temperate evergreen broadleaf forest
	$\times\ (BD/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{tropical}-1)$	= Temperate/boreal deciduous broadleaf forest
	$\times\ (NE/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{tropical}-1)$	= Temperate/boreal evergreen conifers
	$\times\ (ND/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{tropical}-1)$	= Temperate/boreal deciduous conifers
	$\times\ (\sum \text{shrubs}/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times$ tropical	= Raingreen shrubs
	$\times\ (\sum \text{shrubs}/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{tropical}-1)$	= Summergreen shrubs
	$\times\ (\text{natural grasses}/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times$ C4	= C4 natural grasses
	$\times\ (\text{natural grasses}/(\sum \text{trees}+\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{C4}-1)$	= C3 natural grasses
15 Desert	$\times\ (\sum \text{shrubs}/(\sum \text{shrubs}+\text{natural grasses}))$	$\times$ tropical	= Raingreen shrubs
	$\times\ (\sum \text{shrubs}/(\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{tropical}-1)$	= Summergreen shrubs
	$\times\ (\text{natural grasses}/(\sum \text{shrubs}+\text{natural grasses}))$	$\times$ C4	= C4 natural grasses
	$\times\ (\text{natural grasses}/(\sum \text{shrubs}+\text{natural grasses}))$	$\times\ (\text{C4}-1)$	= C3 natural grasses
16 Tundra			= Tundra

Overriding adjustments after initial reclassification:

1) The relative BLUE PFT fractions are reduced to the actual fraction of total grid cell area based on sum of the three abiotic CCI PFTs.

[a] The biome class "Sclerophyll woodland and shrubland" is derived from a temperate BIOME 6000 class and therefore only assigned to the non-tropical BLUE PFTs.

Table S8. Global areal extent ($10^6\ \mathrm{km}^2$) and relative share compared to the vegetated surface cover (%) of the 11 plant functional type (PFT) of the bookkeeping model BLUE for $\mathrm{PFT_{Hansis}}$, $\mathrm{PFT_{ML}}$, $\mathrm{PFT_{MLforest}}$, and $\mathrm{PFT_{MLgrass}}$. $\mathrm{PFT_{Hansis}}$ refers to the default PFT map of BLUE. $\mathrm{PFT_{ML}}$ refers to our best-guess PFT map, whereas $\mathrm{PFT_{MLforest}}$ and $\mathrm{PFT_{MLgrass}}$ refer to our sensitivity maps with maximum forest and grass extent, respectively. The PFT categories are: (a) forest PFTs (sum of IDs 1–6), (b) shrub PFTs (sum of IDs 7–8), (c) grass PFTs (sum of IDs 9–10), (d) tundra (ID 11), and (e) abiotic surface cover.

ID	**BLUE PFT**	**$\mathrm{PFT_{Hansis}}$**		**$\mathrm{PFT_{ML}}$**		**$\mathrm{PFT_{MLforest}}$**		**$\mathrm{PFT_{MLgrass}}$**	
		Area	Share	Area	Share	Area	Share	Area	Share
1	Tropical evergreen forest	17.10	17	14.04	14	14.44	14	13.48	13
2	Tropical deciduous forest	5.93	6	6.68	7	7.99	8	5.42	5
3	Temperate evergreen broadleaf forest	1.57	2	5.99	6	6.73	7	5.21	5
4	Temperate/boreal deciduous broadleaf forest	9.19	9	9.88	10	11.18	11	8.65	8
5	Temperate/boreal evergreen conifers	13.52	13	20.30	20	21.48	21	19.53	19
6	Temperate/boreal deciduous conifers	2.55	3	3.09	3	3.43	3	3.08	3
7	Raingreen shrubs	13.41	13	2.75	3	2.52	2	2.92	3
8	Summergreen shrubs	0.35	0	0.90	1	0.83	1	0.97	1
9	C3 natural grasses	14.37	14	14.30	14	12.43	12	16.37	16
10	C4 natural grasses	19.19	19	18.90	18	16.64	16	21.28	21
11	Tundra	4.16	4	5.63	5	4.80	5	5.54	5
ID	**PFT categories**								
a	Forest PFTs	49.87	49	59.99	59	65.24	64	55.37	54
b	Shrub PFTs	13.76	14	3.64	4	3.34	3	3.89	4
c	Grass PFTs	33.55	33	33.20	32	29.07	28	37.65	37
d	Tundra	4.16	4	5.63	5	4.80	5	5.54	5
e	Abiotic surface cover	47.74	-	46.62	-	46.62	-	46.62	-

References

Amatulli, G., McInerney, D., Sethi, T., Strobl, P., and Domisch, S.: Geomorpho90m, Empirical Evaluation and Accuracy Assessment of Global High-Resolution Geomorphometric Layers, Sci. Data, 7, 162, https://doi.org/10.1038/s41597-020-0479-6, 2020.

Audebert, P., Milne, E., Schiettecatte, L.-S., Dionisio, D., Sinitambirivoutin, M., Pais, C., Proença, C., and Bernoux, M.: Ecological Zoning for Climate Policy and Global Change Studies, Nat. Sustain., 7, 1294–1303, https://doi.org/10.1038/s41893-024-01416-5, 2024.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., and Zernitskaya, V.: Vegetation of Eurasia from the Last Glacial Maximum to Present: Key Biogeographic Patterns, Quat. Sci. Rev., 157, 80–97, https://doi.org/10.1016/j.quascirev.2016.11.022, 2017.

Bonannella, C., Hengl, T., Leal Parente, L., and de Bruin, S.: Current and Future Global Distribution of Potential Biomes under Climate Change Scenarios. (Version 2), Zenodo [data set], https://doi.org/10.5281/zenodo.7822868, 2023.

Brandt, J., Ertel, J., Spore, J., and Stolle, F.: Wall-to-Wall Mapping of Tree Extent in the Tropics with Sentinel-1 and Sentinel-2, Remote Sens. Environ., 292, 113 574, https://doi.org/10.1016/j.rse.2023.113574, 2023.

Buchhorn, M., Smets, B., Bertels, L., de Roo, B., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S.: Copernicus Global Land Service: Land Cover 100m: Collection 3: Epoch 2019: Globe, Zenodo [data set], https://doi.org/10.5281/zenodo.3939050, 2020.

Copernicus Climate Change Service: Land Cover Classification Gridded Maps from 1992 to Present Derived from Satellite Observations, ECMWF [data set], https://doi.org/10.24381/cds.006f2c9a, 2019.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baillie, J. E. M., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., de Souza, N., Pintea, L., Brito, J. C., Llewellyn, O. A., Miller, A. G., Patzelt, A., Ghazanfar, S. A., Timberlake, J., Klöser, H., Shennan-Farpón, Y., Kindt, R., Lillesø, J.-P. B., van Breugel, P., Graudal, L., Voge, M., Al-Shammari, K. F., and Saleem, M.: An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm, BioScience, 67, 534–545, https://doi.org/10.1093/biosci/bix014, 2017.

Elsen, P. R., Saxon, E. C., Simmons, B. A., Ward, M., Williams, B. A., Grantham, H. S., Kark, S., Levin, N., Perez-Hammerle, K.-V., Reside, A. E., and Watson, J. E. M.: Accelerated Shifts in Terrestrial Life Zones under Rapid Climate Change, Glob. Change Biol., 28, 918–935, https://doi.org/10.1111/gcb.15962, 2022.

European Space Agency: CCI-LC User Tools, Version 4.3, European Space Agency [software], 2019.

FAO: Global Ecological Zones for FAO Forest Reporting: 2010 Update, Food and Agriculture Organization of the United Nations, Rome, 2012.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: New 1–Km Spatial Resolution Climate Surfaces for Global Land Areas, Int. J. Climatol., 37, 4302–4315, https://doi.org/10.1002/joc.5086, 2017.

Flantua, S. G., Hooghiemstra, H., Grimm, E. C., Behling, H., Bush, M. B., González-Arango, C., Gosling, W. D., Ledru, M.-P., Lozano-García, S., Maldonado, A., Prieto, A. R., Rull, V., and van Boxel, J. H.: Updated Site Compilation of the Latin American Pollen Database, Rev. Palaeobot. Palynol., 223, 104–115, https://doi.org/10.1016/j.revpalbo.2015.09.008, 2015.

Griffith, D. M., Anderson, T. M., Osborne, C. P., Strömberg, C. A. E., Forrestel, E. J., and Still, C. J.: Biogeographically Distinct Controls on C 3 and C 4 Grass Distributions: Merging Community and Physiological Ecology, Glob. Ecol. Biogeogr., 24, 304–313, https://doi.org/10.1111/geb.12265, 2015.

Hansis, E., Davis, S. J., and Pongratz, J.: Relevance of Methodological Choices for Accounting of Land Use Change Carbon Fluxes, Glob. Biogeochem. Cycles, 29, 1230–1246, https://doi.org/10.1002/2014GB004997, 2015.

Harper, K. L., Lamarche, C., Hartley, A., Peylin, P., Ottlé, C., Bastrikov, V., San Martín, R., Bohnenstengel, S. I., Kirches, G., Boettcher, M., Shevchuk, R., Brockmann, C., and Defourny, P.: A 29-Year Time Series of Annual 300 m Resolution Plant-Functional-Type Maps for Climate Models, Earth Syst. Sci. Data, 15, 1465–1499, https://doi.org/10.5194/essd-15-1465-2023, 2023a.

Harper, K. L., Lamarche, C., Hartley, A., Peylin, P., Ottlé, C., Bastrikov, V., San Martín, R., Bohnenstengel, S. I., Kirches, G., Boettcher, M., Shevchuk, R., Brockmann, C., and Defourny, P.: ESA Land Cover Climate Change Initiative (Land_cover_cci):

Global Plant Functional Types (PFT) Dataset, v2.0.8, NERC EDS Centre for Environmental Data Analysis [data set], https://doi.org/10.5285/26a0f46c95ee4c29b5c650b129aab788, 2023b.

Harrison, S.: BIOME 6000 DB Classified Plotfile Version 1, University of Reading [data set], https://doi.org/10.17864/1947.99, 2017.

Hengl, T.: Potential Distribution of Biomes (Potential Natural Vegetation) at 250 m Spatial Resolution. Version v0.2, Zenodo [data set], https://doi.org/10.5281/zenodo.3526620, 2019.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global Gridded Soil Information Based on Machine Learning, PLOS ONE, 12, e0169 748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hengl, T., Walsh, M. G., Sanderman, J., Wheeler, I., Harrison, S. P., and Prentice, I. C.: Global Mapping of Potential Natural Vegetation: An Assessment of Machine Learning Algorithms for Estimating Land Potential, PeerJ, 6, e5457, https://doi.org/10.7717/peerj.5457, 2018.

Holdridge, L. R.: Life Zone Ecology, Tropical Science Center, San Jose, 1967.

Houghton, R. A., Hobbie, J. E., Melillo, J. M., Moore, B., Peterson, B. J., Shaver, G. R., and Woodwell, G. M.: 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, 235–262, https://doi.org/10.2307/1942531, 1983.

Hoyer, S. and Hamman, J.: Xarray: N-d Labeled Arrays and Datasets in Python, J. Open Res. Softw., 5, 10, https://doi.org/10.5334/jors.148, 2017.

IBGE: Manual Técnico Da Vegetação Brasileira., vol. número 1 of *Manuais Técnicos Em Geociências*, Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro, 2ª edição revista e ampliada edn., ISBN 978-85-240-4272-0, 2012.

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodetskaya, I. V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H. T., Jones, R. G., Krakovska, S., Manzanas, R., Martínez-Castro, D., Narisma, G. T., Nurhati, I. S., Pinto, I., Seneviratne, S. I., van den Hurk, B., and Vera, C. S.: An Update of IPCC Climate Reference Regions for Subcontinental Analysis of Climate Model Data: Definition and Aggregated Datasets, Earth Syst. Sci. Data, 12, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020, 2020.

Luo, X.: Mapping the Global Distribution of C4 Vegetation Using Observations and Optimality Theory, Zenodo [data set], https://doi.org/10.5281/zenodo.10516423, 2024.

Luo, X., Zhou, H., Satriawan, T. W., Tian, J., Zhao, R., Keenan, T. F., Griffith, D. M., Sitch, S., Smith, N. G., and Still, C. J.: Mapping the Global Distribution of C4 Vegetation Using Observations and Optimality Theory, Nat. Commun., 15, 1219, https://doi.org/10.1038/s41467-024-45606-3, 2024.

Marchant, R., Cleef, A., Harrison, S. P., Hooghiemstra, H., Markgraf, V., van Boxel, J., Ager, T., Almeida, L., Anderson, R., Baied, C., Behling, H., Berrio, J. C., Burbridge, R., Björck, S., Byrne, R., Bush, M., Duivenvoorden, J., Flenley, J., de Oliveira, P., van Geel, B., Graf, K., Gosling, W. D., Harbele, S., van der Hammen, T., Hansen, B., Horn, S., Kuhry, P., Ledru, M.-P., Mayle, F., Leyden, B., Lozano-García, S., Melief, A. M., Moreno, P., Moar, N. T., Prieto, A., van Reenen, G., Salgado-Labouriau, M., Schäbitz, F., Schreve-Brinkman, E. J., and Wille, M.: Pollen-Based Biome Reconstructions for Latin America at 0, 6000 and 18 000 Radiocarbon Years Ago, Clim. Past, 5, 725–767, https://doi.org/10.5194/cp-5-725-2009, 2009.

Marinova, E., Harrison, S. P., Bragg, F., Connor, S., de Laet, V., Leroy, S. A., Mudie, P., Atanassova, J., Bozilova, E., Caner, H., Cordova, C., Djamali, M., Filipova–Marinova, M., Gerasimenko, N., Jahns, S., Kouli, K., Kotthoff, U., Kvavadze, E., Lazarova, M., Novenko, E., Ramezani, E., Röpke, A., Shumilovskikh, L., Tanţău, I., and Tonkov, S.: Pollen–Derived Biomes in the Eastern Mediterranean–Black Sea–Caspian–Corridor, J. Biogeogr., 45, 484–499, https://doi.org/10.1111/jbi.13128, 2018.

Mucina, L.: Biome: Evolution of a Crucial Ecological and Biogeographical Concept, New Phytol., 222, 97–114, https://doi.org/10.1111/nph.15609, 2019.

Ni, J., Yu, G., Harrison, S. P., and Prentice, I. C.: Palaeovegetation in China during the Late Quaternary: Biome Reconstructions Based on a Global Scheme of Plant Functional Types, Palaeogeogr. Palaeoclimatol. Palaeoecol., 289, 44–61, https://doi.org/10.1016/j.palaeo.2010.02.008, 2010.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-Learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Pickett, E. J., Harrison, S. P., Hope, G., Harle, K., Dodson, J. R., Kershaw, A. P., Prentice, I. C., Backhouse, J., Colhoun, E. A., D'Costa, D., Flenley, J., Grindrod, J., Haberle, S., Hassell, C., Kenyon, C., Macphail, M., Martin, H., Martin, A. H., McKenzie, M., Newsome, J. C., Penny, D., Powell, J., Ian Raine, J., Southern, W., Stevenson, J., Sutra, J.-P., Thomas, I., van der Kaars, S., and Ward, J.: Pollen–Based Reconstructions of Biome Distributions for Australia, Southeast Asia and the Pacific (SEAPAC Region) at 0, 6000 and 18,000 14 C Yr BP, J. Biogeogr., 31, 1381–1444, https://doi.org/10.1111/j.1365-2699.2004.01001.x, 2004.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

Probst, P., Wright, M. N., and Boulesteix, A.-L.: Hyperparameters and Tuning Strategies for Random Forest, WIREs Data Min. Knowl. Discov., 9, e1301, https://doi.org/10.1002/widm.1301, 2019.

Qin, F., Zhao, Y., and Cao, X.: Biome Reconstruction on the Tibetan Plateau since the Last Glacial Maximum Using a Machine Learning Method, Sci. China Earth Sci., 65, 518–535, https://doi.org/10.1007/s11430-021-9867-1, 2022.

Raynolds, M. K., Walker, D. A., Balser, A., Bay, C., Campbell, M., Cherosov, M. M., Daniëls, F. J., Eidesen, P. B., Ermokhina, K. A., Frost, G. V., Jedrzejek, B., Jorgenson, M. T., Kennedy, B. E., Kholod, S. S., Lavrinenko, I. A., Lavrinenko, O. V., Magnússon, B., Matveyeva, N. V., Metúsalemsson, S., Nilsen, L., Olthof, I., Pospelov, I. N., Pospelova, E. B., Pouliot, D., Razzhivin, V., Schaepman-Strub, G., Šibík, J., Telyatnikov, M. Y., and Troeva, E.: A Raster Version of the Circumpolar Arctic Vegetation Map (CAVM), Remote Sens. Environ., 232, 111 297, https://doi.org/10.1016/j.rse.2019.111297, 2019.

Scheffer, M., Hirota, M., Holmgren, M., van Nes, E. H., and Chapin, F. S.: Thresholds for Boreal Biome Transitions, Proc. Natl. Acad. Sci. U. S. A., 109, 21 384–21 389, https://doi.org/10.1073/pnas.1219844110, 2012.

Schulzweida, U.: CDO User Guide Version 2.3.0, https://doi.org/10.5281/zenodo.10020800, 2023.

Sun, A., Luo, Y., Wu, H., Chen, X., Li, Q., Yu, Y., Sun, X., and Guo, Z.: An Updated Biomization Scheme and Vegetation Reconstruction Based on a Synthesis of Modern and Mid-Holocene Pollen Data in China, Glob. Planet. Change, 192, 103 178, https://doi.org/10.1016/j.gloplacha.2020.103178, 2020.

U.S. EPA: NA_CEC_Eco_Level1. Ecological Regions of North America Level I. 2010 Update, U.S. Environmental Protection Agency [data set], 2010.

Veloso, H. P., Oliveira Filho, L. C., Fonseca Vaz, A. M. S., Lima, M. P. M., Marquete, R., and Braza, J. E. M.: MANUAL Técnico Da Vegetação Brasileira, vol. n. 1 of *Série Manuais Técnicos Em Geociências*, IBGE, Rio de Janeiro, ISBN 85-240-0427-4, 1992.