CLM5 Land Data Documentation

Contents

May 2022

Technical Description of the Land Data of version 5.0 of the Community Land Model (CLM5)

Authors

Peter Lawrence, David Lawrence, Erik Kluzek, William Sacks, Keith Oleson, Johannes Feddema, George Hurtt, Louise Parsons Chini, Bette Otto-Bliesner, Esther Brady, Robert Tomas, Danica Lombardozzi, Katherine Dagon, Rosie Fisher, Jacquelyn Shuman, William Weider, Samuel Levis, Gordon Bonan, Sean Swenson, William Lipscomb, and Fang Li.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1.

INTRODUCTION

The purpose of this document is to fully describe the data that supports version 5.0 of the Community Land Model (CLM5). This data is used to prescribe the nature of the land surface and how that nature changes through time under human land use land cover change (LULCC). The sources, scientific justification, generation methods and evaluation of these parameterizations can be found in the referenced scientific papers (References). This document along with the CLM5 Technote (Lawrence et al. 2018) and the CLM5 User's Guide together provide the user with the scientific description and operating instructions for CLM.

1.1 CLM Land Data History

1.1.1 Inception of CLM

The initial stages of the Community Land Model were the result of the merging of community-developed land modeling focusing on biogeophysics, with an ongoing effort to expand the NCAR Land Surface Model (NCAR LSM, Bonan 1996) to include the carbon cycle, vegetation dynamics, and river routing. The initial effort, known as the Common Land Model, brought together the best features of the three existing land models of: the NCAR LSM (Bonan 1996, 1998); the Institute of Atmospheric Physics, Chinese Academy of Sciences land model (IAP94) (Dai and Zeng 1997); and the Biosphere-Atmosphere Transfer Scheme (BATS) (Dickinson et al. 1993). Documentation for the Common Land Model can be found in Dai et al. (2001 and 2003). A more extensive description of the development history of all versions of CLM can be found in the CLM5 Technote.

The Common Land Model initial land surface mapping was represented through the 28 biome land cover classification scheme of the NCAR LSM (Bonan 1995, 1996, Zeng et al. 2002). The Common Land Model, like the LSM model, required surface input data for each grid cell including: the biome type, which then determined the patch fractions of 12 Plant Functional Types (PFTs); the fraction of the grid cell covered by lakes; the fraction covered by wetlands; monthly climatological Leaf Area Index (LAI); soil texture (percent sand, silt, and clay); and soil color. The biome and land unit maps were generated from the 0.5x0.5 degrees dataset of Olson et al. (1983).

The PFT leaf and stem area indices were specified from monthly values adapted from Dorman and Sellers (1989). The soil texture and soil color maps were taken from LSM, which were in turn taken from BATS (Dickinson et al. 1993, Bonan 1996). The soil properties were originally determined from the Wilson (1984) classification of the FAO Soil Map of the World (FAO/UNESCO, 1974). This classification produced twelve soil texture classes, ranging from very coarse (equivalent to sand) to very fine (equivalent to heavy clay) assigned by averaging the textures from the 1x1 degree data set. Each soil texture class had prescribed sand, silt and clay composition. This classification also produced eight soil color classes ranging from dark to light,

with the scale stretched at the light end. The soil color was responsible for the soil reflectance properties that combined with vegetation properties and snow to determine the surface albedo.

1.1.2 CLM2

Following a period of extensive evaluation and code development the Common Land Model was adopted in May 2002 as the Community Land Model (CLM2) for use with the Community Atmosphere Model (CAM2, the successor to CCM3) and version 2 of the Community Climate System Model (CCSM2). In the CLM2 model, grid cells were divided into five primary land cover types: glacier, lake, wetland, urban, and vegetation. As part of the redevelopment, the biome representation was replaced with explicit specification of PFTs and leaf area index (LAI) from satellite data (Oleson and Bonan 2000, Zeng et al. 2002, Bonan et al. 2002a, b). The change to CLM2 also required modifications to parameterizations for vegetation albedo, leaf physiology, and soil water limitations on photosynthesis.

Vertical heterogeneity in soil texture was implemented for soil temperature and hydrology. A river routing model to prescribe the movement of the fresh water runoff to oceans also was added to the model configuration. Further changes were made in a post release version of the model (CLM2.1) to represent carbon and nitrogen cycling in the model. This involved changing data structures from spatially independent sub-grid patches to one that recognizes three hierarchical scales within a model grid cell: land unit, snow/soil column, and PFT.

The CLM2 surface data were generated from a range of satellite products as described by Zeng et al. (2002) and Bonan et al. (2002). The land unit and vegetation maps were derived from the 1-km International Geosphere–Biosphere Program Data and Information System (IGBP DISCover) dataset (Loveland et al. 2000) and the 1-km University of Maryland tree cover dataset (DeFries et al. 1999, 2000a, b). The vegetation data were generated at 0.5 degrees as maps of the abundance of the seven primary PFTs of needleleaf evergreen or deciduous tree, broadleaf evergreen or deciduous tree, shrub, grass, crop. Temperature and precipitation maps from Willmott and Matsuura (2000) were used to distinguish arctic, boreal, temperate, and tropical plants, C3 and C4 grasses, and evergreen and deciduous shrubs. Monthly leaf area index for each PFT in each 0.5x0.5 degree grid cell was obtained from 1-km Advanced Very High Resolution Radiometer (AVHRR) red and near-infrared reflectances for April 1992–March 1993 (Myneni et al. 1997, Bonan et al. 2002). Stem area index, canopy top height, and canopy bottom height were based on the LSM values prescribed for each PFT (Bonan et al. 2002).

Physiological parameters for the 16 PFTs were obtained from the 12 LSM PFTs (Bonan 1996) so that while the list of PFTs expanded, no new physiologies were introduced. Soil sand, clay and silt texture were updated from the International Geosphere-Biosphere Programme (IGBP) soil dataset (Global Soil Data Task 2000). Soil color mapping however remained the same as the Common Land Model following the methods of Dickinson et al. (1993) and Zeng et al. (2002). Percent lake and wetlands were derived from the 1.0x1.0 degree data of Cogley (1991) for perennial freshwater lakes and swamps/marshes. The river transport network data was developed with water flow direction for each 0.5x0.5 degree grid cell determined as one of eight compass points (north, northeast, east, southeast, south, southwest, west, and northwest) based on the steepest downhill slope determined from the TerrainBase 5 minute Global DTM (digital terrain model) and the CIA World Data Bank II as described in Graham et al. (1999).

1.1.3 CLM3

Following further software developments, evaluation and improvements in biogeophysical parameterizations to correct deficiencies in the coupled model climate, the CLM3 model was released to the community in June 2004 as a component of the Community Climate System Model (CCSM3.0). The climate statistics of CLM3 when coupled to CCSM3.0, and an analysis of selected features of the land hydrological cycle are provided in Dickinson et al. (2006) and Hack et al. (2006). Beyond the updates to the biogeophysical parameterizations the land data of CLM3 remained the same as CLM2.

1.1.4 CLM3.5

Investigations into the hydrologic biases in CLM3, lead to the initiation of a new project in 2004 to improve the hydrology of the model. The outcome of these investigations were a range of improvements that were realized as the development of CLM 3.5 (Oleson et al 2008). An important element of these improvements was the inclusion of new surface datasets based on Moderate Resolution Imaging Spectroradiometer (MODIS) products following the work of Lawrence and Chase (2007). Additional improvements included new parameterizations for canopy integration, canopy interception, frozen soil, soil water availability, soil evaporation and snow fraction (Niu and Yang 2006). The soil hydrology was updated with a new TOPMODEL-based groundwater model for determining water table depth, and surface and subsurface runoff from Niu et al. (2005, 2007). The CLM3.5 model also introduced a factor to simulate nitrogen limitation on plant productivity.

The new surface datasets from Lawrence and Chase (2007) were developed globally at 0.05 degrees for vegetation cover from the MODIS Vegetation Continuous Fields (VCF) data from Hansen et al. (2003) and the IGBP Land Cover data from Friedl et al (2002). The PFT distributions were further refined using the bioclimatic rules of Bonan et al. (2002b) with updated versions of the temperature and precipitation maps of Willmott and Matsuura (2000). To better represent the description of C3 and C4 grasses, the biogeography and climate rules were modified to include the fractional C3/C4 mapping methods of Still et al. (2003), with 2001 – 2003 average monthly MODIS LAI from Myneni et al. (2002) used to describe seasonal variations in vegetation cover as indicator for the C3/C4 growing season. Current day crop distributions for CLM3.5 were prescribed directly from the year 1992 of the historical cropping data set of Ramankutty and Foley (1999). This direct prescription of cropping distributions had the added advantage of providing a consistent method for describing historical cropping in CLM for any year from 1700 to 1992, allowing for land cover change studies in CCSM3.0, such as those performed by Lawrence and Chase (2010).

The 2001 – 2003 average monthly MODIS LAI data of Myneni et al. (2002) were used to produce new monthly PFT LAI values that when combined with the new PFT distributions reproduced the average monthly MODIS LAI values at the 0.05 degree resolution. Soil sand, clay and silt texture remained the same as CLM3 being prescribed from the International Geosphere-Biosphere Programme (IGBP) soil dataset (Global Soil Data Task 2000). The new TOPMODEL-based groundwater model required global topographic characteristic descriptions for calculating the properties of the modeled water table and corresponding surface and subsurface runoff. The topographic characteristics were prescribed from the maximum saturated fraction (*fmax*) defined as the discrete cumulative distribution function (CDF) of the topographic index when the grid cell mean water table depth is zero. For CLM3.5 this was calculated at a spatial resolution of 0.5x0.5 degrees and aggregated for use at the resolution of the model following Niu et al. (2005).

A new soil color map with 20 soil colors was developed to prescribe soil reflectance properties calculated directly from 2001 – 2003 average monthly MODIS albedo data from Schaaf et al. (2003). The wider range of the 20 soil colors was required to cover the soil reflectivity values found in MODIS albedo product that could not be prescribed in the CLM3 eight soil colors scheme. The new soil colors were fitted with CLM soil moisture and albedo targets using the CLM3 surface radiation model for each grid cell for each month to minimize differences between the calculated CLM 3.5 broadband surface albedo and the average monthly MODIS values.

1.1.5 CLM4

The next version of the model, CLM4, was released in June 2010 as a component of the CCSM4.0. The motivation for the new model was to incorporate several recent scientific advances in the understanding and representation of land surface processes, expand model capabilities, and improve surface and atmospheric forcing datasets (Oleson et al. 2010a, Lawrence et al. 2011). For the land surface data a key component was the representation of the transient land use and land cover change (LULCC) of the Couple Model Intercomparison Project version 5 (CMIP5) historical and future Representative Concentration Pathway (RCP) time series (Hurtt et al. 2006, 2011) in the CLM4 land surface data framework as described in Lawrence et al. (2012). The CLM4 surface data update also included the refinement of the global PFT, wetland, and lake distributions.

Other major capabilities added to CLM4 included: a representation of the carbon-nitrogen cycle (CLM4CN) derived in part from the Biome-BGC model (Thornton and Zimmerman 2007, Thornton et al. 2007, White et al. 2000); an urban canyon model to contrast rural and urban energy balance and climate through (CLMU) (Oleson et al. 2010b); a more sophisticated representation of soil hydrology and snow processes; inclusion of organic soil and deep soil into the existing mineral soil treatment to enable more realistic modeling of permafrost (Lawrence and Slater 2008); and an updated biogenic volatile organic compounds (BVOC) model (Pfister et al. 2008). Other modifications of note include more realistic optical properties for grasslands and croplands, new treatments of soil column-groundwater interactions, soil evaporation, aerodynamic parameters for sparse/dense canopies, vertical burial of vegetation by snow, snow cover fraction and aging, black carbon and dust deposition, and vertical distribution of solar energy for snow.

The new CLM4 transient LULCC time series land data were generated as annual 0.5x0.5 degree PFT and wood harvest area maps generated from the Land Use Harmonization (LUH) data sets provided for the CMIP5 Earth system modeling project (Hurtt et al. 2006, 2011). The time series consisted of the Last Millennium 850 – 1849, the Historical period 1850 – 2005, and the four future RCP scenarios of 2.6, 4.5, 6.0 and 8.5 for 2006 – 2100 (Lawrence et al. 2012, Landram et al. 2013). The LUH time series data sets consisted of maps of annual state values for the fraction of the four land units of primary, secondary, crop and pasture, along with the fraction of wood harvest for primary forest and non forest, and secondary mature forest, young forest and non forest.

To capture the annual transient land use data, the CLM4 surface data sets were split into two components. The static, time invariant data were prescribed in the surface data file describing the distribution and character of glaciers, lakes, wetlands, and urban areas, along with soil texture and color, monthly PFT LAI, SAI and canopy height, and the initial PFT distributions of the time series. The time varying annual maps of PFTs along with the annual wood harvest area for each class were prescribed in the land use time series (pftdyn) file. The new transient LULCC and CLM4CN configuration allowed investigations into the climate and carbon cycle impacts of transient LULCC and emissions in CCSM4/CESM1 with prognostic canopy morphology, carbon and nitrogen pools (Lawrence et al. 2012, 2018).

The translation from the LUH time series land units to CLM4 PFTs was performed based on the current day satellite derived grid cell values for each land unit using the CLM4 Land Use Data Tool as described in Lawrence et al. (2012). This process was run for each year of the LUH time series with the land unit PFT and wood harvest fractions from all of the land units being aggregated to the CLM4 grid cell values. Where a land unit was not present in the current day data, nearest neighbor extrapolation was performed using inverse distancing methods from surrounding grid cells. The current day satellite PFT distributions were generated following the methods developed for CLM3.5 by Lawrence and Chase (2007), with updated MODIS and climate products covering the 2001 – 2008 period. New MODIS land cover rules were enforced to prevent low tree fractions from the MODIS VCF data in forests after initial low carbon amounts were found in the CLM4CN model.

The percent glacier remained the same as previous versions of CLM being derived from the IGBP Data and Information System Global 1-km Land Cover Data Set (IGBP DISCover) (Loveland et al. 2000). The percent lake and wetland also remained the same as in previous versions of the model being derived from Cogley (1991). The CLMU urban canyon model required new percent urban mapping for the urban land unit, as well as urban properties for the thermal (e.g., heat capacity and thermal conductivity), radiative (e.g., albedo and emissivity) and morphological (e.g., height to width ratio, roof fraction, average building height, and pervious fraction of the canyon floor) properties of roof/wall/road. The CLMU urban extent was prescribed for the four urban density classes of tall building district (TBD), and high, medium, and low density (HD, MD, LD). Mapping was derived from the LandScan 2004 population density dataset which itself was derived from census data, nighttime lights satellite observations, road proximity and slope (Dobson et al., 2000) as described by Jackson et al. (2010). The average building properties are provided for 33 distinct regions across the globe by Jackson et al. (2010) for each of the four density classes.

New monthly PFT LAI maps were generated with updated 2001 – 2008 average monthly MODIS LAI data and the new current day PFT distributions. Soil texture mapping was updated to include soil organic matter that varied with depth using the methods of Lawrence and Slater, (2008), and the mapping from the International Geosphere-Biosphere Programme (IGBP) soil dataset (Global Soil Data Task 2000). The sand, silt and clay composition of the mineral soil components remained the same as previous versions of the model, also based on the International Geosphere-Biosphere Programme (IGBP) soil dataset. The topographic characteristics for the groundwater model remained the same as CLM3.5 being prescribed from the maximum saturated fraction (*fmax*) from Niu et al. (2005). New soil color maps were generated from updated 2001 – 2008 average monthly MODIS albedo combined with the new PFT and PFT LAI data. The new PFT LAI and soil color data were generated following the methods described in Lawrence and Chase (2007).

1.1.6 CLM4.5

The next version of the model CLM4.5, was released to the community in June 2013 along with the Community Earth System Model version 1.2 (CESM1.2) (Oleson et al. 2013). Specifically the new version of the model included revisions to several parameterizations to reflect new scientific understanding and attempt to reduce biases identified in CLM4 simulations including low soil carbon stocks especially in the Arctic, excessive tropical GPP and unrealistically low Arctic GPP, a dry soil bias in Arctic soils, unrealistically high LAI in the tropics, and a transient 20th century carbon response that was inconsistent with observational estimates.

Major modifications involved updates to canopy processes including a revised canopy radiation scheme and canopy scaling of leaf processes, co-limitations on photosynthesis, revisions to photosynthetic parameters (Bonan et al. 2011), temperature acclimation of photosynthesis, and improved stability of the iterative solution in the photosynthesis and stomatal conductance model (Sun et al. 2012). To address the low soil carbon simulated in CLM4, the CLM4.5 model had an updated vertically resolved soil biogeochemistry scheme with base decomposition rates modified by soil temperature, water, and oxygen limitations, which included vertical mixing of soil carbon and nitrogen due to bioturbation, cryoturbation, and diffusion (Koven et al. 2013). The litter and soil carbon and nitrogen pool structure as well as nitrification and denitrification were modified based on the Century model, with biological fixation revised to distribute fixation more realistically over the year.

The fire model was replaced with a model that included representations of natural and anthropogenic triggers and suppression as well as agricultural, deforestation, and peat fires (Li et al. 2012a, b, Li et al. 2013a). The crop model that had been developed as extensions to the CLM3.5 and CLM4 models (Levis et al. 2009, 2012), was included as a supported option in CLM4.5, with corn, temperate cereals (spring wheat), and soybean explicitly modeled (Levis et al. 2016). The crop model was also extended to include interactive fertilization, organ pools (Drewniak et al. 2013), and irrigation (Sacks et al. 2009).

Finally the model had updated frozen hydrology and snow representations (Swenson et al. 2012, Swenson and Lawrence 2012, 2014, 2015), a completely revised and more realistic lake model (Subin et al. 2012a,b), variable river flow velocities based on mean grid cell slope, and the ability to explicitly model ice sheets when coupled to the Glimmer-CISM model of Lipscomb and Sacks (2012). An explicit surface water store was also added, replacing the wetland land unit and permitting prognostic wetland distribution modeling. The CLMU urban canyon model was updated to simulate multiple urban density classes on the land unit, rather than the single dominant urban density class as done in CLM4. The biogenic volatile organic compounds model was updated to MEGAN2.1, to include speciated monoterpenes, sesquiterpenes, oxygenated VOC as well as isoprene (Guenther et al. 2012). The model also was extended to include a methane production, oxidation, and emissions model (Riley et al. 2011).

While the transient natural vegetation PFT, wood harvest, PFT LAI, SAI and canopy height, and soil color data sets remained unchanged from CLM4, many of the CLM4.5 updates required updates to other elements of the surface data. The coupling of CLM4.5 to the Glimmer-CISM ice sheet model included new high resolution multiple elevation class descriptions for glaciers and ice sheets with the overlap of the two models prescribed through grid overlap files. For both the coupled and uncoupled glacier versions of CLM4.5 the percent glacier and ice sheet where prescribed from the global Randolph Glacier Inventory version 1.0 (RGIv1.0) (Arendt et al. 2012) combined with the Greenland Ice Sheet (Rastner et al. 2012), and the Antarctic Ice Sheet data from the Scientific Committee on Antarctic Research (SCAR) Antarctic Digital Database version 5.0. The ice sheet and Antarctic ice shelf masks were combined with the Global Land One-km Base Elevation Project (GLOBE) topography (Hastings et al. 1999), to generate 10 ice-covered elevation class areas for ice-covered regions. Grid cells flagged as land-ice in the mask but ocean in GLOBE (typically, around ice sheets at high latitudes) were designated land-ice with an elevation of 0 meters. For the uncoupled glacier version of the model, the percent glacier was calculated as the sum of all of the elevation classes for ice sheets and glaciers.

The new lake model from Subin et al. (2012a, b) required the percent lake and area-averaged lake depth to be prescribed globally. New mapping for percent lake was taken from the Global Lake and Wetland Database of Lehner and Doll (2004), with the mean lake depth of each grid

cell calculated based on the global gridded data sets of Kourzeneva (2012). Due to the new surface water store, wetland mapping was removed from surface data sets. The wetlands continued to be represented in non land areas accounting for ocean areas to resolve land masks conflicts between CLM4.5 and other models. The updated CLMU urban model, while simulating multiple urban classes per urban land unit, used the same urban density class mapping and building properties as CLM4.

The CLM4.5 crop model required a new crop land unit to be added to the existing CLM4 land units of glacier, lake, wetland, urban, and vegetated. The crop land unit contained all of the managed Crop Functional Types (CFTs) that were simulated in the CLMCrop model. Unlike the natural vegetated land unit, the CFTs have individual soil columns and thus permit different land management between crops and with the natural vegetation. The CLMCrop model of CLM4.5 represented the three crop types of corn, temperate cereals (spring wheat), and soybean. Other crop areas were retained on the vegetated land unit with the remaining CLM4 crop PFT area. To manage the new CFTs in CLM4.5, the CLM4 PFT list was extended from 16 classes to 25. The new PFT classes split the CLM4 unmanaged crop PFT 15 into unirrigated and irrigated PFTs 15 and 16, and then into unirrigated and irrigated CFTs for the corn, temperate and winter cereals, and soybean crops (17 to 24). The new crop CFTs were prescribed from the present-day crop dataset of Portmann et al. (2010). The managed crops were assigned in the proportions given by Portmann et al. (2010) without exceeding the area previously assigned to the unmanaged crop PFT in the land use time series data. The unirrigated and irrigated fractions of each crop type were prescribed from the current day Global Map of Irrigation Areas from Siebert et al. (2005).

The mineral soil texture mapping of sand, silt and clay remained the same as the CLM4 model being prescribed from the IGBP soil dataset. The soil organic matter data however was updated to include two new sources. The majority of the globe was from ISRIC-WISE (Batjes, 2006), while the high latitudes came from the 0.25 degree version of the Northern Circumpolar Soil Carbon Database (Hugelius et al. 2012). Both datasets report soil carbon down to 1m depth. The maximum saturated fraction (*fmax*) that described topographic characteristics in the ground water model was updated to the 0.125 degree resolution using the 1-km compound topographic indices (CTIs) based on the HYDRO1K dataset (Verdin and Greenlee 1996) from USGS, following the methods of Niu et al. (2005) and Li et al. (2014). The CLM4.5 model also had slope and elevation prescribed from the USGS HYDRO1K 1-km dataset. Slope is used to calculate the variable flow rates of the River Transport Model and for the surface water parameterization. Elevation is used to calculate the grid cell standard deviation of topography for the snow cover fraction parameterization.

The new fire model needed global maps of population density, gross domestic production, peat area fraction, and peak month of agricultural burning. Global population density at 0.5x0.5 degree resolution was developed annually for the 1850-2100 time period from decadal population density data for 1850–1980 from the Database of the Global Environment version 3.1 (HYDEv3.1) (Klein Goldewijk 2011) and population density data for 1990, 1995, 2000, and 2005 from the Gridded Population of the World version 3 dataset (GPWv3) (CIESIN, 2005). Gross Domestic Production (GDP) per capita in 2000 at 0.5x0.5 degree resolution is from van Vuuren et al. (2006), which is the base-year GDP data for IPCC-SRES and derived from country-level World Bank's World Development Indicators (WDI) measured in constant 1995 US dollars (World Bank, 2004) and the UN Statistics Database (UNSTAT, 2005). The peatland area fraction at 0.5x0.5 degree resolution is derived from datasets: peatland data in Indonesia and Malaysian Borneo (Olson et al. 2001); peatland data in Canada (Tarnocai et al. 2011); and bog, fen and mire data in boreal regions (north of 45N) outside Canada provided by the Global Lakes and Wetlands Database

(GLWD) (Lehner and Doll, 2004). The climatological peak month for agricultural waste burning is from van der Werf et al. (2010).

1.2 CLM5 Land Data Representation

While the purpose of this document is to fully describe the data that supports version 5.0 of the Community Land Model (CLM5), the use of the data and how it is represented in the CLM5 model data structures, states, and fluxes is required to fully understand how the data represents the nature of the land surface and how that nature changes through time under human land use and land cover change (LULCC). A full scientific description of the parameterizations and processes that make up the CLM5 model is beyond the scope of the document and can be found in the CLM5 Technote. Here we provide the technical framework, data sources, tools, evaluation and background for the all of the CLM5 land data represented in the model.

1.2.1 CLM5 Background

The CLM5 version of the model builds on the progress made in CLM4.5 with most major components of the model having been updated from the previous version of the model (Lawrence et al. 2019). A schematic of the processes and systems represented in CLM5 is shown in Figure 1.1. From the land data perspective the most notable update is the explicit representation of land use and land cover change (LULCC) to simulate the CMIP6 Land Use Model Intercomparison (LUMIP) historical and projected time series data (Lawrence et al. 2016, Hurtt et al. 2020). The new LULCC data required new time series data generation tools, as well as the compilation and integration of new global vegetation and crop modeling data. Notable changes in the land surface data included new plant functional type (PFT) and crop functional type (CFT) time series mapping, along with new wood harvest, shifting cultivation (not currently implemented in CLM5), irrigation and fertilizer prescription. Other important changes include updates to soil and plant hydrology, snow density, river modeling, carbon and nitrogen cycling and coupling, and crop modeling.

The representation of human management of the land (wood harvest, agriculture and urban) is augmented in several ways. Wood harvest that was prescribed as an area amount in CLM4.5 has been updated with specific PFT wood harvest carbon extraction amounts to address problems in the unknown forest carbon density prior to prescription. The CLM4.5 crop model has been extended to operate globally through the addition of rice and sugarcane as well as tropical varieties of corn and soybean (Badger and Dirmeyer, 2015, Levis et al., 2016). These crop types are added to the existing temperate corn, temperature soybean, spring wheat, and cotton crop types. Fertilization rates and irrigation equipped area updated annually based on crop type and geographic region through an input dataset. The irrigation trigger is updated. Additional minor changes include crop phenological triggers that vary by latitude for selected crop types, grain C and N is now removed at harvest to a 1-year product pool with the carbon for the next season's crop seed removed from the grain carbon at harvest. Through the introduction of the capability to dynamically adjust landunit weights during a simulation, the crop model can now be run coincidentally with prescribed land use, which significantly expands the capabilities of the model (Figures 1.2 and 1.3). Several heat stress indices for both urban and rural areas are calculated and output by default (Buzan et al., 2015). A more sophisticated and realistic building space heating and air conditioning submodel that prognoses interior building air temperature and includes more realistic space heating and air conditioning wasteheat factors is incorporated.
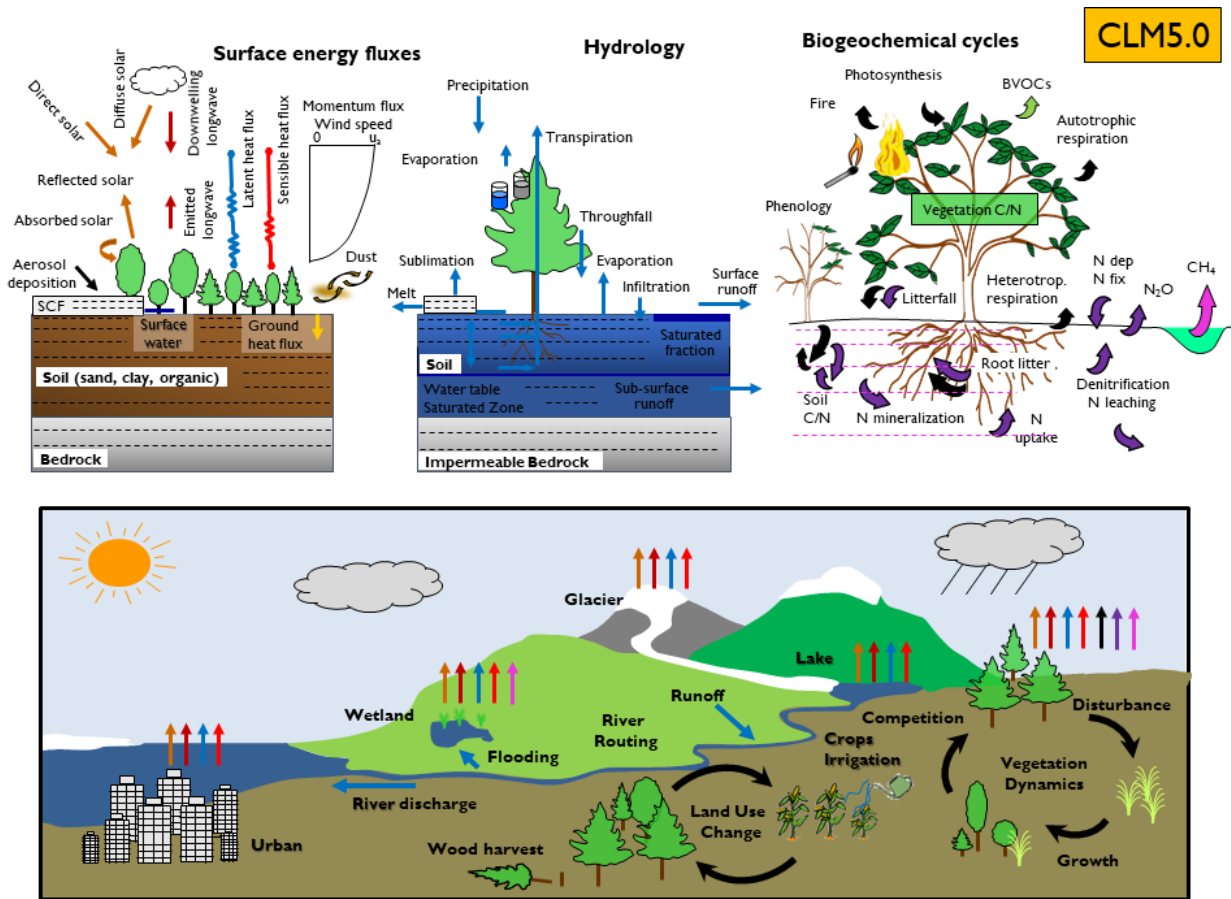
Figure 1.1: Land biogeophysical, biogeochemical, and landscape processes simulated by CLM5 from Lawrence et al. (2019).

Plant nutrient dynamics are substantially updated in CLM5 to resolve several deficiencies with the CLM4 and CLM4.5 nutrient cycling representation. The Fixation and Update of Nitrogen (FUN) model based on the work of Fisher et al. (2010), Brzostek et al. (2014), and Shi et al. (2016) is incorporated. FUN calculates the rate of symbiotic N fixation, with this N passed straight to the plant, not the mineral N pool. Separately, CLM5 also calculates rates of symbiotic (or free living) N fixation as a function of evapotranspiration (Cleveland et al. 1999), which is added to the soil inorganic ammonium (NH4+) pool. The static plant carbon:nitrogen (C:N) ratios utilized in CLM4 and CLM4.5 are replaced with variable plant C:N ratios which allows plants to adjust their C:N ratio, and therefore their leaf nitrogen content, with the cost of N uptake (Ghimire et al. 2016). The implementation of a flexible C:N ratio means that the model no longer relies on instantaneous downregulation of potential photosynthesis rates based on soil mineral nitrogen availability to represent nutrient limitation. Stomatal conductance is now based on the N-limited photosynthesis rather than on potential photosynthesis. Finally, the Leaf Use of Nitrogen for Assimilation (LUNA) (Xu et al., 2012, Ali et al.,2016) model is incorporated. The LUNA model calculates photosynthetic capacity based on optimization of the use of leaf nitrogen under different environmental conditions such that light capture, carboxylation, and respiration are co-limiting.

CLM5 applies a fixed allocation scheme for woody vegetation, rather than a dynamic NPP-based allocation scheme, as was used in CLM4 and CLM4.5. This change was driven by observations that indicate biomass saturates with increasing productivity, which contrast to the behavior in

CLM4 and CLM4.5 where biomass continuously increases with increasing productivity (Negron-Juarez et al., 2015). Soil carbon decomposition processes are unchanged in CLM5, but a new metric for apparent soil carbon turnover times (Koven et al., 2017) suggested parameter changes that produce a weak intrinsic depth limitation on soil carbon turnover rates (rather than the strong depth limitaiton in CLM4.5) and that the thresholds for soil moisture limitation on soil carbon turnover rates in dry soils should be set at a wetter soil moisture level than that used in CLM4.5. The fire model is the same as utilized in CLM4.5 except that a modified scheme is used to estimate the dependence of fire occurrence and spread on fuel wetness for non-peat fires outside cropland and tropical closed forests (Li and Lawrence, 2017) and the dependence of agricultural fires on fuel load is removed.

Several changes are included that are mainly targeted at improving the simulation of surface mass balance over ice sheets. The fresh snow density parameterization is updated to more realistically capture temperature effects and to additionally account for wind effects on new snow density (van Kampenhout et al., 2017). The maximum number of snow layers and snow amount is increased from 5 layers and 1m snow water equivalent to 12 layers and 10m snow water equivalent to allow for the formation of firn in regions of persistent snow-cover (e.g., glaciers and ice sheets) (van Kampenhout et al., 2017). The CISM2 ice sheet model is included for Greenland by default. The ice sheet does not evolve for typical configurations, but ice sheet evolution can be turned on by choosing an appropriate compset. The introduction in CLM5 of the capability to dynamically adjust landunit weights means that a glacier can initiate, grow, shrink, or disappear during a simulation when ice evolution is active. Multiple elevation classes (10 elevation classes by default) and associated temperature, rain/snow partitioning, and downwelling longwave downscaling are used for glacier landunits to account for the strong topographic elevation heterogeneity over glaciers and ice sheets.

The hydrology updates of CLM5 include the introduction of a dry surface layer-based soil evaporation resistance parameterization (Swenson and Lawrence, 2014) and a revised canopy interception parameterization. Canopy interception is now divided into liquid and solid phases, with the intercepted snow subject to unloading events due to wind or above freezing temperatures. The snow-covered fraction of the canopy is used within the canopy radiation and surface albedo calculation. Instead of applying a spatially uniform soil thickness as done in previous versions, soil thickness in CLM5 can vary in space (Brunke et al. 2016, Swenson and Lawrence, 2015) and is set to values within a range of 0.4m to 8.5m depth, derived from a spatially explicit soil thickness data product (Pelletier et al., 2016). The explicit treatment of soil thickness allows for the deprecation of the unconfined aquifer parameterization used in CLM4.5, which is replaced with a zero flux boundary condition and explicit modeling of both the saturated and unsaturated zones. The default model soil layer resolution is increased, especially within the top 3m, to more explicitly represent active layer thickness within the permafrost zone.

A plant hydraulic stress routine is introduced in CLM5 that explicitly models water transport through the vegetation according to a simple hydraulic framework (Kennedy et al., 2019). The water supply equations are used to solve for vegetation water potential forced by transpiration demand and a set of layer-by-layer soil water potentials. Stomatal conductance, therefore, is a function of prognostic leaf water potential. Water stress is calculated as the ratio of attenuated stomatal conductance to maximum stomatal conductance. An emergent feature of the plant hydraulics is soil hydraulic redistribution. In CLM5, maximum stomatal conductance is obtained from the Medlyn conductance model (Medlyn et al., 2011), rather than the Ball-Berry stomatal conductance model that was utilized in CLM4.5 and prior versions of the model. The Medlyn stomatal conductance model is preferred mainly for its more realistic behavior at low humidity

levels (Rogers et al., 2017). The stress deciduous vegetation phenology trigger is augmented with an antecedent precipitation requirement (Dahlin et al. 2015).

Rooting profiles were used inconsistently in CLM4.5 with Zeng (2001) profiles used for water and Jackson et al. (1996) profiles used for carbon inputs. For CLM5, the Jackson et al. (1996) rooting profiles are used for both water and carbon. Roots are deepened for the broadleaf evergreen tropical tree and broadleaf deciduous tropical tree types. Finally, an adaptive time-stepping solution to the Richard's equation is introduced, which improves the accuracy and stability of the numerical soil water solution. The River Transport Model (RTM) is replaced with the Model for Scale Adaptive River Transport (MOSART) (Li et al. 2013b) in which surface runoff is routed across hillslopes and then discharged along with subsurface runoff into a tributary subnetwork before entering the main channel.

Included with the release of CLM5 is a functionally supported version of the Functionally-Assembled Terrestrial Ecosystem Simulator (FATES) (Fisher et al., 2015). A major motivation of FATES is to allow the prediction of biome boundaries directly from plant physiological traits via their competitive interactions. FATES is a cohort model of vegetation competition and co-existence, allowing a representation of the biosphere, which accounts for the division of the land surface into successional stages, and for competition for light between height-structured cohorts of representative trees of various plant functional types. FATES is not active by default in CLM5. The classical dynamic global vegetation model (CLM-DGVM) that has been available within CLM4 and CLM4.5 remains available, though it is largely untested.

1.2.2 CLM5 Surface Data Framework

The land surface in CLM5 is represented as a nested sub-grid hierarchy in which the land component of an Earth system grid cell is composed of a mosaic of land units that represent glaciers, lakes, urban areas, natural vegetation and crops (when the crop model option is turned on) (Figure 1.2). Below the land units, the hierarchy further divides into columns that represent different elements for each corresponding land unit. For the natural vegetation and crop land units the soil columns have a further division to patches that represent plant and crop function types (PFTs and CFTs) that occur on those land units. A single surface data file is used to prescribe the time invariant data describing the land fraction, the distribution and character of land units, along with the glacier elevation classes, lake properties, building properties, soil texture and color, and monthly climatological PFT leaf area index (LAI), stem area index (SAI) and canopy height values. Additionally, the surface data prescribes the initial land unit, PFT, and CFT distributions for a time series describing time evolving land use and land cover change, which is further described in the next section and shown in Figure 1.3. The complete list of data used for generating both the surface data file and the land use time series data files is shown in Table 1.1.

The CLM5 default land fraction is updated from CLM4.5, being prescribed from updated MODIS IGBP Land Cover mapping at 1km resolution described in Chapter 2, which is combined with the default current day glacier land unit extent described in Chapter 4 outside of the MODIS land mask. The CLM5 glacier land unit extent is prescribed by the extent of the 10 elevation classes per grid cell. The initial or default current day distribution of the elevation classes are prescribed in the same manner as in CLM4.5, from ice sheets and Antarctic ice shelf masks combined with the GLOBE topography to generate the 10 ice-covered areas (Arendt et al. 2012, Rastner et al. 2012, Hastings et al. 1999). This organization of glaciers and elevations classes is shown in Figure 1.2, with the mapping further explored in Chapter 4. When coupled to the CISM2 ice sheet model, CLM5 has the capability to dynamically adjust landunit weights which means that a glacier

can initiate, grow, shrink, or disappear into and out of the natural vegetation and crop land units during a simulation when ice evolution is active.

The CLM5 lake land unit is prescribed through a single lake column with average grid cell lake characteristics representing a single average lake model for the grid cell. The distribution and character of lakes are the same as in CLM4.5, taken from the Global Lake and Wetland Database of Lehner and Doll (2004), with the mean lake depth of each grid cell calculated based on the global gridded data sets of Kourzeneva (2012) as described in (Chapter 4). The CLM5 urban land unit is divided into three density classes representing the tall building district, high density, and medium density urban areas. The mapping of the density classes is the same as in CLM4 and CLM4.5. The density classes are mapped from the LandScan 2004 population density dataset with average building properties provided for 33 distinct regions across the globe from Jackson et al. (2010) described in Chapter 4. The urban land units have five columns of roof, sunlit walls and shaded walls, and pervious and impervious canyon floor, with the radiative, thermal (Oleson et al. 2010b).
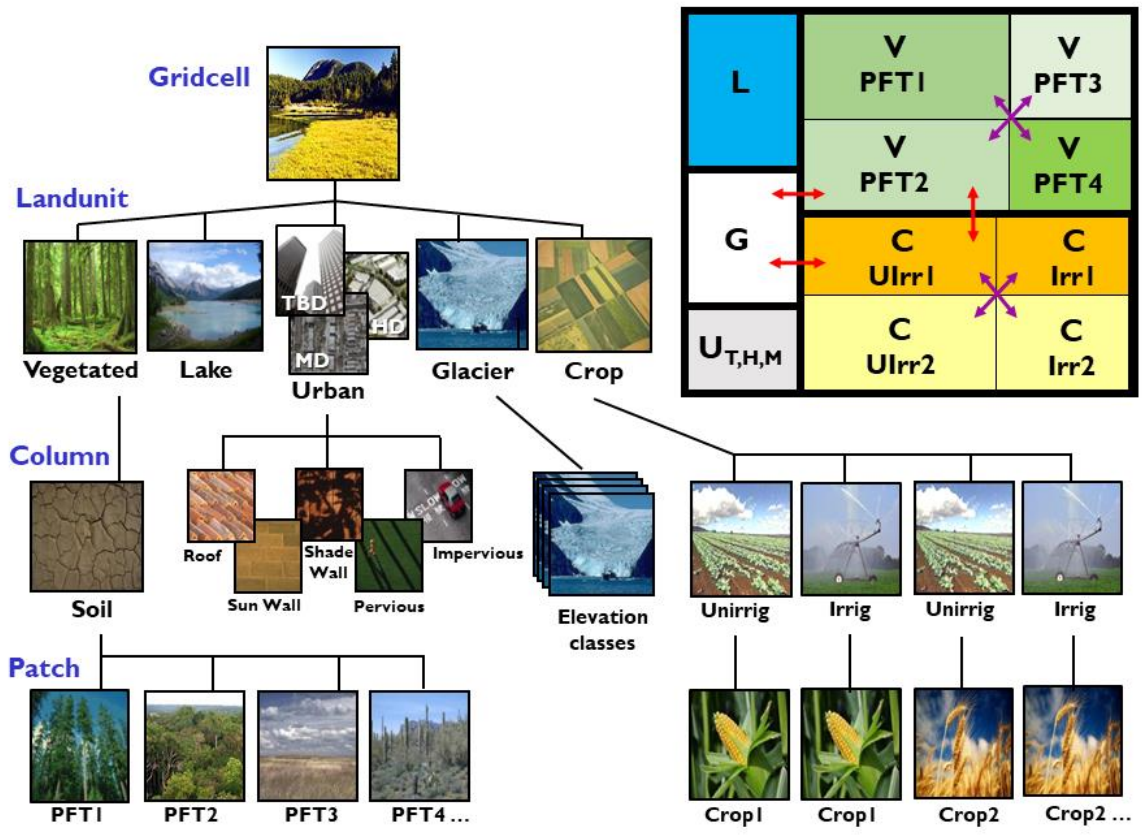
Figure 1.2: Configuration of the CLM5 subgrid hierarchy. Box in upper right shows hypothetical subgrid distribution for a single grid cell. Note that the Crop land unit is only used when the model is run with the crop model active. Abbreviations: TBD – Tall Building District; HD – High Density; MD – Medium Density, G – Glacier, L – Lake, U – Urban, C – Crop, V – Vegetated, PFT – Plant Functional Type, Irr – Irrigated, UIrr – Unirrigated. Red arrows indicate allowed land unit transitions. Purple arrows indicate allowed patch-level transitions.

The naturally vegetated land unit, while capable of having multiple soil columns, by default has a single shared soil column for all natural PFTs in a grid cell. The crop land unit by contrast is split into irrigated and unirrigated columns with a single crop occupying each column. The initial distribution of the natural vegetation and crop land units as well as the patch level PFTs and CFTs are prescribed from the surface data file, along with the soil texture and color of the soil column, and the monthly climatological PFT LAI, SAI and canopy height values. Much of this technote is dedicated to describing the prescription of these properties from satellite and other current day mapping combined with the CMIP6 LUMIP LUH2 land use time series data defined in Lawrence et al. (2016) and Hurtt et al. (2020).

Each soil column in both the naturally vegetated and crop land units can capture variability in the soil and snow state variables for different parts of the grid cell, with independently evolving vertical profiles of carbon, snow, soil water and temperature. The default snow/soil column is represented by 25 layers for ground, with up to 20 of these layers classified as soil layers from depth to bedrock mapping, and the remaining layers classified as bedrock layers. There are additionally up to 10 layers for snow, depending on snow depth. The majority of soil properties remained the same as CLM4.5 and detailed in Chapter 4.

Soil texture for all soil columns is prescribed for mineral content the same way as the CLM4 and CLM4.5 models being derived from the IGBP soil dataset. The soil organic matter data remained the same as CLM4.5 coming from the ISRIC-WISE (Batjes, 2006) global mapping, with the high latitudes coming from the 0.25 degree version of the Northern Circumpolar Soil Carbon Database (Hugelius et al. 2012). The maximum saturated fraction ($fmax$) that described topographic characteristics in the ground water model stayed the same as CLM4.5 at the 0.125 degree resolution using the 1-km compound topographic indices (CTIs) based on the HYDRO1K dataset (Verdin and Greenlee 1996) following the methods of Niu et al. (2005) and Li et al. (2014). Slope and elevation are also obtained from the USGS HYDRO1K 1-km dataset. Slope is used in the surface water parameterization, and elevation is used to calculate the grid cell standard deviation of topography for the snow cover fraction parameterization. Each of these components are detailed in Chapter 4.

The new variable soil thickness capability of CLM5 was prescribed through soil depth to bedrock globally (Brunke et al. 2016, Swenson and Lawrence, 2015) with values derived from the spatially explicit soil thickness data product of Pelletier et al., (2016). New global soil color maps were generated from updated 2003 – 2015 average monthly MODIS albedo combined with the new CLM5 PFT and PFT LAI data, and the two stream radiation canopy model. While not a component of the CLM5 surface data, the new MOSART river transport model is supported by a comprehensive, global hydrography dataset at 0.5x0.5 degrees resolution. The topographic parameters, such as flow direction, channel length, topographic and channel slopes were derived using the Dominant River Tracing (DRT) algorithm (Wu et al., 2011, Wu et al. 2012). The DRT algorithm produces the topographic parameters in a scale-consistent way to preserve/upscale the key features of a baseline high-resolution hydrography dataset at multiple coarser spatial resolutions. In CLM5 the MOSART baseline high-resolution hydrography dataset is the 1km resolution hydrological data and maps based on HydroSHEDS (Lehner and Doll, 2004, Lehner et al., 2008) as described in Li et al. (2015b). Soil properties and MOSART river transport data are detailed in Chapter 4.

The fire model data was prescribed at the grid cell level from the same data as CLM4.5. Global annual population density at 0.5x0.5 degree resolution for the 1850-2100 time period came from the decadal population density data for 1850–1980 from the Database of the Global Environment

version 3.1 (HYDEv3.1) (Klein Goldewijk 2011) and population density data for 1990, 1995, 2000, and 2005 from the Gridded Population of the World version 3 dataset (GPWv3) (CIESIN, 2005). Gross Domestic Production (GDP) per capita in 2000 at 0.5x0.5 degree resolution came from van Vuuren et al. (2006), derived from country-level World Bank's World Development Indicators (WDI) measured in constant 1995 US dollars (World Bank, 2004) and the UN Statistics Database (UNSTAT, 2005). The peatland area fraction at 0.5x0.5 degree resolution was derived from datasets: peatland data in Indonesia and Malaysian Borneo (Olson et al. 2001); peatland data in Canada (Tarnocai et al. 2011); and bog, fen and mire data in boreal regions (north of 45N) outside Canada provided by the Global Lakes and Wetlands Database (GLWD) (Lehner and Doll, 2004). The climatological peak month for agricultural waste burning is from van der Werf et al. (2010). The fire model data are detailed in Chapter 4.

Table 1.1: Complete List of Source Data for all components of CLM5.

Data Set	Source Data	Years	Resolution	Description
Land Sea Mask	MODIS IGBP Land Cover - MCD12Q1 Version 5.1	2001 - 2011	1 km	Data compiled from tiles downloaded from the NASA and USGS Land data Products Distributed Active Archive Center (LP DAAC) website (lpdaac.usgs.gov), mosaicked into global geographic 1km resolution grids. Inland and ocean water bodies used the quality product to identify deep, shallow and coastal ocean water.
	CISM2 Glacier land unit extent	2012	1 km	Glacier extent is taken from the Randolph Glacier Inventory version 1.0 (RGIv1.0: Arendt et al. 2012). Greenland Ice Sheet provided by Frank Paul and Tobias Bolch (University of Zurich: Rastner et al. 2012). Antarctic Ice Sheet data provided by Andrew Bliss (University of Alaska) extracted from the Scientific Committee on Antarctic Research (SCAR) Antarctic Digital Database version 5.0.
Current Day Plant Functional Type (PFT)	MODIS IGBP Land Cover - MCD12Q1 Version 5.1	2001 - 2011	1 km	Data compiled as described above. Used for modifying Tree Cover as well as assignment of Non Tree PFTs into Shrub and Grass.
	MODIS Vegetation Continuous Fields - MOD44B Version 5.1	2000 - 2015	1 km	Data compiled from tiles downloaded from the NASA and USGS Land data Products Distributed Active Archive Center (LP DAAC) website (lpdaac.usgs.gov), mosaicked into global geographic 1km resolution grids. Vegetation Cover Fraction Modified by IGBP Land Cover Type using Google Earth and MODIS LAI correction methods.
	AVHRR Continuous Fields Tree Cover Project	2000	1 km	Tree Morphology data of Defries et al. (2000b) is used to specify the VCF modified tree percentage into Needleleaf and Broadleaf types, as well as the Evergreen and Deciduous components.
	EarthStat 2000, MIRCA 2000 and FAOSTAT products	2005	0.0833 deg	Combined all Crop Functional Type (CFT) Data for 2005 to give Crop Percentage as described later.
	CRU 3.24.01 surface air temperature and precipitation	2000 - 2015	0.25 deg	The PFT physiology and climate rules of Bonan et al (2002b), developed from Nemani and Running (1996), are used to split the tree, shrub and grass PFTs into tropical, temperate and boreal climate groupings.

	MODIS 8 Day Leaf Area Index - MCD15A Version 5.0	2003 - 2015	1 km	MODIS LAI is combined with CRU monthly temperature and precipitation data to determine fractional C3 and C4 grass mapping. MODIS LAI is also used to modify Bare Soil and Non Tree Cover.
Current Day Crop Functional Type (CFT)	MIRCA 2000	2000	0.0833 deg	MIRCA 2000 present-day irrigated and rainfed crop areas dataset of Portmann et al. (2010) consist of global area mapping for all major food crops of the world, as well as fodder grass and other perennial, annual and fibre crops.
	EarthStat 2000	2000	0.0833 deg	EarthStat data are generated for the year 2000, based on the work of Monfreida et al. (2008) and Ray et al. (2012).
	UN FAOSTAT	1961 - 2016	Country	Annual country level UN FAOSTAT crop harvest, yield and irrigation data was used to scale up or down the EarthStat 2000 1 km data to achieve the FAOSTAT target values for each country, while maintaining the relative spatial distributions of each value.
Glaciers	CISM2 Glacier land unit extent	2012	1 km	Data compiled as described above. Floating ice is only provided for the Antarctic and does not include the small area of Arctic ice shelves. High spatial resolution vector data were used to determine the area of glacier, ice sheet and floating ice within 30-second grid cells globally.
	CISM2 Glacier region identifier	2012	1 km	The Glacier land unit is also provided with a region identifier for the CISM2 ice sheet model for: 1. Greenland; 2. Greenland Surrounds for potential ice sheet growth; 3. Antarctica; and 4. All other glaciers.
	CISM2 Glacier elevation class	2012	1 km	Current day distribution of the elevation classes is prescribed from ice sheets and Antarctic ice shelf masks combined with the GLOBE topography to generate the 10 ice-covered areas (Arendt et al. 2012, Rastner et al. 2012, Hastings et al. 1999).
Lakes	Lake extent	2000	1 km	Lake extent is taken from the Global Lake and Wetland Database (GLWD) of Lehner and Doll (2004)within 30-second grid cells globally.
	Lake depth	2010	1 km	Mean lake depth for each grid cell is calculated based on the global gridded data sets of Kourzeneva (2012).
Urban	Urban density class extent	2005	0.05 deg	Present day global urban density class extent is mapped from from the LandScan 2004 population density dataset following from Jackson et al. (2010) for the four classes of tall building district (TBD), as well as high, medium, and low density (HD, MD, LD) urban developments. Low density not used in CLM5.
	Urban region	2005	0.05 deg	The urbane density classes are mapped for 33 distinct regions across the globe as described by Jackson et al. (2010).
	Urban properties	2005	Region	The density classes have average building properties provided for 33 distinct regions across the globe as described by Jackson et al. (2010).
Soils	Soil Depth	2015	0.0833 deg	Spatially explicit soil thickness data from Pelletier et al., (2016) as described in Brunke et al. (2016) and Swenson and Lawrence, (2015).

	Soil Organic Matter	2005	0.0833 deg	Majority of the globe from the ISRIC-WISE (Batjes, 2006) mapping, with high latitudes mapping from the 0.25 degree version of the Northern Circumpolar Soil Carbon Database (Hugelius et al. 2012).
	Soil Mineral Composition	2000	0.0833 deg	IGBP soil dataset (Global Soil Data Task 2000) with the 4931 soil classes mapped to depth varying soil profiles of sand and clay.
Topography	Elevation	1996	Various Res Lat - Lon pts	Average Elevation calculated from HYDRO1K dataset (Verdin and Greenlee 1996) from USGS.
	Slope	1996	Various Res Lat - Lon pts	Average Slope calculated from HYDRO1K dataset (Verdin and Greenlee 1996) from USGS.
	Fmax	1996	0.125 deg	Fmax value is a scalar showing the maximum fraction of inundation. The 1-km compound topographic indices (CTIs) based on the HYDRO1K dataset (Verdin and Greenlee 1996).
Rivers	MOSART River Network	2000	0.5 deg	Dominant River Tracing (DRT) algorithm from Wu et al. (2011) and Wu et al. (2012) using the baseline high-resolution hydrography dataset is the 1 km resolution HydroSHEDS data of Lehner and Doll (2004) and Lehner et al. (2008).
	MOSART Channel Slope	2000	0.5 deg	Dominant River Tracing (DRT) algorithm from Wu et al. (2011) and Wu et al. (2012).
	MOSART Channel Roughness	2000	0.5 deg	Dominant River Tracing (DRT) algorithm from Wu et al. (2011) and Wu et al. (2012).
PFT Leaf and Stem Area Index Values	Current Day Plant and Crop Functional Types (PFTs and CFTs)	2005	1 km	Data compiled as described above. The PFT Monthly LAI and SAI values are generated following updated methods from Lawrence and Chase (2007).
	MODIS 8 Day Leaf Area Index - MCD15A Version 5.0	2003 - 2015	1 km	Data compiled as described above.
	CRU 3.24.01 surface air temperature and precipitation	2000 - 2015	0.25 deg	Data compiled as described above.
PFT Canopy Height	Current Day Plant and Crop Functional Types (PFTs and CFTs)	2005	1 km	Data compiled as described above. The new PFT Canopy Height values are generated following updated methods from Lawrence and Chase (2007).
	ICESAT canopy height	2003 - 2015	1 km	The relative PFT canopy heights used to disaggregate the observed by the ICESAT for the grid cell.
Soil Color	Current Day Plant and Crop Functional Types (PFTs and CFTs)	2005	1 km	Data compiled as described above. The Soil Color values are generated following updated methods from Lawrence and Chase (2007).
	MODIS 8 Day Albedo - MCD43A2 Version 5.1	2003 - 2015	1 km	Monthly Climatology Visible and NIR solar noon albedo were compiled from the USGS DAAC. Albedo values were used to optimize soil colors with offline version of the two stream radiation model with current day PFTs/CFTs,

				snow fraction, CLM4.5 soil moisture, and downwelling solar.
	MODIS 8 Day Snow Fraction - MCD43A3 Version 5.1	2003 - 2015	1 km	Monthly Climatology Snow cover fraction were compiled from the USGS DAAC.
	CLM4.5 Soil Moisture	1985 - 2004	0.9 deg	Monthly Climatology Data generated at 0.9x1.25 degree resolution from CMIP5 historical run interpolated to 1km.
	CLM4.5 Downwelling Solar	1985 - 2004	0.9 deg	Monthly Climatology Data generated at 0.9x1.25 degree resolution from CMIP5 historical run interpolated to 1km.
Fire Model	Population	1850 - 2100	0.5 deg	Annual global population density maps from the Database of the Global Environment version 3.1 (HYDEv3.1) (Klein Goldewijk 2011), the Gridded Population of the World version 3 dataset (GPWv3) (CIESIN, 2005), and CMIP6 Shared Socio-Economic Projection (SSP) global population maps for are used for the period 2015 – 2100.
	Gross Domestic Product	2000	0.5 deg	GDP data for IPCC-SRES and derived from country-level World Bank's World Development Indicators (WDI) measured in constant 1995 US dollars (World Bank, 2004) and the UN Statistics Database (UNSTAT, 2005). Projection (SSP) global population maps for are used for the period 2015 – 2100.
	Peatland Area	2000	0.5 deg	The global peatland data was generated from data in Indonesia and Malaysian Borneo from Olson et al. (2001), data in Canada from Tarnocai et al. (2011), and bog, fen and mire data in boreal regions (north of 45N) outside Canada provided by the Global Lakes and Wetlands Database (GLWD) of Lehner and Döll (2004).
Volatile Organic Compounds	Isoprene Emission Factors	2005	0.5 deg	Isoprene emission factors are globally mapped for PFT groups at the 0.5x0.5 degrees resolution. These global maps are taken from the data of Guenther et al. (2012).
Methane	Monthly Inundated Factors	1993 - 2000	0.25 deg	The Three Inundation Factors (Fi) from Prigent et al. (2007): are Maximum Inundation Factor (F0), the seasonal Surface Runoff Factor (P3), and the Water Table Depth Factor (ZWT0).
Transient Plant Functional Types	Current Day Plant Functional Types (PFTs)	2005	0.25 deg	Current Day PFTs compiled as described above. The Global PFT distributions from 2005 are combined with the LUH2 land use time series to generate time series of CLM5 PFTs for each land use class for each grid cell for each year.
	Historical Land Use Harmonization Time Series States	850 - 2015	0.25 deg	LUH2 historical reconstruction from 850 to 2015 is predominantly based on the History of the Global Environment database (HYDE 3.2), with land use derived from historical population densities as detailed in Klein Goldewijk et al. (2017). For the modern period the mapping is steered towards satellite information using the land cover data of the UNFAO agricultural land use (FAO 2020a), the FAO Global Forest Resources Assessment for 2000 (FAO 2000), and Landsat forest change detection from

|				
				Hansen et al. (2013), as described in Hurtt et al. (2020).
	Transient Holocene Land Use Harmonization	6000bce - 850ce	0.25 deg	LUH2 annual land use time series is generated using the first 10 years of the LUH2 historical time series (850 – 859 CE) combined with the HYDE 3.2 population data for the same period. The population data for the Holocene prior to the LUH2 data is then used to scale land use for each grid based on the population of that year relative to the beginning of the LUH2 time period.
	Transient Share Socio-economic Pathway (SSP) Land Use Harmonization	2015 - 2300	0.25 deg	CMIP6 range of future SSPs generated by ScenarioMIP as evolutions of the SSP narratives and their associated challenges for mitigation and adaptation (Riahi et al. 2017) and (O'Neill et al. 2016). The IAMs responsible produce time series of land use for each of the SSPs as described in Popp et al. (2017) harmonized to a common starting point in 2015 as described in Hurtt et al. (2020).
Transient Crop Functional Types	Current Day Crop Functional Types (CFTs)	1961 - 2016	0.25 deg	Current Day CFTs compiled as described above. The Global CFT distributions from 1961 to 2016 are combined with the LUH2 land use time series to generate time series of CLM5 CFTs for each land use class for each grid cell for each year.
	Transient Holocene, Historical and SSP Land Use Harmonization Time Series States	6000bce - 850ce	0.25 deg	LUH2 transient holocene, historical reconstruction and SSP scenarios for annual crop types as described above and in detail in Hurtt et al. (2020).
Shifting Cultivation	Transient Plant and Crop Functional Types	6000bce - 2300	0.25 deg	Transient PFTs and CFTs compiled as described above. The Global PFT and CFT distributions are combined with the LUH2 land use time series to generate time series of CLM5 Gross Unrepresented PFT changes for each land use class for each grid cell for each year.
Transient CFT Irrigation	Transient Crop Functional Types	6000bce - 2300	0.25 deg	Transient CFTs compiled as described above. The Global CFT distributions from 6000bce to 2300 are combined with the LUH2 annual crop type irrigation mapping to produce CLM5 Irrigated CFTs for each crop class for each grid cell for each year.
	Land Use Harmonization Time Series Management	6000bce - 2300	0.25 deg	LUH2 annual Crop Type Irrigation described in Hurtt et al. (2020).
Transient CFT Fertilizer	Current Day Crop Functional Types (CFTs)	6000bce - 2300	0.25 deg	Transient CFTs compiled as described above. The Global CFT distributions from 6000bce to 2300 are combined with the LUH2 annual crop type fertilizer mapping to produce CLM5 CFT Fertilizer for each crop class for each grid cell for each year.
	Land Use Harmonization Time Series Management	6000bce - 2300	0.25 deg	LUH2 annual Crop Type Fertilizer described in Hurtt et al. (2020).
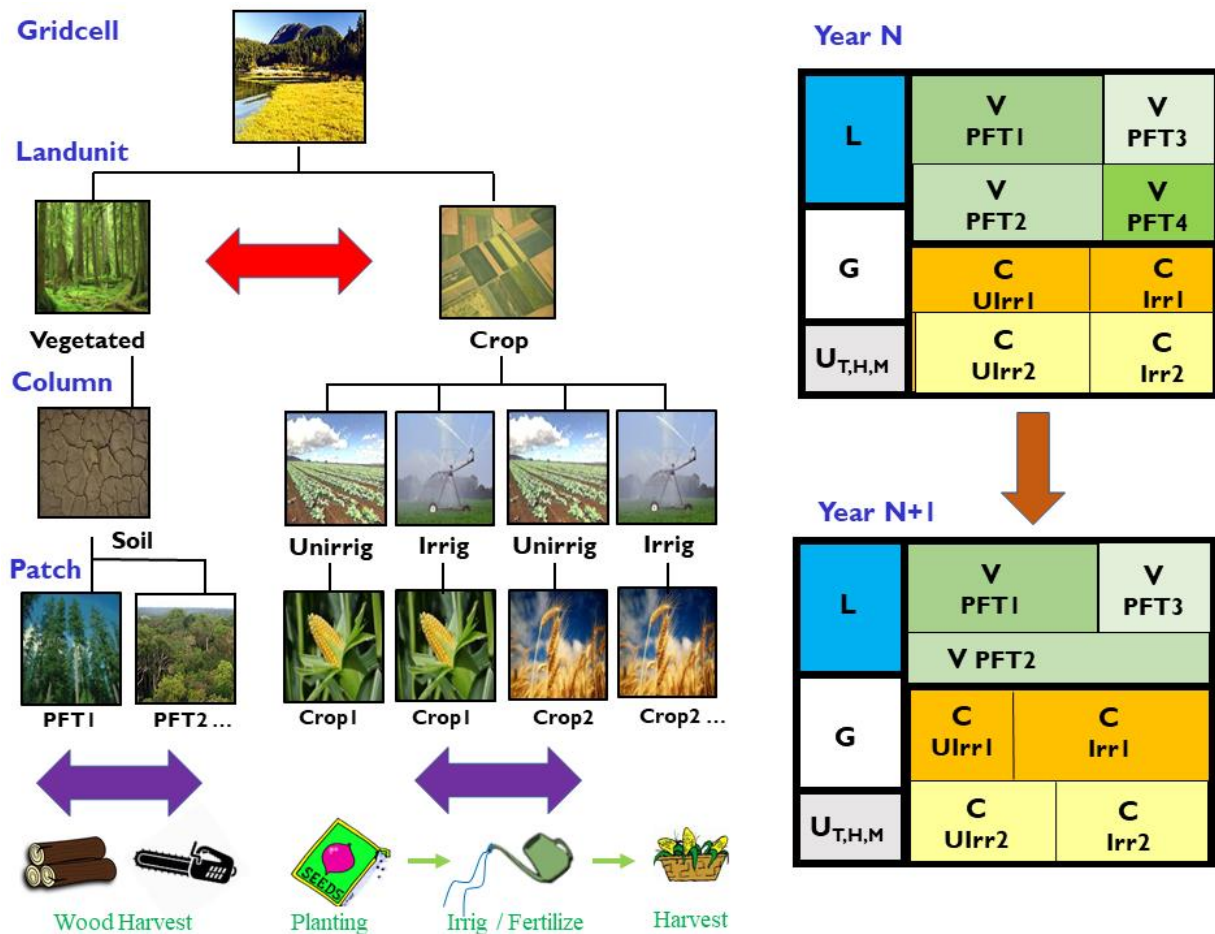
Figure 1.3: CLM5 subgrid hierarchy Land Use and Land Cover Change (LULCC) representations. Boxes on right show hypothetical subgrid transition from a Year N to a Year N + 1. Abbreviations as listed in Figure 1.2. Red arrows indicate allowed land unit transitions. Purple arrows indicate allowed patch-level transitions. Brown arrow indicates combined effective grid cell transitions. Wood harvest and crop model activities shown at bottom.

1.2.3 CLM5 Land Use Land Cover Change Data

The CLM5 introduction of the capability to dynamically adjust land unit weights during a simulation allows for changes in the grid cell area allocated to naturally vegetated and crop land units as well as to changes between PFTs and CFTs within the land units. This capability allows the CLM5 crop model to be run coincidentally with prescribed land use and land cover change as shown in Figure 1.3. This along with the requirement that CESM and CLM5 be able to represent the new CMIP6 LUMIP LUH2 land use time series data (Lawrence et al. 2016, Hurtt et al. 2020) for all CMIP6 simulations led to a full redevelopment of land use and land cover change in the model. The CMIP LUMIP LUH2 land use data is described in detail in Chapter 5.

Annual transient land use data is specified in CLM5 through time varying annual maps of the natural vegetated and crop land units, and the associated PFTs and CFTs of the land units through data on the land use time series file. This is combined with descriptions of the annual land use activities of wood harvest, active crop management through planting, irrigation, fertilizer,

and a not released capability to represent shifting cultivation in the time series file. Figure 1.3 shows the hypothetical two-year transition from year N to N+1. Here the change in land units occur through an increase in the crop and a commensurate decrease in the natural vegetation land unit (red arrows). There are also changes in PFTs within the natural vegetation land unit to slightly increase PFT2 and to remove PFT4 (purple arrow). There are also changes on the crop land unit CFTs with a large increase in the irrigated CFT1 and a smaller increase in the irrigated CFT2 (purple arrow). The combined effective change of the two components gives the areal changes for the grid cell (brown arrow).

The time series file is generated from current day satellite and other current day mapping combined with the CMIP6 LUMIP LUH2 land use time series data detailed in Chapter 5 from Lawrence et al. (2016) and Hurtt et al. (2020). Like the static surface data file much of this technote is devoted to the underlying data and methods required to produce the collection of land use time series files that support LULCC research with the CLM5 and CESM2 models. The generation and evaluation of these time series are the focus of the Chapters 6, 7 and 8. In response to these CMIP6 requirements, CLM5 has new land use transitions for the time periods of: the transient Holocene (6500bce – 849ce), the last millennium (850-1849), the historical period (1850-2014), and eight shared socio-economic pathways (2015-2100), with extension to 2300 for three of them. Each of these time series is evaluated in Chapters 6 and 7 which can be compared with the equivalent LUH2 data from Chapter 5.

1.3 CLM5 Land Data Generation

To support the consistent systematic generation of the CLM5 surface and land use time series data files, a workflow process along with a supporting tools has been developed by NCAR for use with CESM2. This technote details the sources of data and the generation methods used for all components of these files and provides guidance for their applications in modeling the land surface. This section provides an overview of the workflow and tools shown in Figure 1.4. The tools and workflow are further explored through the detailed descriptions found in Chapter 8.

1.3.1 CLM5 Land Use Data Preparation

The first element in the workflow for generating the surface and land use time series files is compiling the best available global current day data descriptions of natural vegetation PFTs, and crop CFTs from satellite and other census products. Collating other representations for glaciers, lakes, soils, drainage, river transport and urban. <to be filled in when complete>

1.3.2 CLM5 Land Use Data Tool

CLM5 Land Use Data Tool takes a CMIP6 LUMIP LUH2 formatted time series data set and processes it with the corresponding descriptions of the LUH2 land use to generate the CLM5 raw data files for each year. Both the LUH2 land use data and the descriptions generated in the data preparation stage can change through the time period. In this manner the amount of c3ann crop, along with the irrigation and fertilizer treatment of the crop can change for a grid cell year to year, but also the CLM5 CFTs that the c3ann crop translates to can change reflecting a hypothetical shift from wheat to cotton. This can also occur with the natural vegetation land unit and PFTs, where a secondary forest can replace pasture, or there can be change in the forest composition from broadleaf deciduous temperate trees to needleleaf evergreen temperate trees. <to be filled in when complete>

1.3.3 CLM5 MkSurfData Tool

Combines the CLM5 raw data files with the glacier, lake and urban data to generate the surface and land use time series files. <to be filled in when complete>
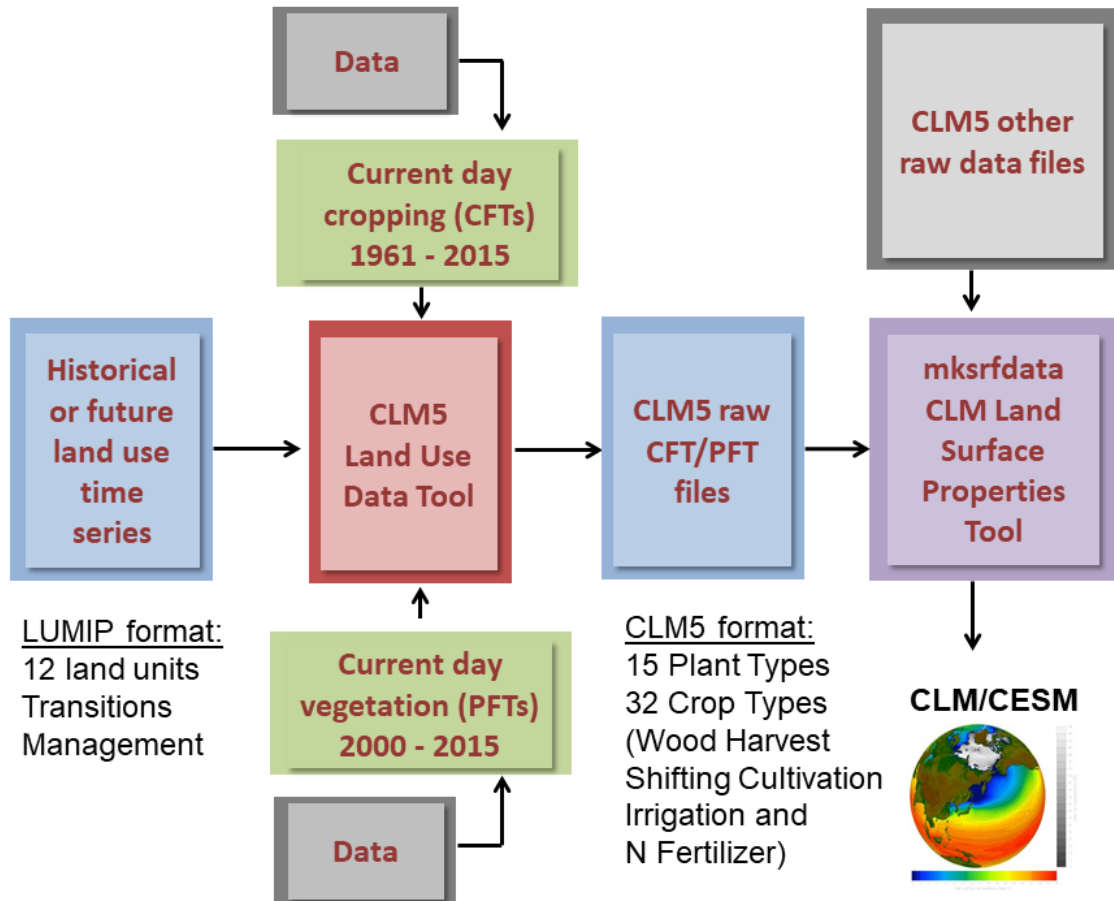

Figure 1.4: CLM5 Land Use and Land Cover Data Generation Tools. Data processing starts with the CMIP6 / LUMIP LUH2 formatted Land Use Time Series which is then combined with PFT and CFT descriptions for Land Units in the CLM5 Land Use Data Tool to generate CLM5 raw data files. These files are combined with other data in the CLM5 mksrfdata tool to generate the surface and land use time series files to be directly read into CLM5.

CHAPTER 2.

CURRENT DAY LAND COVER DATA

2.1 Current Day CLM5 Land Cover Data Overview

Previous versions of CLM developed current day descriptions of the land surface from a range of global satellite, historical cropping and climate products. The methods developed for CLM3.5 by Lawrence and Chase (2007), were updated for CLM4, with new MODIS and climate products, and land cover rules enforced to prevent low tree fractions from the MODIS VCF data in forests after initial low carbon amounts were found in the CLM4CN model. Here we further update the methods of Lawrence and Chase (2007) based on these previous experiences and newly available data products, to produce 1km resolution global CLM5 natural vegetation land unit and plant functional type (PFT) distributions. Updated versions of the CLM5 PFT LAI, canopy height, and soil color data generated in Lawrence and Chase (2007) are detailed in Chapter 4.

Table 2.1: Plant Functional Types (PFTs) included in CLM5.

Patch ID	Plant Function Types (PFTs)	Abbreviations in Technote	
0	Bare Ground	Bare	
1	Needleleaf Evergreen Tree – Temperate	NEM	NdlEvgTemp
2	Needleleaf Evergreen Tree – Boreal	NEB	NdlEvgBorl
3	Needleleaf Deciduous Tree – Boreal	NDB	NdlDecBorl
4	Broadleaf Evergreen Tree – Tropical	BET	BrdEvgTrop
5	Broadleaf Evergreen Tree – Temperate	BEM	BrdEvgTemp
6	Broadleaf Deciduous Tree – Tropical	BDT	BrdDecTrop
7	Broadleaf Deciduous Tree – Temperate	BDM	BrdDecTemp
8	Broadleaf Deciduous Tree – Boreal	BDB	BrdDecBorl
9	Broadleaf Evergreen Shrub – Temperate	SEM	ShrEvgTemp
10	Broadleaf Deciduous Shrub - Temperate	SDM	ShrDecTemp
11	Broadleaf Deciduous Shrub – Boreal	SDB	ShrDecBorl
12	C3 Arctic Grass	G3A	GrsC3Arc
13	C3 Grass	GC3	GrsC3
14	C4 Grass	GC4	GrsC4
15	Unmanaged C3 Crop	Crop	

CLM5 vegetated surfaces are comprised of 15 possible PFTs plus bare ground (Table 2.1) with the individual PFTs identified in the model through the unique Patch IDs. The contribution of each PFT is prescribed as a percentage of the naturally vegetated land unit which itself is calculated as the remainder after glaciers, lakes, urban and crop land units are accounted for from the land fraction of a grid cell. The crop land unit has an additional 64 crop functional types (CFTs) that are added to the Patch ID list when active (Table 3.1). The CLM5 crop model requires detailed individual crop distributions and management data that go beyond the methods developed for

previous representations of the PFTs. A separate but integrated process is explained for the development of the CLM5 current day agriculture through crop CFT descriptions in Chapter 3.

The CLM5 plant types play an important role in describing the land surface as they change the surface properties through changing radiative, hydrological, photosynthetic and surface roughness characteristics. These differences are represented through the PFT level prescription of different leaf and stem optical properties that determine reflection, transmittance, and absorption of solar radiation, root distribution parameters that control the uptake of water from the soil, aerodynamic parameters that determine resistance to heat, moisture, and momentum transfer, and photosynthetic parameters that determine stomatal resistance, photosynthesis, and transpiration. The composition and abundance of PFTs within a grid cell can either be prescribed as time-invariant fields from the surface data set file or can evolve through time with transient land use and land cover change from a land use time series file (Chapters 7 and 8).

2.2 MODIS Land Cover Data

The first stage in the current day land surface mapping of Lawrence and Chase (2007) is to generate the MODIS IGBP global 1km map. The IGBP map determines the common land mask as well as the basis for integrating the other land surface and climate products. Following from the methods of Broxton et al. (2015) we developed the IGBP map as the climatology of the MCD12Q1 Version 5.1 maps for the years 2001 – 2011. The data was compiled from tiles downloaded from the NASA and USGS Land data Products Distributed Active Archive Center (LP DAAC) website (lpdaac.usgs.gov), mosaicked into global geographic 1km resolution grids. To differentiate between inland and ocean water bodies we used the quality product to identify deep, shallow and coastal ocean water.

The IGBP climatology produced a most dominant recorded land cover type for each 1km grid cell for the time period along with a secondary land cover type. Both the primary and secondary land cover types and the fraction of primary land cover from the 11 year climatology are shown in Figure 2.1. The global and IPCC assessment regional areas (described in Chapter 4) for each land cover type are compiled in Table 2.2 and graphed in Figure 2.3a from the grid cell primary land cover type. The fractional compositions of the land cover classes are compiled for MODIS VCF and corrected VCF x AVHRR vegetation cover in Table 2.3 and for current day 1km CLM5 PFTs in Table 2.6.
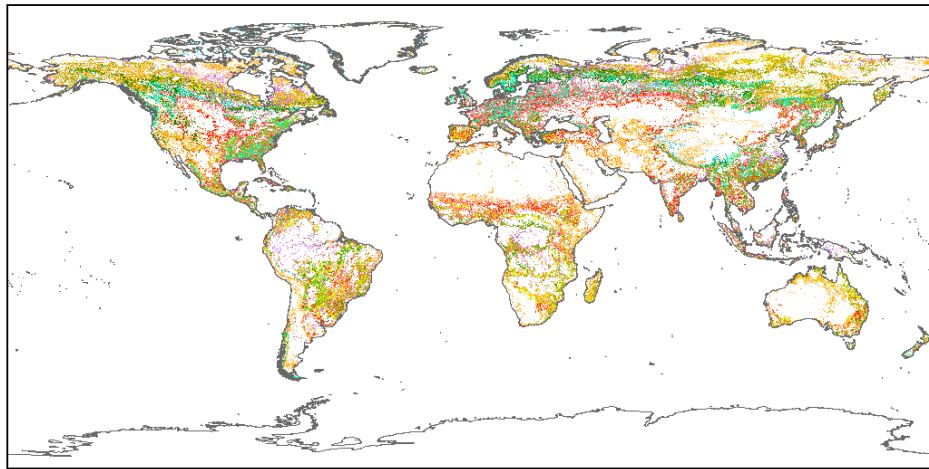
While the analysis of Broxton et al. (2015) provides no area amounts, the compiled global 1km IGBP land cover climatology for CLM5 shows very strong agreement as shown in Figure 2.1a compared to Figure 3 of their paper. This is to be expected given that the same source data is used and the climatology methods are similar, with CLM5 having an additional year of data. The secondary land cover map of Figure 2.1b and the primary fraction of the climatology in Figure 2.1c show that the land cover classification is consistent between the first and second land cover types with strong agreement between the general land cover classes. For example if the primary land cover type is Mixed Forest then the secondary type is often Evergreen Needleleaf Forest, or if the primary land cover type is Crop Mosaic the secondary type is often Cropland.

This level of agreement between the primary and secondary land cover types ensures that the classification disagreement between years has minimal impact on the final CLM5 PFT mapping which includes multiple other data sources. The IGBP land cover mapping is also used for the analysis framework for the MODIS VCF, AVHRR and CLM5 current day PFT mapping.

(a) Primary MODIS IGBP Land Cover

Legend:
- Water
- Evg Ndl For
- Evg Brd For
- Dec Ndl For
- Dec Brd For
- Mixed Forest
- Close Shrub
- Open Shrub
- Woody Sav
- Savanna
- Grassland
- Wetland
- Cropland
- Urban
- Crop Mosaic
- Snow Ice
- Barren
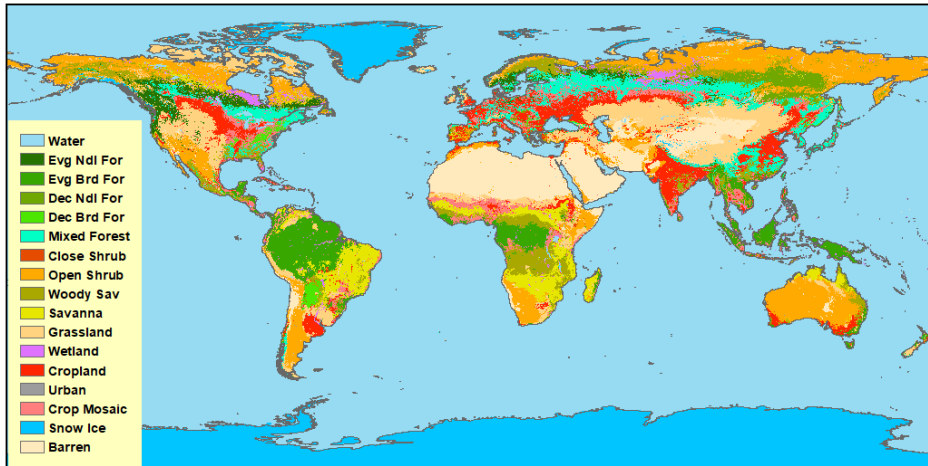
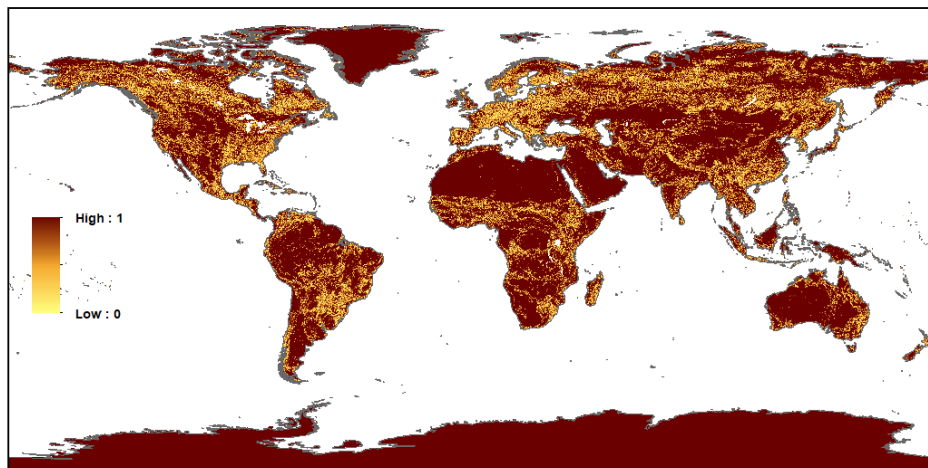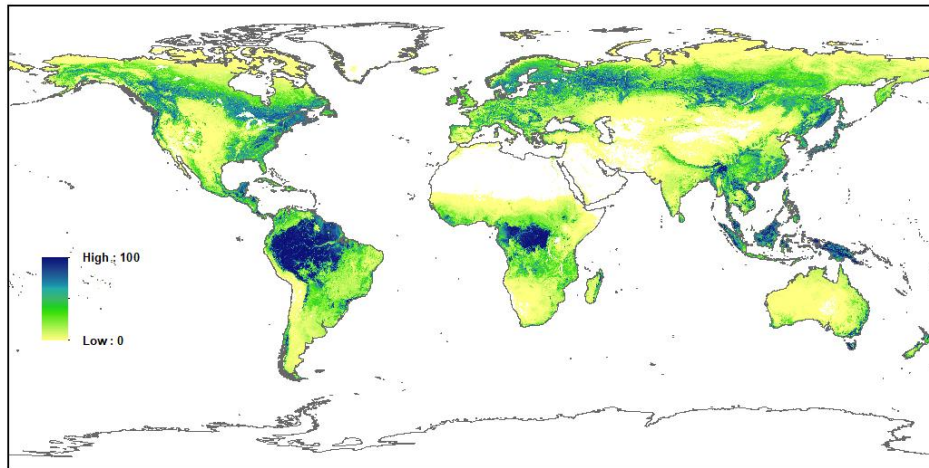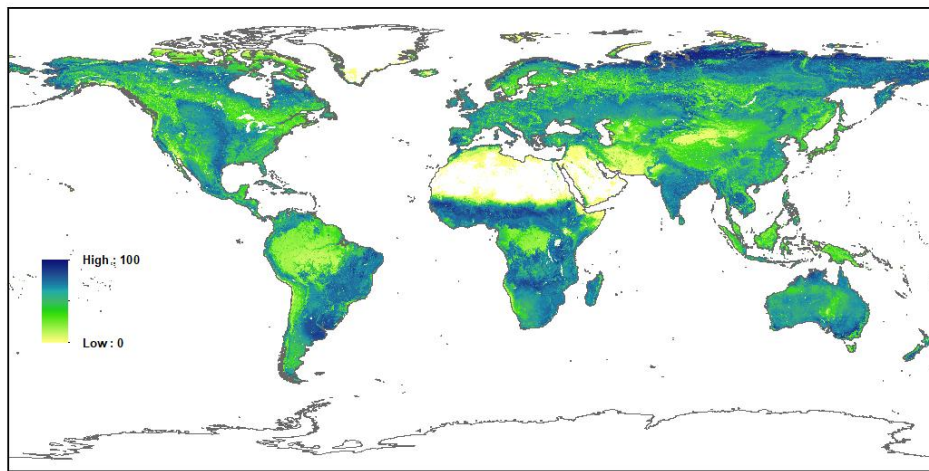## (b) Secondary Land Cover

(c) Primary Fraction Land Cover

High : 1

Low : 0

Figure 2.1: MODIS IGBP Land Cover Climatology Mapping 2001 – 2011 for: (a) Primary Land Cover Class; (b) Secondary Land Cover Class; and (c) Primary Land Cover Class Fraction.

(a) MODIS VCF Percent Tree

(b) MODIS VCF Percent Non Tree (Herbaceous)

(c) MODIS VCF Percent Non Vegetated (Bare)
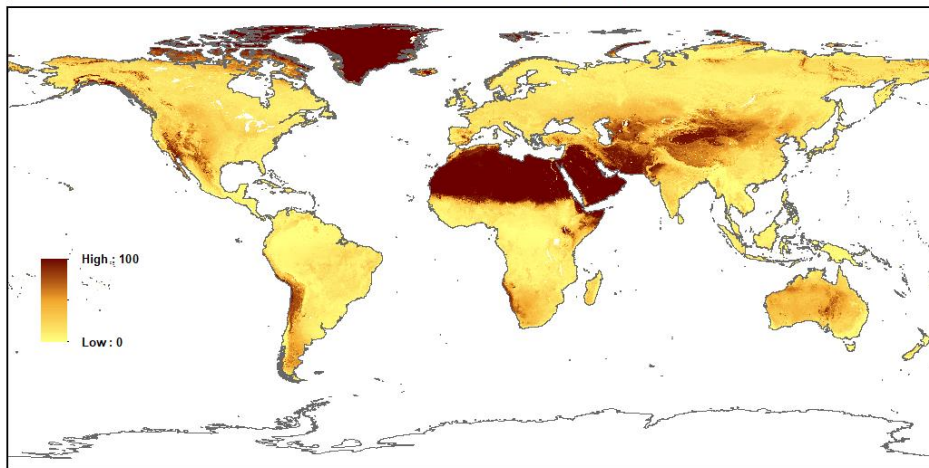
Figure 2.2: MODIS MOD44B Vegetation Continuous Fields (VCF) Climatology for 2000 – 2015 for: (a) Percent Tree; (b) Percent Non-Tree; and (c) Non-Vegetated.

Table 2.2: MODIS IGBP Primary Land Cover Global and IPCC Region Area in millions km^2.

IGBP Class	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Ocean	362.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
Land	148.1	30.0	8.3	11.2	21.2	6.4	20.5	5.2	21.5	5.0	5.1	12.1
Lake	2.8	0.2	0.0	0.1	0.3	0.1	0.1	0.0	0.9	0.0	0.0	0.0
Forest	27.7	2.5	0.6	1.8	5.6	1.4	7.9	0.0	4.4	3.1	0.4	0.0
EvgNdlFor	3.1	0.0	0.0	0.0	0.5	0.4	0.0	0.0	2.1	0.0	0.0	0.0
EvgBrdFor	13.2	2.4	0.3	0.2	0.0	0.0	7.0	0.0	0.0	3.0	0.2	0.0
DecNdlFor	2.0	0.0	0.0	0.0	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DecBrdFor	1.4	0.0	0.0	0.1	0.1	0.1	0.6	0.0	0.4	0.0	0.0	0.0
MixedFor	8.0	0.0	0.3	1.4	3.1	0.9	0.2	0.0	1.8	0.1	0.2	0.0
CloseShrub	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
OpenShrub	20.7	2.7	5.0	0.1	5.8	0.3	2.3	0.6	3.5	0.0	0.5	0.0
WoodySav	10.5	4.0	0.6	0.5	1.2	0.6	1.1	0.0	1.6	0.4	0.4	0.0
Savanna	9.8	4.6	0.7	0.0	0.2	0.0	3.8	0.0	0.4	0.0	0.0	0.0
Grassland	18.3	2.2	0.5	4.0	3.8	0.9	1.8	0.3	4.5	0.0	0.3	0.0
Wetland	2.3	0.2	0.0	0.1	0.7	0.0	0.4	0.0	0.6	0.2	0.0	0.0
Cropland	12.5	1.0	0.6	1.8	1.8	2.1	1.1	0.2	1.7	0.2	2.1	0.0
Urban	0.5	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0
CropMosaic	8.6	2.4	0.0	0.3	1.2	0.8	1.3	0.0	1.2	1.0	0.4	0.0
SnowIce	15.3	0.0	0.0	0.1	0.1	0.1	0.0	0.0	2.6	0.0	0.1	12.1
Barren	19.0	10.2	0.1	2.5	0.6	0.0	0.6	4.1	0.1	0.0	0.8	0.0

The area calculations for the CLM5 MODIS IGBP primary land cover classes are shown in Table 2.2 and Figure 2.3a. The global climatology has a total surface area for the Earth of 510.1 million km^2, with the ocean covering 362.0 million km^2 and the land covering 148.1 million km^2. The ocean and land areas differ slightly from other global estimates due to the 1km mapping resolution and the definition between land ice and sea ice in MODIS relative to other mapping efforts. In that sense the MODIS IGBP land cover mapping is consistent with but not identical to the ESRI WGS84 Countries of the world polygon shape file.

The leading land area in the MODIS IGBP mapping is forests with an area of 27.7 million km^2, which is followed by crop and crop mosaic with a combined area of 21.1 million km^2, shrublands with 20.8 million km^2, savannas with 20.3 million km^2, grassland with 18.3 million km^2, barren with 19.0 million km^2, and snow and ice with 15.3 million km^2. The other land cover types make up the remaining land area with lakes having an area of 2.8 million km^2, wetlands having 2.3 million km^2, and urban having 0.5 million km^2.

Globally forests are dominated by Evergreen Broadleaf Forests with 13.2 million km^2 and Mixed Forests with 8.0 million km^2. There are also smaller areas of Evergreen Needleleaf Forest at 3.1 million km^2, Deciduous Needleleaf Forest at 2.0 million km^2, and Deciduous Broadleaf Forest at 1.4 million km^2. Evergreen Broadleaf Forests are found predominantly in Latin America (Amazon), Southeast Asia, and Africa (Congo). Mixed Forests are found in Eurasia, North America, East Asia, and Europe. Cropland and Crop Mosaic are distributed throughout the regions. Shrublands are predominantly found in Eurasia, Australia and the Pacific, North America, Africa and Latin America. Savannas are found predominantly in Africa, Latin America, and North America. Grasslands are predominantly found in North America (Prairies), East Asia, Eurasia (Steppes), Africa and Latin America (Pampas).
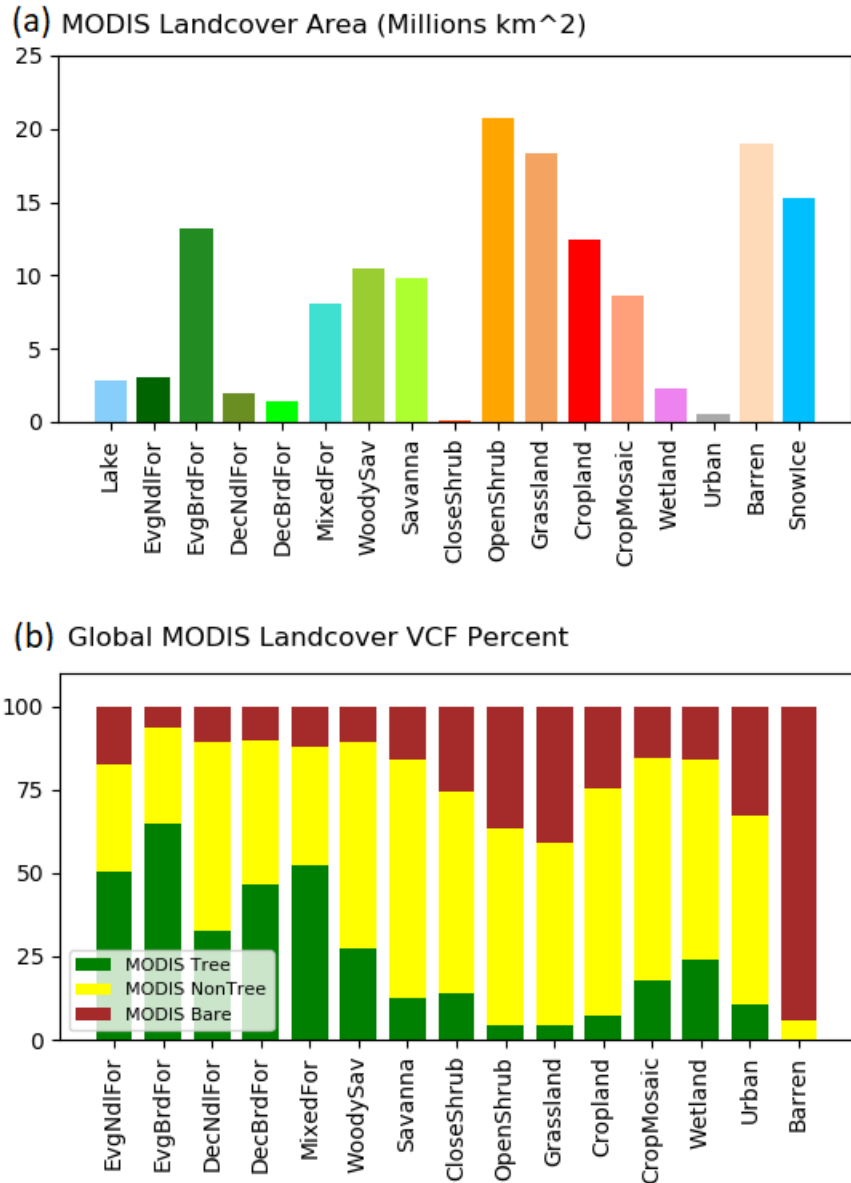
Figure 2.3: (a) Global area of MODIS IGBP Land Cover Mapping 2003 – 2008; and (b) Global average composition of IGBP classes in Tree, Non-Tree and Bare from MODIS Vegetation Continuous Fields 2003 – 2011.

2.3 MODIS Vegetation Continuous Fields Data

The next stage in the current day land surface mapping of Lawrence and Chase (2007) is to generate the global 1km percentages of trees, non-tree (herbaceous), and bare ground maps. In Lawrence and Chase (2007) this is done directly from the MODIS Vegetation Continuous Fields global 1km map. As previously noted from CLM4, new land cover rules were required to prevent low tree fractions from the MODIS VCF data in forests for use in CLM PFT mapping. For CLM5 these rules have been further developed through tree percentage corrections developed from Google Earth inspections of 290 FLUXNET sites. This process identified a systematic biases

between the site investigation and the MODIS VCF product that varied by land cover type as described in the next section.

For the CLM5 current day land cover mapping, the initial global percentage tree, non-tree and bare ground mapping were developed as climatologies of the MOD44B Version 5.1 VCF annual maps for the years 2000 – 2015. The global maps for the percentage of each class are shown in Figure 2.2, with their average composition for each MODIS land cover type shown in Table 2.3 and Figure 2.3b. The data was compiled from tiles downloaded from the NASA and USGS Land data Products Distributed Active Archive Center (LP DAAC) website (lpdaac.usgs.gov), mosaicked into global geographic 1km resolution grids. The VCF maps do not include Antarctica, which is considered all Snow and Ice in the land cover mapping. While there is active land cover change over this period, there is also inter-annual differences in atmospheric conditions and surface conditions from fire and other disturbances that impact the VCF mapping. The VCF climatology period overlaps the MODIS land cover mapping period making them directly comparable as average annual surface descriptions from the early 2000s.

Table 2.3: MODIS IGBP by Percentage Raw VCF, alpha and beta parameters for Tree scaling, and Corrected VCF with MODIS Land Cover and AVHRR Tree Morphology.

IGBP Class	Tree	Herb	Bare	α	β	Ndl Evg	Brd Evg	Ndl Dec	Brd Dec	Herb	Bare
EvgNdlFor	50	32	17	0.66	66.0	97	0	0	0	3	0
EvgBrdFor	65	29	7	0.66	66.0	0	99	0	0	1	0
DecNdlFor	32	57	11	0.66	66.0	0	0	88	0	12	0
DecBrdFor	46	43	10	0.66	73.0	0	0	0	97	3	0
MixedFor	52	36	12	0.66	66.0	64	3	0	31	3	0
CloseShrub	14	60	26	1.0	0.0	7	1	0	6	60	26
OpenShrub	4	59	37	1.0	0.0	2	0	1	2	59	37
WoodySav	27	62	11	1.0	25.0	17	1	1	34	48	0
Savanna	13	72	16	1.0	10.0	2	0	0	20	71	6
Grassland	4	55	41	1.0	0.0	2	0	0	2	55	41
Wetland	24	60	16	1.0	0.0	9	3	1	10	60	16
Cropland	7	68	25	1.0	0.0	2	0	0	5	68	25
Urban	11	57	33	1.0	0.0	4	0	0	6	57	33
CropMosaic	18	67	16	1.0	0.0	4	1	0	13	67	16
SnowIce	0	3	97	1.0	0.0	0	0	0	0	3	97
Barren	0	6	94	1.0	0.0	0	0	0	0	6	94

The global mapping of MODIS land cover in Figure 2.1 and MODIS VCF in Figure 2.2 show good general agreement. The percentage trees correspond well with high tree cover in the tropical forests of the Amazon and the Congo, and to a lesser extent in Southeast Asia. There is also higher tree cover for the needleleaf, broadleaf deciduous, and mixed forests of the boreal and mid-latitude forests of North America, Europe, Eurasia and China. The bare ground mapping also corresponds well with the land cover mapping, with very high values for the Sahara, the Arabian peninsula, Central Asia, Australia and the Southwest of North America.

Following on from the CLM4 findings, closer analysis reveals that there are systematic differences between the MODIS VCF products in forests and savannas than are experienced with detailed inspection of FLUXNET sites with the Google Earth as shown in Figure 2.4 and Figure 2.5. There is also a difference between what is being mapped by the VCF product in terms of projected foliage cover, and what is being modeled by CLM5 in terms of the amount of a particular

vegetation type. These systematic biases result in low average tree values for all of the forest and savanna land cover classes as shown in Table 2.3.

The low tree values in forests are accompanied by high bare ground and non-tree vegetation. For Evergreen Needleleaf Forest this results in an average tree cover of 50%, and for Deciduous Needleleaf Forest a tree cover of just 32%. For Evergreen Broadleaf Forest this is slightly higher at 65%, and for Deciduous Broadleaf Forest 46%. Mixed Forest also has a low tree cover at only 52%. The low tree percentage may be in part the result of different representations for calculating projected foliage cover versus the use of the tree PFTs in CLM5 to represent their relative contribution to the naturally vegetated land unit. A more complete evaluation of the tree representation in the MODIS VCF product and how it is applied in Earth system modeling is beyond this project but would be of great value to the community.

2.4 MODIS VCF Google Earth Correction

Given the limited resources available for correcting the MODIS VCF product we used the Google Earth product to interactively investigate 290 FLUXNET sites based on their latitude and longitude. The MODIS land cover type and VCF fields were extracted from the 1km global maps for each site. This was compared with the highest resolution Google Earth satellite imagery available for the site as well as the street view images at the nearest available location. The percentage tree, non-tree and bare ground percentages along with land cover class were estimated and recorded in a site level database. The imagery along with the satellite and estimated tree cover is shown for selected sites in Figure 2.4. A comparison between the satellite data and the estimated tree percentages separated by land cover class is shown in Figure 2.5a.

Both of these analyses identify that there are systematic biases in the MODIS VCF tree cover data for these sites. The biases vary by land cover type in both their relative slope and offset from the one to one line in Figure 2.5a. To address this systematic bias, the correction Equation 2.1 was applied to all of sites with the α and β parameters specified based on the MODIS land cover class found at the site. The α and β parameters for all land cover classes are listed in Table 2.3.

$$TreePCT_{LC\ CORRECTED} = \alpha_{LC}\ TreePCT_{VCF} + \beta_{LC} \tag{2.1}$$

The result of the correction equation is shown for each land cover class in Figure 2.5b. As would be expected, the corrected tree cover percentages now correspond to the one to one line when compared to the estimated values from the FLUXNET sites. The correction equation was applied to the global MODIS VCF tree cover data using the MODIS land cover class layer to prescribe the α and β parameters. To offset the increase in tree percent, reductions in bare ground and non-tree vegetation were made to ensure that the three values always summed to 100%. The impacts of the tree correction on the average composition of the MODIS IGBP land cover types are shown in Table 2.3 and Figure 2.6a.

The largest changes in average tree composition were in Deciduous Needleleaf Forest where the tree percentage was increased from 32% to 88%. This was followed by Deciduous Broadleaf Forest where the tree percentage was increased from 46% to 97%, and Mixed Forest where the tree percentage was increased from 52% to 95%. The increase in Evergreen Broadleaf Forest was smaller but still substantial going from 65% to 99%. Changes in tree percentage in Woody Savanna was also substantial going from 27% to 51% and Savanna going from 13% to 22%. All other land cover classes were left at the original MODIS VCF values. The impact of the increases in tree cover was to remove bare ground from all forests and the Woody Savanna.
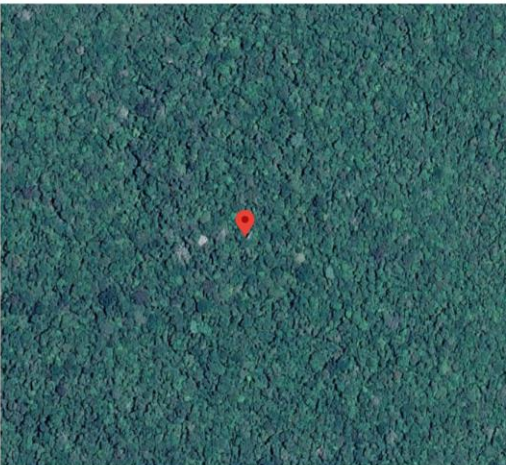
(a) Saskatchewan, Canada - CA-Obs (53.987 -105.118)
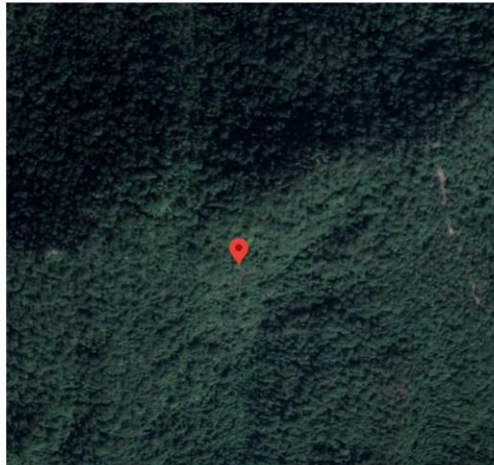IGBP: NdlEvgFor Raw VCF: 56% Corrected VCF: Tree 80%
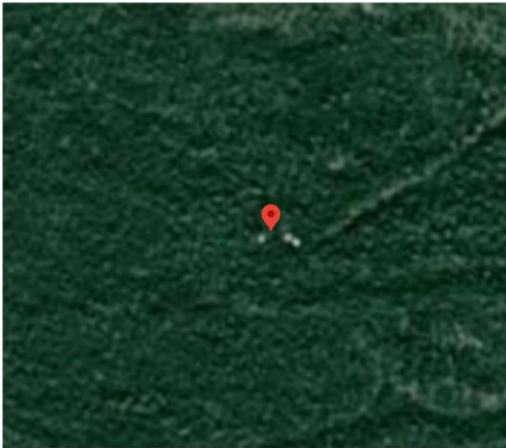
(b) Björklinge, Sweden - SE-Nor (60.086 17.480)
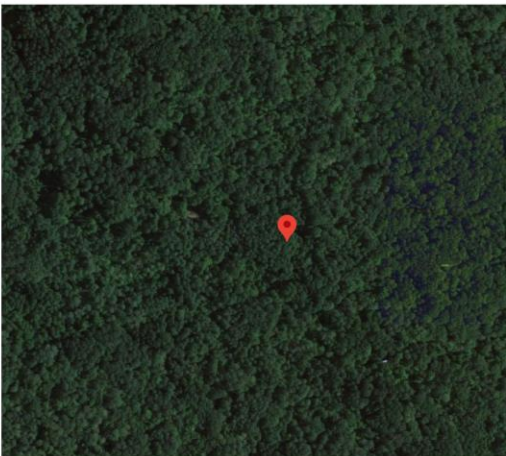IGBP: NdlEvgFor Raw VCF: Tree 60% Corrected VCF: Tree 80%
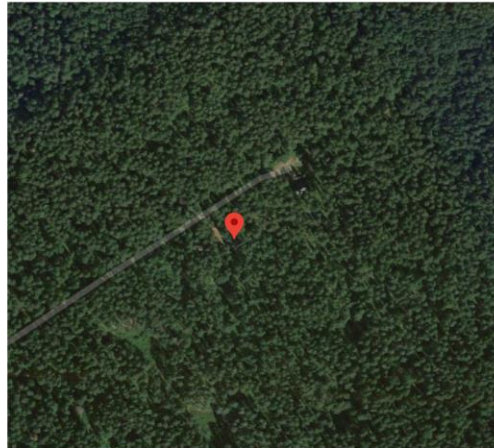
(c) Manaus - State of Amazonas, Brazil - BR-Ma2 (-2.609,-60.209)
IGBP:EvgBrdFor Raw VCF: Tree 77% Corrected VCF: Tree 100%

(d) Dinghu Mountain Nature Reserve, China - CN-Din (23.167,112.533)
IGBP: EvgBrdFor Raw VCF: Tree 53% Corrected VCF: Tree 100%

(e) Munro Township, Michigan, USA - US-UMB (45.560 -84.716)
IGBP: DecBrdFor Raw VCF: Tree 60% Updated VCF: Tree 100%

(f) Athabasca County, Alberta, Canada - CA-WP1 (54.954 -112.467)
IGBP: MixedFor Raw VCF: Tree 45% Corrected VCF: Tree 95%
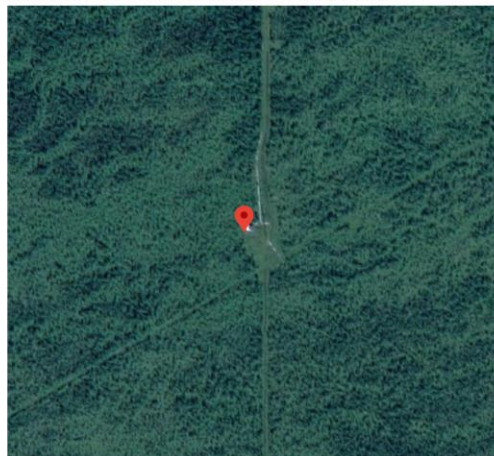
Figure 2.4: Google Earth FLUXNET Images for forest types: (a) Needleleaf Evergreen CA-OBS; (b) Needleleaf Evergreen SE-NOR; (c) Evergreen Broadleaf BR-MA2; (d) Evergreen Broadleaf CN-DIN; (e) Deciduous Broadleaf US-UMB; and (f) Mixed CA-WP1.
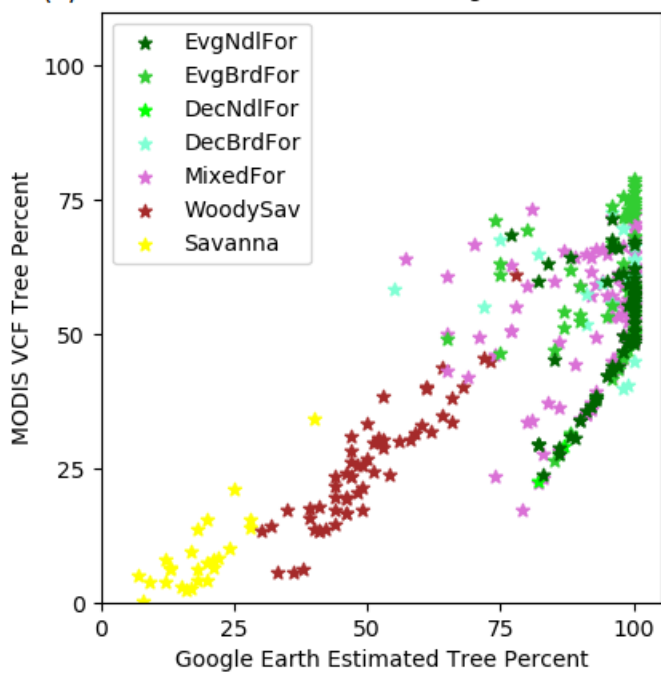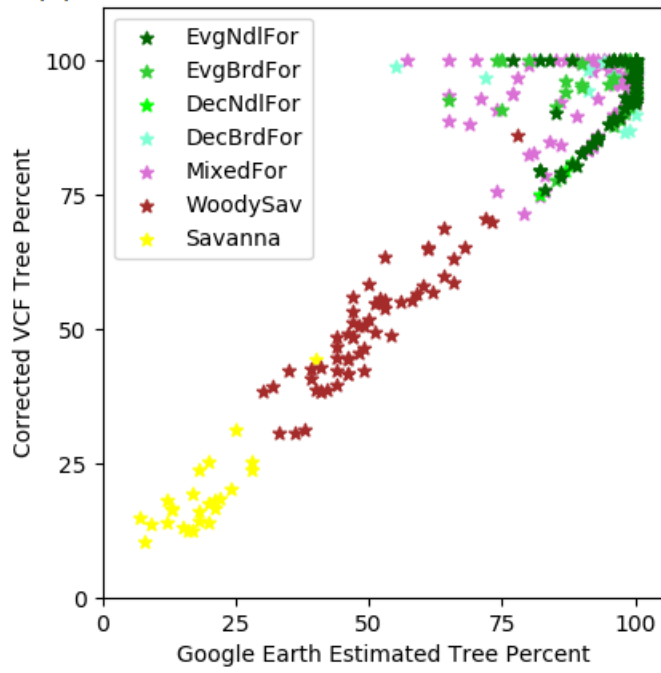
Figure 2.5: Google Earth Estimated Tree Percent for Forest and Savanna Land Cover Types compared to: (a) MODIS VCF Tree Percent; and (b) Corrected VCF Tree Percent.
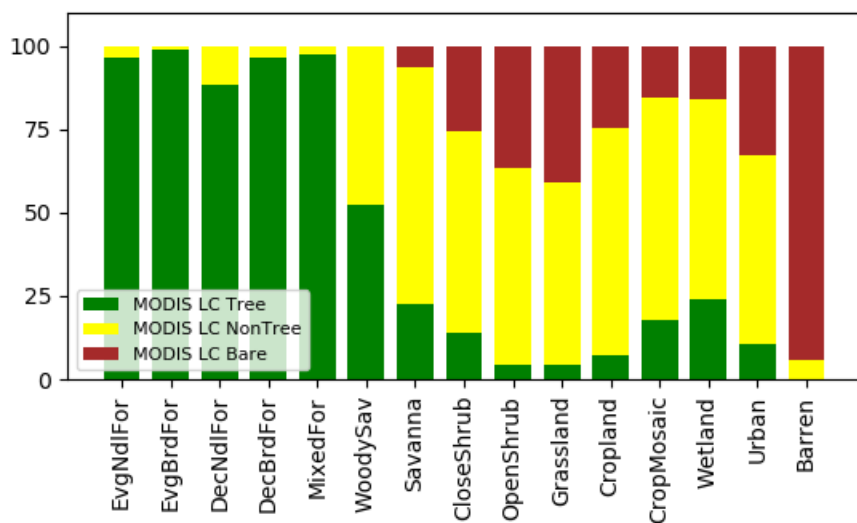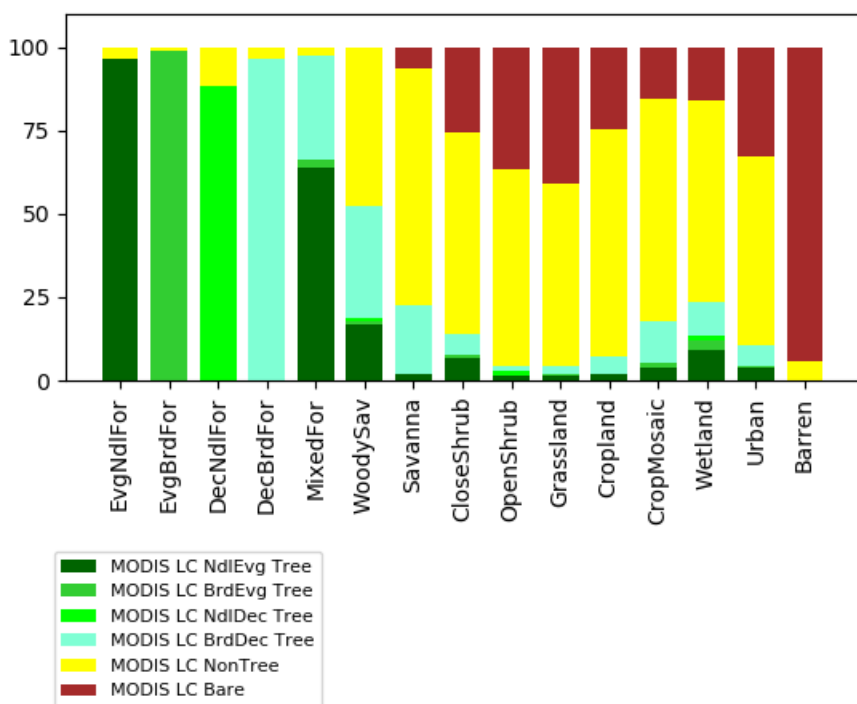
Figure 2.6: Global average composition of IGBP classes for: (a) Tree, Non-Tree and Bare from Corrected Vegetation Continuous Fields data; and (b) Needleleaf Evergreen Tree, Broadleaf Evergreen Tree, Needleleaf Deciduous Tree, Broadleaf Deciduous Tree, Non-Tree and Bare when combined with Land Cover Type and AVHRR Vegetation Continuous Fields data.

2.5 AVHRR Tree Morphology Data

The third stage in the current day land surface mapping of Lawrence and Chase (2007) is to combine the MODIS VCF percentages of trees, non-tree, and bare ground mapping with tree morphology from the AVHRR Continuous Fields Tree Cover Project data of Defries et al. (2000b). The tree morphology data is used to specify the tree percentage into Needleleaf and Broadleaf types, as well as the Evergreen and Deciduous components. For the CLM5 current day land cover mapping we used the corrected VCF data in the place of the MODIS VCF product, and only used the AVHRR morphology data where the tree morphology was not explicit prescribed in the MODIS IGBP land cover class. In this manner, all trees in Evergreen Needleleaf Forest, Evergreen Broadleaf Forest, Deciduous Needleleaf Forest, and Deciduous Broadleaf Forest were allocated the corresponding tree morphology of the land cover class. Only land cover classes outside of these classes were allocated the AVHRR components.

The resulting global maps from combining the corrected VCF, MODIS IGBP land cover, and AVHRR tree morphology data are shown in Figure 2.7 and Figure 2.8. The combined impact on the composition of Evergreen, Deciduous, Needleleaf and Broadleaf tree types, as well as non-tree and bare ground percentages are listed by IGBP land cover type in Table 2.3, and shown in Figure 2.6b. The global mapping shows that the new tree morphology mapping corresponds very well with the MODIS land cover mapping of Figure 2.1.

The Needleleaf Evergreen Tree mapping (Figure 2.7a) corresponds directly with the boreal and mid-latitude Evergreen Needleleaf Forest and Mixed Forest land cover classes. The Broadleaf Evergreen Tree mapping (Figure 2.7b) corresponds directly to the Evergreen Broadleaf Forest of the Amazon, Congo and Southeast Asia. The Needleleaf Deciduous Tree mapping (Figure 2.7c) corresponds to the Deciduous Needleleaf Forest of Siberia. The Broadleaf Deciduous Tree mapping (Figure 2.8a) corresponds with boreal, mid-latitude and tropical Deciduous Broadleaf Forest and Mixed Forest, as well as tropical Woody Savanna and Savanna. These relationships are also reflected in the land cover class compositions in Table 2.3 and in Figure 2.6b.

2.6 Non Tree Vegetation Data

The remaining Non Tree Vegetation (Figure 2.8b) corresponds with the MODIS land cover classes of Cropland, Shrubland, Grassland, and Savannas. Following Lawrence and Chase (2007), the Non Tree Vegetation area is mapped to crop, shrub, and grass depending on the IGBP land cover map and the annual crop area from an independent cropping data set. For CLM3.5 the current day cropping came from the year 1992 of the historical cropping data of Ramankutty and Foley (1999). For CLM4.0 and 4.5 the cropping data set was updated annually from the CMIP5 historical and future RCP time series of Hurtt et al. (2006, 2011).

For CLM5 the initial current day cropping data is taken from the total area of crops combined from the EarthStat 2000, MIRCA 2000 and FAOSTAT products, which also prescribe the current day rainfed and irrigated CFT distributions as detailed in Chapter 3. For the CLM5 with active land use and land cover change, the current day PFTs and CFTs are combined with the CMIP6 historical and projected LUH2 time series data of Hurtt et al. (2020) as described in Chapter 5. For Closed Shrubland the remaining Non Vegetated percentage after accounting for the Crop PFT area is allocated to shrub at a ratio of 0.85 and grass at 0.15. For Open Shrubland the remaining Non Vegetated percentage is allocated to shrub at a ratio of 0.65 and grass at 0.35. For all other land cover types the remaining Non Vegetated percentage is allocated to grass. For all land cover types bare ground is the remainder after all of the other vegetation is accounted.

(a) MODIS VCF x AVHRR Needleleaf Evergreen Tree (%)
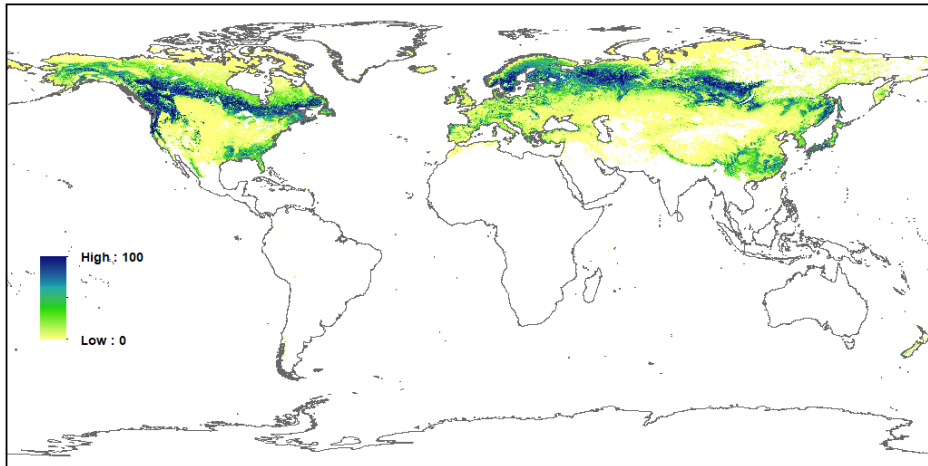
(b) MODIS VCF x AVHRR Broadleaf Evergreen Tree (%)
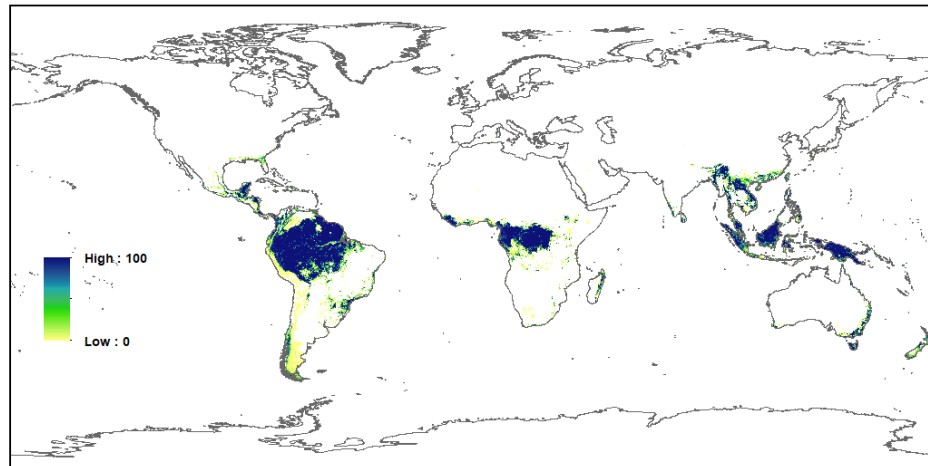
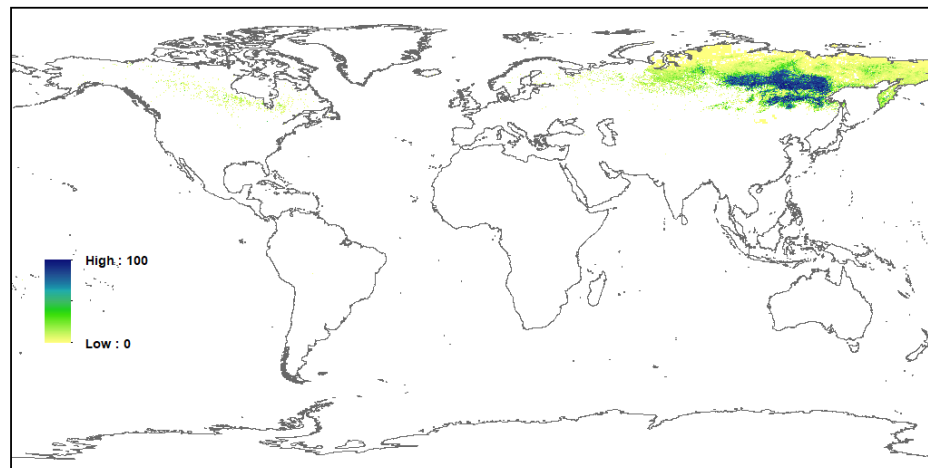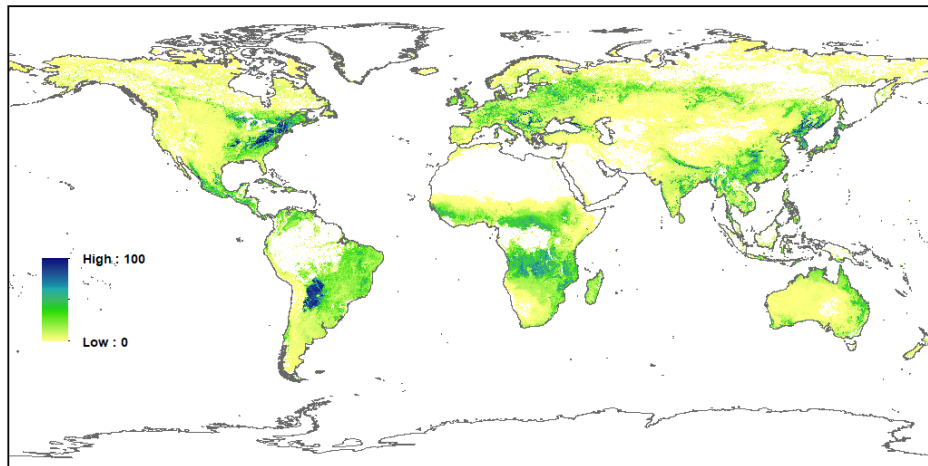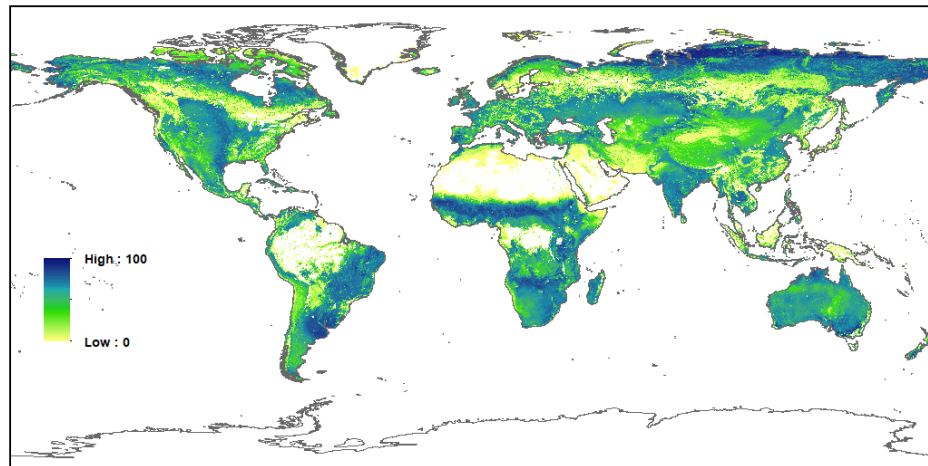## (c) MODIS VCF x AVHRR Needleleaf Deciduous Tree (%)

Figure 2.7: MODIS IGBP Land Cover x Corrected VCF x AVHRR Vegetation Mapping Part 1: (a) Needleleaf Evergreen Tree; (b) Broadleaf Evergreen Tree; and (c) Needleleaf Deciduous Tree.

(a) MODIS VCF x AVHRR Broadleaf Deciduous Tree (%)
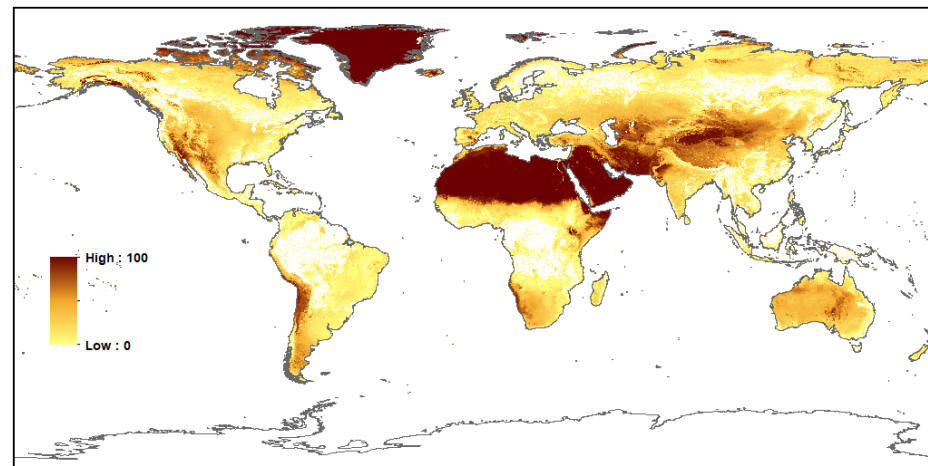
(b) MODIS VCF Non Tree Vegetation (%)

(c) MODIS VCF Non Vegetated (Bare) (%)

Figure 2.8: MODIS IGBP Land Cover x Corrected VCF x AVHRR Vegetation Mapping Part 2: (a) Broadleaf Deciduous Tree; (b) Non Tree Vegetation; and (c) Non Vegetated (Bare).

The current day CLM5 Crop PFT mapping was from the total area of crops from the combined EarthStat 2000, MIRCA 2000 and FAOSTAT products mapped to CLM5 CFTs for the year 2005, as described in Chapter 3. For CLM5 without the crop model option on, the crop area is directly prescribed as the Unmanaged C3 Crop PFT as described in this chapter.

2.7 CRU Temperature and Precipitation Data

Having generated the tree percentage maps for Needleleaf, Broadleaf, Evergreen and Deciduous components along with the, Crop, Shrub, Grass and Bare Ground maps the next stage from Lawrence and Chase (2007) is to develop the bioclimatic PFT mapping. The PFT physiology and climate rules of Bonan et al (2002b), developed from Nemani and Running (1996), are used to split the tree, shrub and grass PFTs into tropical, temperate and boreal climate groupings. The climate rules for each PFT are listed in Table 2.4. For previous versions of CLM these climate rules were calculated using the climatological monthly surface air temperature and precipitation surfaces from Willmott and Matsuura (2000).
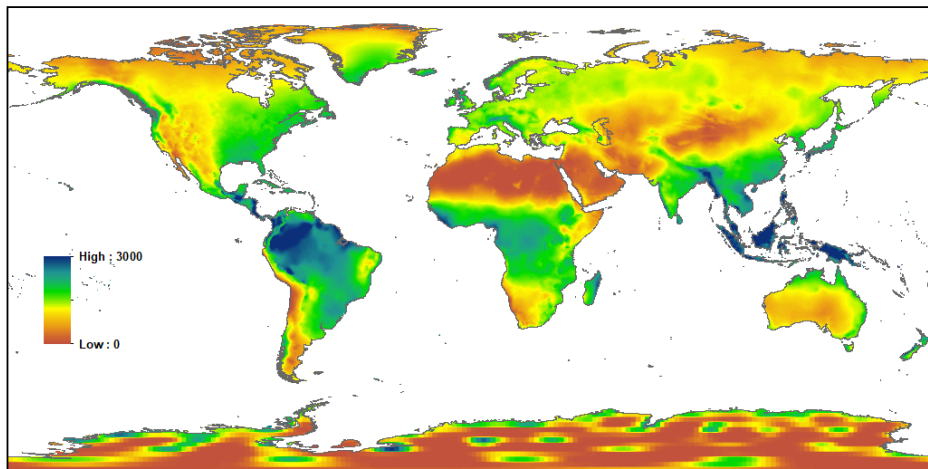
Table 2.4: Plant Functional Type (PFT) Climate and Leaf Area Index (LAI) Rules.

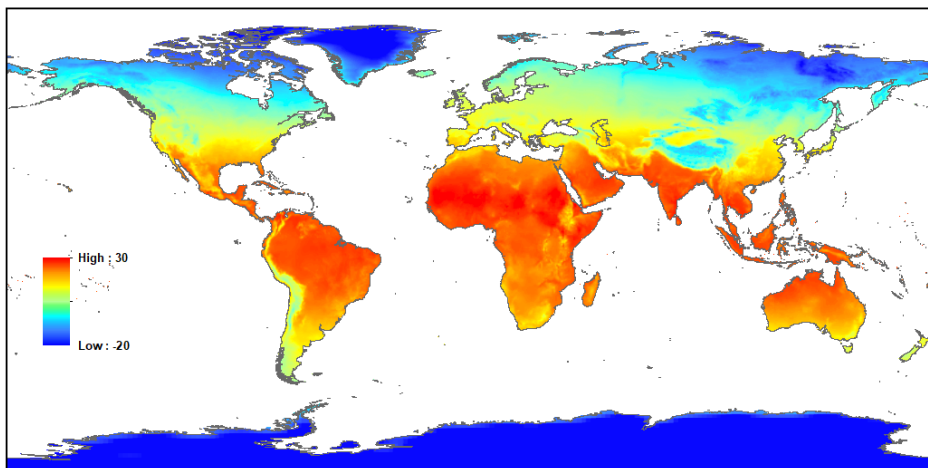| CLM5 Class	Climate Rules	Leaf Area Index (LAI) Rules
Bare		$<= (1.0 - 1.5*LAI_{MAX})*100.0$
NdlEvgTemp	$T_c > -19°C$ and GDD > 1200	
NdlEvgBorl	$T_c <= -19°C$ or GDD <= 1200	
NdlDecBorl		$<= (1.0 - 0.25*LAI_{MIN})*100.0$
BrdEvgTrop	$T_c > 15.5°C$	
BrdEvgTemp	$T_c <= 15.5°C$	
BrdDecTrop	$T_c > 15.5°C$	$<= (1.0 - 0.25*LAI_{MIN})*100.0$
BrdDecTemp	$-15°C < T_c <= 15.5°C$ and GDD > 1200	$<= (1.0 - 0.25*LAI_{MIN})*100.0$
BrdDecBorl	$T_c <= -15°C$ or GDD <= 1200	$<= (1.0 - 0.25*LAI_{MIN})*100.0$
ShrEvgTemp	$T_c > -19°C$ and GDD > 1200 and ($P_{ann} > 520$ mm and $P_{win} > 2/3 P_{ann}$)	
ShrDecTemp	$T_c > -19°C$ and GDD > 1200 and ($P_{ann} <= 520$ mm or $P_{win} <= 2/3 P_{ann}$)	
ShrDecBorl	$T_c <= -19°C$ or GDD <= 1200	
GrsC3Arc	GDD <= 1000	
GrsC3	GDD > 1000; C3 Season T <= 22°C and P > 25mm	$= LAI_{C3\ SUM} / LAI_{ANN\ SUM} * Grass$
GrsC4	GDD > 1000; C4 Season T > 22°C and P > 25mm	$= LAI_{C4\ SUM} / LAI_{ANN\ SUM} * Grass$
Crop		

T_c, temperature of coldest month. GDD, growing-degree days above 5°C. P_{ann}, annual precipitation. P_{win}, winter precipitation. LAI_{MIN}, minimum monthly LAI. LAI_{MAX}, maximum monthly LAI. $LAI_{C3\ SUM}$, cumulative sum of monthly LAI for all months in the C3 growing season. $LAI_{C4\ SUM}$, cumulative sum of monthly LAI for all months in the C4 season. $LAI_{ANN\ SUM}$, cumulative sum of monthly LAI for all months in the year.

For CLM5 the monthly climatology surface air temperature and precipitation have been updated to use the CRU 3.24.01 (Harris et al. 2014) 2000 – 2015 monthly surface air temperature and precipitation data, with Antarctica and other missing land values filled with NCEP/NCAR reanalysis (Kalnay et al. 1996). To use the climate data with the other fine resolution data, the coarse native grids of both products were spatially interpolated to the 1km grid and land mask of the MODIS land cover product.

(a) CRU Average Annual Precipitation (mm/year)   2000 - 2015

High : 3000

Low : 0

(b) CRU Average Annual Temperature (Deg C)   2000 - 2015

High : 30

Low : -20

(c) CRU Climate - Whittaker Biomes   2000 - 2015

Ocean
Tropical Rainforest
Tropical Forest
Tropical Savanna
Sub Tropical Desert
Temperate Rainforest
Temperate Forest
Woodland Shrubland
Temperate Grassland
Taiga
Tundra
Ice and Rock
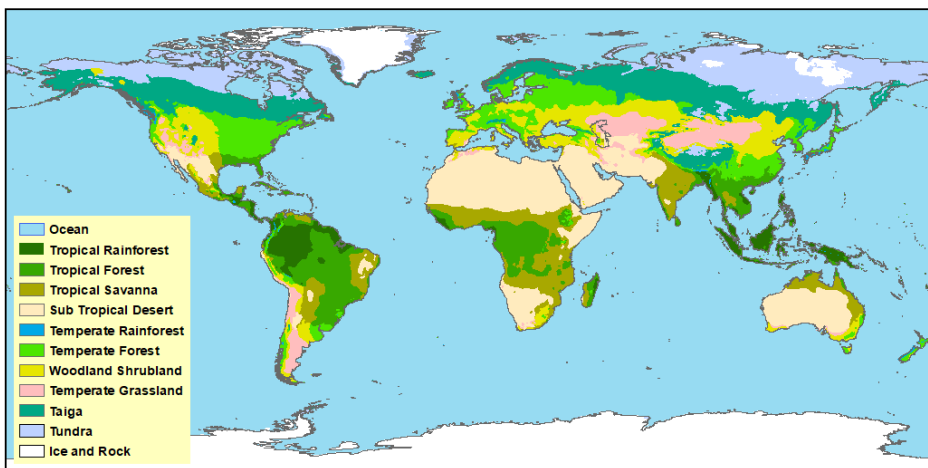
Figure 2.9: CRU 3.24.01 2000 – 2015 Average Annual Climate: (a) Precipitation; (b) Surface Air Temperature; and (c) Whittaker (1970) Biomes from Annual Precipitation and Air Temperature.

The CLM5 current day land cover mapping required a number of climate and LAI products to be generated using the climatological monthly surface air temperature and precipitation data. The first two of these were T_{avg}, the annual average temperature and T_c, the temperature of the coldest month. The third was GDD, Growing-Degree Days above 5°C with the monthly days multiplied by the temperature above 5°C for all months with an average temperature over 5°C. The last two were P_{ann}, annual average precipitation and P_{win}, average winter precipitation with winter defined as November through April in the Northern Hemisphere and May through October in the Southern Hemisphere. The LAI products are detailed in the next section.

The CRU 3.24.01 average annual precipitation and surface air temperature are shown in Figures 2.9a and 2.9b. These two surfaces are combined with climate rules to produce the Whittaker (1970) bioclimatic analysis framework used throughout this Technote. The global mapping of the Whittaker biomes is shown in Figure 2.9c, with analysis plots for the biome climate space and the CLM5 PFTs shown in Figures 2.20, 2.21 and 2.22. The correspondence between the MODIS land cover and Whittaker biome maps shows the strong influence that temperature and precipitation have on the vegetation distributions of the world supporting the use of the Nemani and Running (1996) climate biome rules to specify the physiologically distinct PFT types. The PFT Whittaker biome analysis in Figures 2.19, 2.20 and 2.21 show there are strong correspondences between the Whittaker biomes from average annual temperature and precipitation, and the more nuanced climate rules shown in Table 2.4. The inclusion of the coldest monthly temperature, as well as seasonal precipitation and growing degree days in the PFT rules, however results in PFT overlaps in Whittaker space for vegetation of the same form but different climate representation.

The 2000 – 2015 climate period used for generating the CRU 3.24.01 climate data also has distinct impacts on the boundaries generated from bioclimatic rules in both the Whittaker biome mapping of Figure 2.9c and the CLM5 PFT mapping in general. An alternative climate period of 1901 – 1916 was used with the CRU 3.24.01 climate data to assess the impacts of historical climate change on the biome and PFT mapping (not shown). The two major impacts of using the 2000 – 2015 climate were the poleward shift of the Taiga/Tundra and Temperate Forest/Taiga boundaries from high latitude warming, and the expansion of Savannas into Subtropical Desert regions in northern Australia with increased precipitation. Other changes were smaller but observable. It should be noted that the satellite observations coincide with the 2000 – 2015 climate period the bioclimatic mapping making it valid for current day climate, but large differences in climate will require a new application of the bioclimatic rules with the new climate to ensure the correct prescription of PFTs is maintained.

2.8 MODIS Leaf Area Index (LAI) Data

The current day land surface mapping of Lawrence and Chase (2007) uses the methods of Still et al. (2003) to determine the fractional C3 and C4 grass mapping using average monthly MODIS LAI combined with monthly precipitation and surface air temperature to describe seasonal variations in vegetation cover as indicator for the C3 and C4 growing season. For CLM5, the use of minimum and maximum average monthly MODIS LAI has been extended to provide additional constraints on the specification of bare ground and tree PFT deciduousness. The MODIS LAI rules used in the PFT mapping are detailed in Table 2.4.

(a) MODIS Leaf Area Index (LAI) Climatology January   2003 - 2015
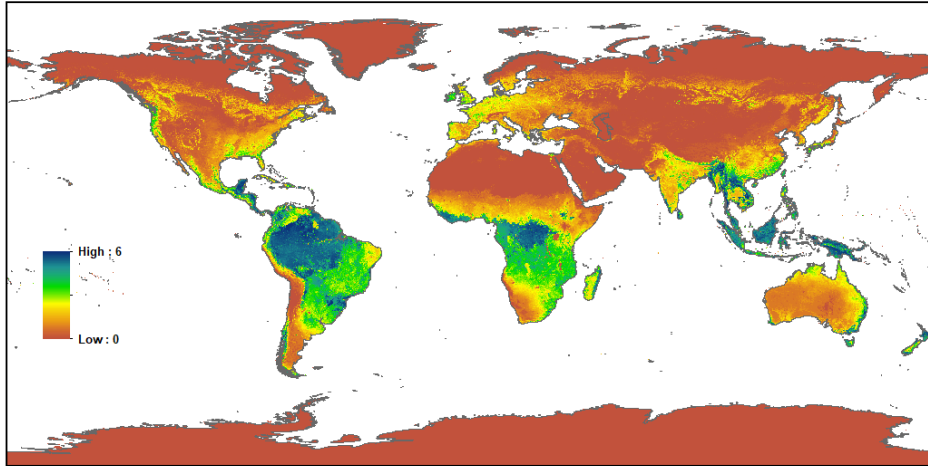
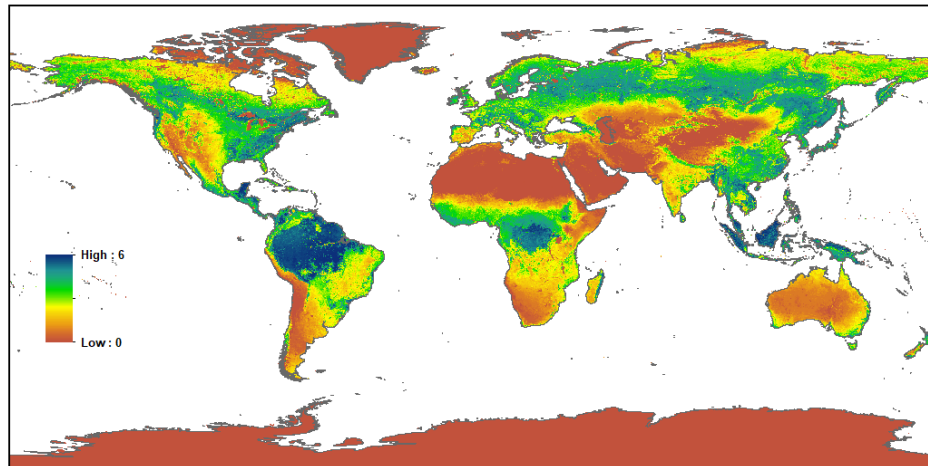(b) MODIS Leaf Area Index (LAI) Climatology July   2003 - 2015

Figure 2.10: MODIS MCD15A2 Monthly LAI climatology 2003 – 2015: (a) January LAI; and (b) July LAI.

For the CLM5 current day land cover mapping, average monthly MODIS LAI maps were developed as monthly climatologies of the MCD15A2 Version 5.0 8-day LAI maps for the years 2003 – 2015. The data was compiled from tiles downloaded from the NASA and USGS Land data Products Distributed Active Archive Center (LP DAAC) website (lpdaac.usgs.gov), mosaicked into global geographic 1km resolution grids. Quality flags were used to ensure only the highest quality cloud free grid cells were used from the 12-year period. Grid cells that had no good LAI retrievals for the entire period for a given 8-day date were filled by temporal inverse time weighted interpolation from days either side of the missing date. A full description of the monthly climatological LAI generation process can be found in Chapter 4. The global maps for the January and July climatological monthly LAI values are shown in Figure 2.10.

(a) MODIS CRU Cumulative Monthly LAI C3 Season   2003 - 2015
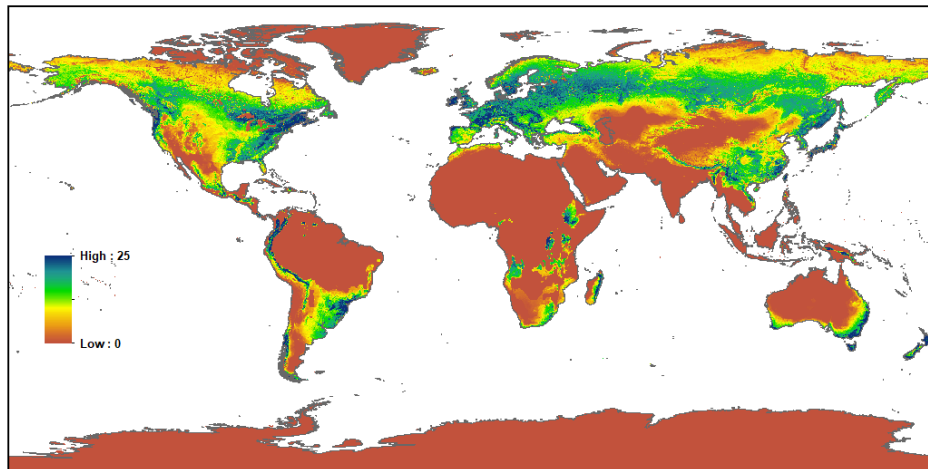
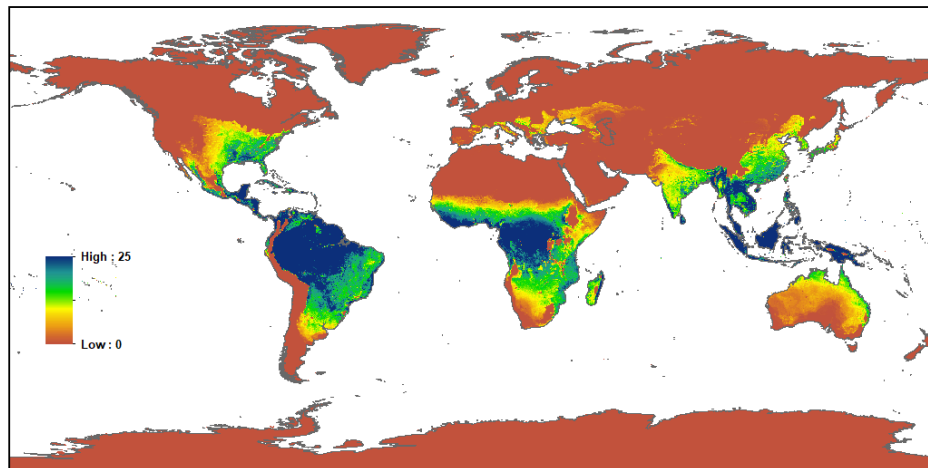(b) MODIS CRU Cumulative Monthly LAI C4 Season   2003 - 2015

Figure 2.11: MODIS MCD15A2 Monthly LAI climatology 2003 – 2015 and CRU 3.24.01 Monthly Temperature and Precipitation climatology 2000 – 2015: (a) Cumulative Monthly LAI C3 Growing Season; and (b) Cumulative Monthly LAI C4 Growing Season.

The maximum values of the monthly LAI maps shows there are areas with seasonal LAI that are mapped as having large amounts of bare ground in the MODIS VCF data shown in Figure 2.2c and the corrected values in Figure 2.8c. These areas are most evident in central Asia, Australia, North America and other areas where vegetation has strong response to seasonal precipitation or temperature. For the MODIS VCF product, the high bare ground reflects sparse vegetation averaged over the entire year. For CLM5, which has the seasonal cycle of vegetation cover represented through leaf and stem area indexes, either prescribed as detailed in Chapter 4 or explicitly simulated through the PFT Biogeochemistry, the dormancy of vegetation does not reflect the absence of that vegetation. To address this we use the maximum observed LAI to be a limit on the prescription of bare ground in CLM5 as detailed in Table 2.4. The assumption is that a maximum observed monthly LAI of 0.66 represents the point beyond which bare ground should not be prescribed in CLM5, rather vegetation should be prescribed and simulated with the LAI cycling to capture the periods of the year when the projected vegetation cover is very low.

Following from the work of Dahlin et al. (2015), the drought deciduous tree phenology in CLM4 was found to have major deficiencies in representing the LAI cycles of Tropical Broadleaf Deciduous trees. Part of this issue was due to the phenology representation and soil moisture triggers of the PFT in CLM4. Another part of the problem, however, was due to differences in definition of what a Tropical Broadleaf Deciduous Tree was in mapping versus represented in CLM. For dry tropical forests and savannas, the deciduous trees mapped may have strong LAI seasonality but not lose all of their leaves as represented in CLM.

The minimum monthly MODIS LAI values demonstrate the issues that arise with tropical deciduous tree mapping, with LAI still present at the lowest point of the year despite the MODIS land cover and AVHRR vegetation mapping identifying them as 100% deciduous. To address this inconsistency, the deciduous tree components are mapped as a mix of deciduous and evergreen when the minimum LAI suggests there are leaves present all year. To prevent confusion between different tree types, the evergreen is allocated to the Needleleef tree PFTs before the Broadleaf tree PFTs. The LAI deciduous tree constraints are detailed in Table 4. The constraint assumes that a minimum LAI of 4.0 represents 100% evergreen tree contribution. The impact of the deciduous tree constraint is apparent in the PFT mapping of the next section, in particular the much wider extent of Broadleaf Evergreen Tropical trees in Figure 2.15b compared to the Broadleaf Evergreen trees in Figure 2.7c.

The final use of MODIS monthly for the CLM5 current day PFT mapping is to separate the non-Arctic grasses into the C3 and C4 photosynthetic pathways using the methods of Still et al. (2003). For this NDVI is replaced with monthly LAI following Lawrence and Chase (2007) and then cumulative summed for all months that fall within the C3 or C4 grass growing seasons. The growing seasons are defined by having more than 25 mm of rain and a mean monthly air temperature above 22°C (warm) for C4 grass and below 22°C (cool) for C3 grass. The cumulative LAI maps for both C3 and C4 grasses are shown in Figure 2.11. The effectiveness of this method is demonstrated in the CLM5 C4 Grass PFT map of Figure 2.18b, which is a very close match to the total fraction of C4 vegetation map shown in Figure 4 of Still et al. (2003).

2.9 CLM5 Current Day Land Unit and Plant Function Types (PFTs)

The final step in the current day land cover mapping of Lawrence and Chase (2007) is to generate the CLM Land Unit and PFT maps. For the CLM5 1 km resolution current day data this process involve combining the Ocean, Land, Glacier, Lake, and Urban base maps, with the tree percentage maps for Needleleaf, Broadleaf, Evergreen and Deciduous, along with the, Crop, Shrub, Grass and Bare Ground maps, and the precipitation, temperature and LAI maps combined with the rules of Table 2.4. The global area for each of the generated CLM5 components is shown in Figure 2.12a with the MODIS IGBP land cover PFT composition shown in Figure 2.12b for each land cover class. The global and IPCC regional area for each component are listed in Table 2.5, with the MODIS IGBP land cover composition listed in Table 2.6. The global maps for each of the CLM5 components are shown in Figures 2.13, 2.14, 2.15, 2.16, 2.17, and 2.18. Whittaker bioclimatic analysis for each of the CLM5 components is shown in Figure 2.19, 2.20, and 2.21.

Table 2.5: Current Day CLM5 Land Unit and Plant Functional Types (PFTs) derived from MODIS Land Cover x VCF x LAI x AVHRR x FAOEarthStat x CRU Climate. Global and IPCC Region Areas in millions km^2.

CLM5 Class	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Ocean	362.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
Land	148.1	30.0	8.3	11.2	21.2	6.4	20.5	5.2	21.5	5.0	5.1	12.1
Glacier	15.3	0.0	0.0	0.1	0.1	0.1	0.0	0.0	2.6	0.0	0.1	12.1
Lake	2.8	0.2	0.0	0.1	0.3	0.1	0.1	0.0	0.8	0.0	0.0	0.0
Urban	0.5	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0
Tree	37.5	5.5	1.2	1.8	6.9	2.1	10.0	0.0	6.2	3.3	0.7	0.0
Shrub	9.8	1.2	2.1	0.0	3.4	0.1	0.9	0.1	1.8	0.0	0.1	0.0
Grass	47.7	12.1	3.1	4.7	5.6	3.5	7.7	0.7	5.4	1.7	3.1	0.0
Bare	26.5	11.0	1.9	3.0	1.7	0.1	1.3	4.3	1.9	0.0	1.1	0.0
NdlEvgTemp	5.1	0.0	0.2	0.7	1.3	0.9	0.0	0.0	1.8	0.0	0.1	0.0
NdlEvgBorl	6.1	0.0	0.0	0.3	2.5	0.4	0.0	0.0	2.8	0.0	0.0	0.0
NdlDecBorl	2.1	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BrdEvgTrop	15.9	4.0	0.2	0.0	0.0	0.0	8.4	0.0	0.0	3.0	0.3	0.0
BrdEvgTemp	1.4	0.1	0.5	0.1	0.0	0.0	0.5	0.0	0.0	0.1	0.1	0.0
BrdDecTrop	2.3	1.3	0.1	0.0	0.0	0.0	0.7	0.0	0.0	0.1	0.1	0.0
BrdDecTemp	3.3	0.0	0.2	0.4	0.5	0.7	0.2	0.0	1.2	0.0	0.1	0.0
BrdDecBorl	1.1	0.0	0.0	0.2	0.5	0.0	0.1	0.0	0.2	0.0	0.0	0.0
ShrEvgTemp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ShrDecTemp	4.8	1.2	2.1	0.0	0.1	0.0	0.9	0.1	0.3	0.0	0.1	0.0
ShrDecBorl	5.0	0.0	0.0	0.0	3.2	0.1	0.1	0.0	1.6	0.0	0.0	0.0
GrsC3Arc	8.0	0.0	0.0	1.5	3.2	0.3	0.3	0.0	2.6	0.0	0.1	0.0
GrsC3	16.2	1.5	0.8	1.8	4.5	2.0	1.6	0.4	3.1	0.0	0.2	0.0
GrsC4	18.7	8.6	2.0	0.3	0.3	0.1	4.8	0.0	1.0	0.7	0.8	0.0
Crop	12.7	1.9	0.3	2.6	0.8	1.4	1.3	0.2	1.2	1.0	2.1	0.0

As expected, the CLM5 global area, ocean area, and land area, and Glacier and Urban land unit area were all directly consistent with the MODIS IGBP land cover mapping as they were directly prescribed from these types as listed in Table 2.2. The Lake land unit area was slightly reduced from the land cover mapping globally and for the IPCC regions. This was the result of the FAOEarthStat Crop mapping and other land units having areas where they overwrote the 1 km land cover map to reduce the Lake land unit. In general, the differences in lake areas were small.

The CLM5 Glacier land unit mapping of Figure 2.13a shows the MODIS land cover had the expected coverage over Antarctica, Greenland and other Arctic islands, the Himalayas, Alaska, and north-western Canada. The Whittaker bioclimatic analysis of Figure 2.19b shows the majority of Glacier is colder or wetter than Tundra, but there is mapping of Glacier all the way into the Taiga biome as well. This inconsistency can be tracked to the original resolution difference between the land cover data and the CRU climate, where topographic differences at 1 km are not captured in the average climate of the 0.5 degree grids of CRU. For the default CLM5 surface data sets, the raw Glacier data is overwritten with the percent glacier and ice sheet from the global Randolph Glacier Inventory version 1.0 (RGIv1.0) (Arendt et al. 2012) combined with the Greenland Ice Sheet (Rastner et al. 2012), and the Antarctic Ice Sheet data from the Scientific Committee on Antarctic Research (SCAR) Antarctic Digital Database version 5.0 as described in Chapter 4.
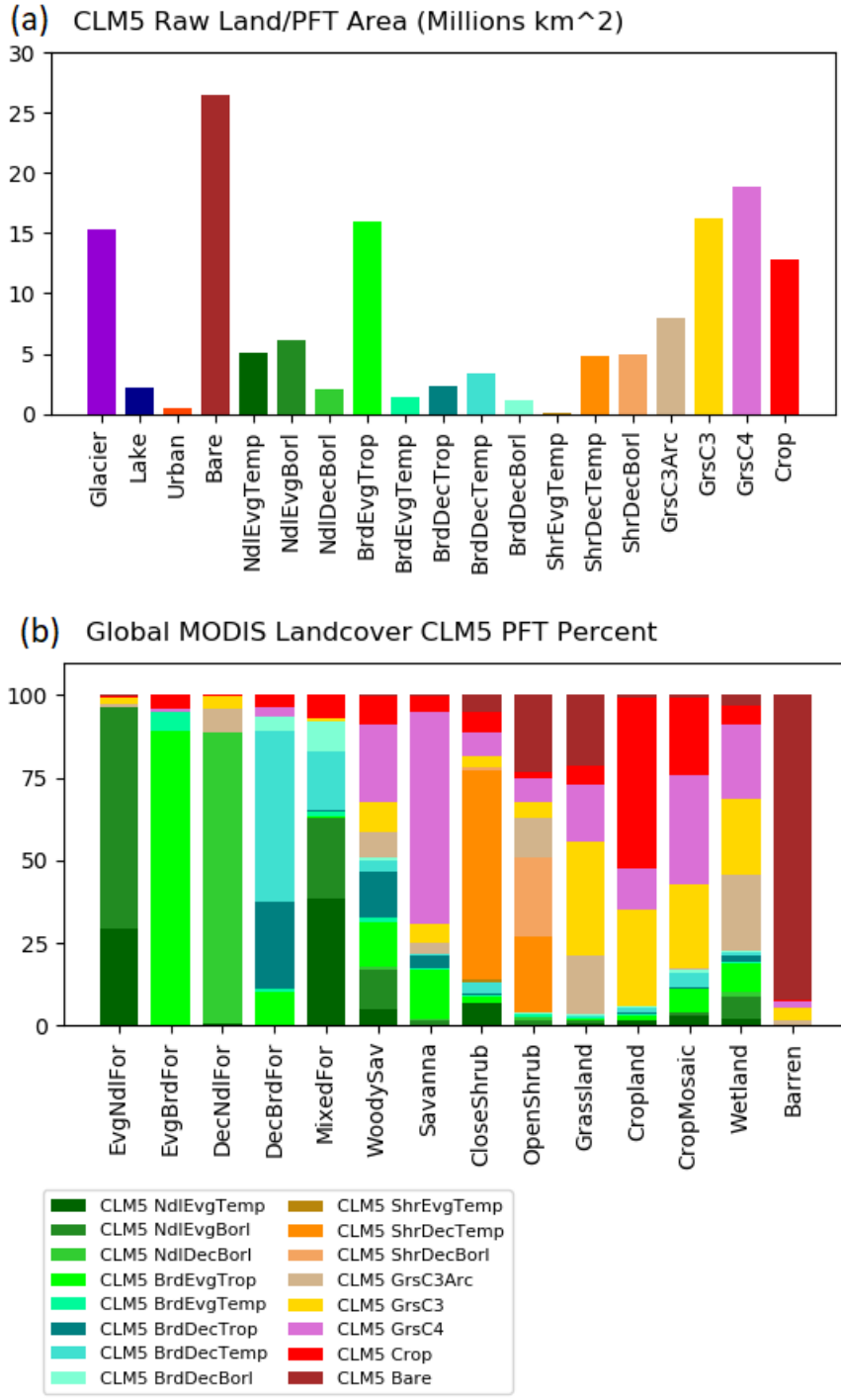
Figure 2.12: Global Current Day 1 km resolution CLM5 PFT mapping: (a) Total area; and (b) MODIS land cover type compositions.

The CLM5 Lake land unit mapping of Figure 2.13b shows the global distribution of lakes dominated by North America with the Great Lakes and then widespread smaller lakes through Canada. There also were large contributions from Scandinavia and the Caspian Sea, which is included as a lake in this mapping. The mapping also shows iconic lakes such as Lake Victoria

and Lake Baikal. The Whittaker bioclimatic analysis of Figure 2.19b shows that there are three main climate groupings for lakes. These are the cold wet areas of the northern higher latitudes, the arid subtropical areas of central Asia, and the tropical savanna regions. Interestingly there are very few lakes in high precipitation temperate and tropical lands. For the default CLM5 surface data sets the raw Lake data is overwritten with the percent lake from the Global Lake and Wetland Database of Lehner and Doll (2004), with the mean lake depth of each grid cell calculated based on the global gridded data sets of Kourzeneva (2012), as described in Chapter 4.

Table 2.6: MODIS IGBP Land Cover Type by Percentage CLM5 PFT Composition.

IGBP Class	bare	nem	neb	ndb	bet	bem	bdt	bdm	bdb	sem	sdm	sdb	gca	gc3	gc4	crop
EvgNdlFor	0	29	67	0	0	0	0	0	0	0	0	0	1	2	0	1
EvgBrdFor	0	0	0	0	89	6	0	0	0	0	0	0	0	0	1	4
DecNdlFor	0	1	0	88	0	0	0	0	0	0	0	0	7	4	0	0
DecBrdFor	0	0	0	0	10	1	26	52	4	0	0	0	0	0	3	4
MixedFor	0	38	24	0	0	1	1	18	9	0	0	0	0	1	0	7
CloseShrub	5	7	0	0	2	1	0	3	0	1	63	1	0	3	7	7
OpenShrub	23	0	1	1	0	1	0	0	0	0	23	24	12	5	7	2
WoodySav	0	5	12	1	14	1	14	4	1	0	0	0	8	9	23	9
Savanna	0	0	2	0	15	0	4	0	0	0	0	0	3	6	64	5
Grassland	21	1	1	0	0	1	0	1	0	0	0	0	17	35	17	6
Wetland	3	2	7	1	9	0	2	1	1	0	0	0	23	23	22	5
Cropland	1	1	0	0	1	1	0	2	0	0	0	0	0	29	12	52
CropMosaic	1	3	1	0	7	0	0	5	1	0	0	0	0	25	33	24
Barren	92	0	0	0	0	0	0	0	0	0	0	0	2	4	2	0

The CLM5 Urban land unit mapping of Figure 2.13c shows the global distribution of urban areas are concentrated in the four main regions of the eastern United States, Europe, China, and south-eastern Brazil. Interestingly the MODIS land cover map has very little urban area mapped in India, Africa or Southeast Asia. The Whittaker bioclimatic analysis of Figure 2.19c show two climate groupings for the Urban land unit in Temperate Woodlands and Seasonal Forests, and in Tropical Seasonal Forests and Savannas. For the default CLM5 surface data sets the raw Urban data is overwritten with the four urban density classes of tall building district (TBD), and high, medium, and low density (HD, MD, LD) derived from the LandScan 2004 population density dataset as described by Jackson et al. (2010), and detailed in Chapter 4.

At the highest level, the CLM5 PFT mapping was dominated by Grasses with 47.7 million km^2, then trees at 37.5 million km^2, Bare Ground at 26.5 million km^2, Crop at 12.7 million km^2, and finally Shrubs at 9.8 million km^2. The Grasses were found predominantly in Africa, with smaller amounts in Latin America, Eurasia, North America, and then other regions. This pattern reflects the MODIS land cover mapping of Grassland and Savannas. Trees where found predominantly in Latin America, Eurasia, North America and Africa reflecting the land cover mapping of Forests and Woody Savannas. Bare Ground was found predominantly in Africa, Middle East, and Eurasia reflecting the land cover mapping of Barren deserts such as the Sahara, the Arabian, and the Tibetan Plateau. Crops were found predominantly in East Asia (China), Southern Asia (India, Nepal, Pakistan and Bangladesh), and Africa, with smaller amounts in Europe, Latin America, North America, and other regions.

(a) MODIS LC CLM5 Glacier   (%)
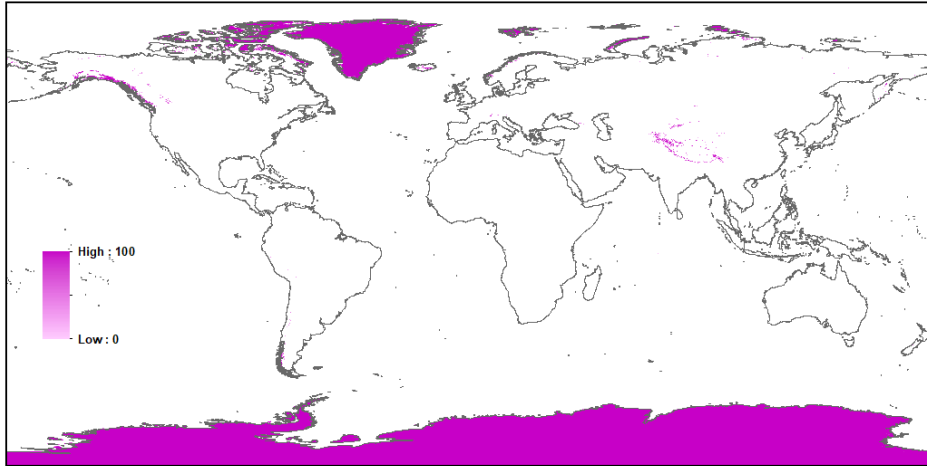
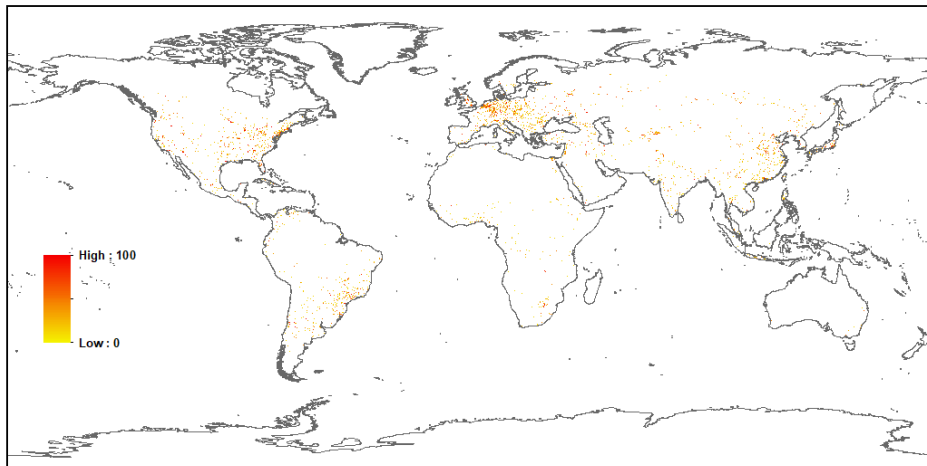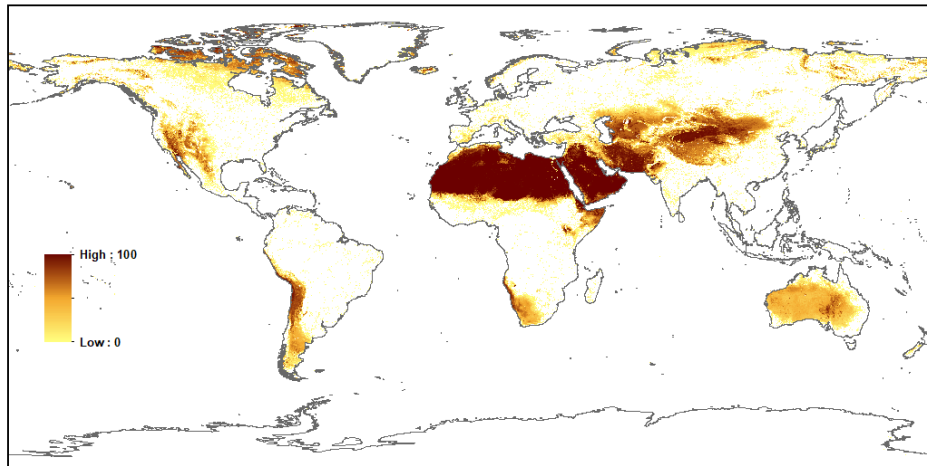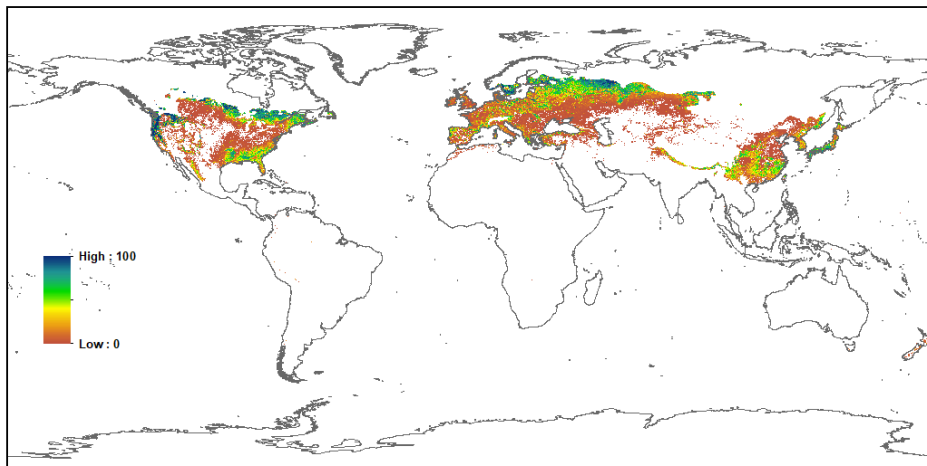(b) MODIS LC CLM5 Lake   (%)
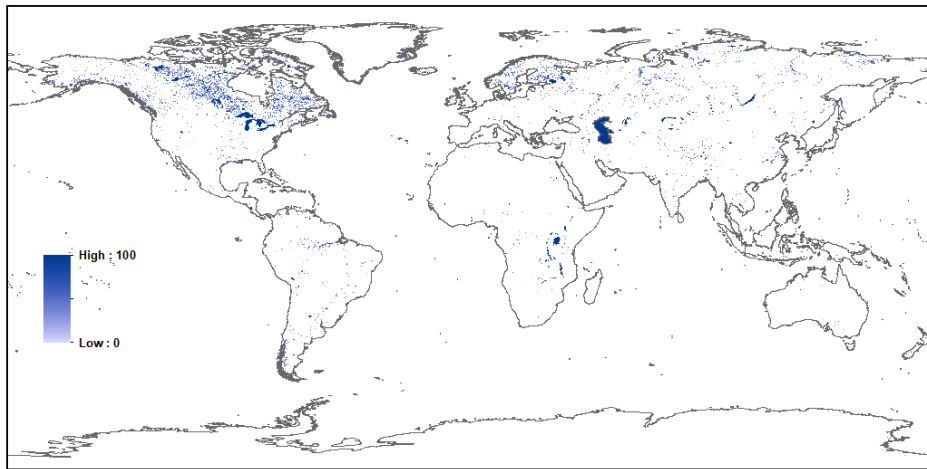
(c) MODIS LC CLM5 Urban   (%)

Figure 2.13: Global Current Day 1 km resolution CLM5 Land Unit Mapping: (a) Percent Glacier; (b) Percent Lake; and (c) Percent Urban.

(a) MODIS LC/VCF/LAI CLM5 Bare Soil   (%)

(b) MODIS LC/VCF/AVHRR CLM5 Needleleaf Evergreen Temperate   (%)

(c) MODIS LC/VCF/AVHRR CLM5 Needleleaf Evergreen Boreal   (%)
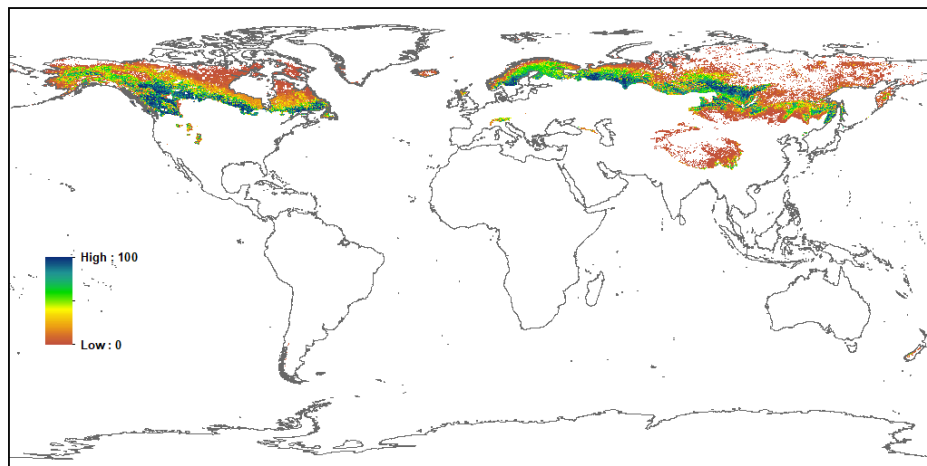
Figure 2.14: Global Current Day 1 km resolution CLM5 PFT Mapping: (a) Percent Bare Ground; (b) Percent Needleleaf Evergreen Temperate Tree; and (c) Percent Needleleaf Evergreen Boreal Tree.

(a) MODIS LC/VCF/AVHRR CLM5 Needleleaf Deciduous Boreal   (%)
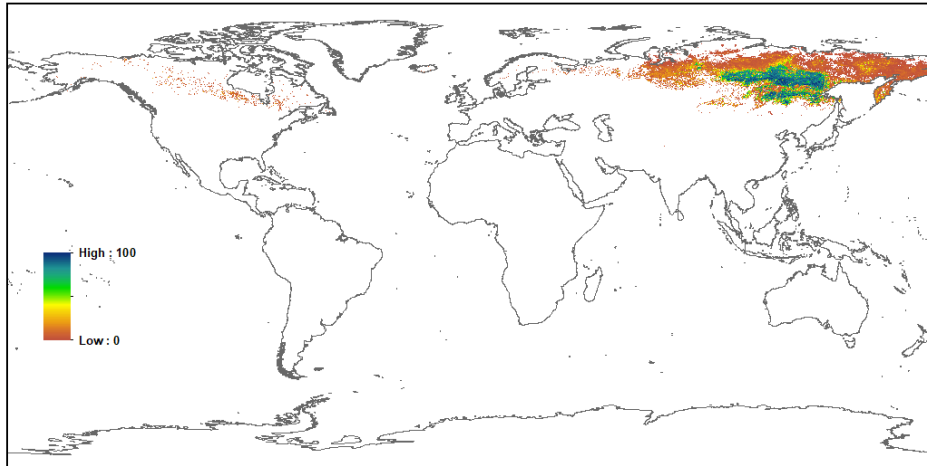
(b) MODIS LC/VCF/AVHRR/LAI CLM5 Broadleaf Evergreen Tropical   (%)
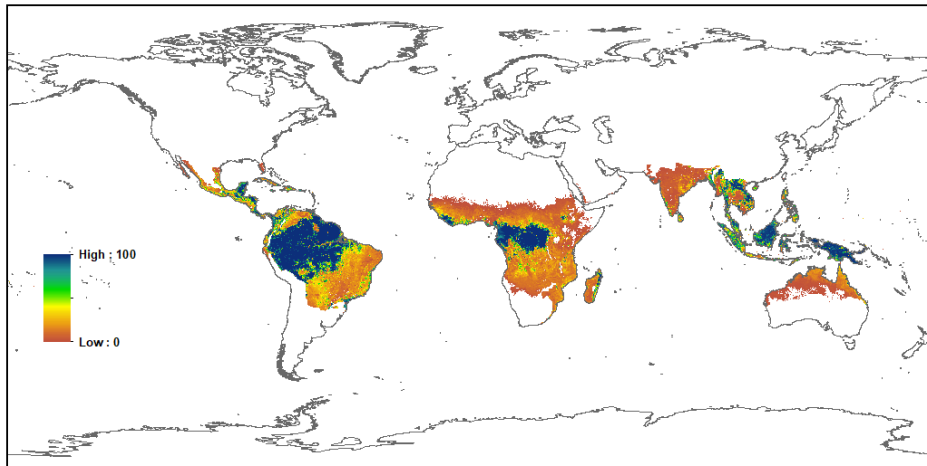
(c) MODIS LC/VCF/AVHRR/LAI Broadleaf Evergreen Temperate   (%)
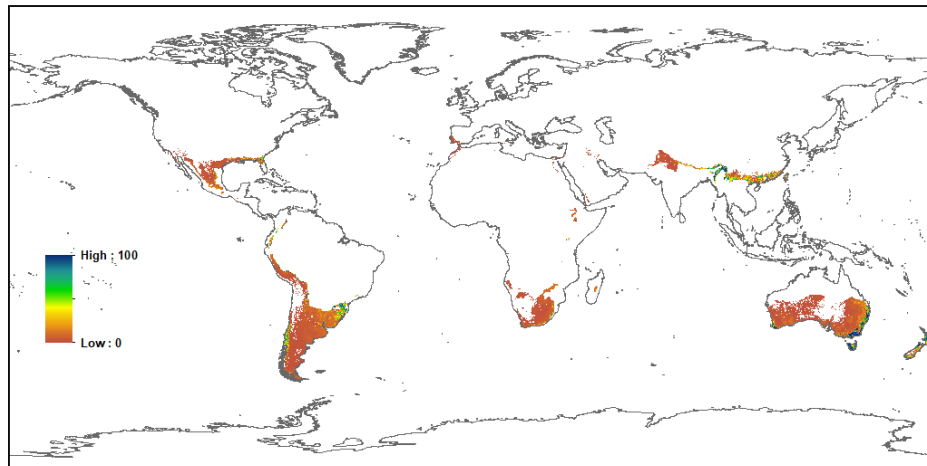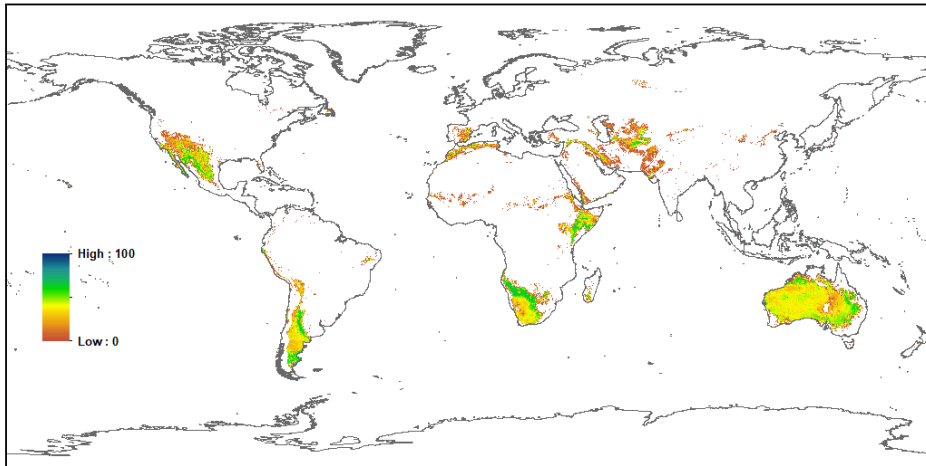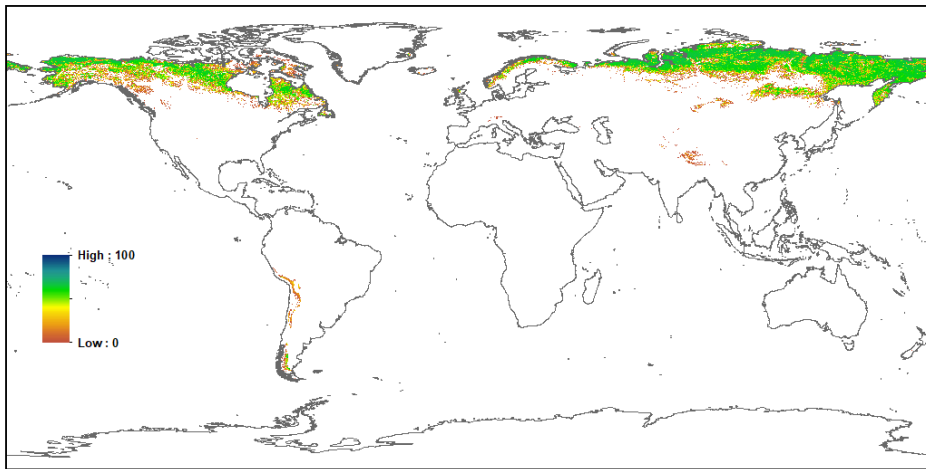
Figure 2.15: Global Current Day 1 km resolution CLM5 PFT Mapping: (a) Percent Needleleaf Deciduous Boreal Tree; (b) Percent Broadleaf Evergreen Tropical Tree; and (c) Percent Broadleaf Evergreen Temperate Tree.

(a) MODIS LC/VCF/AVHRR/LAI CLM5 Broadleaf Deciduous Tropical   (%)
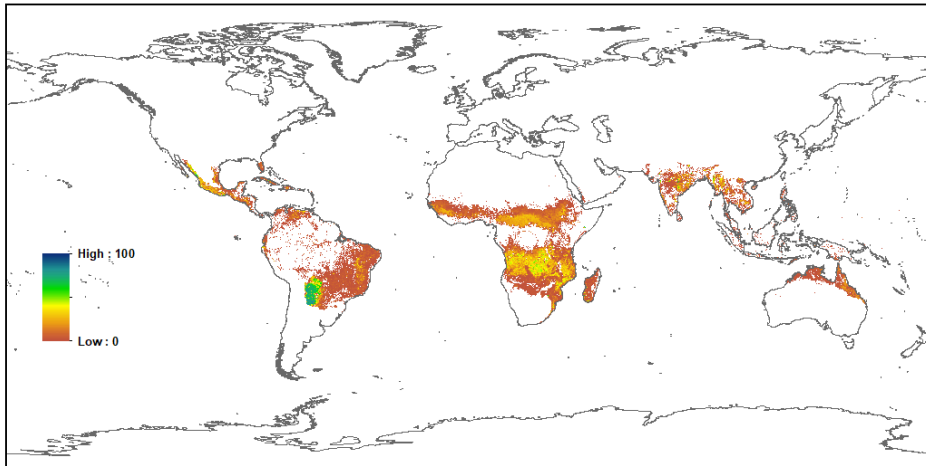
(b) MODIS LC/VCF/AVHRR/LAI CLM5 Broadleaf Deciduous Temperate   (%)
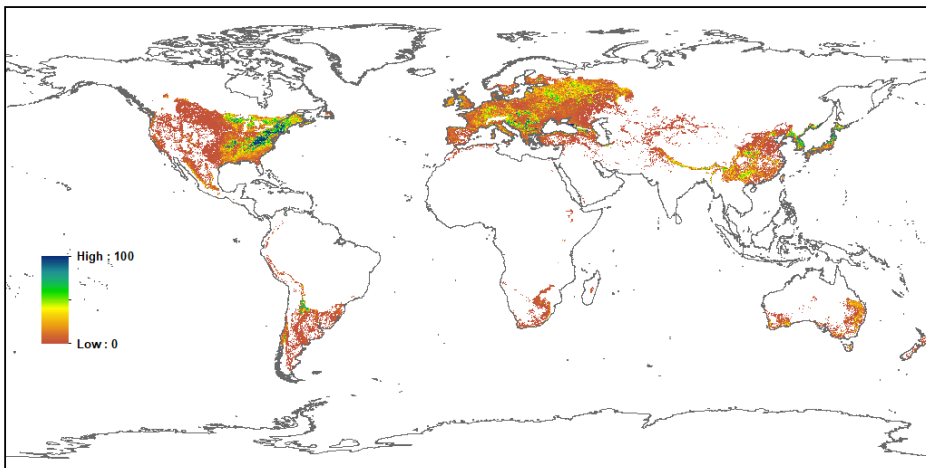
(c) MODIS LC/VCF/AVHRR/LAI CLM5 Broadleaf Deciduous Boreal   (%)
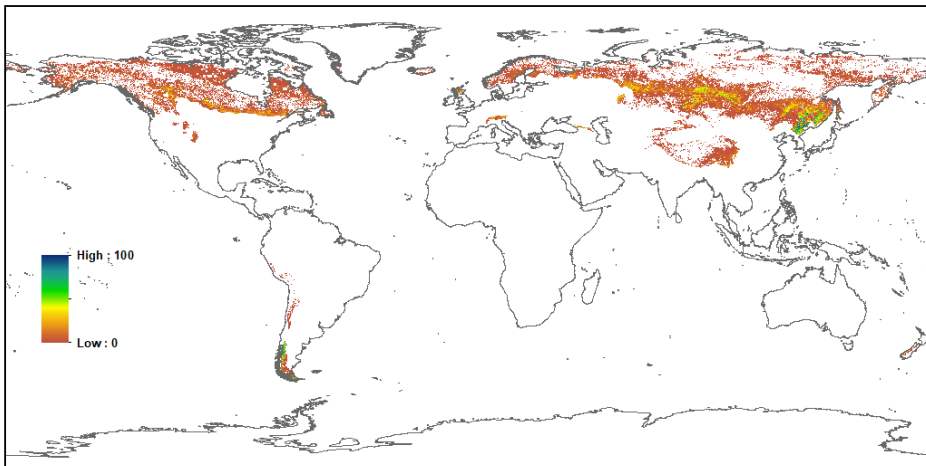
Figure 2.16: Global Current Day 1 km resolution CLM5 PFT Mapping: (a) Percent Broadleaf Deciduous Tropical Tree; (b) Percent Broadleaf Deciduous Temperate Tree; and (c) Percent Broadleaf Deciduous Boreal Tree.

(a) MODIS LC/VCF/LAI CLM5 Shrub Deciduous Temperate   (%)

(b) MODIS LC/VCF/LAI CLM5 Shrub Deciduous Boreal   (%)

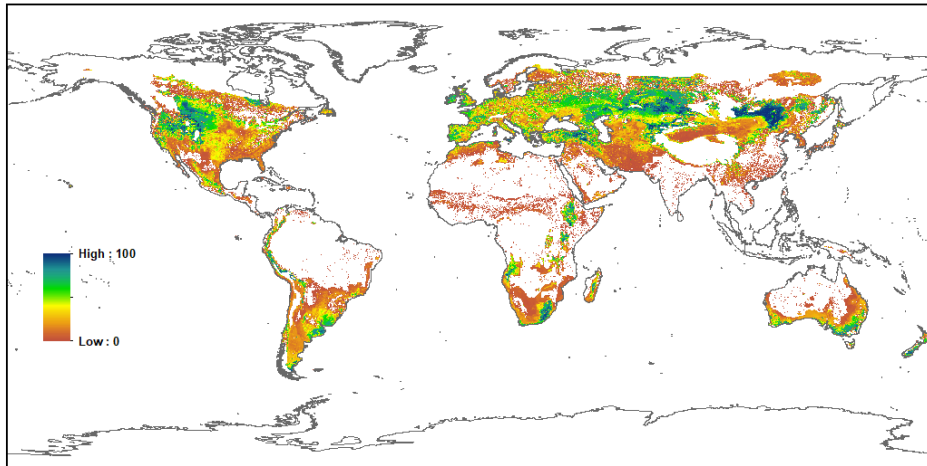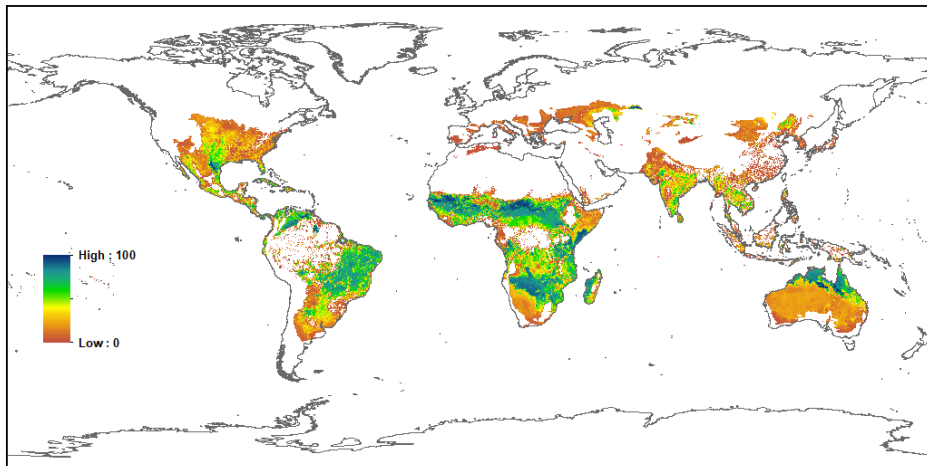(c) MODIS LC/VCF/LAI CLM5 Grass C3 Arctic   (%)
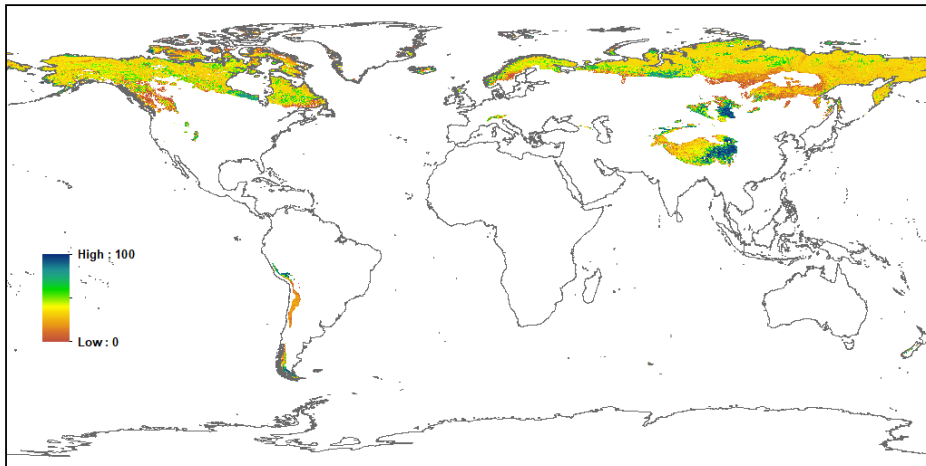
Figure 2.17: Global Current Day 1 km resolution CLM5 PFT Mapping: (a) Percent Deciduous Temperate Shrub; (b) Percent Deciduous Boreal Shrub; and (c) Percent C3 Arctic Grass.

(a) MODIS LC/VCF/LAI CLM5 Grass C3 (%)

(b) MODIS LC/VCF/LAI CLM5 Grass C4 (%)
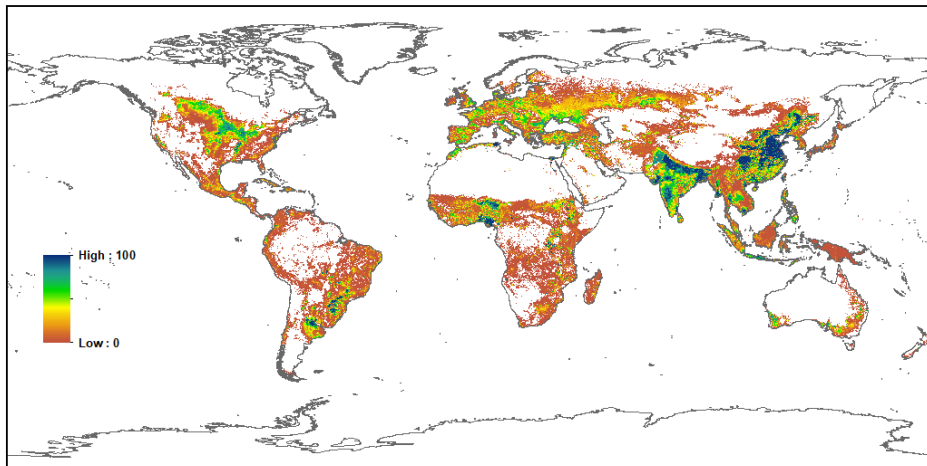
(c) EARTHSTAT/FAOSTAT CLM5 Crop (%)

Figure 2.18: Global Current Day 1 km resolution CLM5 PFT Mapping: (a) Percent C3 Grass; (b) Percent C4 Grass; and (c) Percent Crop.

The Bare Ground mapping of Figure 2.14a reflects the regional analysis revealing all of the iconic deserts of the world. For the CLM5 land units and other PFTs, Bare Ground is the biggest contributor as shown in Figure 2.12a. The impact of the MODIS land cover tree correction and the maximum MODIS LAI correction are apparent when compared to the original MODIS VCF Non Vegetated map of Figure 2.2c. This is reflected in the PFT contribution to MODIS land cover classes listed in Table 2.5 and shown in Figure 2.12b, where globally forest and savanna land cover classes now have little to no Bare Ground, while Barren, Grassland and Open Shrubland all remain with very high Bare Ground contributions. The Whittaker bioclimatic analysis shown in Figure 2.19e highlights the strong relationship between Bare Ground and precipitation with almost all of the mapping below 500 mm/year and becoming the exclusive class below 250 mm/year.

The CLM5 Needleleaf Tree PFT mapping was split into Needleleaf Evergreen Temperate at 5.1 million km^2, Needleleaf Evergreen Boreal at 6.1 million km^2, and Needleleaf Deciduous Boreal at 2.1 million km^2. The global mapping of each component is shown in Figure 2.14b and c, and Figure 2.15a, showing their distinct climatic patterns. Evergreen Boreal trees correspond strongly with the Taiga, while Evergreen Temperate trees correspond with warmer Needleleaf and Mixed Forests, and Needleleaf Deciduous trees correspond with the Larch forests of eastern Siberia.

The PFT composition of the MODIS land cover classes of Table 2.6 and Figure 2.12a, show that globally the Needleleaf Evergreen Temperate Trees were found predominantly in Mixed Forests contributing 38%, with a smaller contribution in Evergreen Needleleaf Forests at 29%. Needleleaf Evergreen Boreal Trees were found predominantly in Evergreen Needleleaf Forests contributing 67%, with smaller contributions to Mixed Forests at 24%. The Needleleaf Deciduous Boreal Trees were almost exclusively in the Deciduous Needleleaf Forests contributing 88% of their PFTs. The Whittaker analysis of Figure 2.19f, g and h, shows the same distinct bioclimatic divisions. The Needleleaf Evergreen Temperate Trees were predominantly found above 2°C, the Needleleaf Evergreen Boreal Trees were predominantly found between -8°C and 3°C, and the Needleleaf Deciduous Boreal Trees were predominantly found below -2°C. All Needleleaf Trees required more than 250 mm/year precipitation, except in extremely cold climates, and were in general restricted to less than 2000 mm/year.

The Broadleaf Evergreen Tree mapping was split into the Tropical component at 15.9 million km^2, and Temperate component at 1.4 million km^2. The Broadleaf Evergreen Tropical Tree mapping of Figure 2.15b shows the distinct tropical rainforests of the Amazon, Congo and Southeast Asia, while the Broadleaf Evergreen Temperate Tree mapping in Figure 2.15c is mainly in the Southern Hemisphere in the Evergreen Broadleaf Forests and Woodlands of southern Australia, Africa and South America. The mapping also shows the effect of the minimum LAI constraint on deciduous trees with Broadleaf Evergreen trees extending beyond the MODIS land cover Evergreen Broadleaf Forests into Deciduous Broadleaf Forests and Savannas reflecting the fraction of trees that retain leaves throughout the year.

The PFT composition of the MODIS land cover classes in Table 2.6 and Figure 2.12a, show that globally the Broadleaf Evergreen Tropical Trees made up 89% of PFTs in Evergreen Broadleaf Forests, 10% of Deciduous Broadleaf Forests, and 15% of Savannas. Broadleaf Evergreen Temperate Trees however, made up only 6% Evergreen Broadleaf Forests, and 1% or less of other land cover types. The Whittaker analysis of Figure 2.20a and b show that the Broadleaf Evergreen Tropical Trees occurred predominately between 20°C and 30°C, and between 1000 and 4000 mm/year, while Broadleaf Evergreen Temperate Trees occurred predominately between 10°C and 20°C, and between 400 and 2000 mm/year.

The Broadleaf Deciduous Tree mapping was split into the Tropical component at 2.3 million km^2, the Temperate component at 3.3 million km^2, and the Boreal component at 1.1 million km^2. The global Broadleaf Deciduous Tropical Tree mapping of Figure 2.16a highlights the contribution that they make in South America to the Deciduous Broadleaf Forests of the Gran Chaco into the Pantanal, with lower values through the Savannas of the Cerrado into the Caatinga. In Africa, India, Southeast Asia and northern Australia, the mapping highlights the Whittaker Dry Tropical Forests and Savannas regions outside of the Moist Tropical Forests. The Broadleaf Deciduous Temperate Tree mapping in Figure 2.16b and the Broadleaf Deciduous Boreal Tree mapping in Figure 2.16c show that in they were predominantly in the Northern Hemisphere, where they contributed to the Deciduous Broadleaf Forests of the eastern United States, Europe, the Korean peninsula, and Japan, with lower contributions for Mixed Forests and Savanna.

The PFT composition of land cover classes in Table 2.6 and Figure 2.12a show that globally Broadleaf Deciduous Tropical Trees made up 26% of Deciduous Broadleaf Forests, with 15% of Woody Savannas, and 4% of Savannas. Broadleaf Deciduous Temperate Trees made up 52% of Deciduous Broadleaf Forests and 18% of Mixed Forests. Broadleaf Deciduous Boreal Trees made up only 4% of Deciduous Broadleaf Forests and 9% of Mixed Forests. The Whittaker analysis of Figure 2.20c, d and e show there are three distinct climate regimes, with the Broadleaf Deciduous Tropical Trees between 20°C and 30°C, and between 500 and 1500 mm/year, the Broadleaf Deciduous Temperate Trees between 3°C and 20°C, and between 300 and 1500 mm/year, and the Broadleaf Deciduous Boreal Trees below 3°C, and between 250 and 1000 mm/year.

The Shrub PFT mapping was evenly split between Deciduous Temperate Shrubs with 4.8 million km^2 and Deciduous Boreal Shrubs with 5.0 million km^2. Evergreen Temperate Shrubs have almost no contribution globally which can be traced to the climate rules from Table 2.4 that require that these shrubs need to have more than 520 mm/year precipitation and for two thirds of that precipitation to fall in the winter. Because of this, the Evergreen Temperate Shrubs are not shown globally as they can not be adequately viewed on a global map. The global map of Deciduous Temperate Shrubs of Figure 2.17a shows five main areas in inland Australia, southern Africa, southern South America, southwestern North America and western Eurasia. The global map of Deciduous Boreal Shrubs of Figure 2.17b shows that they predominantly occur in high northern latitudes corresponding with Tundra in North America, Scandinavia, and Siberia.

The PFT composition of land cover classes in Table 2.6 and Figure 2.12a show that globally Evergreen Temperate Shrubs contributed 1% to Closed Shrubland with no major contribution to any other land cover class. The Deciduous Temperate Shrubs contributing 63% to Closed Shrubland and 23% to Open Shrubland, and the Deciduous Boreal Shrubs contribution 1% to Closed Shrubland and 24% to Open Shrubland. The MODIS land cover mapping from Table 2.2 shows there was very little Closed Shrubland globally, so the vast majority of shrub mapping was from Open Shrubland. The Whittaker analysis of Figure 2.20f, g and h reflect the other analysis with Evergreen Temperate Shrubs having very limited area for shrubs above 7°C and above 520 mm/year as required by the climate rules. Below 520 mm/year the shrubs were split into Deciduous Temperate Shrubs above 7°C, and into Deciduous Boreal Shrubs below 7°C.
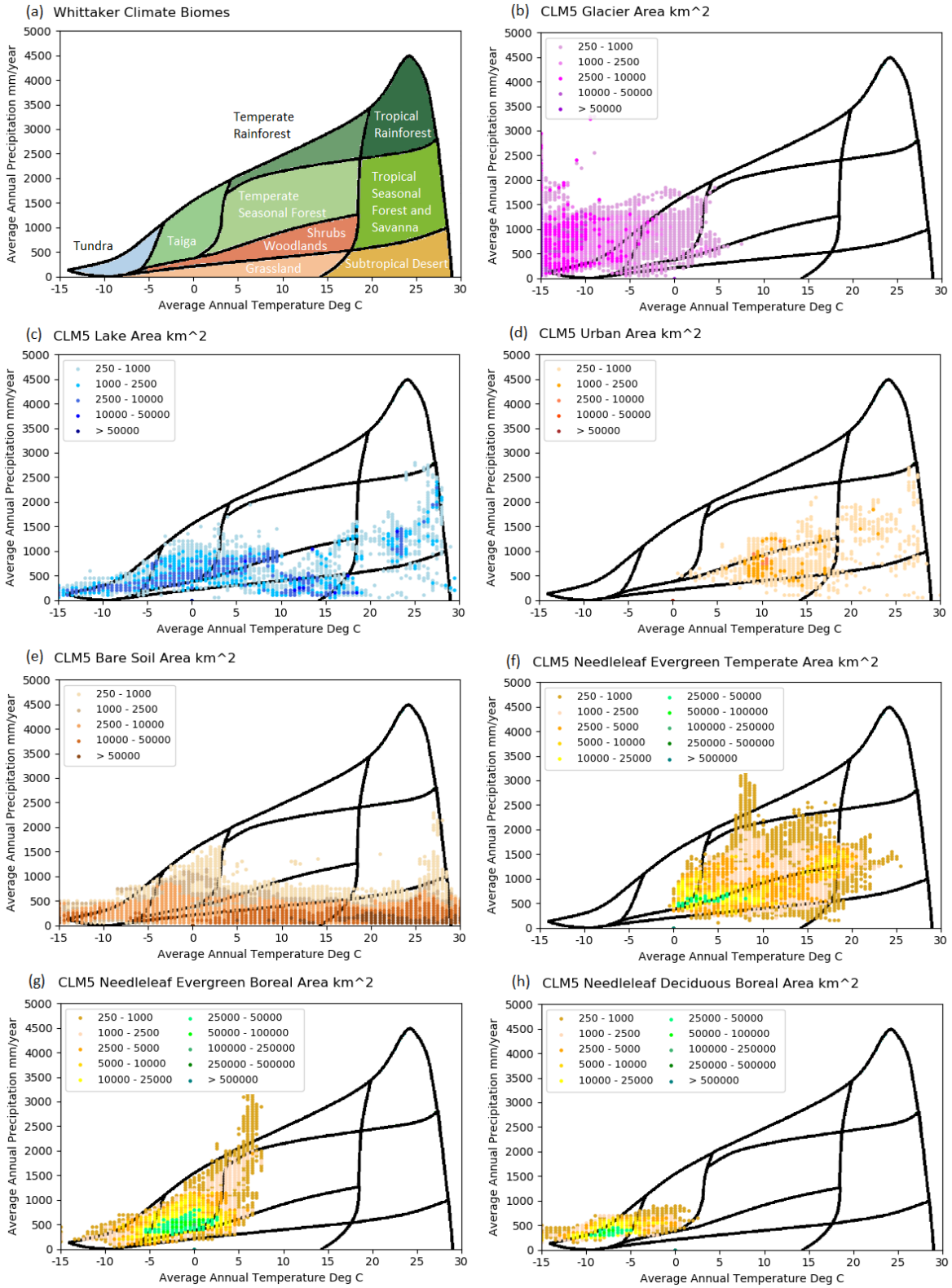
Figure 2.19: Global CLM5 Whittaker (1970) analysis: (a) Biomes; (b) Glacier; (c) Lake; (d) Urban; (e) Bare Ground; (f) NdlEvgTemp; (g) NdlEvgBorl; and (h) NdlDecBorl.
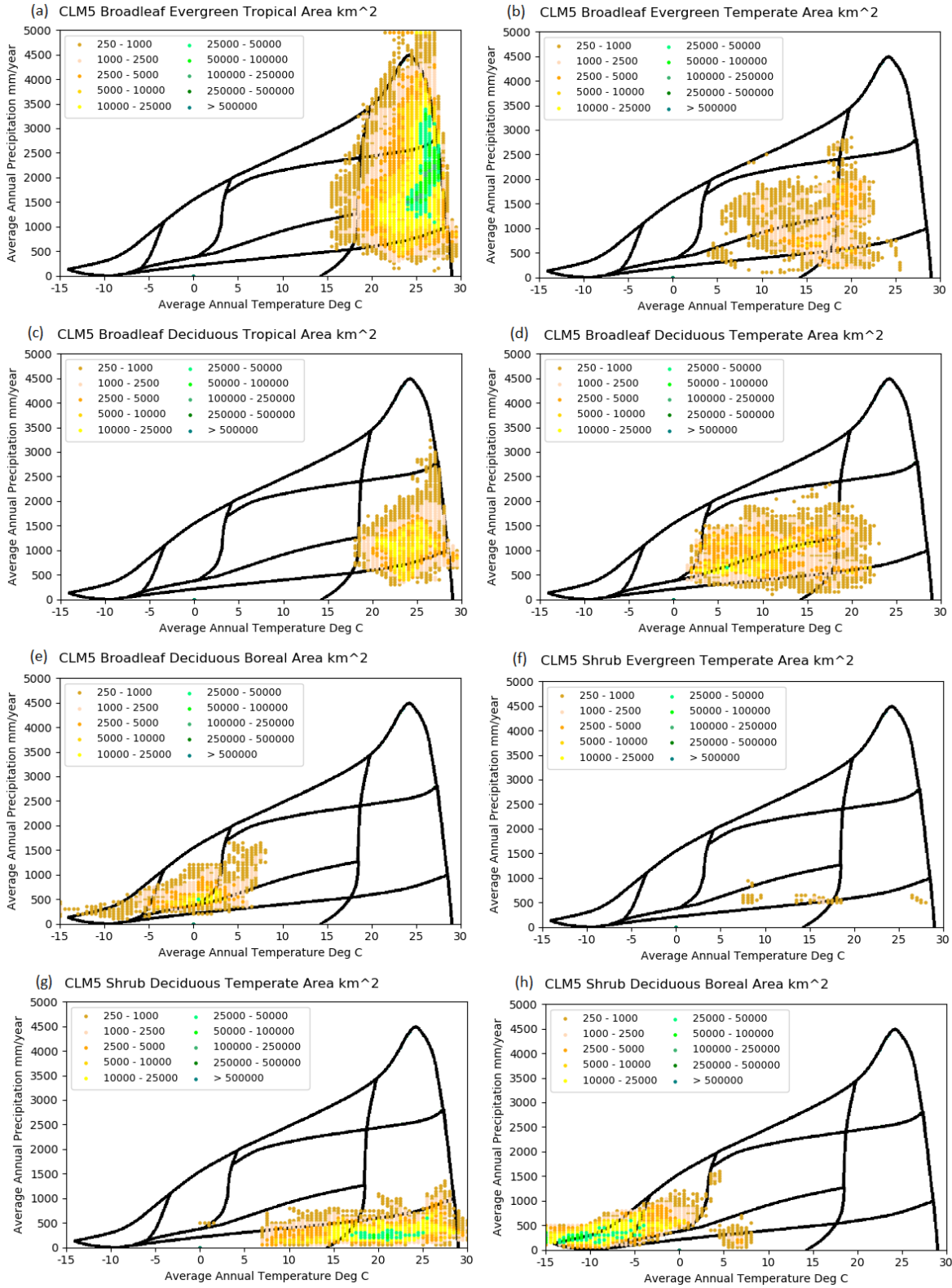
Figure 2.20: Global CLM5 Whittaker (1970) analysis: (a) BrdEvgTrop; (b) BrdEvgTemp; (c) BrdDecTrop; (d) BrdDecTemp; (e) BrdDecBorl; (f) ShrEvgTemp; (g) ShrDecTemp; and (h) ShrDecBorl.
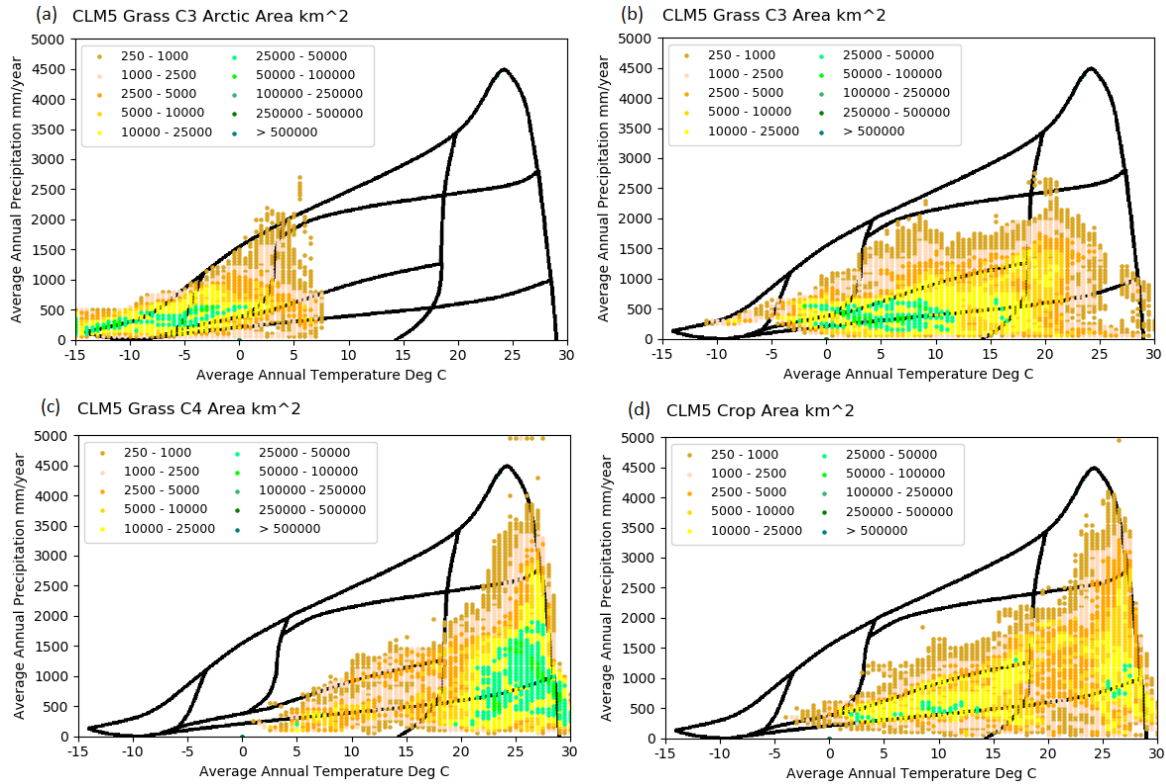
Figure 2.21: Global CLM5 Whittaker (1970) analysis: (a) GrsC3Arc; (b) GrsC3; (c) GrsC4; and (d) Crop.

The Grass PFT mapping was split into the C3 Arctic Grass component at 8.0 million km^2, the C3 Grass component at 16.2 million km^2, and the C4 Grass component at 18.7 million km^2. The global map of C3 Arctic Grass of Figure 2.17c shows the same spatial pattern as the Deciduous Boreal Shrub, with the only exceptions in the Tibetan Plateau and Mongolia which where mapped as Grasslands in the MODIS land cover map. The global map of C3 Grass of Figure 2.18a shows they predominantly occupy the temperate and sub-tropical Grasslands. In the Northern Hemisphere this includes the Great Plains of North America, Europe, and the Eurasian Steppes. In the Southern Hemisphere this included the Pampas of South America, and southern areas of Africa and Australia. The global map of C4 Grass of Figure 2.18b show they predominantly occupy the tropical Savannas of the Cerrado in South America, drier areas of Africa, and northern Australia. They are also the dominant grasses in India, Southeast Asia, and Central America.

The PFT composition of land cover classes in Table 2.6 and Figure 2.12a show that globally C3 Arctic Grass contributed 17% to Grassland, 23% to Wetland and 12% to Open Shrubland. This reflected the MODIS land cover mapping of Tundra where these grasses occurred. Globally C3 Grass contributed 35% to Grassland and 23% to Wetland, with major contributions to Cropland at 29% and Cropland Mosaic at 25%. Globally C4 Grass contributed 64% to Savanna, 23% to Woody Savanna, and 22% to Wetland. C4 Grass also contributed 12% to Cropland and 33% to Cropland Mosaic. The high contribution of C3 Grass and C4 Grass to Cropland reflected the heterogeneity of this land cover type as well as differences in the mapping methods for crops between MODIS remote sensing and the EarthStat 2000, MIRCA 2000 and FAOSTAT products used to generate CLM5 CFTs as described in Chapter 3.

The Whittaker analysis of grasses shown in Figure 2.21a, b and c shows three distinct climate regimes for the three Grass PFTs. The C3 Arctic Grass climate matches the Deciduous Boreal Shrub climate fairly closely with temperatures below 0°C. The C3 Grass climate ranges from around -1°C to 20°C, and C4 Grass above 20°C. There is some overlap in the climate range between C3 Grass and C4 Grass between 5°C and 25°C following the growing season rules of Still et al. (2003). This is most evident in global maps in North America and other sub-tropical Grasslands.

The current day Crop PFT mapping was 12.7 million km^2. The global map of the Crop PFT of Figure 2.18c shows strong correspondence with the MODIS IGBP Crop and Crop Mosaic land cover types. The largest areas of Crop PFT are in India and China, with other areas found in the Midwest of the United States into Canada, Brazil into Argentina, Nigeria, Europe and southern Australia. The PFT composition of land cover classes in Table 2.6 and Figure 2.12a show that globally the Crop PFT contributed 52% to Cropland and 24% to the Cropland Mosaic. The CLM5 Crop PFT mapping from EarthStat 2000, MIRCA 2000 and FAOSTAT products also had substantial contribution to other land cover types highlighting the differences in mapping between purely remotely sensed MODIS land cover data and those guided by country and regional inventories as discussed in Chapter 3. The Whittaker analysis of the Crop PFT in Figure 2.21d shows there are two climate hotspots for crops. The first is in temperate climates between 2°C and 18°C with precipitation between 400 and 1200 mm/year. The second is between 22°C and 28°C with precipitation between 500 and 1500 mm/year.

The 1 km high resolution CLM5 PFT data can be directly used for generating current day CLM5 surface data using the mksurfdata tool as described in Chapter 8. For CMIP6 and most other supported configurations of CLM5 however, the high resolution PFT data is used in combination with the high resolution CFT data described in Chapter 4, and the LUMIP LUH2 land use time series data of Hurtt et al. (2020), described in Chapter 5, to generate CLM5 surface and land use time series data sets. All of these component data are brought together using the CLM5 Land Use Data Tool as detailed in Chapter 6, 7 and 8, with the final raw data sets being used in the mksurfdata tool.

CHAPTER 3.

CURRENT DAY AGRICULTURE DATA

3.1 Current Day CLM5 Agriculture Overview

The CLM5 model has the option to represent agricultural land use explicitly through the CLM5 Crop model. A full description of the Crop model is provided in the CLM5 Technote of Lawrence et al. (2018), with operationally details in Lombardozzi et al. (2019), and more background and details in Levis et al (2016). As introduced in Chapter 1, CLM5 with the Crop model on explicitly simulates crops on a managed Crop land unit that coexists with the natural vegetation of a grid cell. The managed crops are represented through the 32 Crop Functional Types (CFTs) listed in Table 3.1, with each CFT represented in both Rainfed and Irrigated forms. The original CFT list was derived from the crops of the MIRCA 2000 crop mapping of Portmann et al (2010). In CLM5 the individual CFTs of the managed Crop land unit each have a separate soil column, unlike the default shared soil column of the Naturally Vegetated land unit. The individual soil columns permit different annual land management practices on the CFTs including the application of irrigation and nitrogen fertilizer.

CLM5 includes eight actively managed crop types (Temperate Soybean, Tropical Soybean, Temperate Corn, Tropical Corn, Spring Wheat, Cotton, Rice, and Sugarcane) that are chosen based on the availability of corresponding algorithms in AgroIBIS and as extended by Badger and Dirmeyer (2015), and further described in Levis et al. (2016). The representations of Sugarcane, Rice, Cotton, Tropical Corn, and Tropical Soy are new in CLM5. Sugarcane and Tropical Corn are both C4 plants and are therefore represented using the Temperate Corn functional form. Tropical Soybean uses the Temperate Soybean functional form with tropical parameters, while Rice and Cotton use the Spring Wheat functional form with new parameters. The CLM5 list of CFTs also includes 23 inactive crops that do not yet have the associated parameters required for active management. Each of the inactive crop types is simulated using the parameters of the spatially closest associated crop type that is most similar to the functional type (e.g., C3 or C4), which is required to maintain similar phenological parameters based on climate thresholds as described at the end of this chapter.

To support the use of the CLM5 Crop model global 1 km data sets have been developed for the current day from the MIRCA 2000 irrigated and rainfed crop mapping of Portmann et al (2010), combined with the EarthStat 2000 global harvested area mapping of 175 crops based on the work of Monfreida et al. (2008) and Ray et al. (2012). These global 1 km data sets have then been combined with the United Nations Food and Agricultural Organization (UNFAO) FAOSTAT (2016) countries of the world database to produce an annual time series of crop mapping for the period 1961 to 2016 in the MIRCA crop classes shown in Table 3.2.

Table 3.1: Crop Functional Types (CFTs) included with the CLM5 BGC CROP option.

Patch ID	Crop Function Types (CFTs)	MIRCA ID	Managed	Crop Parameters Used
15	Unmanaged Crop – Rainfed	24, 26, 27	None	Not Applicable
16	Unmanaged Crop – Irrigated	24, 26, 27	None	Not Applicable
17	Temperate Corn - Rainfed	2*	Active	Temperate Corn – Rainfed
18	Temperate Corn – Irrigated	2*	Active	Temperate Corn – Irrigated
19	Spring Wheat – Rainfed	1	Active	Spring Wheat – Rainfed
20	Spring Wheat – Irrigated	1	Active	Spring Wheat – Irrigated
21	Winter Wheat – Rainfed	-	Substitute	Spring Wheat – Rainfed
22	Winter Wheat – Irrigated	-	Substitute	Spring Wheat – Irrigated
23	Temperate Soybean - Rainfed	8*	Active	Temperate Soybean – Rainfed
24	Temperate Soybean – Irrigated	8*	Active	Temperate Soybean – Irrigated
25	Barley – Rainfed	4	Substitute	Spring Wheat – Rainfed
26	Barley – Irrigated	4	Substitute	Spring Wheat – Irrigated
27	Winter Barley – Rainfed	-	Substitute	Spring Wheat – Rainfed
28	Winter Barley – Irrigated	-	Substitute	Spring Wheat – Irrigated
29	Rye – Rainfed	5	Substitute	Spring Wheat – Rainfed
30	Rye – Irrigated	5	Substitute	Spring Wheat – Irrigated
31	Winter Rye – Rainfed	-	Substitute	Spring Wheat – Rainfed
32	Winter Rye – Irrigated	-	Substitute	Spring Wheat – Irrigated
33	Cassava – Rainfed	11	Substitute	Rice – Rainfed
34	Cassava – Irrigated	11	Substitute	Rice – Irrigated
35	Citrus – Rainfed	18	Substitute	Spring Wheat – Rainfed
36	Citrus – Irrigated	18	Substitute	Spring Wheat – Irrigated
37	Cocoa – Rainfed	22	Substitute	Rice – Rainfed
38	Cocoa – Irrigated	22	Substitute	Rice – Irrigated
39	Coffee – Rainfed	23	Substitute	Rice – Rainfed
40	Coffee – Irrigated	23	Substitute	Rice – Irrigated
41	Cotton – Rainfed	21	Active	Cotton – Rainfed
42	Cotton – Irrigated	21	Active	Cotton – Irrigated
43	Datepalm – Rainfed	19	Substitute	Cotton – Rainfed
44	Datepalm – Irrigated	19	Substitute	Cotton – Irrigated
45	Foddergrass – Rainfed	25	Substitute	Spring Wheat – Rainfed
46	Foddergrass – Irrigated	25	Substitute	Spring Wheat – Irrigated
47	Grapes – Rainfed	20	Substitute	Spring Wheat – Rainfed
48	Grapes – Irrigated	20	Substitute	Spring Wheat – Irrigated
49	Groundnuts – Rainfed	16	Substitute	Rice – Rainfed
50	Groundnuts – Irrigated	16	Substitute	Rice – Irrigated
51	Millet – Rainfed	6	Substitute	Tropical Corn – Rainfed
52	Millet – Irrigated	6	Substitute	Tropical Corn – Irrigated
53	Oilpalm – Rainfed	14	Substitute	Rice – Rainfed
54	Oilpalm – Irrigated	14	Substitute	Rice – Irrigated
55	Potatoes – Rainfed	10	Substitute	Spring Wheat – Rainfed
56	Potatoes – Irrigated	10	Substitute	Spring Wheat – Irrigated
57	Pulses – Rainfed	17	Substitute	Spring Wheat – Rainfed
58	Pulses – Irrigated	17	Substitute	Spring Wheat – Irrigated
59	Rapeseed – Rainfed	15	Substitute	Spring Wheat – Rainfed
60	Rapeseed – Irrigated	15	Substitute	Spring Wheat – Irrigated
61	Rice – Rainfed	3	Active	Rice – Rainfed
62	Rice – Irrigated	3	Active	Rice – Irrigated
63	Sorghum – Rainfed	7	Substitute	Tropical Corn – Rainfed

64	Sorghum – Irrigated	7	Substitute	Tropical Corn – Irrigated
65	Sugarbeet – Rainfed	13	Substitute	Spring Wheat – Rainfed
66	Sugarbeet – Irrigated	13	Substitute	Spring Wheat – Irrigated
67	Sugarcane – Rainfed	12	Active	Sugarcane – Rainfed
68	Sugarcane – Irrigated	12	Active	Sugarcane – Irrigated
69	Sunflower – Rainfed	9	Substitute	Spring Wheat – Rainfed
70	Sunflower – Irrigated	9	Substitute	Spring Wheat – Irrigated
71	Miscanthus - Rainfed	-	Substitute	Tropical Corn – Rainfed
72	Miscanthus – Irrigated	-	Substitute	Tropical Corn – Irrigated
73	Switchgrass – Rainfed	-	Substitute	Tropical Corn – Rainfed
74	Switchgrass – Irrigated	-	Substitute	Tropical Corn – Irrigated
75	Tropical Corn – Rainfed	2*	Active	Tropical Corn – Rainfed
76	Tropical Corn – Irrigated	2*	Active	Tropical Corn – Irrigated
77	Tropical Soybean – Rainfed	8*	Active	Tropical Soybean – Rainfed
78	Tropical Soybean – Irrigated	8*	Active	Tropical Soybean – Irrigated

*Temperate varieties were assigned poleward of 30N and 30S. Tropical varieties were assigned equatorward.

3.2 MIRCA 2000 Current Crop Mapping

The MIRCA 2000 present-day irrigated and rainfed crop areas dataset of Portmann et al. (2010) consist of global area mapping for all major food crops of the world, as well as fodder grass and other perennial, annual and fibre crops. The MIRCA 2000 data set consists of global maps for the 27 crop types listed in Table 3.2 on a five-arc-minute (~10 km) grid. The MIRCA 2000 maps are generated from the 175 crop type global mapping of Monfreida et al. (2008).

The original data for both of these products was the gridded maps of global croplands for the year 2000 by Ramankutty et al. (2008), combined with census based inventories and crop calendars for thousands of individual administrative units throughout the world from a wide range of national and international organizations. To use the data with the 1 km resolution MODIS, AVHRR, and CRU derived CLM5 land cover mapping, the 10 km data sets were spatially interpolated to the MODIS 1 km grid and land mask.

For CLM 4.5 the MIRCA crop data was used to directly prescribe the Crop model crop functional types (CFTs) using the relationships shown in Table 3.1. For CLM5 the mapping from MIRCA 2000 crop types to CLM Crop CFTs was maintained however additional data was included to generate yield, updated mapping, and time series data from the EarthStat 2000 and UN FAOSTAT (2016) data sets as described in the following sections. The MIRCA crop types are directly comparable with the EarthStat 2000 and FAOSTAT global data sets as shown in Table 3.2, with the global area of each crop type compared to these products for the years 1961, 2000 and 2016 in Table 3.3 and Figure 3.1.

Table 3.2: MIRCA, EarthStat and FAOSTAT Crop lists.

ID	MIRCA Crop	EarthStat List	FAOSTAT List
1	Wheat	Wheat	Wheat
2	Maize	Greencorn; Maize; Popcorn	Maize; Maize Green
3	Rice	Rice	Rice
4	Barley	Barley	Barley
5	Rye	Rye	Rye
6	Millet	Millet	Millet
7	Sorghum	Sorghum	Sorghum
8	Soybeans	Soybean	Soybeans
9	Sunflower	Sunflower	Sunflower
10	Potatoes	Potato	Potatoes
11	Cassava	Cassava	Cassava; Cassava Leaves
12	Sugarcane	Sugarcane; Sugarnes	Sugar Cane; Sugar nes
13	Sugarbeet	Sugarbeet	Sugar Beet
14	Oilpalm	Oilpalm	Oil Palm
15	Rapeseed	Rapeseed	Rapeseed
16	Groundnuts	Groundnut	Groundnuts
17	Pulses	Bean; Broadbean; Chickpea; Cowpea; Greenbean; Greenbroadbean; Greenpea; Legumenes; Lentil; Lupin; Pea; Pigeonpea; Pulsenes; Stringbean	Beans Dry; Beans Green; Broad Beans; Chick Peas; Cow Peas; Lentils; Lupins; Peas Dry; Peas Green; Pigeon Peas; Pulses nes; String Beans
18	Citrus	Citrusnes; Grapefruitetc; Lemonlime; Orange; Tangetc	Fruit Citrus; Grapefruit; Lemons; Oranges; Tangerines
19	Datepalm	Date	Dates
20	Grapes	Grape	Grapes
21	Cotton	Cotton	Seed Cotton
22	Cocoa	Cocoa	Cocoa
23	Coffee	Coffee	Coffee
24	Other Perennial	Almond; Apple; Apricot; Areca; Avocado; Banana; Berrynes; Blueberry; Brazil; Carob; Cashewapple; Cashew; Castor; Cherry; Chestnut; Cinnamon; Clove; Coconut; Cranberry; Currant; Fig; Fruitnes; Gooseberry; Gums; Hazelnut; Karite; Kiwi; Kolanut; Mango; Mate; Mustard; Nutmeg; Nutnes; Olive; Papaya; Peachetc; Pear; Pepper; Persimmon; Pistachio; Plantain; Plum; Quince; Raspberry; Sourcherry; Spicenes; Stonefruitnes; Taro; Tea; Tropicalnes; Tung; Vanilla; Walnut; Yautia	Almonds; Apples; Apricots; Areca; Avocados; Bananas; Berries; Blueberries; Brazil; Carobs; Cashew; Cashewapple; Castor; Cherries; Cherries Sour; Chestnut; Cinnamon; Cloves; Coconuts; Cranberries; Currants; Figs; Fruit Fresh; Fruit Pome; Fruit Stone; Fruit Tropical; Gooseberries; Gums; Hazelnuts; Jojoba; Karite; Kiwi; Kola; Mangoes; Mate; Mustard; Nutmeg; Nuts; Olives; Papayas; Peaches; Pears; Pepper; Persimmons; Pistachios; Plantains; Plums; Quinces; Raspberries; Spices; Tallowtree Seed; Taro; Tea; Tung Nuts; Vanilla; Walnuts; Yautia
25	Foddergrass	Oats	Oats

26	Other Annual	Aniseetc; Artichoke; Asparagus; Bambara; Buckwheat; Cabbage; Canaryseed; Carrot; Cauliflower; Cerealnes; Chicory; Chilleetc; Cucumberetc; Eggplant; Fonio; Garlic; Ginger; Greenonion; Hempseed; Hop; Lettuce; Linseed; Melonetc; Melonseed; Mixedgrain; Mushroom; Oilseednes; Okra; Onion; Peppermint; Pimento; Pineapple; Poppy; Pumpkinetc; Pyrethrum; Quinoa; Rootnes; Safflower; Sesame; Spinach; Strawberry; Sweetpotato; Tobacco; Tomato; Triticale; Vegetablenes; Vetch; Watermelon; Yam	Anise; Artichokes; Asparagus; Bambara; Buckwheat; Cabbages; Canary Seed; Carrots; Cauliflowers; Cereals; Chicory; Chillies Dry; Chillies Green; Cucumbers; Eggplants; Fonio; Garlic; Ginger; Grain Mixed; Hempseed; Hops; Leeks; Lettuce; Linseed; Melons; Melonseed; Mushrooms; Oilseeds; Okra; Onions Dry; Onions Green; Peppermint; Pineapples; Poppy; Pumpkins; Pyrethrum; Quinoa; Roots and Tubers; Safflower; Sesame; Spinach; Strawberries; Sweet Potatoes; Tobacco; Tomatoes; Triticale; Vegetables Fresh; Vegetables Legume; Vetches; Watermelons; Yams
27	Other Fibre	Abaca; Agave; Coir; Fibrenes; Flax; Hemp; Jute; Jutelikefiber; Kapokfiber; Kapokseed; Ramie; Rubber; Sisal	Agave; Bastfibres; Coir; Fibre Crops nes; Flax; Hemp; Jute; Kapok Fibre; Kapok Fruit; Kapokseed; Manila Fibre; Ramie; Rubber; Sisal

3.3 EarthStat 2000 Current Crop Mapping

To allow for a more comprehensive cropping data set with extended and updated global agricultural data including yield and production data, the EarthStat 2000 Harvested Area and Yield maps for 175 crops from from the EarthStat website are included in the CLM5 CFT mapping. The EarthStat data are generated for the year 2000, based on the work of Monfreida et al. (2008) and Ray et al. (2012). The EarthStat project is a is a collaboration between the Global Landscapes Initiative (GLI) at the University of Minnesota's Institute on the Environment and the Land Use and Global Environment (LUGE) lab at the University of British Columbia.

The EarthStat 2000 data sets were created by combining national, state, and county level census statistics with a recently updated global data set of croplands on a five-arc-minute (~10 km) grid. The resulting land use data sets depict circa the year 2000 the area (harvested) and yield of the 175 distinct crops of the world. The crop types are directly related to the crops listed in the MIRCA mapping and the FAOSTAT crop inventories of the next section. The compatibility of these data sets is to be expected given their shared origins from Monfreida et al. (2008) and FAOSTAT for the year 2000.

Like the MIRCA 2000 data the 10 km data sets were spatially interpolated to the MODIS 1 km grid and land mask so they could be used with the 1 km resolution MODIS, AVHRR, and CRU derived CLM5 land cover mapping. To use the data with the CLM5 CFTs, the 175 crops from the EarthStat 2000 data set were also combined into the 28 MIRCA crop types as shown in Table 3.2. All eight of the MIRCA 2000 crop types that are actively modeled in CLM5 can be directly mapped from the EarthStat 2000 data. Another 17 of the MIRCA 2000 crop types can be directly mapped with only the five remaining general crop types needing to be aggregated from sub crops types. The global area of the EarthStat 2000 crops reclassified to the MIRCA crop types are compared to the original MIRCA 2000 mapping and the FAOSTAT database for the years 1961, 2000 and 2016 in Table 3.3 and in Figure 3.1.
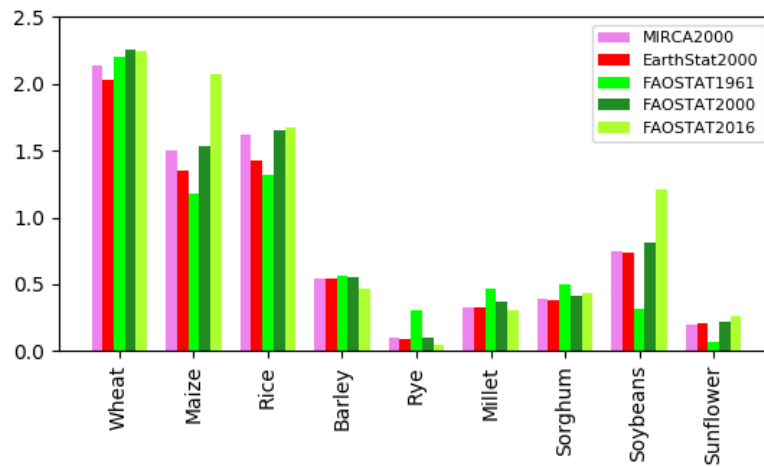
Table 3.3: Global Crop Area from MIRCA 2000, EarthStat 2000 and UN FAOSTAT 1961, 2000 and 2016 in millions of km^2.

MIRCA Crop	MIRCA 2000	EarthStat 2000	FAOSTAT 1961	FAOSTAT 2000	FAOSTAT 2016
Wheat	2.14	2.04	2.21	2.26	2.25
Maize	1.51	1.35	1.18	1.53	2.08
Rice	1.62	1.42	1.32	1.65	1.68
Barley	0.55	0.54	0.56	0.55	0.47
Rye	0.10	0.09	0.31	0.10	0.04
Millet	0.33	0.33	0.47	0.37	0.31
Sorghum	0.40	0.39	0.50	0.42	0.44
Soybeans	0.74	0.74	0.32	0.81	1.21
Sunflower	0.20	0.20	0.07	0.22	0.27
Potatoes	0.19	0.19	0.23	0.23	0.22
Cassava	0.15	0.15	0.10	0.17	0.22
Sugarcane	0.20	0.19	0.09	0.20	0.27
Sugarbeet	0.06	0.06	0.07	0.06	0.05
Oilpalm	0.09	0.09	0.03	0.10	0.20
Rapeseed	0.24	0.24	0.07	0.30	0.37
Groundnuts	0.22	0.22	0.17	0.26	0.29
Pulses	0.66	0.69	0.69	0.71	0.86
Citrus	0.07	0.07	0.02	0.08	0.11
Datepalm	0.01	0.01	0.00	0.01	0.01
Grapes	0.07	0.07	0.09	0.07	0.08
Cotton	0.33	0.30	0.34	0.33	0.30
Cocoa	0.07	0.07	0.04	0.08	0.10
Coffee	0.10	0.10	0.10	0.11	0.11
Otherperennial	0.72	0.64	0.27	0.74	0.93
Foddergrass	1.04	0.13	0.39	0.13	0.09
Otherannual	1.08	0.82	0.84	1.06	1.27
Otherfibre	0.72	0.11	0.10	0.11	0.14
All Crops	13.60	11.24	10.59	12.66	14.36
Irrigated	3.18	2.14	1.41	2.51	2.80

3.4 UN FAOSTAT Annual Harvested Crop Areas

To provide time series descriptions that capture recent changing agricultural practice, the MIRCA 2000 and EarthStat 2000 spatial mapping were combined with the annual country level cropping area and yield data from the FAOSTAT (2016) database for the years 1961 through 2016, to support the generation of annual CLM5 CFT mapping. The annual country level crop harvest, yield and irrigation data was used to scale up or down the EarthStat 2000 1 km data to achieve the FAOSTAT target values for each country, while maintaining the relative spatial distributions of each value. Like the EarthStat 2000 data, the crop and yield data for the 178 crop classes of the FAOSTAT database were aggregated into the MIRCA crop types using the relationships listed in Table 3.2.

(a) Global Crop Areas Part 1 (Millions km^2)

(b) Global Crop Areas Part 2 (Millions km^2)
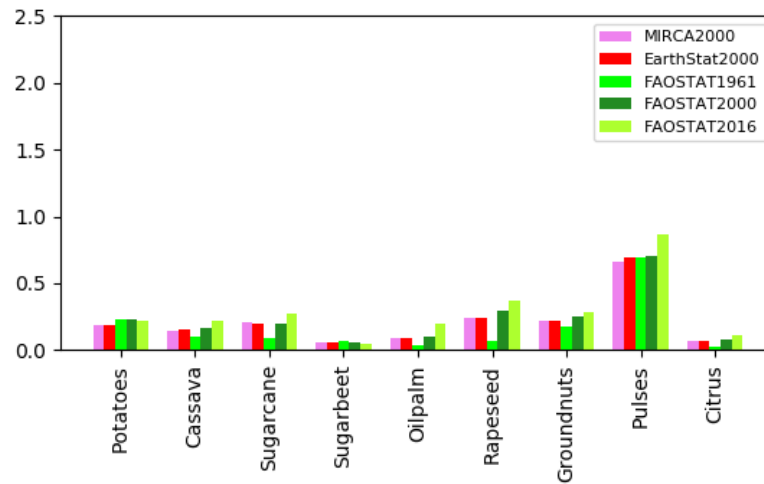
(c) Global Crop Areas Part 3 (Millions km^2)
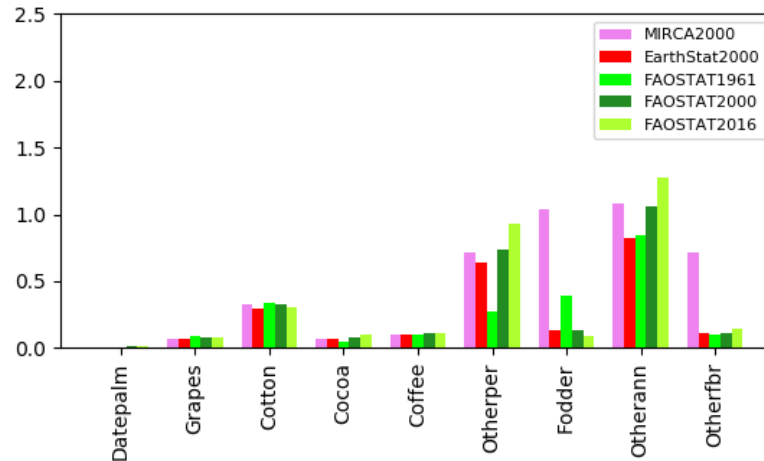
Figure 3.1: Global Crop Area from MIRCA 2000, EarthStat 2000 and FAOSTAT for the years 1961, 2000 and 2016.

The FAOSTAT scaled EarthStat 2000 product is referred here as the FAOEarthStat data for a given year. The global areas of the FAOEarthStat data in MIRCA crop types are shown for the years 1961, 2000 and 2016 in Table 3.3 and Figure 3.1 compared to the same MIRCA crop types from the MIRCA 2000 and EarthStat 2000 data sets. The total FAOEarthStat crop area for the year 2000 is 12.7 million km^2, which differs from both the MIRCA 2000 harvested area at 13.6 million km^2, and the EarthStat 2000 harvested area at 11.2 million km^2. The disagreement between the different products can be partially traced to differences in the definitions of harvested versus physical area in all of the products. The EarthStat 2000 harvested area includes multiple harvests and multi-cropping on the same physical land, while the MIRCA 2000 physical area takes into account cropping calendars to address the multi-cropping.

For the FAOEarthStat country crop area values mapped on to the spatial distributions of the EarthStat 2000 maps, strict area limits were applied to ensure that the physical area under all crops could not exceed 100% of a grid cell. This requirement reflected the crop representations in CLM5 which does not allow for multi-cropping, with a single growing season for each crop of a physical area of land. This difference in harvested area versus physical irrigated and rainfed crop land is also important when comparing to the CMIP6 / LUH2 crop mapping in Chapter 5, 6 and 7.

The other major source of differences between the products is from the Fodder Grass and Other Fibre crop classes, which are much smaller for EarthStat 2000, and FAOSTAT 2000 areas when compared to the MIRCA 2000. As Fodder Grass and Other Fibre crops do not have the same physiological life cycle as the crops simulated in the CLM5 Crop model, the lower contribution of these crops in the global data sets is beneficial, as these areas will be represented through natural grass and tree PFTs rather than remapped CFTs.

Of the major crops, Wheat is consistently the largest crop globally at around 2.0 million km^2 in all three data sets with a small increase in the FAOSTAT time series from 1961 to 2000 and then a small decline from 2000 to 2016. The next largest crop globally was Rice at around 1.6 million km^2 with the FAOSTAT time series showing a similar expansion from 1961 to 2000 and then leveling off from 2000 to 2016. The third largest crop globally was Maize at around 1.5 million km^2 for the year 2000. The FAOSTAT time series shows the rapid expansion of Maize from 1961 to 2016 nearly doubling over that time. The fourth largest crop globally was Soybeans at around 0.8 million km km^2 for the year 2000. Like Maize, Soybeans expanded rapidly from 1961 to 2016 increasing fourfold in that time.

Other Annual and Other Perennial crops also had large contributions collectively with 1.1 and 0.7 million km^2 respectively in the year 2000. The FAOSTAT data showed both of these general crop classes had large increases from 1961 to 2016. Smaller contributions came from Barley, Sorghum, Millet and Rye, with all of these crops showing large declines in area from 1961 to 2016. Minor crops such as Sunflower, Cassava, Sugarcane, Oilpalm, Rapeseed and Groundnuts showed large relative increases in area over the period. The remaining minor crops showed relative small changes in global area over the period.

3.5 CLM5 Annual Crop Functional Type (CFT) Generation

The global irrigated and rainfed crop mapping from the FAOEarthStat data was used to generate CLM5 annual 1 km resolution Crop Functional Type (CFT) mapping for each years in the time series 1961 to 2016. The MIRCA crop types were mapped to the CLM5 CFT based on the rules listed in Table 3.1. These rules also included the re-classification to temperate and tropical

varieties of Corn and Soybeans based on latitude relative to 30N and 30S, with temperate varieties poleward and tropical varieties equatorward.

Table 3.4: Current Day CLM5 Crop Functional Type (CFT) area for the year 2005 derived from EarthStat and FAOSTAT data in the MIRCA crop classes. Global and IPCC Region Areas in millions km^2.

CLM5 CFT	Global	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS
Unmanaged	1.91	0.38	0.01	0.58	0.05	0.21	0.11	0.03	0.04	0.26	0.24
Temp Corn	0.86	0.01	0.00	0.32	0.02	0.13	0.03	0.00	0.32	0.00	0.02
Wheat	2.21	0.09	0.14	0.30	0.41	0.43	0.09	0.10	0.30	0.00	0.33
Temp Soybean	0.58	0.00	0.00	0.13	0.01	0.01	0.12	0.00	0.31	0.00	0.00
Barley	0.55	0.05	0.05	0.01	0.11	0.22	0.01	0.04	0.05	0.00	0.01
Rye	0.07	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00
Cassava	0.17	0.11	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00
Citrus	0.08	0.01	0.00	0.03	0.00	0.01	0.02	0.00	0.00	0.00	0.01
Cocoa	0.08	0.06	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00
Coffee	0.10	0.02	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.02	0.00
Cotton	0.34	0.05	0.00	0.06	0.03	0.01	0.02	0.00	0.06	0.00	0.11
Datepalm	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fodder	0.11	0.00	0.01	0.00	0.04	0.04	0.01	0.00	0.02	0.00	0.00
Grapes	0.07	0.00	0.00	0.01	0.00	0.04	0.00	0.00	0.00	0.00	0.00
Groundnuts	0.25	0.09	0.00	0.06	0.00	0.00	0.01	0.00	0.01	0.02	0.07
Millet	0.35	0.20	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.12
Oilpalm	0.12	0.04	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.08	0.00
Potatoes	0.20	0.01	0.00	0.07	0.03	0.04	0.01	0.00	0.01	0.00	0.02
Pulses	0.74	0.19	0.02	0.08	0.01	0.04	0.07	0.01	0.04	0.04	0.24
Rapeseed	0.30	0.00	0.01	0.10	0.00	0.05	0.00	0.00	0.06	0.00	0.07
Rice	1.57	0.08	0.02	0.44	0.00	0.01	0.06	0.01	0.01	0.42	0.52
Sorghum	0.46	0.28	0.01	0.01	0.00	0.00	0.03	0.01	0.02	0.00	0.09
Sugarbeet	0.05	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.01	0.00	0.00
Sugarcane	0.20	0.01	0.00	0.02	0.00	0.00	0.09	0.00	0.00	0.02	0.04
Sunflower	0.24	0.01	0.00	0.02	0.06	0.08	0.02	0.00	0.01	0.01	0.03
Trop Corn	0.73	0.27	0.00	0.08	0.00	0.00	0.22	0.00	0.00	0.08	0.07
Trop Soybean	0.37	0.01	0.00	0.03	0.00	0.00	0.25	0.00	0.00	0.01	0.08
All Crop	12.71	1.96	0.28	2.40	0.82	1.39	1.27	0.22	1.28	1.01	2.07
Irrigated	2.52	0.09	0.03	0.54	0.11	0.23	0.15	0.10	0.24	0.21	0.82

The Current Day CLM5 CFT areas for the year 2005 are listed globally and for IPCC regions in Table 3.4. Global CLM5 CFT areas are also graphed in Figure 3.2, with the active CFTs mapped in Figures 3.3, 3.4, and 3.5, and the other crops which are substituted to active CFTs mapped in Figures 3.6, 3.7, 3.8, 3.9, 3.10, 3.11 and 3.12. The total global crop area of the CLM5 CFT mapping for 2005 was 12.7 million km^2, which is nearly identical to the area of FAOEarthStat for the year 2000. Unlike the FAOEarthStat data, the new CLM5 CFT mapping was impacted by other CLM5 land unit constraints from the Urban, Lakes and Glacier mapping that were developed in Chapter 2. This results in slightly lower values than the absolute areas for each land grid cell as defined from the MIRCA 2000 and EarthStat 2000 mapping. This is also the same total crop area reported in Chapter 2 for the current day Crop land unit. The Global Irrigated CFT area was 2.5 million km^2 was nearly identical to the FAOEarthStat year 2000 area as well. This represented just under 20% of all crops.

(a) CLM5 Raw CFT Area (Millions km^2)

(b) CLM5 Raw Other CFT Area (Millions km^2)
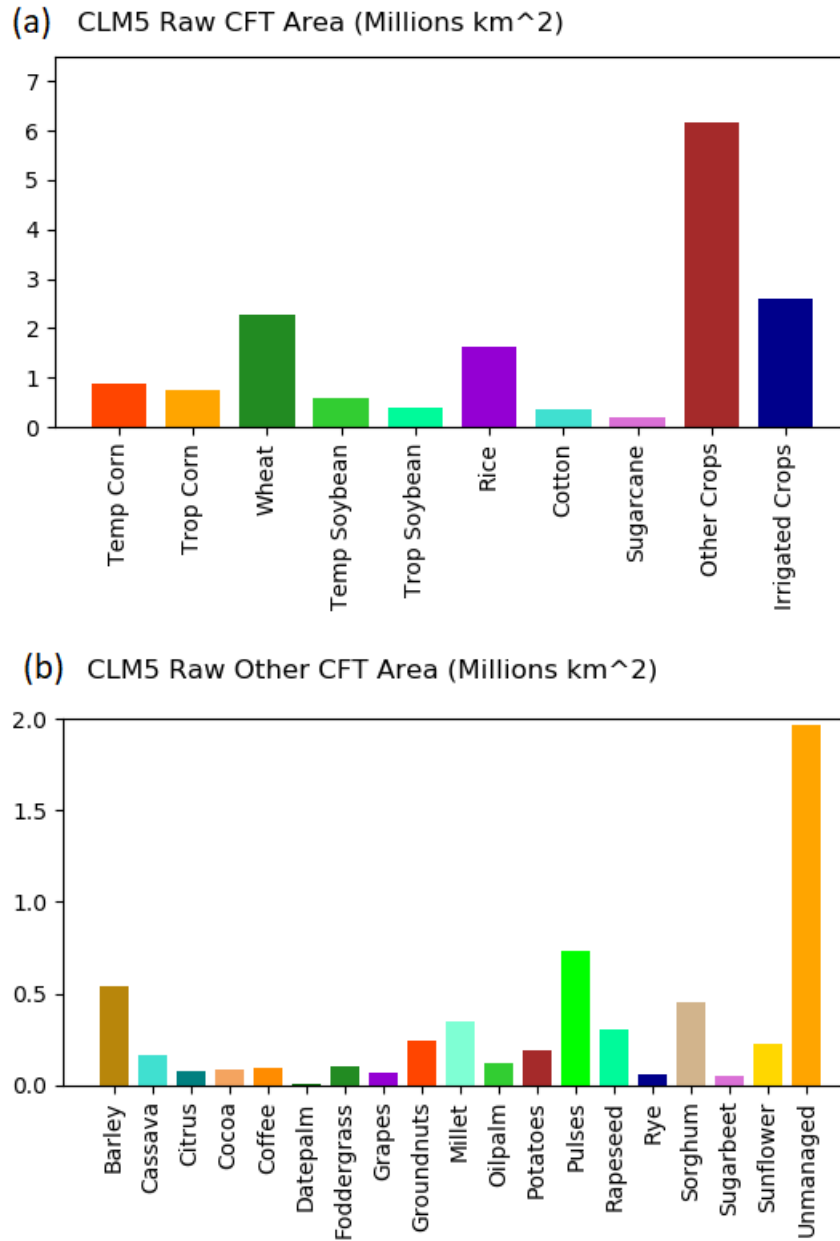
Figure 3.2: Global CLM5 Crop Functional Type (CFT) Area for the year 2005 for: (a) Crops actively represented; and (b) Other crops not currently represented but simulated with substituted parameters or unmanaged.

The largest area of CLM5 CFT crops for the year 2005 was in Eastern Asia, which is dominated by China, with 2.4 million km^2. The next largest area of crops was in Southern Asia, which is dominated by India, Pakistan and Bangladesh, at 2.1 million km^2. The third largest crop area was in Africa at 2.0 million km^2. The next three regions for cropping areas were Europe, North America and Latin America with all just under 1.4 million km^2. The CLM5 active crops made up just under 6.9 million km^2, or around 54% of crops. The remaining crops that needed to be substituted with active crop parameters made up the remaining 46%. These distributions directly reflect the CLM5

Crop land unit mapping described in Chapter 2 and shown in Figure 2.18 as they are from the same data.

The CLM5 CFT irrigated areas were largest in Southern Asia with 0.82 million km, followed by Eastern Asia at 0.54 million km^2 and North America at 0.24 million km^2. The remaining regions of Europe, Southeast Asia and Latin America all were around 0.2 million km^2. The percentage of irrigated crop was largest in the Middle East at around 45% of all crops. The next highest was Southern Asia at just under 40% going down regionally to Africa which had just under 5% irrigation for all crops. The global map of irrigated areas are shown in Figure 3.5c. The map shows the distinct regions of irrigation for the Ganges and Indus river plains of Southern Asia, as well as the Yangtze and Yellow river plains of Eastern Asia, and the Ogallala aquifer of North America.

The current day 1961 to 2016 CLM5 1 km CFT data can be directly combined with the land unit and PFT mapping from Chapter 2 to produce annual current day 1km resolution surface data for CLM5 for the period. Alternatively, the current day CFT data can be used with the LUMIP / CMIP6 historical and future LUH2 land use time series data to describe the five LUH2 crop land use types, which are then used in the CLM5 Land Use Data Tool as described in Chapter 6, 7 and 8. This second use with the LUH2 time series data is the most common application.

3.6 CLM5 Active Crop Functional Types (CFTs)

The largest active CLM5 CFT area was Wheat at 2.2 million km^2, reflecting the 2.3 million km^2 in the FAOEarthStat data for the year 2005. The largest areas of Wheat were in Europe followed by Eurasia and Southern Asia, with smaller areas in North America and East Asia. The global distribution of Wheat is shown in Figure 3.3c. The map shows the large areas of Wheat in the Indus and Ganges river plains, eastern China, the Steppes of Russia, across Europe, the Great Plains of North America, and south west and south east of Australia.

The second largest active cropping area in the FAOEarthStat data for the year 2005 was Maize at just over 1.6 million km^2. In the CLM5 CFT mapping this was split into Temperate Corn with 0.9 million km^2 and Tropical Corn with 0.7 million km^2. The largest areas of Temperate Corn were found in North America and East Asia, with smaller amounts in Europe. The largest areas of Tropical Corn were found in Africa then Latin America, with smaller amounts in other tropical regions. The CLM5 Temperate Corn CFT is mapped in Figure 3.3a, with the Tropical Corn CFT mapped in Figure 3.3b. The Temperate Corn mapping shows the main areas were in the Midwest of the United States, Europe, and northeastern China. The Tropical Corn mapping shows the wide distribution of corn throughout the tropics with large areas in South America, Africa, India and Southeast Asia.

The third largest active cropping area in the FAOEarthStat data for the year 2005 was Rice at just over 1.6 million km^2. In the CLM5 CFT mapping this resulted in a global area of just under 1.6 million km^2. The largest areas were in Southern Asia, East Asia, and Southeast Asia, with much smaller areas in other regions. The CLM5 Rice CFT is mapped in Figure 3.4c. The global mapping show that the Rice CFT was concentrated in eastern India into Bangladesh, Thailand, Indonesia and eastern China with large amounts of the Rice CFT grown around these areas. Outside of these areas, the Rice CFT was relatively sparse as a crop.

(a) EARTHSTAT / FAOSTAT CLM5 Temperate Corn   (%)
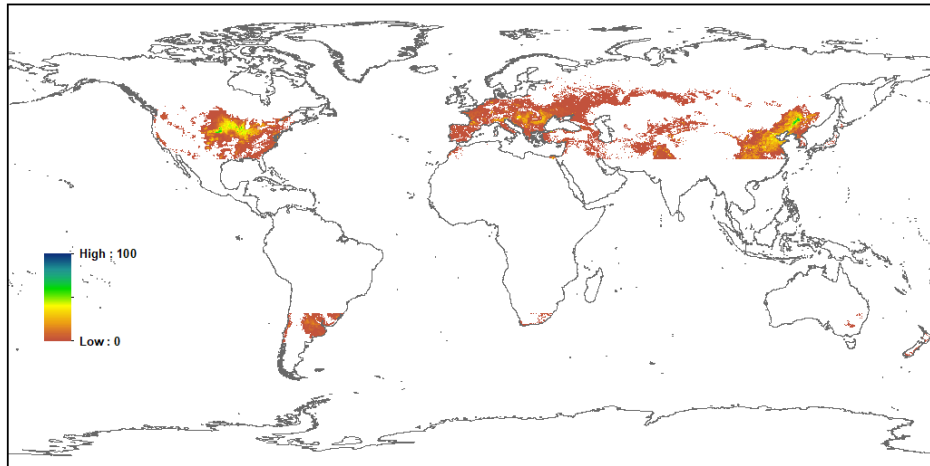
(b) EARTHSTAT / FAOSTAT CLM5 Tropical Corn   (%)
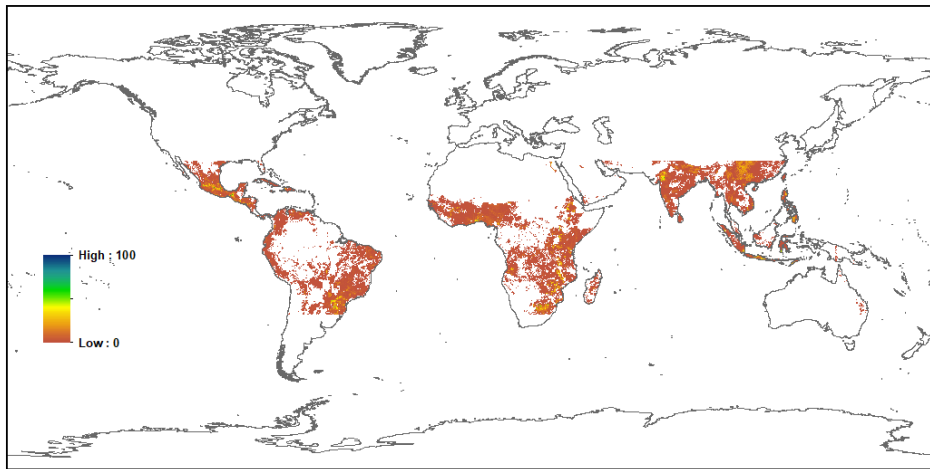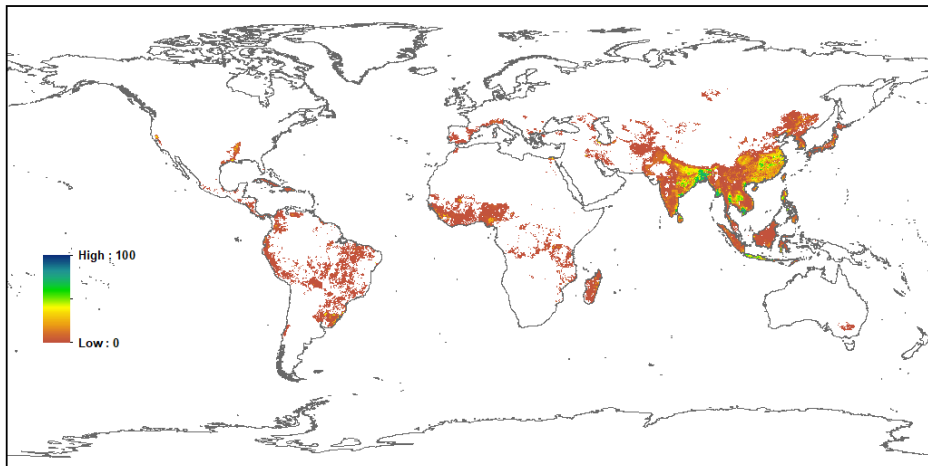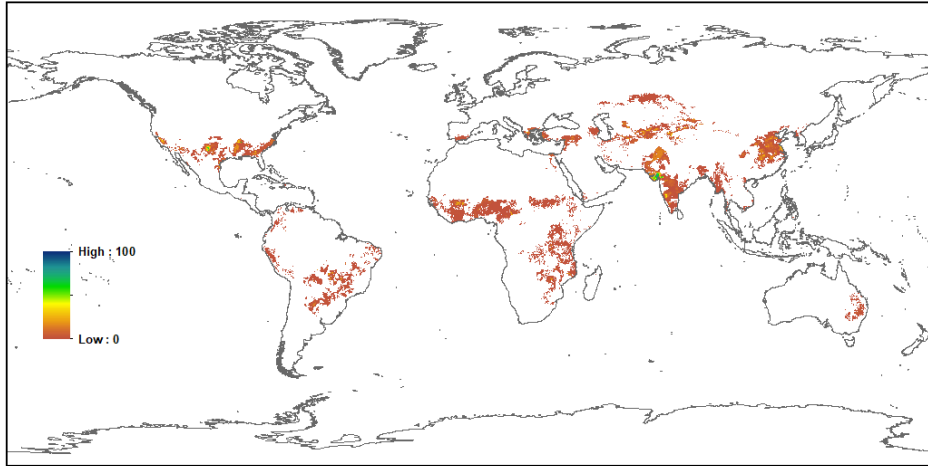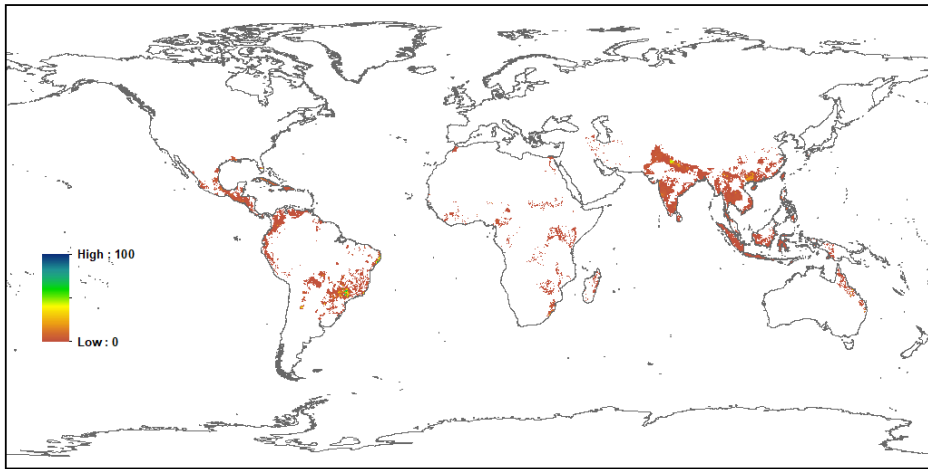
(c) EARTHSTAT / FAOSTAT CLM5 Wheat   (%)
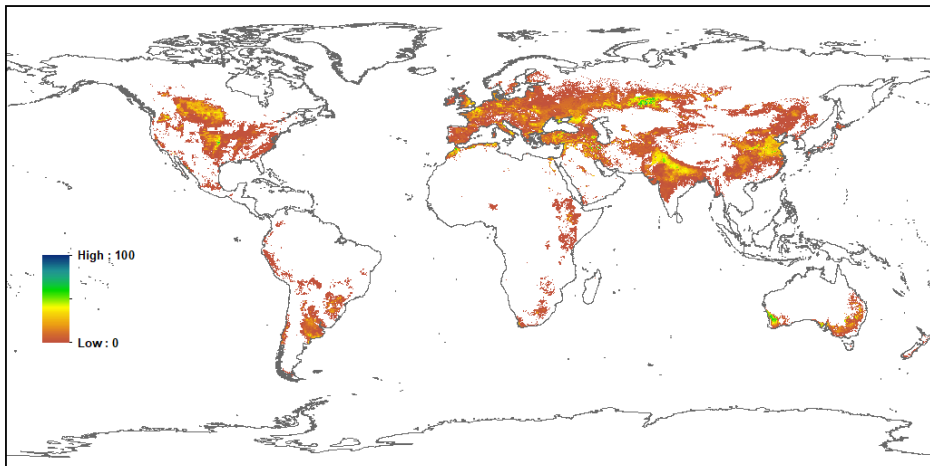
Figure 3.3: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Temperate Corn; (b) Percent Tropical Corn; and (c) Percent Wheat.

(a) EARTHSTAT / FAOSTAT CLM5 Temperate Soybean   (%)
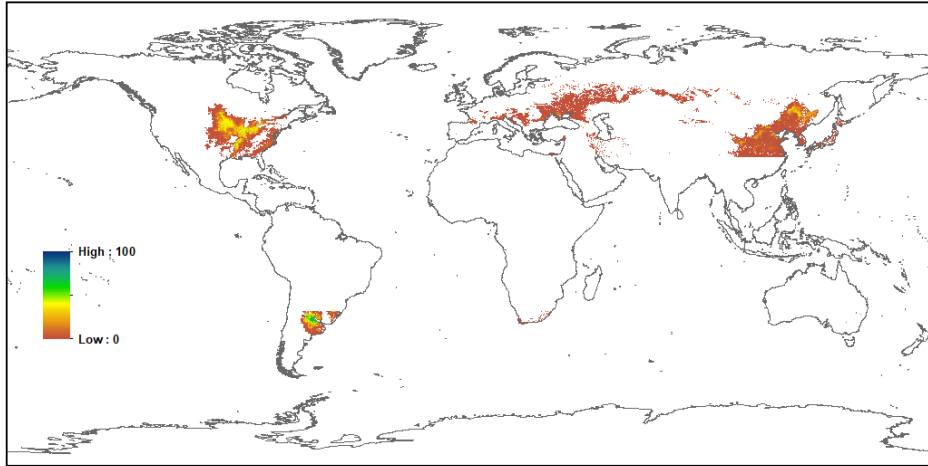
(b) EARTHSTAT / FAOSTAT CLM5 Tropical Soybean   (%)
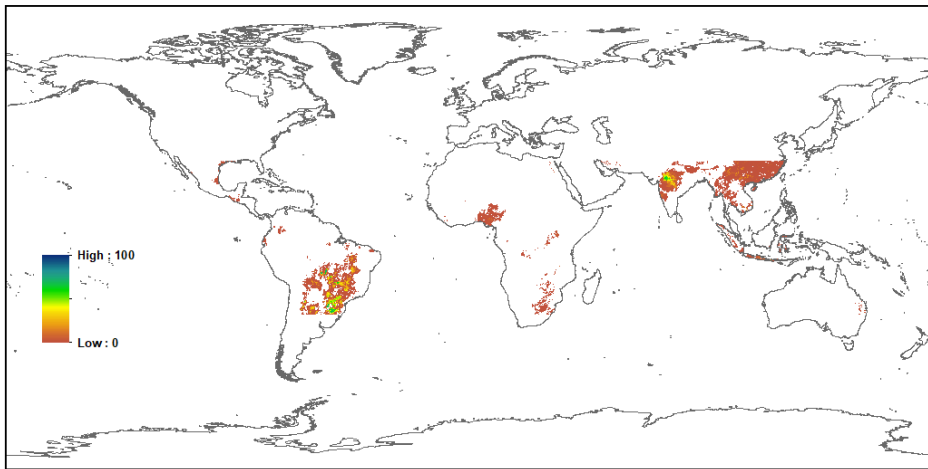
(c) EARTHSTAT / FAOSTAT CLM5 Rice   (%)

Figure 3.4: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Temperate Soybean; (b) Percent Tropical Soybean; and (c) Percent Rice.

(a) EARTHSTAT / FAOSTAT CLM5 Cotton (%)

(b) EARTHSTAT / FAOSTAT CLM5 Sugarcane (%)

(c) EARTHSTAT / MIRCA 2000 / FAOSTAT CLM5 Irrigated Crop (%)
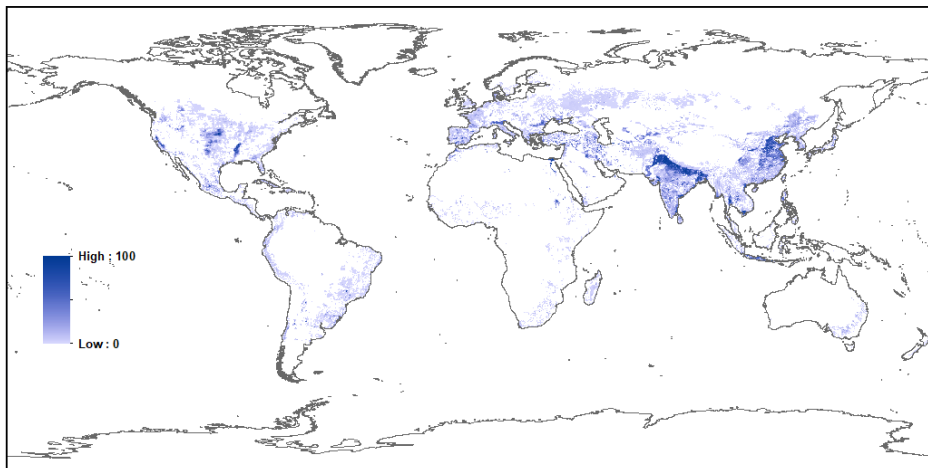
Figure 3.5: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Cotton; (b) Percent Sugarcane; and (c) Percent Irrigated Crops.

The fourth largest active cropping area in the FAOEarthStata data for the year 2005 was Soybeans at just under a 1.0 million km^2. In the CLM5 CFT mapping this was split into Temperate Soybean at just under 0.6 million km^2, and Tropical Soybean at just under 0.4 million km^2. The largest areas of the Temperate Soybean CFT were in North America, East Asia and Latin America. The largest areas of the Tropical Soybean CFT were in Latin America with smaller areas in Southern Asia and East Asia. The CLM5 Temperate Soybean CFT is mapped in Figure 3.4a, with the Tropical Soybean CFT mapped in Figure 3.4b. The Temperate Soybean mapping shows the main areas were in the Midwest of the United States, northeastern China and northeastern Argentina. The Tropical Soybean mapping shows concentrations in southeastern Brazil, northwestern India, and southern China.

The last two areas of active crops were Cotton and Sugarcane, which had just under 0.4 and 0.2 million km^2 respectively in both the FAOEarthStat and CLM5 CFT mapping for the year 2005. The Cotton CFT had largest areas in Southern Asia, East Asia, North America, and Africa. The Sugarcane CFT had largest areas in Latin America, Southern India, Southeast Asia and Eastern Asia. The CLM5 Cotton CFT is mapped in Figure 3.5a, with the Sugarcane CFT mapped in Figure 3.5b. The Cotton mapping shows concentrations in northeastern and central India, eastern China, the southern United States, with smaller area distributed globally. The Sugarcane mapping shows concentrations in southeastern Brazil, central America into Columbia, Venezuela and Ecuador, India into Southeast Asia, and northeastern Australia.

3.7 CLM5 Other Crop Functional Types (CFTs)

The largest area of the other crop CFTs was the Unmanaged Crop at 1.9 million km^2 for the year 2005. This was composed of the Other Perennial, Other Annual, and Other Fibre in the MIRCA crop types, which were themselves combined from over a hundred individual crops in the FAOSTAT and EarthStat 2000 data. The Unmanaged Crop CFT accounted for 15% of all crop area from the FAOEarthStat data for 2005. The Unmanaged Crop CFT had the largest areas in East Asia, Africa, Southeast Asia, Southern Asia, and Europe, with other regions having substantially smaller contributions from the residual crop type.

The CLM5 Unmanaged Crop CFT mapping is shown in Figure 3.12a. The mapping shows the largest concentrations of Unmanaged Crop CFT were in China, India, Southeast Asia, Africa and Europe. The remaining Unmanaged Crop mapping was distributed globally reflecting the wide range of crops not included explicitly in the MIRCA crop types. Due to the large range of crops combined in this CFT and the wide spatial and climate range over which they occur, there are no substitutions for the parameters and physiology of the Unmanaged Crop CFT and they are simulated as natural grasses outside of the CLM5 Crop model.

The second largest other crop CFT area was Pulses which had just over 0.7 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The largest areas of the Pulses CFT were in Southern Asia then Africa, with smaller areas in the other regions. The CLM5 Pulses CFT mapping is shown in Figure 3.10a. The mapping shows concentrations of Pulses in India, northeastern China, Africa and Canada. The remaining Pulses CFT mapping is distributed widely at lower concentrations. The Pulses CFT is simulated in the CLM5 Crop model using parameters and physiology from the Spring Wheat CFT.

The third largest other crop CFT area was Barley which had just under 0.6 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The largest areas of the Barley CFT were in Europe, then Eurasia, with smaller areas in North America, Africa, Asia Pacific and the Middle

East. The CLM5 Barley CFT mapping is shown in Figure 3.6a. The mapping shows the distribution of Barley across temperate latitudes in the Northern Hemisphere concentrated on Europe into Eurasia and with limited area in the Southern Hemisphere predominantly in Australia. The Barley CFT is simulated in the CLM5 Crop model using parameters and physiology from the Spring Wheat CFT.

The fourth largest other crop CFT area was Sorghum which had just under 0.5 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Sorghum CFT areas were primarily found in Africa and Southern Asia, with smaller areas in Latin America and North America. The CLM5 Sorghum CFT mapping is shown in Figure 3.11a. The mapping shows the Sorghum CFT was concentrated in Nigeria, Ethiopia, Sudan and central India, with smaller areas in the central United States and northeastern China. The Sorghum CFT is simulated in the CLM5 Crop model using parameters and physiology from the Tropical Corn CFT.

The fifth largest other crop CFT area was Millet which had just under 0.4 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Millet CFT areas were primarily found in Africa and Southern Asia, with a small area in East Asia. The CLM5 Millet CFT mapping is shown in Figure 3.9a. The mapping shows the Millet CFT was concentrated in Nigeria and India with small areas found in northeastern China, Ukraine and Russia. The Millet CFT is simulated in the CLM5 Crop model using parameters and physiology from the Tropical Corn CFT.

The sixth largest other crop CFT area was Rapeseed which had just over 0.3 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Rapeseed CFT areas were primarily in East Asia, with smaller areas in Southern Asia, North America and Europe. The CLM5 Rapeseed CFT mapping is shown in Figure 3.10b. The mapping show the Rapeseed CFT was concentrated in northern India, China, Europe and Canada, with a small area in southwestern and southeastern Australia. The Rapeseed CFT is simulated in the CLM5 Crop model using parameters and physiology from the Spring Wheat CFT.

The seventh largest other crop CFT area was Groundnuts which had just under 0.3 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Groundnuts CFT was found primarily in Africa, Southern Asia and East Asia. The CLM5 Groundnuts CFT mapping is shown in Figure 3.8c. The mapping shows low concentrations through Africa, India, China and Indonesia. The Groundnut CFT is simulated in the CLM5 Crop model using parameters and physiology from the Rice CFT.

The eighth largest other crop CFT area was Sunflower which had just over 0.2 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Sunflower CFT was found primarily in Europe and Eurasia with smaller areas in Southern Asia, Latin America and East Asia. The CLM5 Sunflower CFT mapping is shown in Figure 3.11c. The mapping shows large concentrations in Ukraine into Russia, France and Spain, southern India, north eastern China, Argentina and the central North America. The Sunflower CFT is simulated in the CLM5 Crop model using parameters and physiology from the Spring Wheat CFT.

The ninth largest other crop CFT area was Potatoes which had just over 0.2 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Potatoes CFT was found primarily in Eastern Asia, with smaller areas in Europe, Eurasia and Southern Asia. The CLM5 Potatoes CFT mapping is shown in Figure 3.9c. The mapping shows the three major areas of China, Europe into Russia and northern India. The Potatoes CFT is simulated in the CLM5 Crop model using parameters and physiology from the Spring Wheat CFT.

(a) EARTHSTAT / FAOSTAT CLM5 Barley　　　　　　　　　　　(%)

High : 100

Low : 0
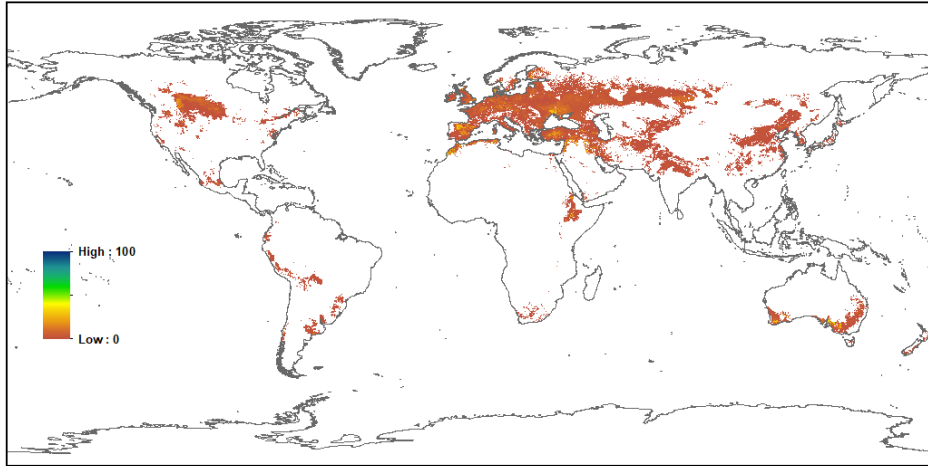
(b) EARTHSTAT / FAOSTAT CLM5 Cassava　　　　　　　　　　(%)

High : 100

Low : 0
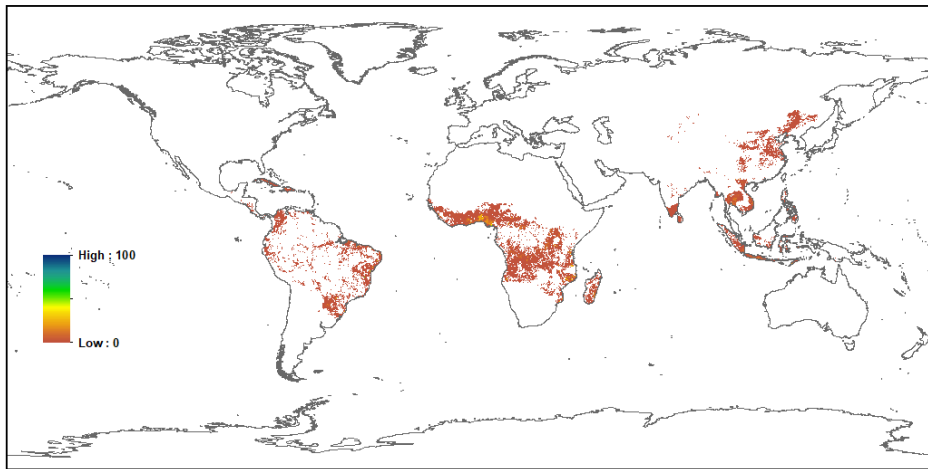
(c) EARTHSTAT / FAOSTAT CLM5 Citrus　　　　　　　　　　(%)
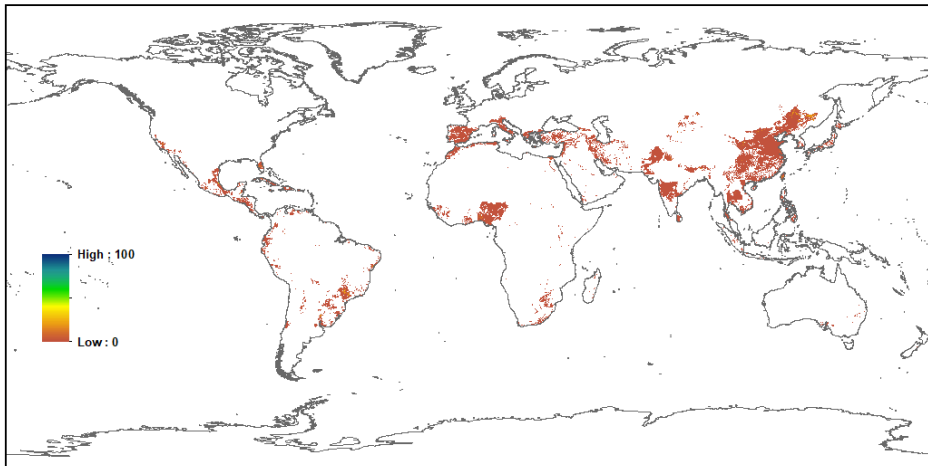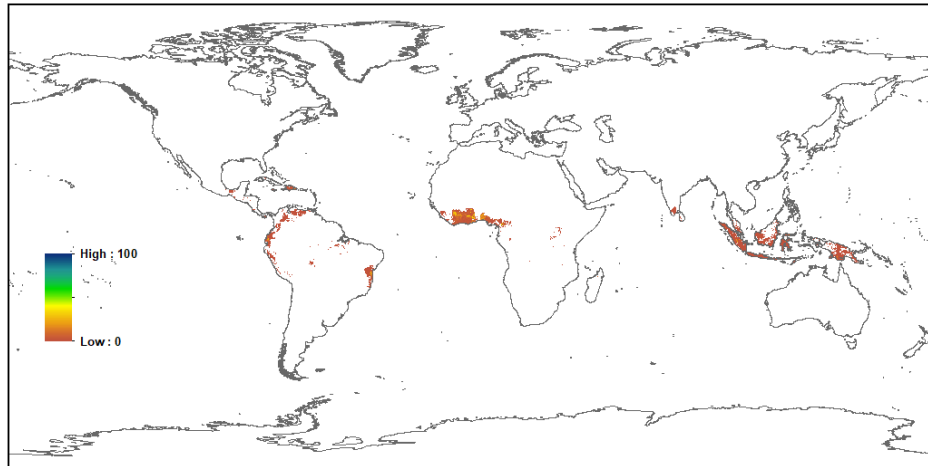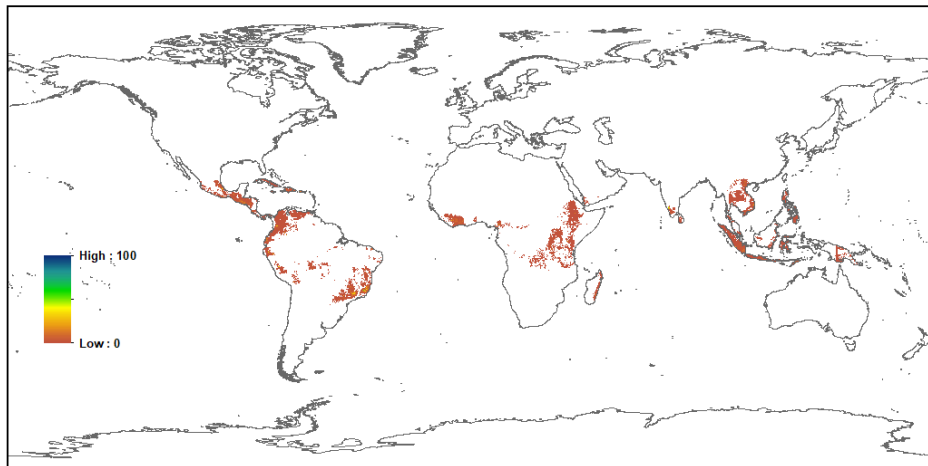
High : 100

Low : 0

Figure 3.6: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Barley; (b) Percent Cassava; and (c) Percent Citrus.

(a) EARTHSTAT / FAOSTAT CLM5 Cocoa   (%)

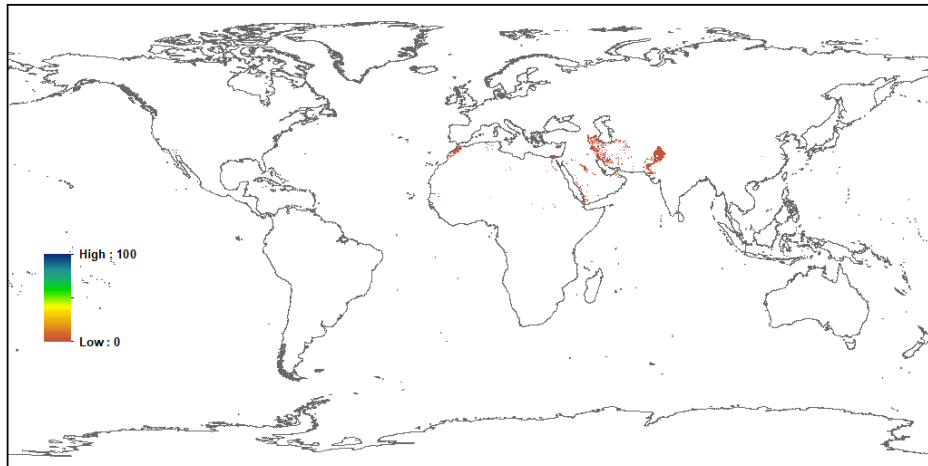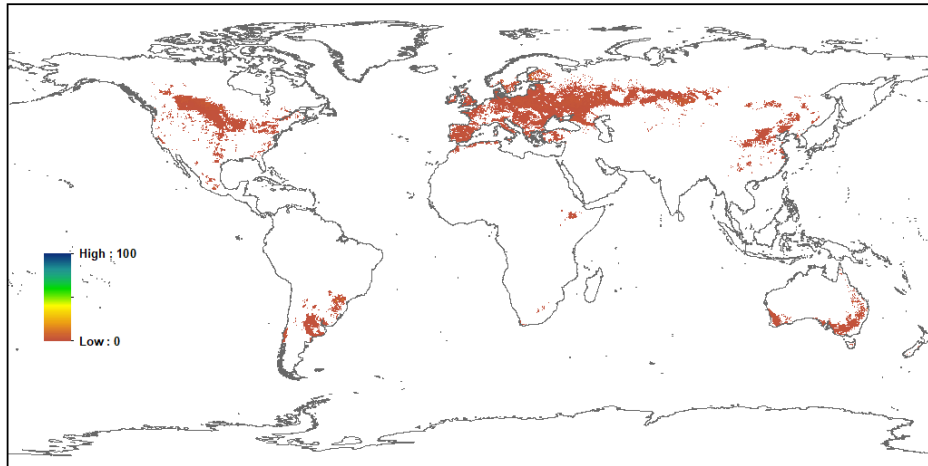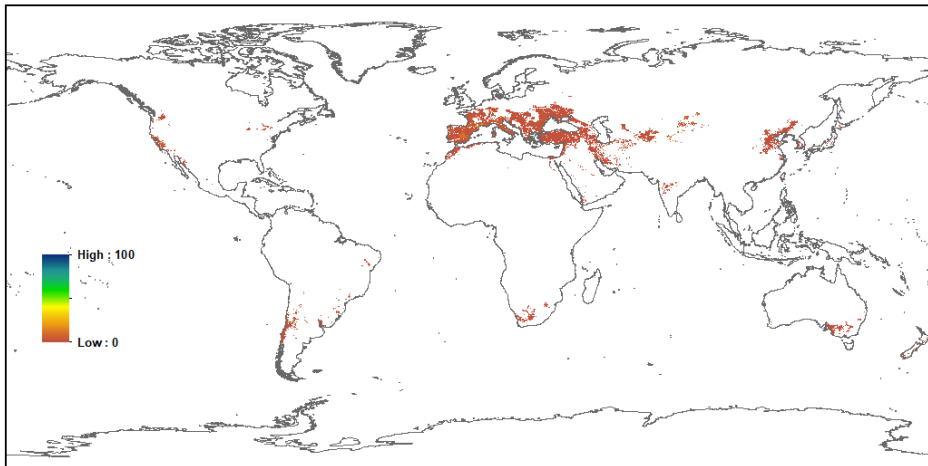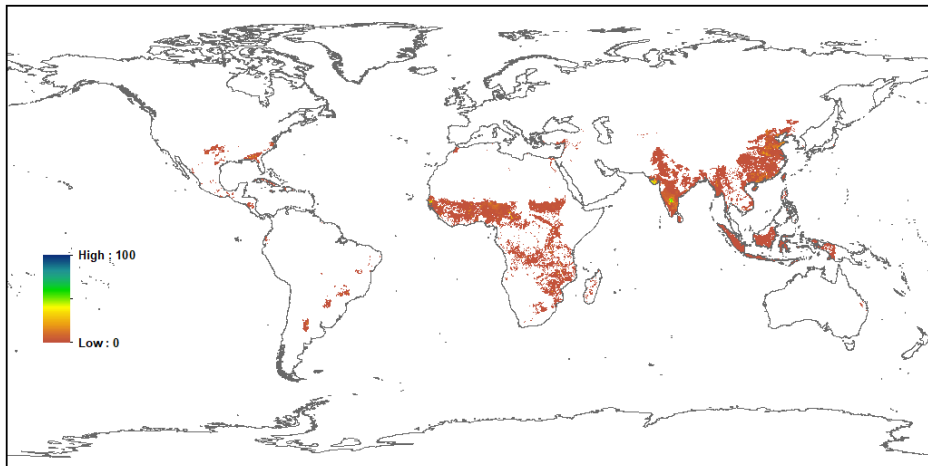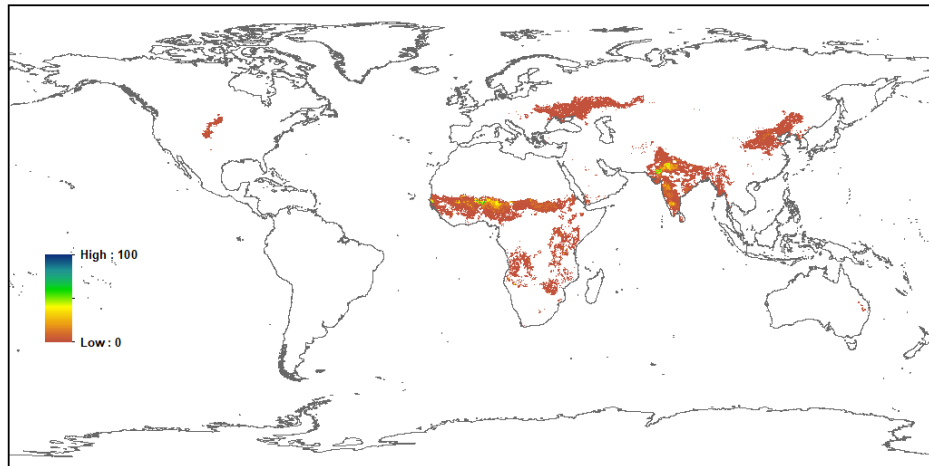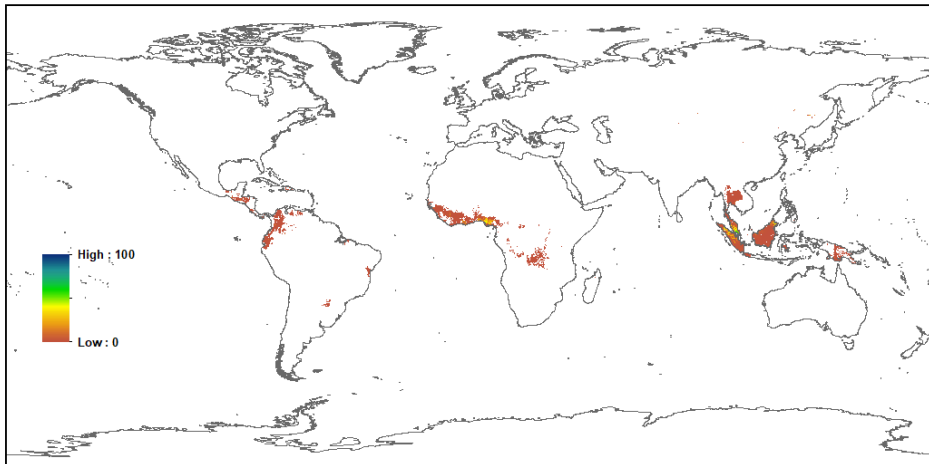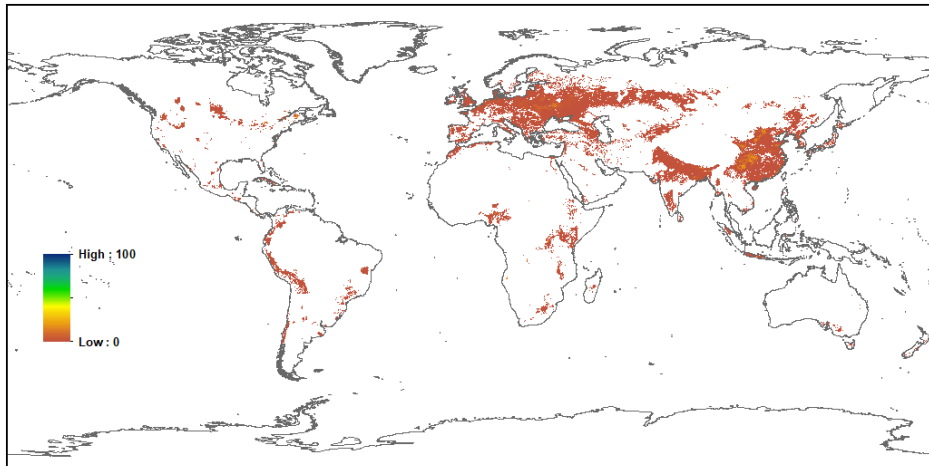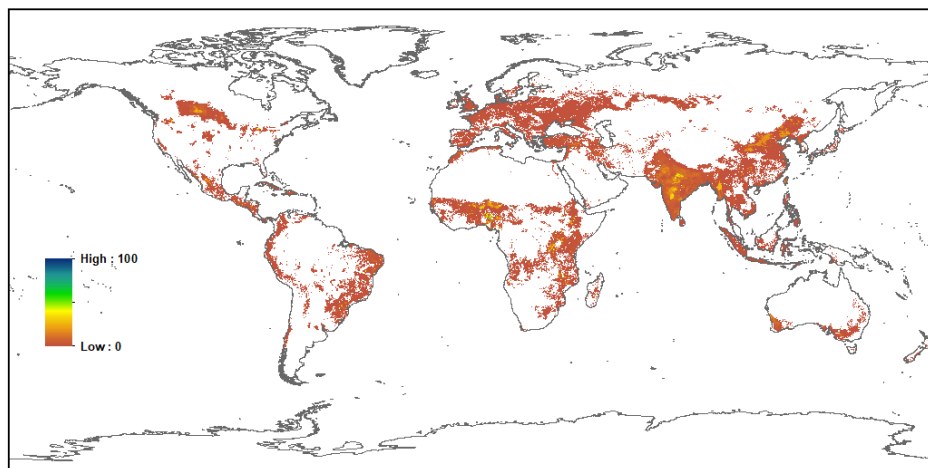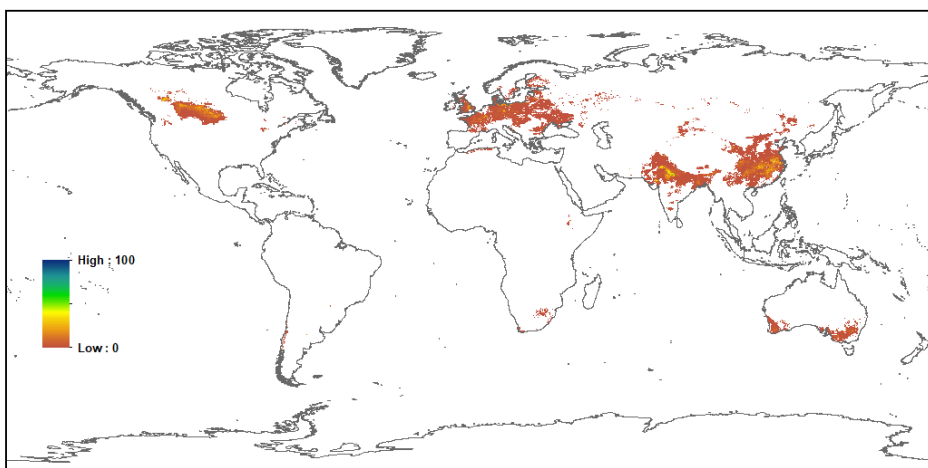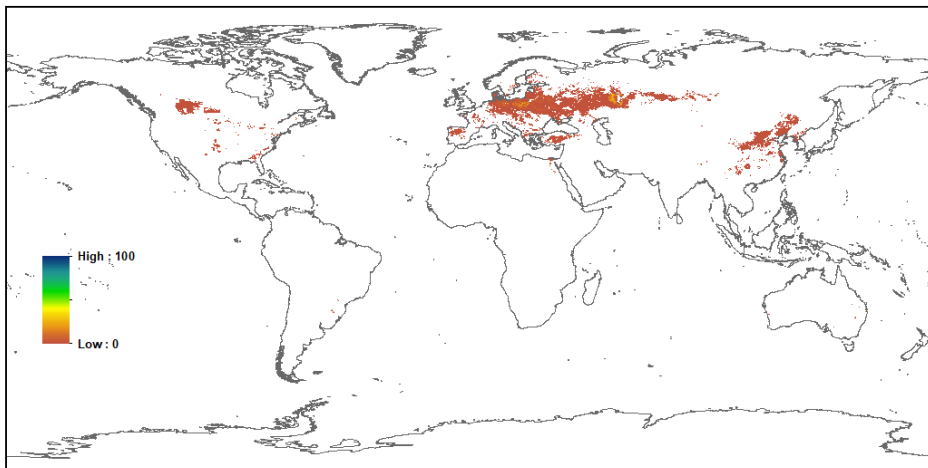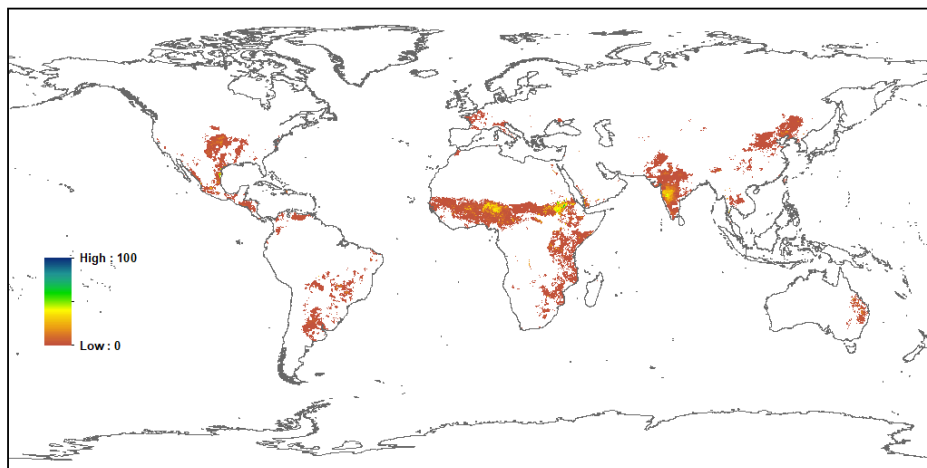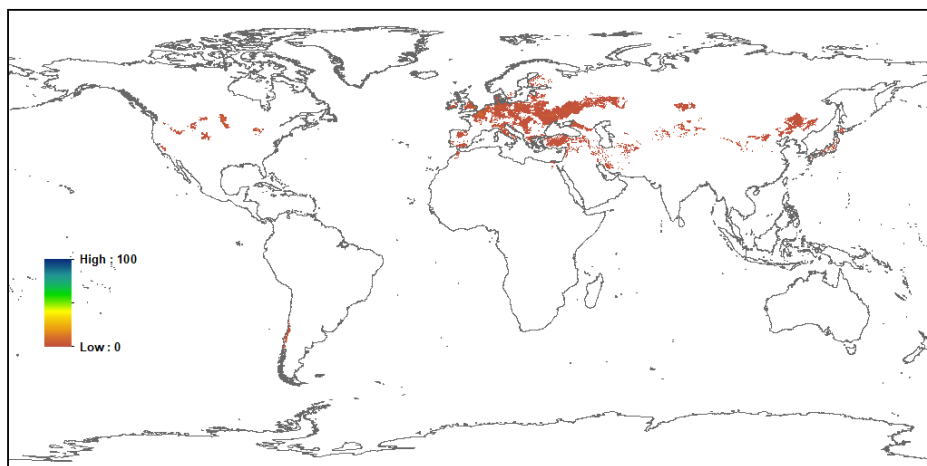(b) EARTHSTAT / FAOSTAT CLM5 Coffee   (%)
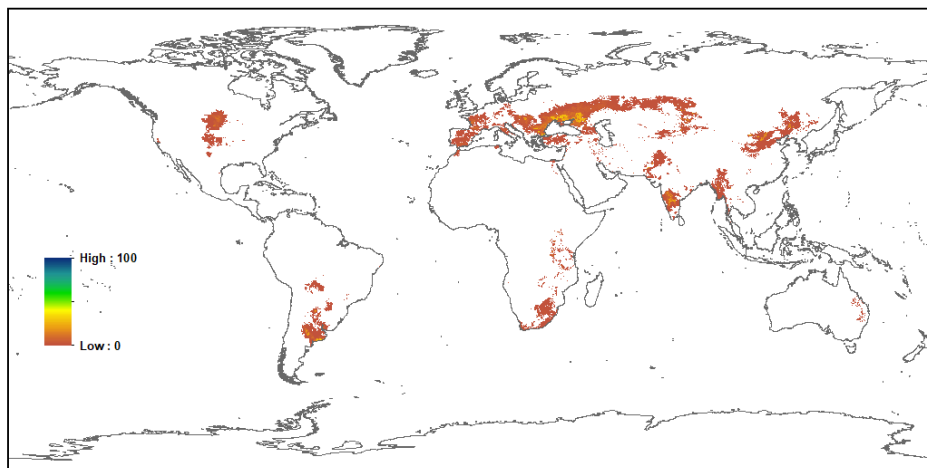
(c) EARTHSTAT / FAOSTAT CLM5 Datepalm   (%)

Figure 3.7: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Cocoa; (b) Percent Coffee; and (c) Percent Datepalm.

(a) EARTHSTAT / FAOSTAT CLM5 Fodder Grass   (%)

(b) EARTHSTAT / FAOSTAT CLM5 Grapes   (%)

(c) EARTHSTAT / FAOSTAT CLM5 Groundnuts   (%)

Figure 3.8: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Fodder Grass;
(b) Percent Grapes; and (c) Percent Groundnuts.

(a) EARTHSTAT / FAOSTAT CLM5 Millet   (%)

(b) EARTHSTAT / FAOSTAT CLM5 Oil Palm   (%)

(c) EARTHSTAT / FAOSTAT CLM5 Potatoes   (%)

Figure 3.9: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Millet; (b) Percent Tropical Oil Palm; and (c) Percent Potatoes.

(a) EARTHSTAT / FAOSTAT CLM5 Pulses   (%)

(b) EARTHSTAT / FAOSTAT CLM5 Rapeseed   (%)

(c) EARTHSTAT / FAOSTAT CLM5 Rye   (%)

Figure 3.10: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Pulses; (b) Percent Rapeseed; and (c) Percent Rye.

(a) EARTHSTAT / FAOSTAT CLM5 Sorghum   (%)

(b) EARTHSTAT / FAOSTAT CLM5 Sugarbeet   (%)

(c) EARTHSTAT / FAOSTAT CLM5 Sunflower   (%)

Figure 3.11: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Sorghum; (b) Percent Sugarbeet; and (c) Percent Sunflower.

(a) EARTHSTAT / FAOSTAT CLM5 Remaining Unmanaged Crops   (%)
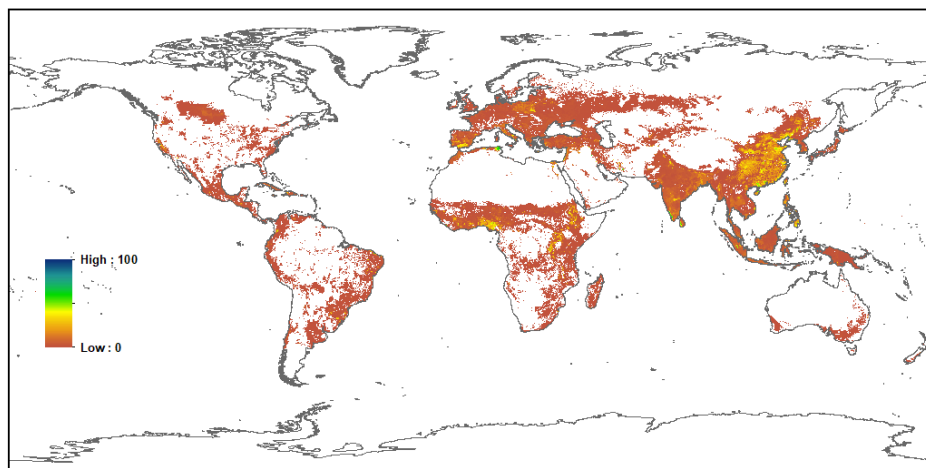
Figure 3.12: Global Current Day 1 km resolution CLM5 CFT Mapping: (a) Percent Remaining Unmanaged Crops

The tenth largest other CFT area was Cassava which had just under 0.2 million km² in both the CLM5 and FAOEarthStat data for the year 2005. The Cassava CFT was found primarily in Africa, with smaller areas in Latin America and Southeast Asia. The CLM5 Cassava CFT mapping is shown in Figure 3.6b. The mapping shows a major concentration in Nigeria, with widespread cultivation across Africa. There are also areas of cultivation across Brazil, southern India, Southeast Asia, and China. The Cassava CFT is simulated in the CLM5 Crop model using parameters and physiology from the Rice CFT.

The ninth smallest CFT area was Oilpalm which had just over 0.1 million km² in both the CLM5 and FAOEarthStat data for the year 2005. The Oilpalm CFT was found primarily in Southeast Asia and Africa, with a smaller area in Latin America. The CLM5 Oilpalm CFT mapping is shown in Figure 3.9b. The mapping shows the major concentrations in Malaysia, Indonesia, and Nigeria. There are smaller concentrations in Columbia and other parts of Africa. The Oilpalm CFT is simulated in the CLM5 Crop model using the parameters and physiology from the Rice CFT.

The eighth smallest CFT area was Fodder Grass which had just over 0.1 million km² in both the CLM5 and FAOEarthStat data for the year 2005. The Fodder Grass CFT was found primarily in Europe and Eurasia, with smaller areas in North America, Latin America and Asia Pacific. The CLM5 Fodder Grass CFT mapping is shown in Figure 3.8a. The mapping shows the main areas of Fodder Grass were from Europe into Russia, Canada into the northern United States, northeast Argentina and southern Australia. The Fodder Grass CFT is simulated in the CLM5 Crop model using the parameters and physiology from the Spring Wheat CFT.

The seventh smallest CFT area was Coffee which had around 0.1 million km² in both the CLM5 and FAOEarthStat data for the year 2005. The Coffee CFT was found primarily in Latin America, with smaller areas in Africa and Southeast Asia. The CLM5 Coffee CFT mapping is shown in Figure 3.7b. The mapping shows the distinct growing regions with Central America into Columbia, southeast Brazil, eastern Africa, and Southeast Asia. The Coffee CFT is simulated in the CLM5 Crop using the parameters and physiology from Rice.

The sixth smallest CFT area was Citrus which had 0.08 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Citrus CFT was found primarily in East Asia and Latin America with smaller areas in Europe, Africa and Southern Asia. The CLM5 Citrus CFT mapping is shown in Figure 3.6c. The mapping shows the widespread cultivation across China, with concentrations southeast Brazil, Spain, Nigeria, Pakistan and central India. The Citrus CFT is simulated in CLM5 Crop using the parameters and physiology from Spring Wheat.

The fifth smallest CFT area was Cocoa which had 0.08 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Cocoa CFT was found predominantly in Africa with smaller areas in Latin America and Southeast Asia. The CLM5 Cocoa CFT mapping is shown in Figure 3.7a. The mapping shows the intense cultivation in the Cote d'Ivoire and Ghana, and wider cultivation in the Caribbean, Columbia, eastern Brazil, Indonesia and Malaysia. The Coco CFT is simulated in CLM5 Crop using the parameters and physiology from Rice.

The fourth smallest CFT area was Grapes which had 0.07 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Grapes CFT was found Europe with a smaller area in East Asia. The CLM5 Grapes CFT mapping is shown in Figure 3.8b. The mapping shows the intense cultivation in Spain, France and Italy into Turkey, Ukraine and Russia. There is also Grapes cultivation in northeastern China and the wine growing regions of Chile, Argentina, South Africa, southern Australia and the United States. The Grapes CFT is simulated in CLM5 Crop using the parameters and physiology from Spring Wheat.

The third smallest CFT area was Rye which had 0.07 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The area of Rye dropped drastically through the FAOSTAT period starting at 0.31 million km^2 in 1961 and ending 0.04 million km^2 in 2016. The Rye CFT was found predominantly in Europe and Eurasia. The CLM5 Rye CFT mapping is shown in Figure 3.10c. The mapping shows the restricted area cultivated from eastern Europe into Russia, northeastern China, and Canada. The Rye CFT is simulated in CLM5 Crop using the parameters and physiology of Spring Wheat.

The second smallest CFT area was Sugarbeet which had 0.05 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Sugarbeet CFT was found predominantly in Europe with smaller areas in Eurasia and North America. The CLM5 Sugarbeet CFT mapping is shown in Figure 3.11b. The mapping shows the limited area of cultivation from eastern Europe into Russia, northeast China, and limited areas of Canada and the northern United States. The Sugarbeet CFT is simulated in CLM5 Crop using the parameters and physiology of Spring Wheat.

The smallest CFT area was Datepalm which had 0.01 million km^2 in both the CLM5 and FAOEarthStat data for the year 2005. The Datepalm CFT was found predominantly in the Middle East, Africa and Southern Asia. The CLM5 Dateplam CFT mapping is shown in Figure 3.7c. The mapping shows the extremely limited growing area through Iran, Pakistan and Morocco. The Datepalm CFT is simulated in CLM5 Crop using the parameters and physiology of Cotton.
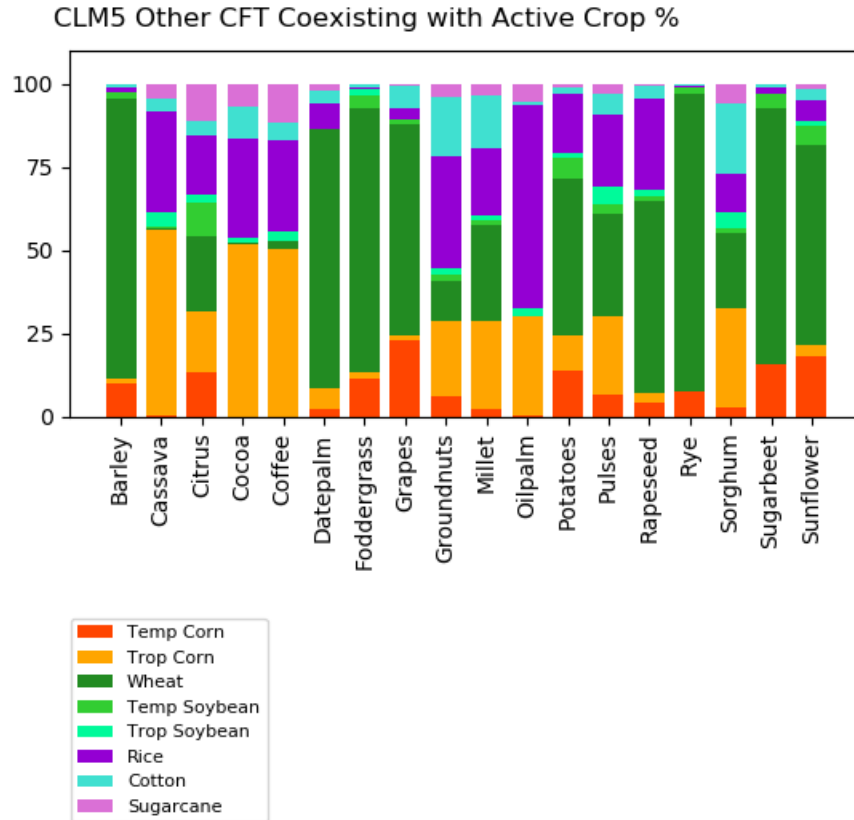
Figure 3.13: Global CLM5 Crop Functional Type (CFT) Composition of crops not currently represented in CLM5 that are coexisting with crops that are actively represented in the model.

3.8 CLM5 Other to Active Crop Mapping

For CLM5, the 23 inactive crop CFTs that have not yet had the parameters and physiology required for active management developed are assigned the parameters and physiology of the spatially closest associated active crop CFT that corresponds to their functional type (e.g., C3 or C4). The intention is to allow the crops to be simulated in the CLM5 Crop model with parameters that are appropriate for the climates and agricultural practices in which they are found. While this substitution leaves the representation of these CFTs with varying degrees of accuracy, it does allow the use of the Crop model over all of the agricultural areas of the planet which results in major improvements in carbon cycle and surface climate simulation compared with representing crops as grasses, as done without the Crop model, see Lombardozzi et al. (2019).

The substitution rules from inactive CFTs to active CFT parameters are listed in Table 3.1. These rules are generated predominantly based on the coexistence of the active CFT in the same location as the inactive CFT. The contribution from each coexisting active CFT for each inactive CFT is shown in Figure 3.13. For the mapped CFTs: Barley, Rye, Citrus, Fodder Grass, Grapes, Potatoes, Pulses, Rapeseed, Sugarbeet and Sunflower are all substituted to Spring Wheat; Cassava, Cocoa, Coffee, Groundnuts and Oilpalm are substituted to Rice; Millet and Sorghum are substituted to Tropical Corn; and Datepalm is substituted to Cotton. Of these, only Datepalm does not follow these rules, which is a historical carryover from the original parameter set up.

CHAPTER 4.

OTHER CURRENT DAY DATA

4.1 Other Current Day CLM5 Data Overview

To represent the complexity of the Current Day land surface, CLM5 prescribes a range of other global surface data, in addition to the Current Day Land Cover and Agriculture data of Chapter 2 and Chapter 3. Many of the data products facilitate sub-models within CLM5 such as the Lake, Urban, Fire, Volatile Organic Compounds, or Methane models. Other data products support external models coupled to CLM5, such as the CISM2 Ice Sheet model, or the MOSART river routing model.

Other current day data products support the prescription of components of the CLM5 land surface not included directly by the land cover and agriculture data. This data includes the soil properties of depth, organic and mineral composition. It also includes topographic properties such as elevation, slope and drainage, and surface radiation properties determined by soil color with associated dry and saturated visible and near infrared reflectances.

The final component of the other global surface data is the Satellite Phenology (SP) data. This data prescribes MODIS derived monthly Leaf and Stem Area Index (LAI and SAI) data, and ICESAT derived canopy height top and bottom for each Plant Functional Type (PFT). The SP data is used as an alternative to the prognostic morphology and phenology simulated in CLM5 with active Biogeochemistry selected. This chapter details the nature and sources of each of these other global data.

4.2 Glaciers CISM2

When coupled to the CISM2 land ice model, CLM5 has dynamically adjusting Glacier land units allowing the Glacier area to initiate, grow, shrink, or disappear during a simulation, representing active ice evolution, as detailed in Lipscomb and Sacks (2012). Beyond the simple prescription of the Glacier land unit CLM5 describes the distribution of ice caps, ice sheets, ice shelves and mountain glaciers, with multiple elevation classes (10 elevation classes by default) for each ice type. The elevation classes simulate the associated temperature, rain/snow partitioning, and downwelling longwave downscaling to account for the strong topographic elevation heterogeneity over glaciers and ice sheets.

While the CLM5 Land Cover data from Chapter 2 determines the Glacier land unit cover from the MODIS land cover product, this is only a placeholder for the land use and land cover data generation process, which is then replaced with the more detailed representation of Glacier for consistent use in both CLM5 and the CISM2 ice sheet model. The CISM2 Glacier extent is mapped in Figure 4.1a. As would be expected the Glacier extent is highly consistent with the current day MODIS land cover mapped Glacier extent shown in Figure 2.13a of Chapter 2.

The biggest difference in the CISM2 Glacier extent is the inclusion of Antarctic ice shelves, in particular the Ross and Ronne-Filchner, that cover nearly 1 million km^2, but are considered ocean extent by the MODIS land cover data. Table 4.1 shows that on land the CISM2 Glacier extent also differed from the MODIS mapping with 16.2 million km^2 compared to the 15.3 million km^2 of Table 2.5. Regionally this difference was greatest in North America where the CISM2 Glacier extent is 2.0 million km^2 compared to the 2.6 million km^2 of MODIS. This difference between the products mainly derives from the classification between active glaciers and permanent snow fields which MODIS is incapable of differentiating.

The CISM2 Glacier land unit extent is taken from the first globally complete glacier inventory, the Randolph Glacier Inventory version 1.0 (RGIv1.0: Arendt et al. 2012). Vector data for the Greenland Ice Sheet were provided by Frank Paul and Tobias Bolch (University of Zurich: Rastner et al. 2012). Antarctic Ice Sheet data were provided by Andrew Bliss (University of Alaska) and were extracted from the Scientific Committee on Antarctic Research (SCAR) Antarctic Digital Database version 5.0.

Floating ice is only provided for the Antarctic and does not include the small area of Arctic ice shelves. High spatial resolution vector data were used to determine the area of glacier, ice sheet and floating ice within 30-second grid cells globally. The Glacier land unit is also provided with a region identifier for the CISM2 ice sheet model for: 1. Greenland; 2. Greenland Surrounds for potential ice sheet growth; 3. Antarctica; and 4. All other glaciers as shown in Figure 4.1b.

To develop the elevation classes of each glacier type, the 30-second glacier, ice sheet and Antarctic ice shelf masks were draped over equivalent-resolution GLOBE topography (Global Land One-km Base Elevation Project, Hastings et al. 1999) to extract approximate ice-covered elevations of ice-covered regions. Grid cells flagged as land-ice in the mask but ocean in GLOBE (typically, around ice sheets at high latitudes) were designated land-ice with an elevation of 0 meters. The distributions by elevation are used to divide each glacier land unit into columns based on elevation class.

The glacier land unit extent is prescribed by the initial or default total extent which is then broken down into and the 10 elevation classes per grid cell. The current day distribution of the elevation classes is prescribed from ice sheets and Antarctic ice shelf masks combined with the GLOBE topography to generate the 10 ice-covered areas (Arendt et al. 2012, Rastner et al. 2012, Hastings et al. 1999). The 10 elevation classes are generated from the USGS 3 minute topographic map of the world and are shown in Figure 4.1c.

Table 4.1: Other Current Data Global and IPCC Region Area in millions km^2.

Land Unit	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Glacier	16.22	0.00	0.00	0.06	0.07	0.05	0.04	0.00	2.04	0.00	0.04	12.16
Lake	2.90	0.30	0.03	0.11	0.40	0.16	0.28	0.02	1.12	0.04	0.04	0.00
Urban	0.80	0.07	0.02	0.16	0.02	0.12	0.06	0.02	0.04	0.07	0.22	0.00
TBD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
High	0.16	0.01	0.01	0.04	0.00	0.01	0.02	0.01	0.00	0.01	0.05	0.00
Medium	0.64	0.06	0.01	0.12	0.02	0.11	0.04	0.01	0.04	0.05	0.18	0.00

(a) CISM2 Glacier   (%)
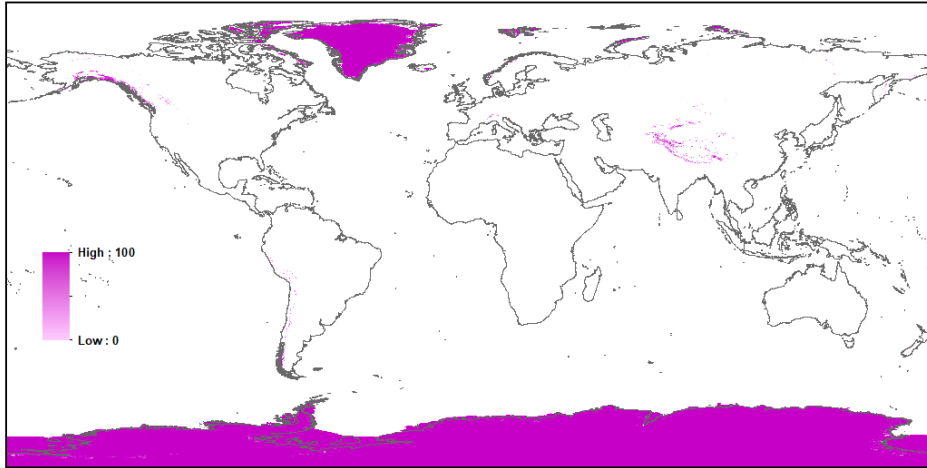
(b) CISM2 Glacier Ice Sheet Regions
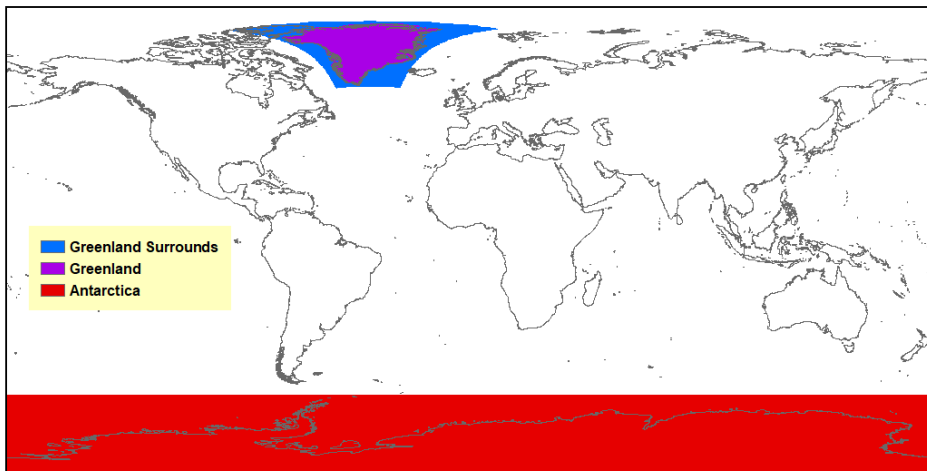
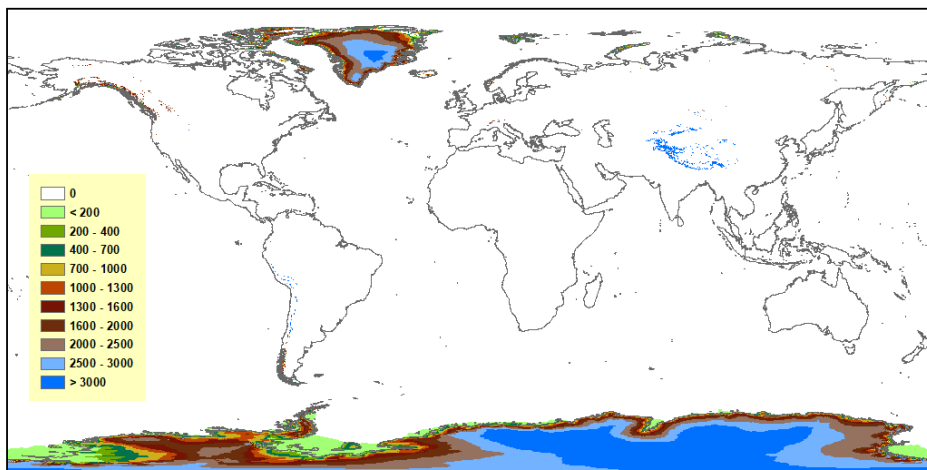## (c) CISM2 Glacier Mean Elevation Class (MEC)   (meters)

Figure 4.1: Global Current Day CLM5 CISM2 Land Ice Model Surface Data for: (a) Glacier Extent; (b) Glacier Region; and (c) Glacier Mean Elevation Class (MEC).

4.3 Lakes

The CLM5 Lake sub model, described in Subin et al. (2012a), uses both spatial lake extent and lake depth to simulate lake water temperature, vertical mixing, ice, snow, and energy and moisture fluxes to the atmosphere for the Lake land unit. Like the Glacier land unit, the CLM5 Lake land unit mapping from the MODIS land cover product described in Chapter 2, is only a placeholder for the land use and land cover data generation process. The MODIS land cover data generated lake extent is replaced with the more detailed representation of the CLM5 Lake model.

(a) CLM5 GLWD Lake   (%)

(b) CLM5 Lake Depth   (meters)
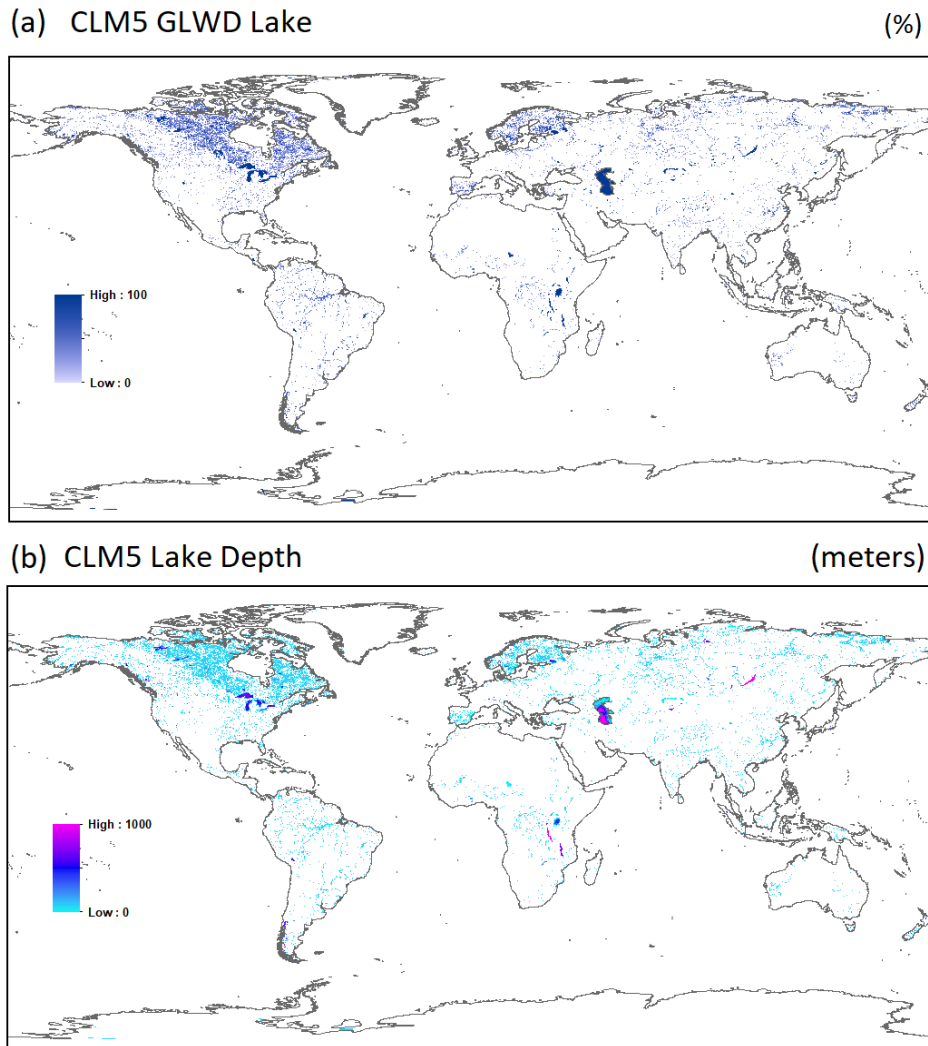
Figure 4.2: Global Current Day CLM5 Lake Model Surface Data for: (a) Lake Extent; and (b) Lake Depth.

The CLM5 Lake extent is taken from the Global Lake and Wetland Database (GLWD) of Lehner and Doll (2004), shown in Figure 4.2a. The total global area of lakes in the GLWD data is at 2.9 million km^2 as shown in Table 4.1. This is slightly higher than the MODIS derived lake area of 2.8 million km^2 shown in Table 2.5. While the total global area is similar, the differences at the regional level are relatively large. North America and Africa have 50% more lake are extent in the GLWD

data, while Eurasia has 35% more. The MODIS data by contrast has higher coastal lake areas where the land mask boundary separates water bodies between ocean and lake classifications. The CLM5 mean lake depth for each grid cell is calculated based on the global gridded data sets of Kourzeneva (2012), shown in Figure 4.2b. Here the difference between shallow lakes over much of the world are differentiated from the deeper Great Lakes of North America, the Caspian Sea, and the very deep Lake Baikal, Lake Tanganyika and Lake Malawi. Due to the new surface water store in CLM5 wetland mapping was removed from surface data sets. The wetlands continued to be represented in non-land areas accounting for ocean areas to resolve land masks conflicts between CLM5 and other models.

4.4 Urban CLMU

The CLM5 Urban sub model (CLMU) allows the simulation of urban environments within CESM, with a particularly focus on investigating the urban climate where people live as described in Oleson et al. (2010b). As such, the CLMU model allows scientific study of how climate change affects the urban heat island and possible urban planning and design strategies. Urban areas in CLM5 are represented by up to three urban landunits per gridcell according to density class. The urban landunit is based on the "urban canyon" concept of Oke (1987) in which the canyon geometry consists of roofs, walls, and canyon floor. The canyon floor is divided into pervious (e.g., to represent residential lawns, parks) and impervious (e.g., to represent roads, parking lots, sidewalks) fractions.

The CLM5 Urban model uses a static representation of present day global urban extent and urban properties developed by Jackson et al. (2010). Urban extent is defined for the four classes of tall building district (TBD), as well as high, medium, and low density (HD, MD, LD) urban developments. For CLM5 only the TBD, HD, and MD classes are used as the LD class is highly rural and likely better modeled as a vegetated surface. The density classes are mapped from the LandScan 2004 population density dataset with average building properties provided for 33 distinct regions across the globe as described by Jackson et al. (2010). The urban land units have five columns of roof, sunlit walls and shaded walls, and pervious and impervious canyon floor, with the radiative, thermal (Oleson et al. 2010b).

Like the Glacier and Lake land unit mapping, the new Urban density class mapping used with the CLM5 Urban model replaces the urban mapping from the MODIS land cover product generated with the Current Day land cover data in Chapter 2. The combined Urban extent from the TBD, HD and MD classes is shown in Figure 4.3a. The mapping shows the CLMU combined urban extent is much larger in India and China, and marginally smaller in Europe, North America and South America, than the MODIS derived urban area shown in Figure 2.13c.

Table 4.1 shows globally the CLMU combined urban spatial extent is 0.8 million km^2, which is 60% larger than the 0.5 million km^2 of the MODIS mapping. Given the total land area of 148.1 million km^2, this still represents only around 0.5% of the land area being represented by the CLMU model. The MD urban class makes up much of the Urban extent at 0.64 million km^2. The HD urban class makes up the remaining 0.16 million km^2, while the TBD makes up only 618 km^2 globally.

(a) CLM5 All Urban Areas CLMU   (%)
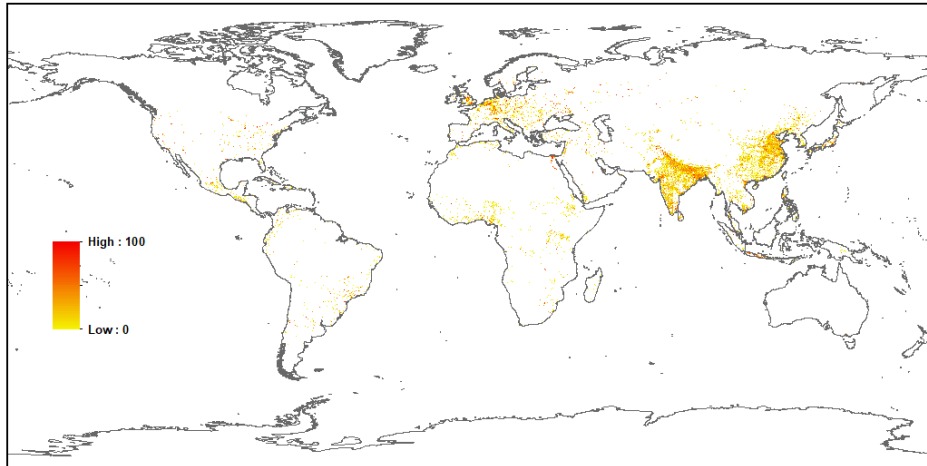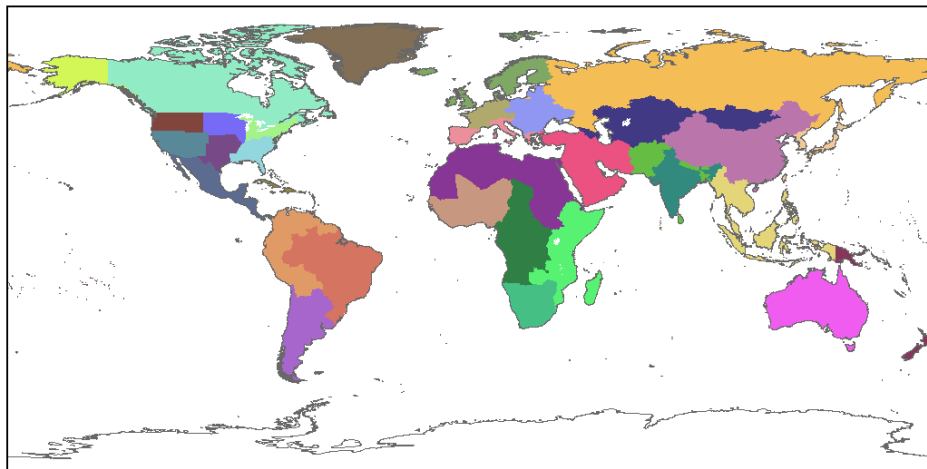

(b) CLM5 Urban Regions CLMU

Figure 4.3: Global Current Day CLM5 Urban Model Surface Data for: (a) All Combined Urban Class Extent; and (b) Urban Class Regions.

The regional density classes combine with the global mapping to provide urban properties for the thermal (e.g., heat capacity and thermal conductivity), radiative (e.g., albedo and emissivity) and morphological (e.g., height to width ratio, roof fraction, average building height, and pervious fraction of the canyon floor) properties of roof/wall/road. The 33 regions developed by Jackson et al. (2010) are shown in Figure 4.3b.

4.5 Soil Depth, Texture, and Organic Content

The CLM5 soil thermal and hydrological properties are prescribed globally for all land units through soil depth, organic content, and sand and clay content as described in Lawrence et al. (2018). The new representation of variable soil thickness in CLM5 is prescribed through soil depth to bedrock globally as described in Brunke et al. (2016) and Swenson and Lawrence, (2015), with values derived from the spatially explicit soil thickness data product of Pelletier et al., (2016). The soil organic matter data is produced for the majority of the globe from the ISRIC-WISE (Batjes, 2006) mapping, while the high latitudes mapping comes from the 0.25 degree version of the

Northern Circumpolar Soil Carbon Database (Hugelius et al. 2012). Both datasets report soil carbon down to 1m depth. The soil mineral sand and clay texture in CLM5 are prescribed from the International Geosphere-Biosphere Programme (IGBP) soil dataset (Global Soil Data Task 2000) with the 4931 soil classes mapped to depth varying soil profiles of sand and clay in the model.

The CLM5 variable soil depth data is mapped globally in Figure 4.4a. The mapping shows very large spatial heterogeneity with soil depth varying from 0.029 in barren rocky areas such as the Tibetan plateau to 50.0 meters in deep sandy such as central Australia, or other organic or clay soils such as found in North America. Table 4.2 shows the global average soil depth is 15.6 meters with a range of average soil depths across the IPCC regions. The deepest average soils are found in the Middle East, Asia Pacific and Africa regions. The shallowest average soils are found in the Other, East Asia, and Eurasia regions.

Table 4.2: CLM5 Global and IPCC Regional Average Soil Properties. Depth in meters, and soil properties for the top layers (0 – 0.09m), middle layers (0.1 – 1.3m) and bottom layers (1.3 – 3.4m). Organic Matter in kgOM/m^3, and Sand and Clay as percentage of mineral soil.

Soil Property	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Depth m	15.6	18.1	19.1	8.9	12.3	12.9	16.2	20.8	17.0	13.5	16.9	0.9
Organic Top	36.1	15.6	19.2	25.2	75.3	44.6	24.0	10.1	61.9	33.6	19.0	32.4
Organic Mid	10.7	6.2	6.2	7.1	19.7	11.4	8.4	4.1	16.6	12.3	6.7	9.5
Sand % Top	49.2	55.6	58.4	50.9	41.2	41.5	47.8	52.8	48.4	44.9	45.6	41.4
Sand % Mid	47.0	51.7	51.3	48.3	41.9	44.4	43.6	48.2	49.4	42.7	42.5	39.2
Sand % Bot	47.0	50.1	49.8	50.1	43.6	45.5	43.1	46.2	49.6	44.4	41.5	40.2
Clay % Top	22.5	21.5	18.6	21.9	21.5	24.0	27.9	17.5	20.4	27.2	26.2	27.6
Clay % Mid	25.8	25.5	23.0	24.6	22.5	24.0	33.3	21.5	22.6	32.1	30.2	32.3
Clay % Bot	25.3	26.3	22.8	22.7	20.6	21.5	33.9	20.3	21.6	32.5	30.0	33.0

The CLM5 soil organic matter is mapped globally for the top 0.09 meters in Figure 4.4b and the middle 0.1 – 1.3 meters in Figure 4.4c. Soil organic matter below the 1.3 meter soil layer is prescribed at 0.0 km OM/m^3. The mapping shows that for both the top and middle soil layers organic matter varies greatly both spatially and with depth. Spatially the mapping shows that the largest amounts of organic matter are in high northern latitudes in North America, Scandinavia and Russia following the expansive peatlands of these regions. Table 4.2 shows the strong vertical differences in soil organic matter is consistent globally with the top 0.09 meters having 36.1 kg OM/m^3, which is almost four times the content of middle soil layer from 0.1 – 1.3 meters at 10.7 kg OM/m^3. The regional organic matter content reflects the global mapping with the highest content in Eurasia, followed by North America and Europe.

The CLM5 sand percentage mapping from the IGBP soil data set is shown globally for the top 0.09 meters in Figure 4.5a, the middle 0.1 – 1.3 meters in Figure 4.5b, and the bottom 1.3 – 3.4 meters in Figure 4.5c. The three maps show that the sand content varies spatially, but there is strong correspondence between all soil layers with sandy soils persisting from the top to bottom soil layers, with a general reduction in sand content with depth. The global maps show the soils with the highest sand content correspond with the deserts of northern and southern Africa, the Middle East, central Asia and Australia. Table 4.2 shows that globally soils were almost 50% sand content through the soil column with a small decrease in sand content with depth. Regionally the largest average sand content was found in the Asia Pacific, then Africa and Middle East regions.

(a) CLM5 Soil Depth (meters)
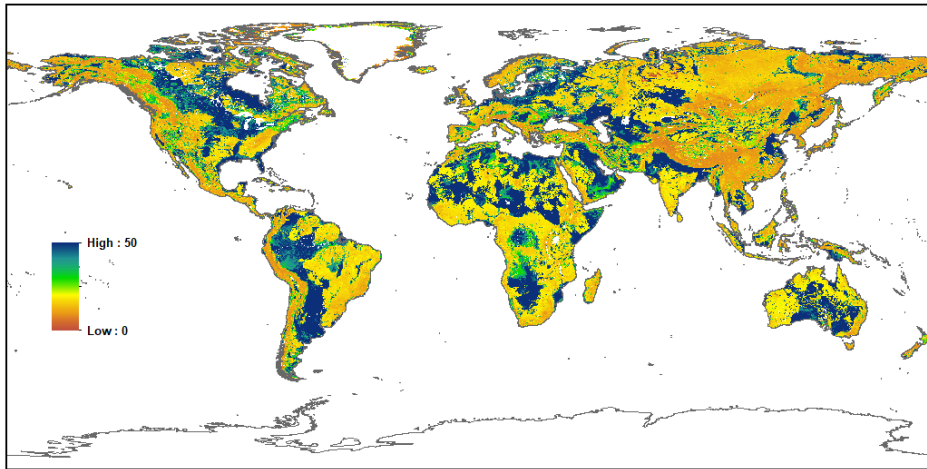
(b) CLM5 Organic Content Top (avg 0 - 0.09m) (kgOM / m3)
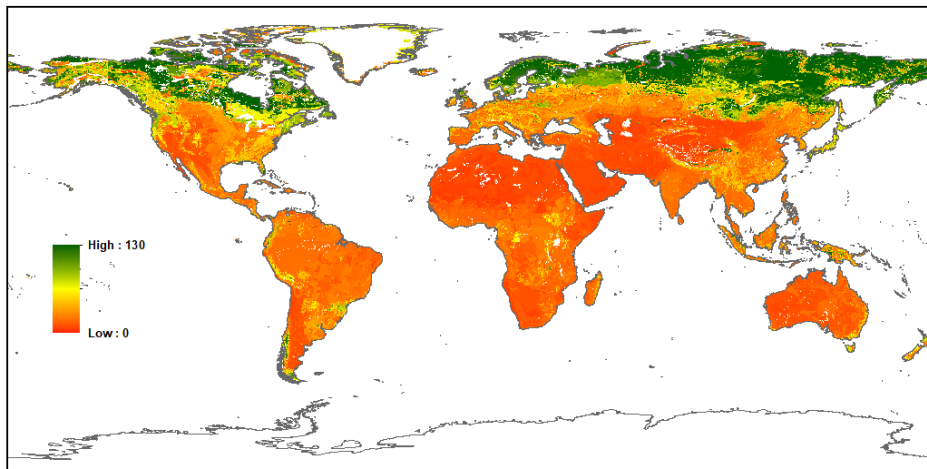
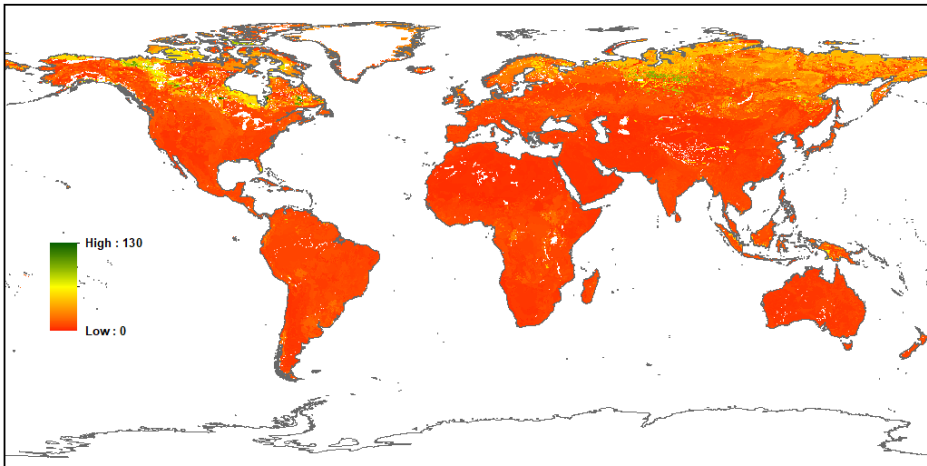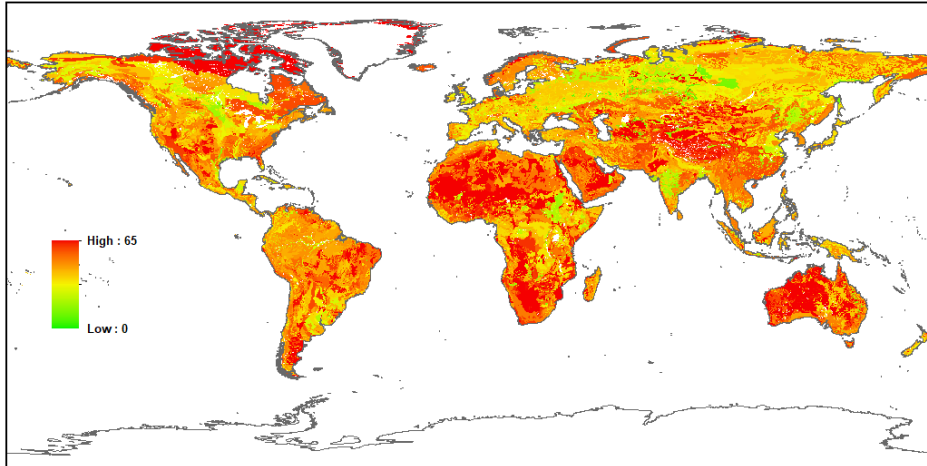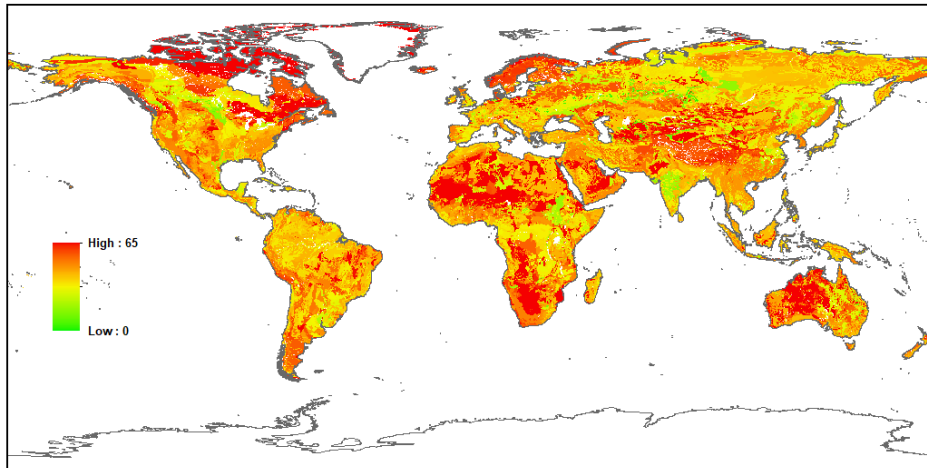(c) CLM5 Organic Content Middle (avg 0.1 - 1.3m) (kgOM / m3)

Figure 4.4: Global Current Day CLM5 Soil Properties Data for: (a) Soil Depth to Bedrock; (b) Soil Organic Content for the top 0.09m; and (c) Soil Organic Content for the next 0.1 – 1.3m

(a) CLM5 Sand Content Top (avg 0 - 0.09m) (%)

(b) CLM5 Sand Content Middle (avg 0.1 - 1.3m) (%)
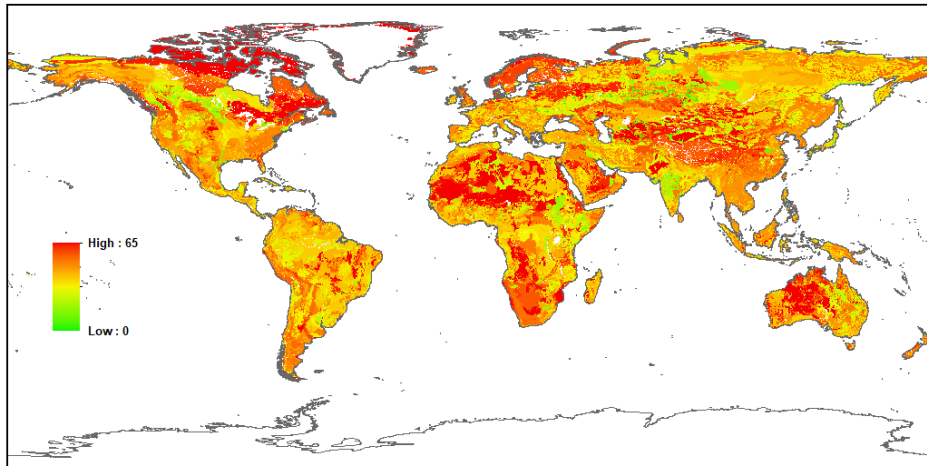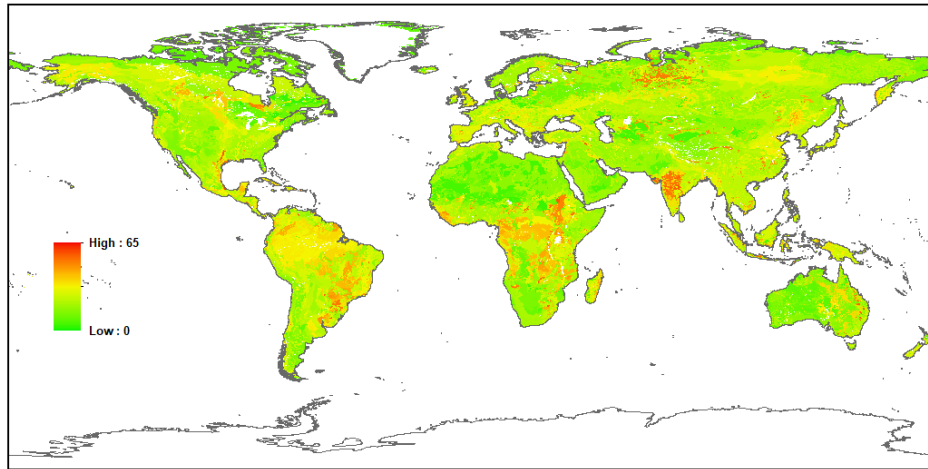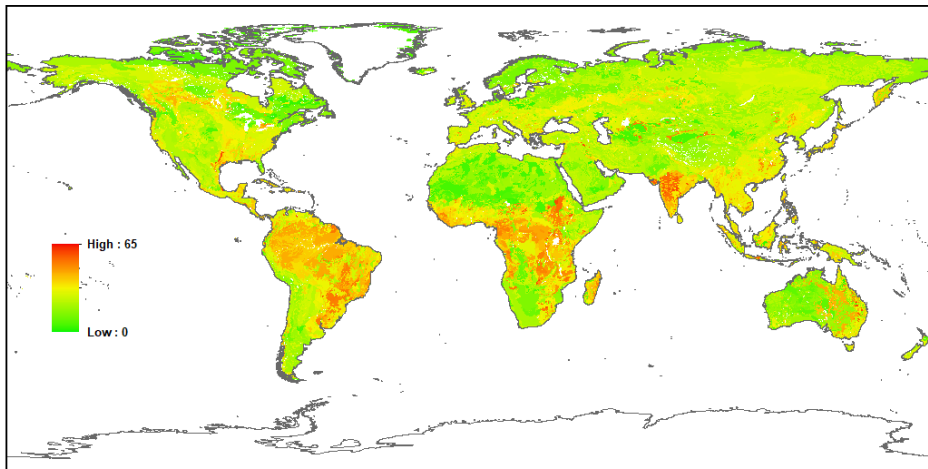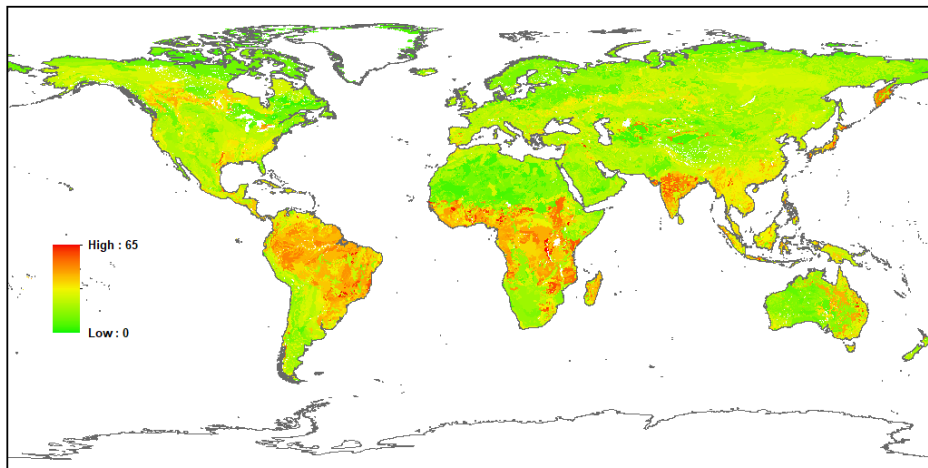
(c) CLM5 Sand Content Bottom (avg 1.3 - 3.4m) (%)

Figure 4.5: Global Current Day CLM5 Soil Sand Content for: (a) the top 0.09m; (b) the next 0.1 – 1.3m; and (c) the bottom 0.1 – 1.3m.

(a) CLM5   Clay Content Top (avg 0 - 0.09m)   (%)

(b) CLM5   Clay Content Middle (avg 0.1 - 1.3m)   (%)

(c) CLM5   Clay Content Bottom (avg 1.3 - 3.4m)   (%)

Figure 4.6: Global Current Day CLM5 Soil Clay Content for: (a) the top 0.09m; (b) the next 0.1 – 1.3m; and (c) the bottom 0.1 – 1.3m.

The CLM5 clay percentage mapping is shown globally for the top 0.09 meters in Figure 4.6a, the middle 0.1 – 1.3 meters in Figure 4.6b, and the bottom 1.3 – 3.4 meters in Figure 4.6c. Like sand, clay content mapping varies spatially with a strong correspondence between all soil layers, but with clay content increasing with depth instead of decreasing. The three major areas of clay soils are the Amazon into the Atlantic forests of South America, central Africa into Madagascar, and India into Southeast Asia. Table 4.2 shows that globally the average clay content was around 25% with the highest content found in Latin America, Southeast Asia, and Southern Asia.

4.5 Topography, Runoff and Drainage

Soil water in CLM5 is predicted from a multi-layer model, in which the vertical soil moisture transport is governed by infiltration, surface and sub-surface runoff, gradient diffusion, gravity, and canopy transpiration through root extraction. The moisture input at the grid cell surface is the sum of liquid precipitation reaching the ground and melt water from snow. The moisture flux is partitioned between surface runoff, surface water storage, and infiltration into the soil. The simple TOPMODEL-based (Beven and Kirkby 1979) runoff model (SIMTOP) described by Niu et al. (2005) is implemented to parameterize surface and sub-surface runoff.

A surface water store has been added to CLM5 to represent wetlands and small, sub-grid scale water bodies. As a result, the wetland land unit has been removed as of CLM4.5. Surface water storage and outflow are functions of fine spatial scale elevation variations called microtopography. The microtopography is assumed to be distributed normally around the grid cell mean elevation. As no global datasets exist for microtopography, the CLM5 surface water store uses a parameterization is a simple function of slope. The CLM5 model has elevation and slope prescribed from the USGS HYDRO1K 1-km dataset. Slope is used in the surface water parameterization, and elevation is used to calculate the grid cell standard deviation of topography for the snow cover fraction parameterization. The CLM5 elevation is mapped globally in Figure 4.7a, with the slope mapped in Figure 4.7b. The elevation map show the key topographic features of the Himalayas, the Tibetan plateau, the Rocky Mountains of North America, the Andes, Greenland and Antarctica. The slope map shows the orographic transitions at the edges of these features.

Beyond elevation and slope, a key concept underlying the CLM5 SIMTOP runoff and drainage approach is the concept of fractional saturated area. The saturated fraction in CLM5 is determined by the topographic characteristics and soil moisture state of a grid cell. The saturated portion of a grid cell contributes to surface runoff by the saturation excess mechanism (Dunne runoff), which is combined with the infiltration excess (Hortonian) runoff to generate surface runoff.

The maximum saturated fraction (fmax) that describes topographic characteristics in the ground water model is prescribed at the 0.125 degree resolution using the 1-km compound topographic indices (CTIs) based on the HYDRO1K dataset (Verdin and Greenlee 1996) from USGS, following the methods of Niu et al. (2005) and Li et al. (2014). The CLM5 global map of fmax is shown in Figure 4.7c. The fmax value is a scalar showing the maximum fraction of inundation. The map shows there are four main areas of maximum inundation in the Amazon through the Atlantic forests of South America, the central and north west regions of Africa, the northern parts of North America and the Southeast Asian islands into Papua New Guinea.

(a) CLM5 Elevation (meters)
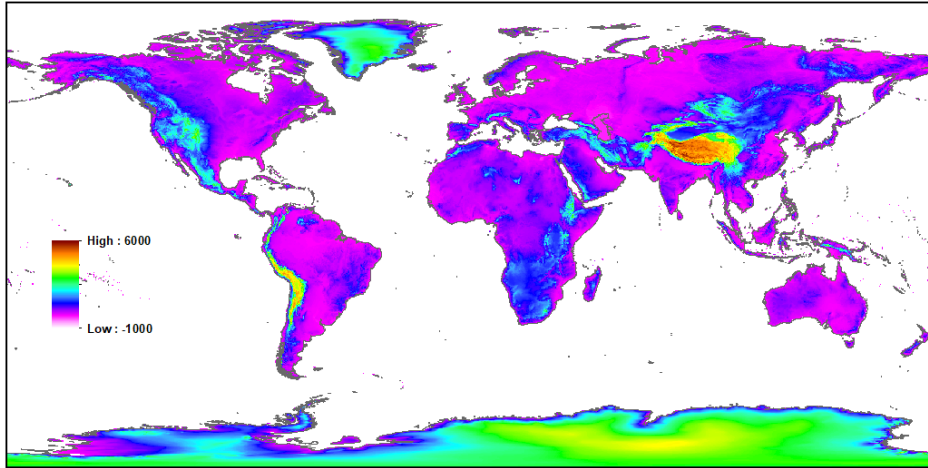
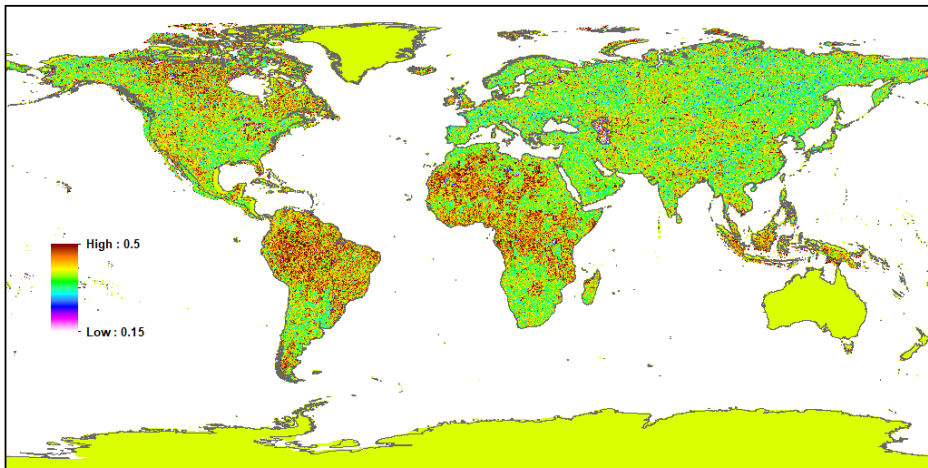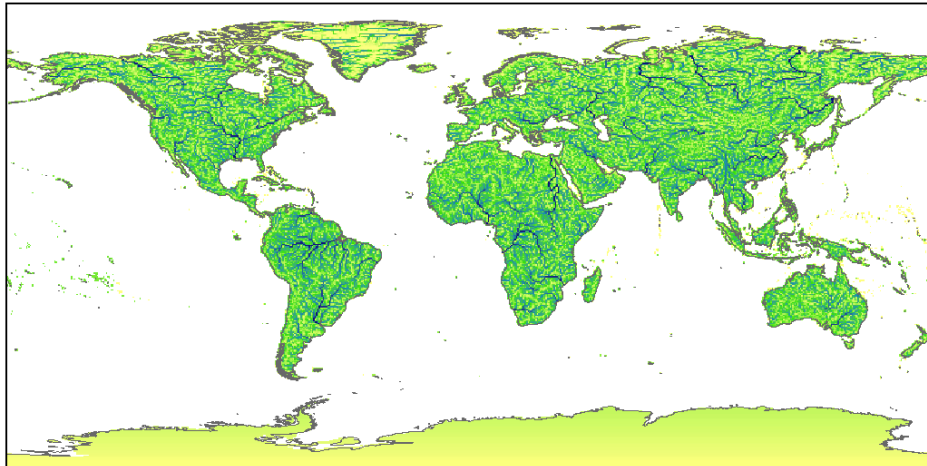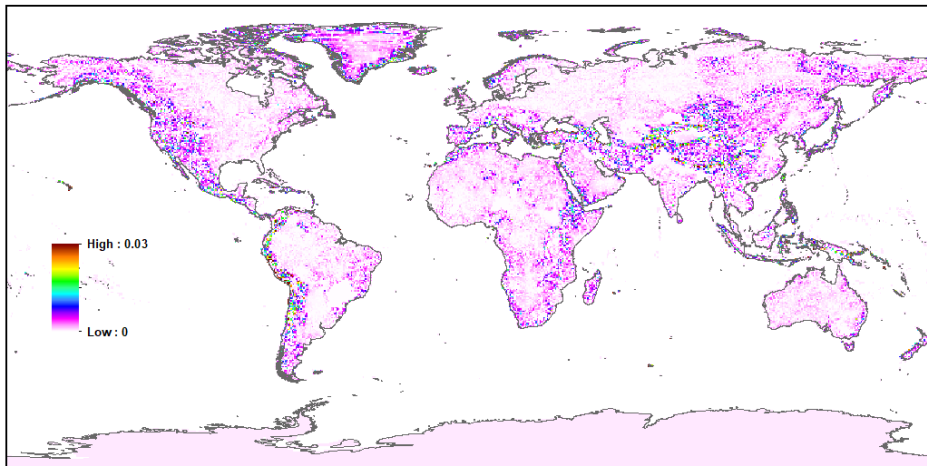(b) CLM5 Slope (Degrees)
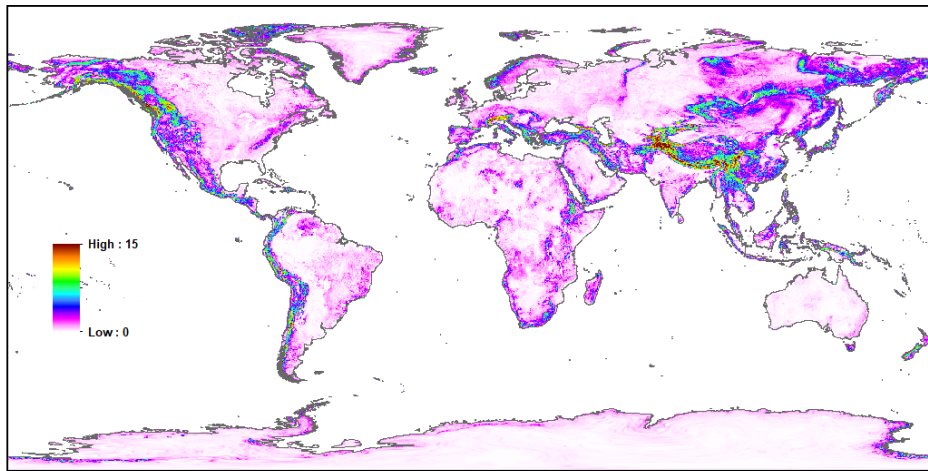
(c) CLM5 FMAX

Figure 4.7: Global Current Day CLM5 Hydrology Surface Data for: (a) Elevation; (b) Slope; and (c) Maximum Saturated Fraction (fmax)

(a) CLM5 MOSART River Network 0.5 Degrees

(b) CLM5 MOSART Channel Slope (Fraction)

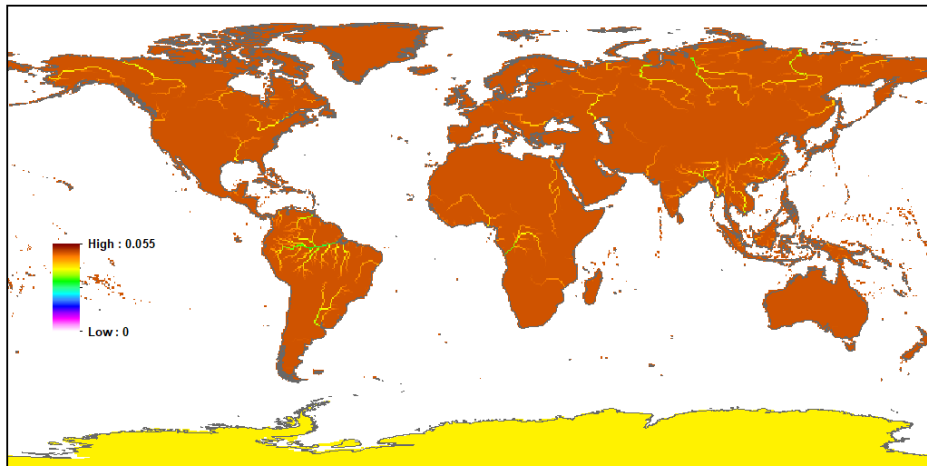(c) CLM5 MOSART Channel Roughness (Manning Coefficient)

Figure 4.8: Global Current Day CLM5 MOSART River Transport Model Data for: (a) River Network; (b) Channel Slope; and (c) Channel Roughness.

4.7 River Transport MOSART

CLM5 uses the Model for Scale Adaptive River Transport (MOSART) (Li et al. 2013b) to transport surface runoff is routed across hillslopes and then discharged along with subsurface runoff into a tributary subnetwork before entering the main channel. The MOSART model is supported by a comprehensive, global hydrography dataset at 0.5 degree resolution. The topographic parameters (such as flow direction, channel length, topographic and channel slopes, etc.) were derived using the Dominant River Tracing (DRT) algorithm from Wu et al. (2011) and Wu et al. (2012) using the baseline high-resolution hydrography dataset is the 1 km resolution HydroSHEDS data of Lehner and Doll (2004) and Lehner et al. (2008). The CLM5 MOSART River Network, Channel Slope, and Channel Roughness are mapped globally in Figure 4.8.

4.8 MODIS Satellite Phenology Leaf and Stem Area Index

CLM5 updates the Satellite Phenology (SP) representations of Plant Functional Type (PFT) monthly Leaf Area Index (LAI) and Stem Area Index (SAI) data from CLM4.5. The new PFT LAI values are derived from MODIS MCD15A2 version 5 8-day data from 2003 to 2015, combined with the current day PFT distributions from Chapter 2, and CRU monthly climatology surface temperature data. The new PFT LAI and SAI values are generated following updated methods from Lawrence and Chase (2007). The new PFT LAI and SAI data are combined with the ICESAT derived PFT Canopy Height, described in section 4.10, to provide CLM5 prescribed monthly vegetation phenology and morphology.

The first step in the generation of the monthly PFT LAI is creating a global 8-day satellite climatology of LAI from the 12 years of the raw MCD15A2 data at the 1 km resolution of the source data. The 8-day climatology was generated using all high-quality data for that day of the year for the 12 years. Any grid cells that had no high-quality data for a particular day were temporally filled using an inverse time weighted interpolation from the previous and next climatology values for the grid cell. The 8-day climatology values were then averaged for each month for the grid cell to produce a monthly climatology of LAI at the 1 km resolution.

Table 4.4: CLM5 Plant Functional Type (PFT) Leaf and Stem Area Index (LAI and SAI), and Canopy Height Satellite Phenology Parameters.

	LAIMax	Evg	K Evg	Cold	Base T	S GDD	SAI α	SAIMin	Bottom	Top
NdlEvgTemp	7	1	0.15	0	5	200	0.25	1	8.5	17
NdlEvgBorl	7	1	0.15	0	5	200	0.25	1	8.5	17
NdlDecBorl	6	0	0	1	2	100	0.25	1	7	14
BrdEvgTrop	7	1	0.15	0	5	200	0.25	1	1	35
BrdEvgTemp	7	1	0.15	0	5	200	0.25	1	1	35
BrdDecTrop	6	0	0	1	5	200	0.25	1	10	18
BrdDecTemp	6	0	0	1	5	200	0.25	1	11.5	20
BrdDecBorl	6	0	0	1	5	200	0.25	1	11.5	20
ShrEvgTemp	4	0	0	0	5	200	0.25	1	0.1	0.5
ShrDecTemp	4	0	0	0	5	200	0.25	1	0.1	0.5
ShrDecBorl	4	0	0	0	5	200	0.25	1	0.1	0.5
GrsC3Arc	4	0	0	0	5	200	0.25	1	0.01	0.5
GrsC3	4	0	0	0	5	200	0.25	1	0.01	0.5
GrsC4	4	0	0	0	5	200	0.25	1	0.01	0.5
Crop	6	0	0	0	5	200	0	0.1	0.01	0.5

The monthly climate LAI climatology was combined with the current day PFT distributions of Chapter 2, PFT LAI maximum values taken from Lawrence and Chase (2007), CRU monthly climatology surface air temperature data, and evergreen and cold deciduous phenology rules to produce monthly PFT LAI values for each grid cell at 1 km. The PFT LAI maximum values along with the evergreen minimum LAI extinction and cold deciduous parameters are shown in Table 4.4. The initial PFT LAI allocation from the grid cell level LAI is performed conservatively using the fraction of potential maximum LAI for the grid cell as shown in Equations 4.1 and 4.2. The initial allocation ensures that the when the individual PFT percentages are combined with the PFT LAI for a month, then they exactly recreate the MODIS MCD15A2 climatology LAI for that month for that grid cell.

After the initial allocation of the monthly PFT LAI, the phenology parameters of Table 4.4 were used to calculate the monthly Evergreen Minimum PFT LAI values using Equation 4.3, and the monthly Cold Deciduous Maximum PFT LAI values using Equation 4.4. The Evergreen phenology of Equation 4.3 is a smoother version of the Evergreen Tree LAI phenology minimum values generated with the methods of Zeng et al. (2002). The Cold Deciduous Maximum PFT LAI methods of Equation 4.4 are taken directly from the Lund-Potsdam-Jena (LPJ) dynamic vegetation model, as described by Sitch et al. (2003). Both phenology methods were first introduced in CLM3.5 as described by Lawrence and Chase (2007).

The final monthly individual PFT LAI values were calculated at the 1 km resolution, accounting for the monthly PFT minimum and maximum values, using Equation 4.2 with the updated potential and remaining unallocated MODIS LAI for both the constrained and unconstrained PFTs. This calculation allowed for the non-conservative nature of the phenology rules so that when combined with PFT percentages they could exceed or not meet the MODIS LAI for a grid cell, accounting individual PFT phenology. This was important for snow covered regions at low sun angles in Boreal or Austral winter where the MODIS data has poor LAI retrievals for many months.

$$Potential\ LAI = \sum_{i=All\ PFTs} PFT\ Frac_i \times PFT\ LAI\ MAX_i \tag{4.1}$$

$$PFT\ LAI_{i\ month} = \frac{MODIS\ LAI_{month}}{Potential\ LAI} \times PFT\ LAI\ MAX_i \tag{4.2}$$

$$PFT\ LAI\ Evg\ Min_{i\ month} = PFT\ LAI_{i\ maxmonth} \times \exp(-K\ Evg_i \times Months_{i\ frommax}) \tag{4.3}$$

$$GDD_{i\ month} \geq S\ GDD_i \rightarrow PFT\ LAI\ Decid\ Max_{i\ month} = PFT\ LAI\ MAX_i \tag{4.4}$$
$$GDD_{i\ month} < S\ GDD_i \rightarrow PFT\ LAI\ Decid\ Max_{i\ month} = 0$$

The results of the CLM5 monthly PFT LAI mapping are shown for all combined PFT LAIs for January in Figure 4.9a, and for July in Figure 4.10a. Comparison of the January all combined PFT LAI to the January climatology MODIS LAI mapping of Figure 2.10a shows the higher winter LAI for Needleleaf Evergreen forests on North America, northern Europe and Eurasia. Outside of these regions the comparison shows the consistency between the CLM5 combined PFT LAI data and the original MODIS LAI data for January. The July MODIS LAI mapping of Figure 2.10b shows the same agreement between the CLM5 combined PFT LAI data and the original MODIS data for most areas. The contributions from Evergreen Trees relative to the Deciduous Trees can be seen for January in Figure 4.10 and for July in Figure 4.11. reflecting the Evergreen and Cold Deciduous phenology rules and their relative PFT LAI maximum values.

The average monthly MODIS, All Vegetation, and individual PFT LAI values are calculated globally and regionally for January and July in Table 4.5. At a global level the average total combined PFT LAI and the MODIS climatology data are the same at 0.9 m2/m2 in January and 1.5 m2/m2 in July reflecting the North Hemisphere Boreal summer green up. The regional analysis show Southeast Asia had the highest average LAI at 3.7-3.8 m2/m2 followed by Latin America at 2.6-2.7 m2/m2 for January and July. Europe, Eurasia, North America and East Asia all have strong seasonal LAI with average LAI 1.5-2.0 m2/m2 in Boreal summer drop to less than 0.5 m2/m2 in Boreal winter.

Table 4.5: MODIS CLM5 Global and IPCC Regional Plant Functional Type (PFT) Monthly Climatology Leaf Area Index (LAI) in m^2/m^2. MODIS data is generated as the monthly climatology from 2003 – 2015 from 8-day 1 km MCD15A2 version 5 data.

|	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
January												
MODIS	0.9	1.0	0.7	0.3	0.1	0.5	2.7	0.0	0.2	3.7	0.7	0.0
AllVeg	0.9	1.0	0.7	0.3	0.3	0.6	2.7	0.0	0.3	3.7	0.7	0.0
NdlEvgTemp	1.5	1.3	2.4	1.5	1.4	1.4	1.4	1.2	1.5	2.5	1.6	1.6
NdlEvgBorl	1.2	0.8	2.6	1.2	1.4	1.2	1.0	0.0	0.9	1.0	0.9	0.8
NdlDecBorl	0.0	0.8	1.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	1.1
BrdEvgTrop	3.3	2.8	1.4	3.2	0.0	0.0	3.9	0.9	2.3	4.2	1.8	3.0
BrdEvgTemp	2.2	2.0	1.7	3.0	2.4	2.2	2.4	0.9	1.8	4.5	2.1	2.9
BrdDecTrop	1.7	1.6	1.2	1.0	0.0	0.0	1.9	0.4	0.9	1.6	1.1	1.2
BrdDecTemp	0.4	1.6	1.0	0.3	0.0	0.2	1.5	0.2	0.1	1.8	0.7	1.3
BrdDecBorl	0.1	3.7	2.2	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0
ShrEvgTemp	0.4	0.4	0.5	0.0	0.1	0.3	0.7	0.2	0.3	0.0	0.0	0.0
ShrDecTemp	0.4	0.4	0.4	0.1	0.1	0.3	0.4	0.2	0.2	0.6	0.2	0.5
ShrDecBorl	0.0	0.3	0.7	0.1	0.0	0.0	0.4	0.0	0.0	0.1	0.1	0.0
GrsC3Arc	0.0	0.7	1.2	0.1	0.0	0.0	0.7	0.0	0.0	0.1	0.0	1.1
GrsC3	0.4	0.9	0.6	0.1	0.0	0.4	1.1	0.3	0.1	1.5	0.3	1.1
GrsC4	0.9	0.9	0.5	0.2	0.1	0.3	1.5	0.3	0.2	1.8	0.5	2.4
Crop	1.2	1.3	0.7	0.4	0.0	0.4	2.3	0.3	0.2	3.4	0.8	0.0
July												
MODIS	1.5	0.9	0.7	1.4	2.0	1.9	2.6	0.1	1.5	3.8	0.9	0.0
AllVeg	1.5	0.9	0.7	1.4	2.0	2.0	2.6	0.1	1.6	3.8	0.9	0.0
NdlEvgTemp	3.4	1.8	3.9	3.3	3.7	3.3	1.9	3.1	3.6	2.7	2.1	2.1
NdlEvgBorl	2.8	0.5	1.8	3.0	3.5	2.8	1.0	0.0	2.2	1.5	1.6	0.4
NdlDecBorl	2.8	0.0	0.0	3.6	2.9	2.1	0.2	0.0	1.9	0.0	1.9	0.0
BrdEvgTrop	3.2	2.6	1.1	3.9	0.0	0.0	3.9	1.1	3.4	4.3	1.8	3.6
BrdEvgTemp	1.7	1.2	1.4	4.0	2.9	2.5	1.6	1.3	4.0	3.9	2.2	3.5
BrdDecTrop	1.6	1.6	0.9	2.0	0.0	0.0	1.5	0.9	1.9	2.2	1.5	1.5
BrdDecTemp	2.6	0.6	1.6	2.6	3.2	2.7	0.9	2.2	3.2	2.3	1.9	1.8
BrdDecBorl	2.8	0.0	0.4	2.9	3.3	2.9	0.0	0.0	2.4	1.5	1.4	0.0
ShrEvgTemp	0.3	0.3	0.4	0.0	0.4	0.4	0.1	0.3	0.6	0.0	0.4	0.0
ShrDecTemp	0.4	0.3	0.3	1.0	0.5	0.4	0.4	0.4	0.6	0.7	0.4	0.5
ShrDecBorl	1.3	0.2	0.1	0.4	1.4	1.5	0.2	0.0	1.1	0.7	0.6	0.0
GrsC3Arc	1.1	0.3	0.2	0.7	1.5	1.4	0.2	0.0	1.0	0.6	0.5	0.5
GrsC3	1.1	0.6	0.8	1.1	1.5	1.3	0.8	0.4	1.5	2.0	0.6	1.3
GrsC4	1.0	0.9	0.4	1.1	0.7	1.1	1.3	0.4	1.5	2.1	0.6	2.8
Crop	1.9	1.3	1.5	2.3	2.3	2.0	2.0	0.5	2.3	3.6	1.0	0.0
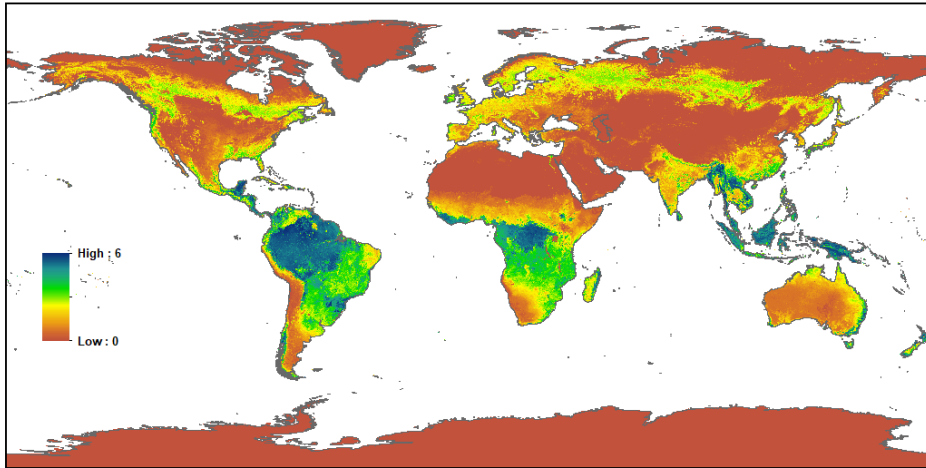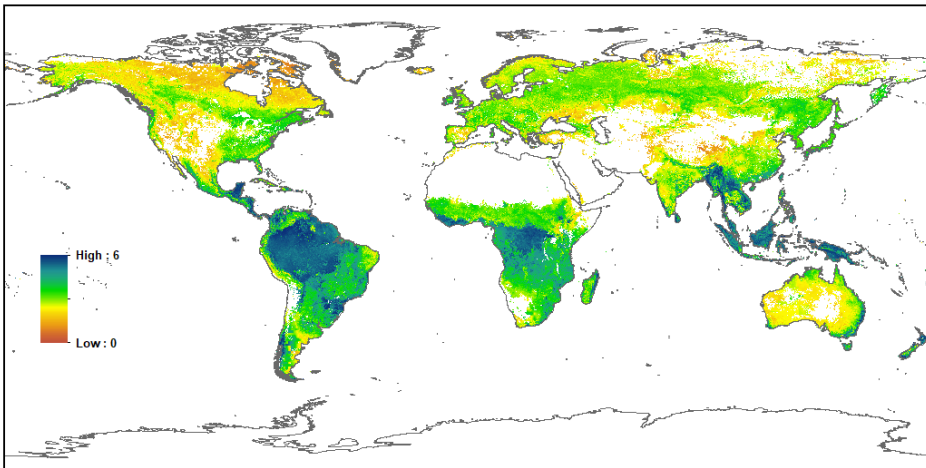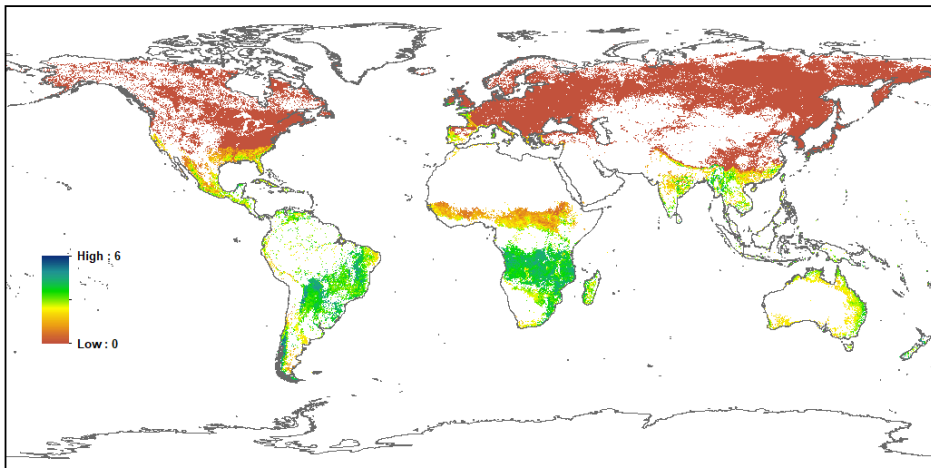
Figure 4.9: Global Current Day CLM5 MODIS January PFT LAI Data for: (a) All Vegetation Combined; (b) Evergreen Trees; and (c) Deciduous Trees.
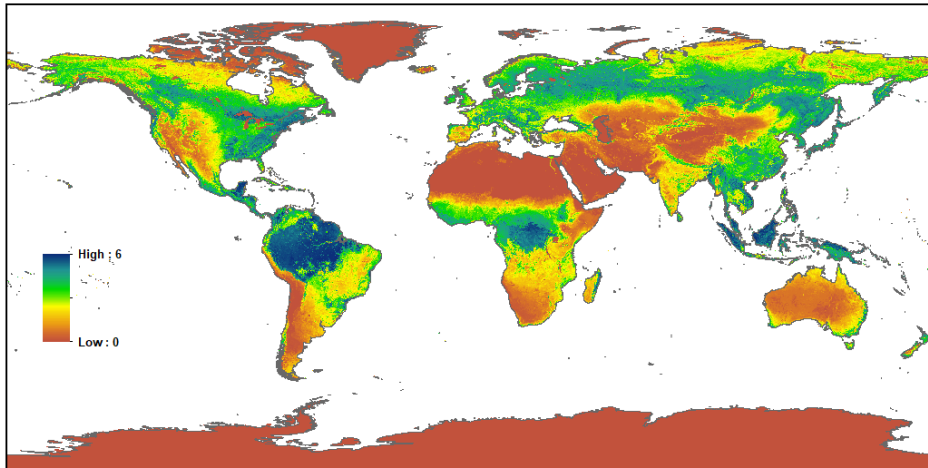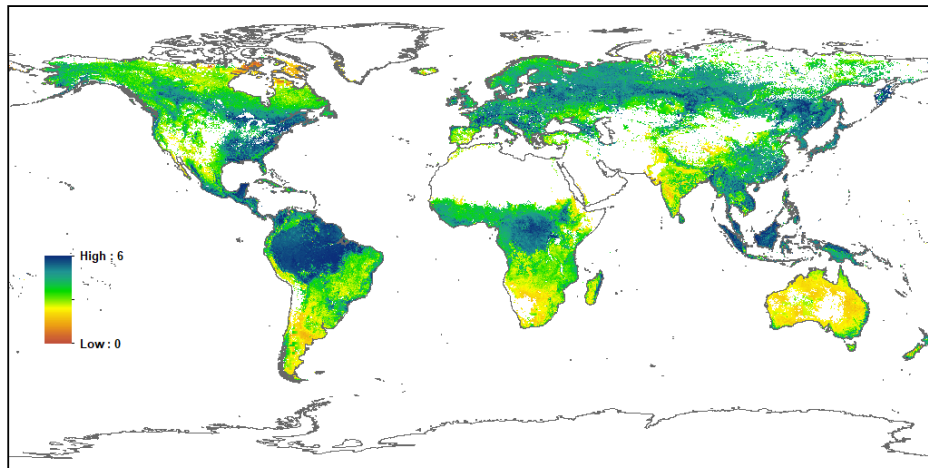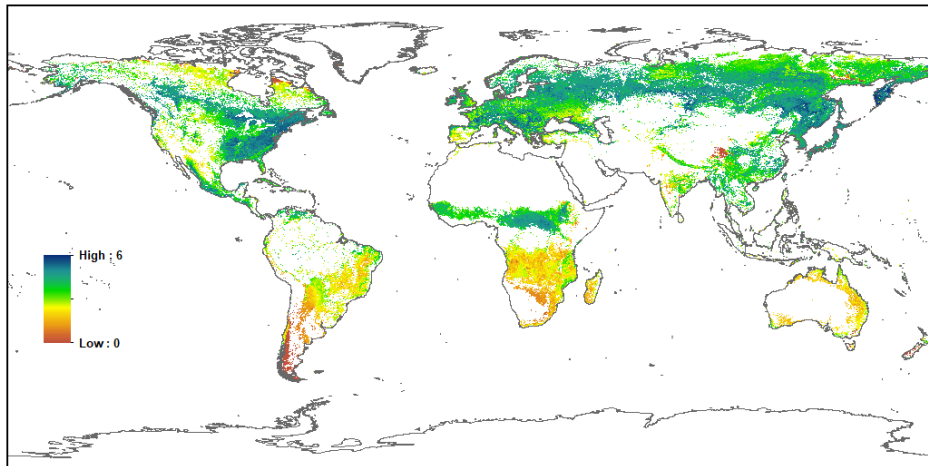
Figure 4.10: Global Current Day CLM5 MODIS July PFT LAI Data for: (a) All Vegetation Combined; (b) Evergreen Trees; and (c) Deciduous Trees.

The highest PFT LAI values are for Broadleaf Evergreen Trees, with average values up to 4.5 m2/m2 in Southeast Asia, and up to 4.0 m2/m2 in Latin America, East Asia and North America. In Africa, Broadleaf Evergreen Trees have substantially lower average values of 2.6-2.8 m2/m2 due to the larger areas of Savannas relative to Tropical Forests in this region. Needleleaf Evergreen Trees also have high average PFT LAI values in Boreal summer but with a strong seasonal cycle despite the evergreen phenology rules enforced in the allocation process. The highest values for Needleleaf Evergreen Trees are 3.9 m2/m2 in the Asia Pacific Developed region which includes Japan. Eurasia, North America, East Asia and Europe all have Boreal summer values of around 3.5 m2/m2.

For non-tree PFTs the shrub and grass PFT LAIs are all lower than the tree PFT LAIs. Average Temperate Shrub PFT LAI values are globally around 0.4 m2/m2 for both January and July showing no consistent seasonal cycle. Boreal Shrubs however have a strong seasonal cycle with an average global value of 1.3 m2/m2 in July during Boreal summer, and 0.0 m2/m2 in January during Boreal winter. Average C4 Grass PFT LAI also shows little seasonal cycle globally at around 1.0 m2/m2 in both January and July reflecting their tropical and sub-tropical dominance. The non-Arctic C3 Grass PFT LAI by contrast does have a seasonal cycle with the global average value ranging from 0.4 m2/m2 in Boreal winter to 1.1 m2/m2 in Boreal summer.

The Arctic C3 Grass PFT LAI shows an even stronger seasonal cycle with the global average ranging from 0.0 m2/m2 in Boreal winter to 1.1 m^2/m^2 in Boreal summer. The regional analysis of Grass PFT LAI shows that these values have a wide range of geographic spread in terms of both seasonality and LAI density. Finally, the Crop PFT LAI has substantially higher average values than those of Shrub and Grass PFTs, with a maximum in July of 1.9 m2/m2 in Boreal summer, and wide range of regional values reflecting the widespread distribution of the Crop PFT across the world.

The precise allocation of the monthly climatology MODIS LAI to CLM5 PFT LAI is shown for six FLUXNET sites in Figures 4.11 and 4.12. In these plots the North America sites of Canada White Pine 2 (CA-WP2), US Duke 1 (US-Dk1), and US Tonzi (US-Ton), all show the persistence the Needleleaf Evergreen Tree PFT LAI values that maintain canopy even when the MODIS monthly climatology LAI drops rapidly or even approaches zero. For the CA-WP2 site in northern Alberta, the Needleleaf Evergreen Tree PFT LAI persists all the way through winter until spring, while the Broadleaf Deciduous Tree and Grass PFT LAI values go to 0.0 m^2/m^2 during September and don't return until April.

At the more southern US-Dk1 site in North Carolina, the Broadleaf Deciduous Tree and Crop PFT LAI values persist until November and return earlier in February. For the US-Ton site in northern California the PFT LAI values are more directly driven by the Mediterranean climate with the wet winter and resulting in all the PFT LAI values peaking in April during the growing season and then decreasing through the hot dry summer until the winter rains start the green up in January again. The differences between the PFTs in each of these cases reflect their underlying phenology.

The tropical South American sites of Brazil Santarem 1 (BR-Sa1), and Brazil Pandeiros (BR-Pan) have very different representations to the North American sites. The BR-Sa1 in the eastern Amazon has only Broadleaf Evergreen Tree PFTs and as such has a direct translation from the MODIS climatology monthly LAI to the PFT LAI. The MODIS LAI has a maximum value of 6.0 m^2/m^2 in August which corresponds to the end of the rainy season, while there is a minimum in value of 3.5 m^2/m^2 in March corresponding with the height of the rainy season. The rainy season minimum indicates that there is strong cloud interference during this time period that is not corrected by the MODIS LAI algorithms or the PFT LAI phenology rules.
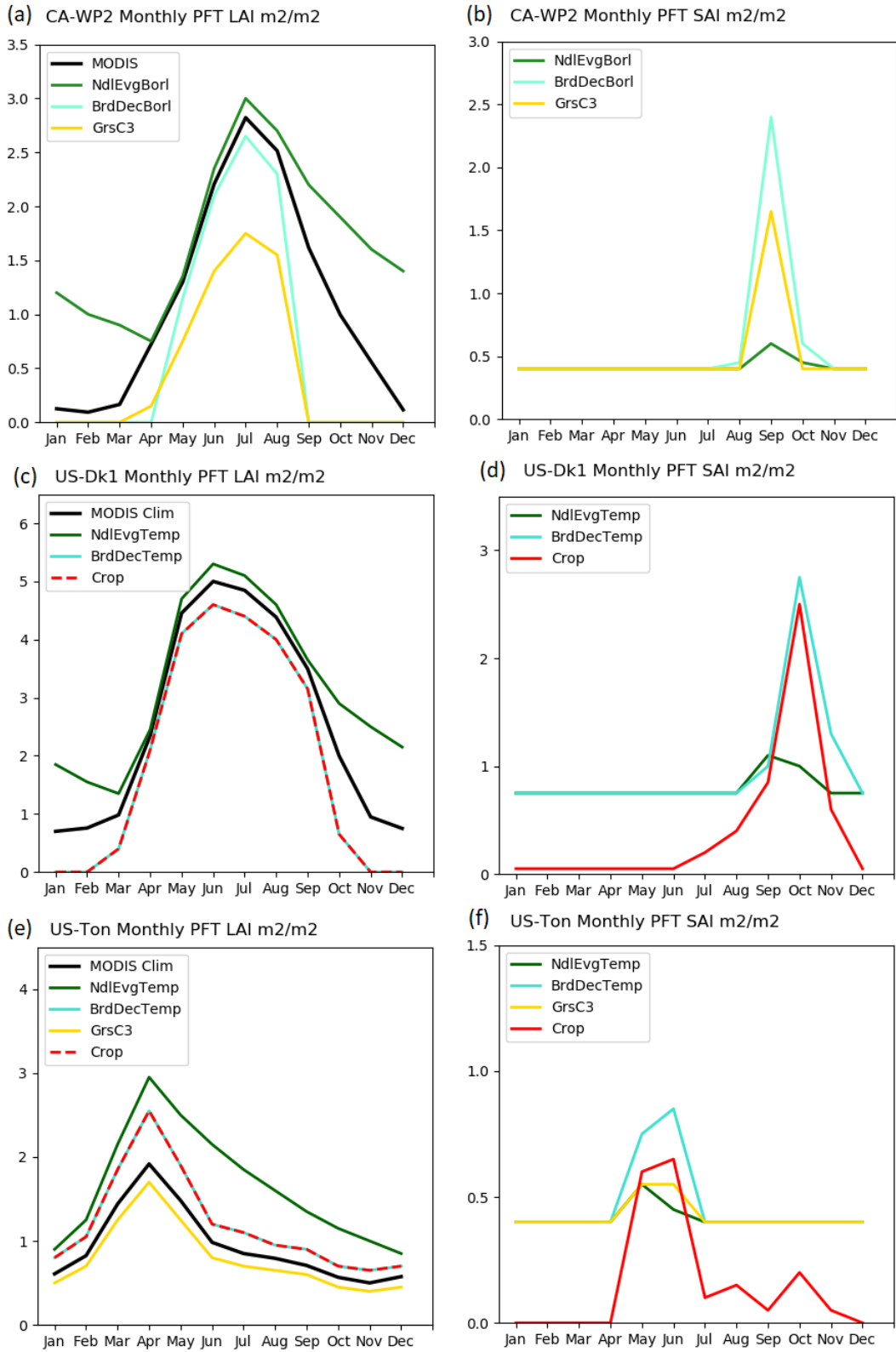
Figure 4.11: CLM5 MODIS Monthly PFT LAI and SAI for FLUXNET Sites of: (a) Canada White Pine 2 (CA-WP2); (b) US Duke 1 (US-Dk1); and (c) US Tonzi (US-Ton).
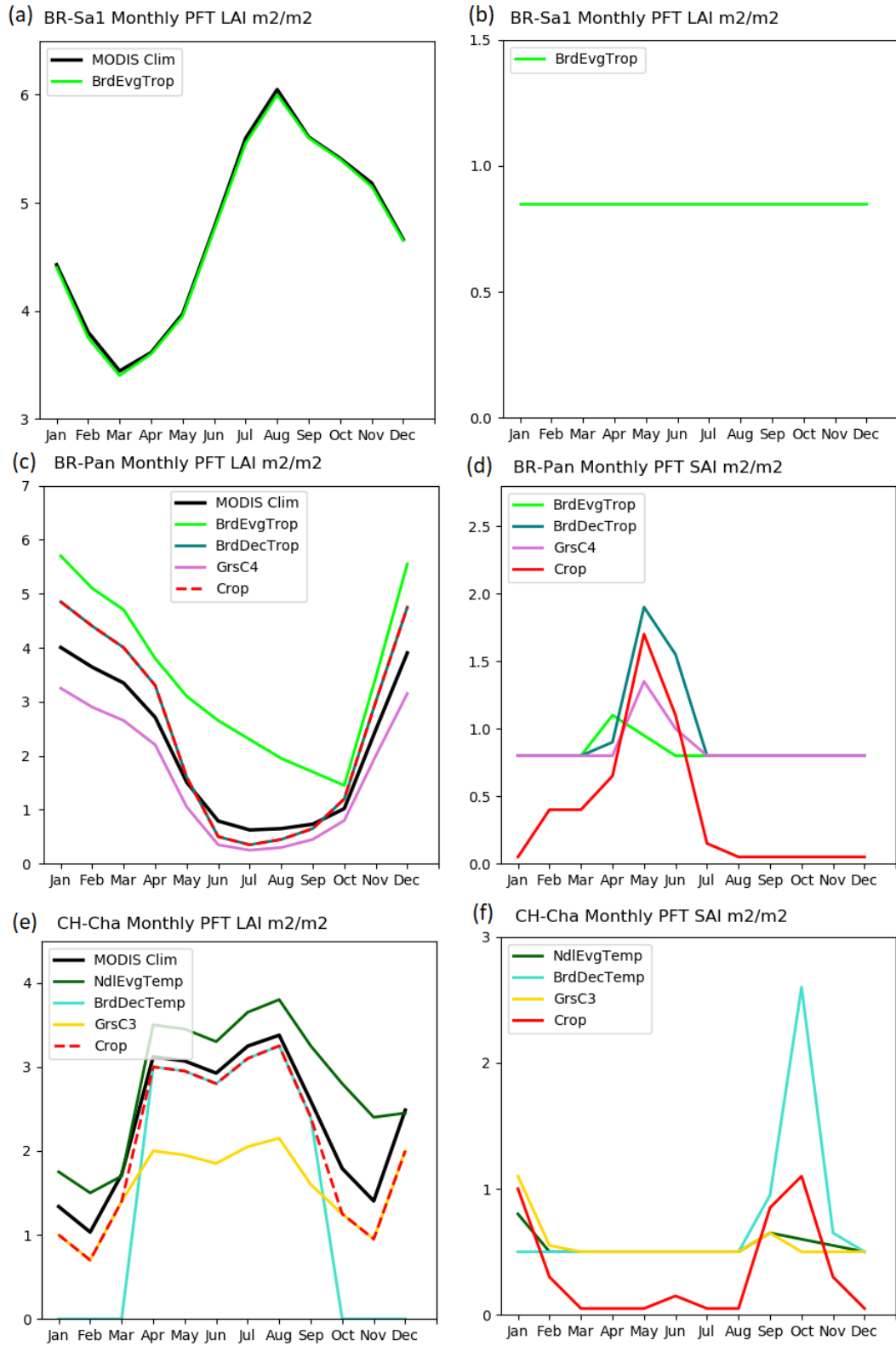
Figure 4.12: CLM5 MODIS Monthly PFT LAI and SAI for FLUXNET Sites of: (a) Brazil Santarem 1 (BR-Sa1); (b) Brazil Pandeiros (BR-Pan); and (c) Switzerland Chamau (CH-Cha).

The BR-Pan site in the southeast of Brazil is a complex site with the 15% Broadleaf Evergreen Trees, 21% Broadleaf Deciduous Trees, 62% C4 Grass, and 2% Crop composition. This complexity reflects the Savanna ecotone of the landscape surrounding the site. The MODIS climatology monthly LAI has a maximum value of 4.0 m^2/m^2 in January at the height of the rainy season and a minimum in from June to October of around 0.9 m^2/m^2 during the dry season.

The relative PFT LAI maximum values of the PFTs result in Broadleaf Evergreen Trees having a maximum PFT LAI of 5.9 m^2/m^2, while Broadleaf Deciduous Trees and Crops have a maximum PFT LAI of 4.9 m^2/m^2, and C4 Grass has a maximum PFT LAI of 3.2 m^2/m^2. The evergreen phenology rules ensure that the Broadleaf Evergreen Tree PFT LAI is maintained through the dry season with a minimum value of 1.6 in October at the end of the dry season. The Broadleaf Deciduous Trees, Crops and C4 Grass PFT LAI values all drop to less 0.5 m^2/m^2 for the dry season.

The final site of Switzerland Chamau (CH-Cha), is a complex European site with 17% Needleleaf Evergreen Tree, 69% Broadleaf Deciduous Tree, 2% C3 Grass, and 12% Crop composition. The Needleleaf Evergreen Tree PFT LAI is consistently higher than the other PFTs with a summer maximum of 3.8 m^2/m^2 and winter minimum 1.5 m^2/m^2. The Broadleaf Deciduous Tree PFT LAI shows the impact of the cold deciduous phenology being senescent from October to March. The Crop and Non-Arctic C3 Grass both follow the MODIS monthly LAI with peaks in August, but with a much larger peak for Crop PFT LAI at 3.0 m^2/m^2, compared to C3 Grass PFT LAI at 2.0 m^2/m^2.

The CLM5 Satellite Phenology (SP) mode also prescribes monthly PFT Stem Area Index (SAI) generated from the current day PFT distributions of Chapter 2 and the monthly PFT LAI values above. The prescribed monthly PFT SAI values are produced globally at the same 1 km resolution as the PFT and monthly PFT LAI data. The monthly PFT SAI values are calculated using updated versions of the methods of Lawrence and Chase (2007), using Equations 4.5, 4.6, 4.7 and 4.8, with PFT parameters from Table 4.4. This model for generating PFT SAI from PFT LAI and PFT SAI minimum values was originally developed by Zeng et al. (2002).

$$PFT\ Dead\ Leaf\ SAI_{i\ month} = max\big(0, PFT\ LAI_{i\ prev} - PFT\ LAI_{i\ month}\big) \qquad (4.5)$$

$$PFT\ SAI_{i\ month} = \alpha\ PFT\ SAI_{i\ prev} + PFT\ Dead\ Leaf\ SAI_{i\ month} \qquad (4.6)$$

$$PFT\ SAI\ Grid\ Min_i = \frac{PFT\ LAI_{i\ maxmonth}}{PFT\ LAI\ MAX_i} \times PFT\ SAI\ Min_i \qquad (4.7)$$

$$PFT\ SAI_{i\ month} = max(PFT\ SAI\ Grid\ Min_i, PFT\ SAI_{i\ month}) \qquad (4.8)$$

In this model the PFT SAI elements represents all the non-photosynthesizing components of the CLM5 canopy. The PFT SAI elements are important for simulating radiation interception, canopy temperature, canopy precipitation and dew interception, as well as canopy roughness, with the associated turbulent energy, moisture and momentum fluxes from these processes. To achieve this the CLM5 monthly PFT SAI represents both the dead leaves, and the stems and branches of the PFT.

The new dead leaf contribution to PFT SAI is calculated by Equation 4.5 as the difference in green leaf PFT LAI from one month to the next. The monthly dead leaf turnover in the canopy is represented by Equation 4.6 using the decay fraction SAI α from the previous month as listed in Table 4.4. To ensure that the stems and branches of the PFTs are always represented, a minimum PFT SAI is set for each PFT. To account for the wide range of possible vegetation densities biogeographically, the minimum PFT SAI value is set locally for a grid cell using Equation 4.7 with the PFT SAI minimum value from Table 4.4, scaled by the grid cell maximum monthly PFT LAI relative to the PFT LAI maximum value also from Table 4.4.

Table 4.6: CLM5 Global and IPCC Regional Plant Functional Type (PFT) Monthly Climatology Stem Area Index (SAI) in m^2/m^2.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
January												
AllVeg	0.3	0.3	0.2	0.2	0.4	0.4	0.5	0.0	0.3	0.6	0.2	0.0
NdlEvgTemp	0.5	0.3	0.6	0.5	0.5	0.5	0.3	0.5	0.5	0.5	0.4	0.3
NdlEvgBorl	0.4	0.1	0.4	0.4	0.5	0.4	0.2	0.0	0.3	0.2	0.2	0.1
NdlDecBorl	0.5	0.1	0.1	0.6	0.5	0.3	0.1	0.0	0.3	0.0	0.3	0.2
BrdEvgTrop	0.6	0.6	0.2	0.6	0.0	0.0	0.7	0.2	0.5	0.7	0.4	0.6
BrdEvgTemp	0.4	0.3	0.3	0.7	0.5	0.4	0.4	0.2	0.6	0.8	0.4	0.5
BrdDecTrop	0.5	0.5	0.2	0.4	0.0	0.0	0.5	0.2	0.3	0.6	0.5	0.4
BrdDecTemp	0.5	0.3	0.4	0.5	0.6	0.6	0.3	0.4	0.5	0.7	0.5	0.3
BrdDecBorl	0.5	0.6	0.4	0.5	0.5	0.5	0.3	0.0	0.4	0.3	0.2	0.0
ShrEvgTemp	0.2	0.2	0.1	0.0	0.2	0.2	0.2	0.1	0.2	0.0	0.1	0.0
ShrDecTemp	0.1	0.2	0.1	0.2	0.2	0.1	0.2	0.1	0.2	0.2	0.1	0.2
ShrDecBorl	0.3	0.1	0.2	0.1	0.3	0.4	0.1	0.0	0.3	0.2	0.2	0.0
GrsC3Arc	0.3	0.2	0.3	0.2	0.4	0.3	0.2	0.0	0.2	0.2	0.1	0.3
GrsC3	0.3	0.3	0.3	0.3	0.4	0.4	0.3	0.1	0.4	0.6	0.2	0.4
GrsC4	0.4	0.4	0.2	0.3	0.3	0.3	0.5	0.2	0.4	0.6	0.3	0.7
Crop	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.0	0.1	0.2	0.1	0.0
July												
AllVeg	0.3	0.3	0.2	0.2	0.4	0.5	0.5	0.0	0.3	0.6	0.2	0.0
NdlEvgTemp	0.5	0.4	0.6	0.5	0.6	0.5	0.3	0.5	0.5	0.5	0.4	0.4
NdlEvgBorl	0.4	0.1	0.5	0.4	0.5	0.4	0.2	0.0	0.3	0.2	0.3	0.1
NdlDecBorl	0.7	0.1	0.1	0.6	0.7	0.3	0.1	0.0	0.3	0.0	0.3	0.2
BrdEvgTrop	0.6	0.6	0.2	0.6	0.0	0.0	0.7	0.2	0.5	0.7	0.4	0.6
BrdEvgTemp	0.4	0.3	0.3	0.7	0.5	0.5	0.4	0.2	0.6	0.8	0.4	0.6
BrdDecTrop	0.5	0.5	0.2	0.4	0.0	0.0	0.5	0.3	0.3	0.5	0.4	0.4
BrdDecTemp	0.5	0.3	0.4	0.5	0.6	0.6	0.3	0.5	0.5	0.6	0.5	0.4
BrdDecBorl	0.5	0.6	0.4	0.5	0.5	0.5	0.3	0.0	0.4	0.3	0.3	0.0
ShrEvgTemp	0.2	0.2	0.1	0.0	0.3	0.2	0.2	0.1	0.2	0.0	0.2	0.0
ShrDecTemp	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.1	0.2
ShrDecBorl	0.3	0.2	0.3	0.1	0.3	0.4	0.1	0.0	0.3	0.2	0.2	0.0
GrsC3Arc	0.3	0.2	0.4	0.2	0.4	0.3	0.2	0.0	0.2	0.2	0.1	0.3
GrsC3	0.3	0.3	0.3	0.3	0.4	0.5	0.3	0.2	0.4	0.6	0.2	0.4
GrsC4	0.4	0.4	0.2	0.3	0.3	0.3	0.5	0.2	0.4	0.6	0.3	0.7
Crop	0.2	0.2	0.1	0.1	0.2	0.4	0.2	0.1	0.1	0.2	0.1	0.0

(a) CLM5 All Vegetation PFT SAI Climatology January   2003 - 2015
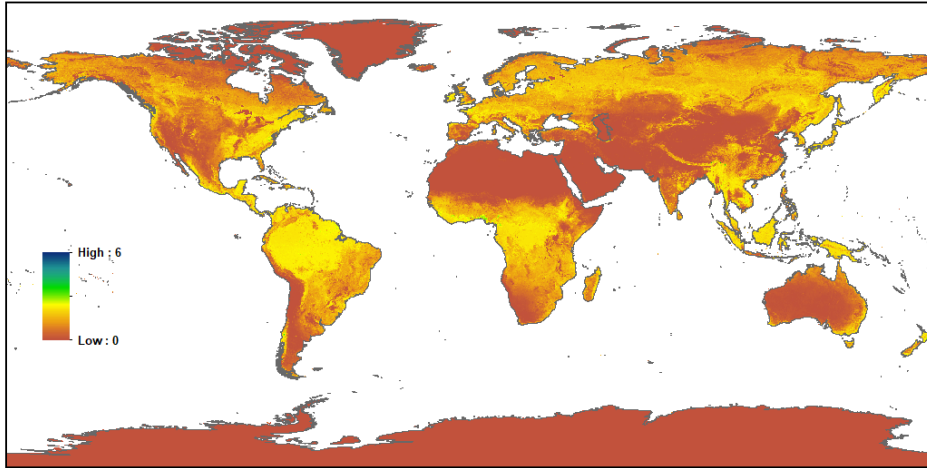
(b) CLM5 Tree Evergreen PFT SAI Climatology January   2003 - 2015
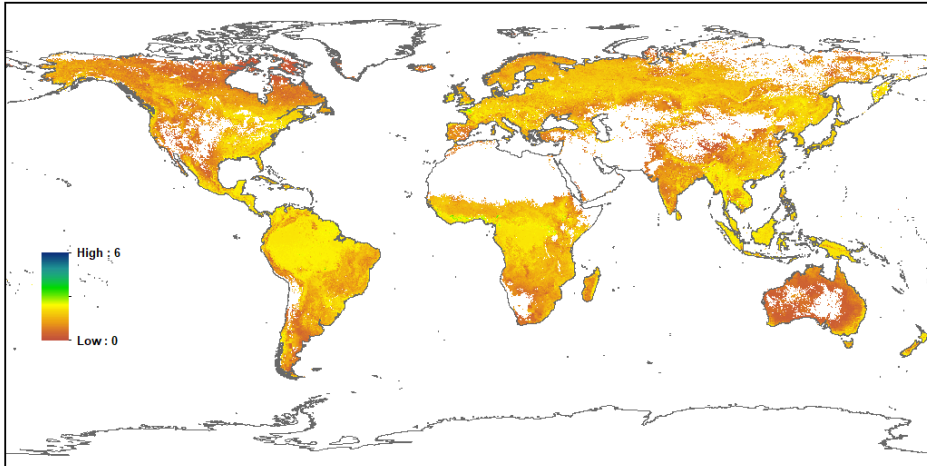
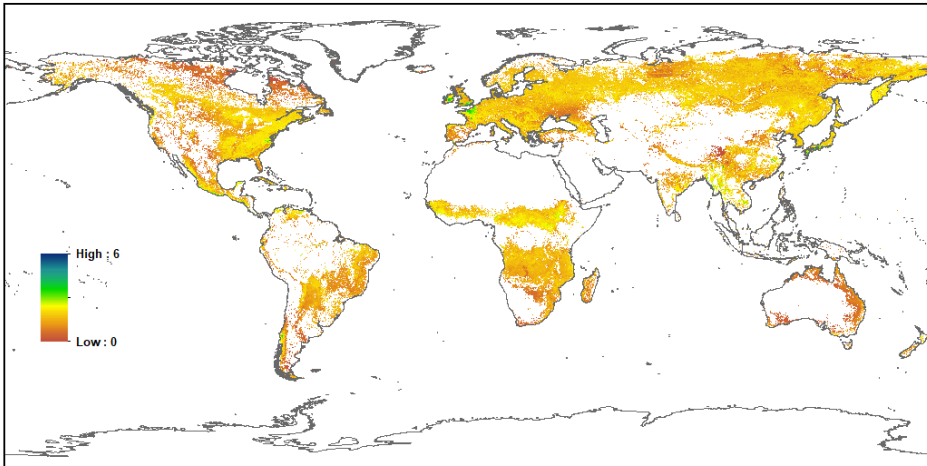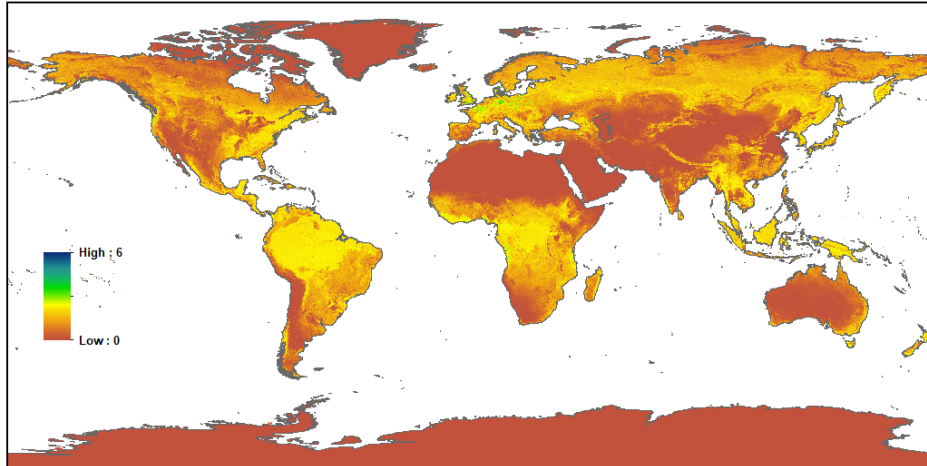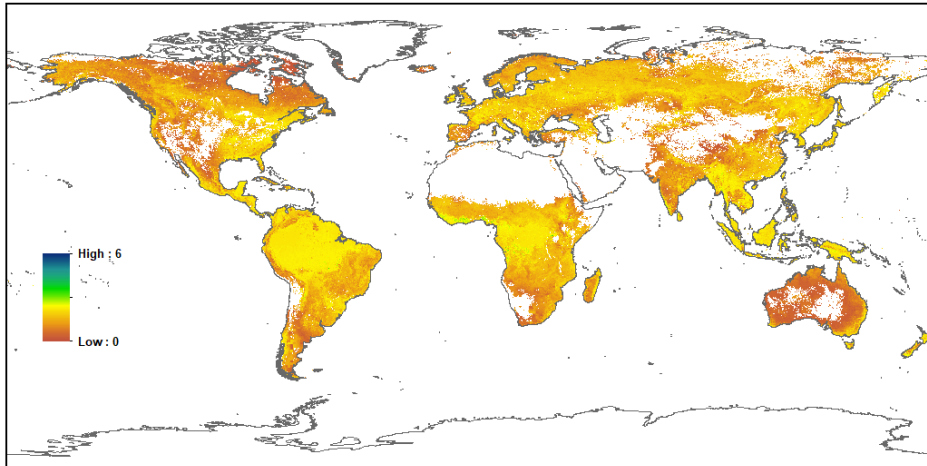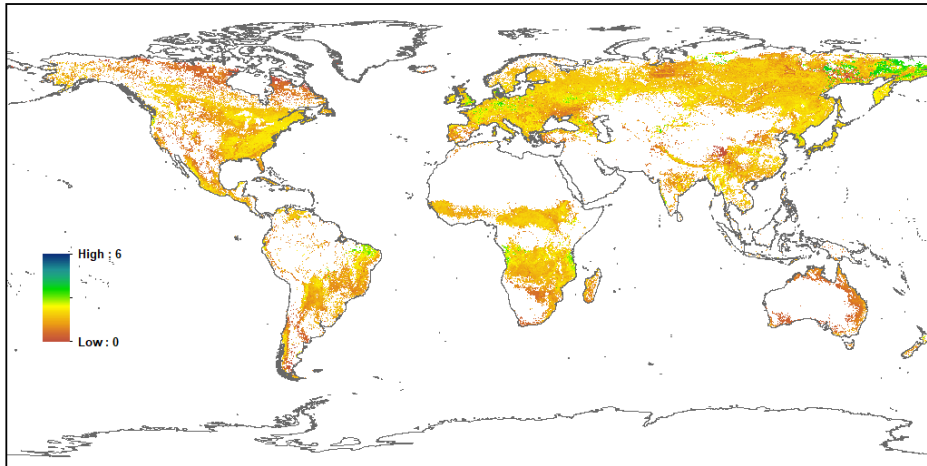(c) CLM5 Tree Deciduous PFT SAI Climatology January   2003 - 2015

Figure 4.13: Global Current Day CLM5 MODIS January PFT SAI Data for: (a) All Vegetation Combined; (b) Evergreen Trees; and (c) Deciduous Trees.

(a) CLM5 All Vegetation PFT SAI Climatology July   2003 - 2015

(b) CLM5 Tree Evergreen PFT SAI Climatology July   2003 - 2015

(c) CLM5 Tree Deciduous PFT SAI Climatology July   2003 - 2015

Figure 4.14: Global Current Day CLM5 MODIS July PFT SAI Data for: (a) All Vegetation Combined; (b) Evergreen Trees; and (c) Deciduous Trees.

The results of the CLM5 monthly PFT SAI mapping are shown for all combined PFT SAIs for January in Figure 4.13a, and for July in Figure 4.14a. Comparison of the all combined PFT SAI to the all combined PFT LAI mapping of Figure 4.10a and 4.11a, shows that SAI is consistently lower than LAI for both months. There are areas that have short periods of higher SAI such as with Broadleaf Deciduous Temperate and Boreal trees during autumn that can reach values over 4.0 m^2/m^2 for a month before dropping the dead leaves from the canopy.

The Broadleaf Evergreen trees of Figure 4.13b and Figure 4.14b maintain constant PFT SAI values close to the SAI minimum values of Table 4.4 in forested areas such as the Amazon, Congo and Southeast Asia. In more sparsely vegetated areas such as savannas the Broadleaf Evergreen tree PFT SAI values are lower reflecting the lower PFT LAI maximum values of these ecosystems. This can be seen in the eastern parts of South America, north and south of the Congo in Africa and inland Australia of both figures.

The Needleleaf Evergreen trees also maintain the PFT SAI minimum values through the Boreal and Temperate forests but have lower values further north into the Tundra. The Deciduous tree PFT SAI of Figure 4.13c and Figure 4.14c both show the dynamic range of these PFTs with small areas in both January and July reaching values as high as 4.0 m^2/m^2. These higher PFT SAI values are more widely spread in September and October in the Boreal autumn when the PFT LAI values drop rapidly for the Deciduous tree PFTs.

The average monthly All Vegetation, and individual PFT SAI values are calculated globally and regionally for January and July in Table 4.6. The average global values show the consistency between Boreal winter and summer for all PFT SAI values. The regional analysis shows the highest PFT SAI values were for Southeast Asia at 0.6 m^2/m^2, with Eurasia, Europe and Latin America all around 0.5 m^2/m^2. The individual average monthly PFT SAI values showed some variety across regions and globally, but all average values were less than 1.0 for both months.

The calculation of the CLM5 monthly PFT SAI from PFT LAI values is shown for the six FLUXNET sites in Figures 4.11 and 4.12. The three North America sites of CA-WP2, US-Dk1, and US-Ton, all show very different timings for the Deciduous tree and Herbaceous monthly PFT SAI values. The northern site of CA-WP2 has a strong peak in PFT SAI for Broadleaf Deciduous Boreal Tree and C3 Grass in September with a small increase in Needleleaf Evergreen Boreal Tree. The more southern US-DK1 site has a peak in PFT SAI a month later in October for Broadleaf Deciduous Temperate Tree and Crop. The northern California site of US-Ton by contrast has a peak in PFT SAI in June in Broadleaf Deciduous Temperate Tree and Crop with smaller peaks in C3 Grass and Needleleaf Evergreen Temperate Tree. These peaks in PFT SAI correspond with the beginning of the hot dry Boreal summer of the Mediterranean climate of the region.

The South American FLUXNET sites in Figure 4.12 again show very different annual cycles for PFT SAI compared to the North American sites. For the tropical rainforest site at BR-Sa1 the Broadleaf Evergreen Tropical Tree PFT SAI is constant at the local PFT SAI minimum value of 0.9 m^2/m^2. The drier savanna site of BR-Pan has a peak in PFT SAI in May corresponding with the dry season. The peak in PFT SAI is found in Broadleaf Deciduous Tropical Tree, C4 Grass and Crop, but with a much smaller peak in Broadleaf Evergreen Tropical Tree. The Swiss site at CH-Cha has a similar PFT SAI seasonality to the US-Dk1 site, with a peak in PFT SAI in October for both the Broadleaf Deciduous Temperate Tree and Crop PFTs. The FLUXNET sites also show the difference in minimum PFT SAI for Crop compared to other PFTs.

For the final CLM5 PFT LAI and SAI input data files the values for each PFT are extrapolated globally from the distributions shown in Figures 4.9, 4.10, 4.13 and 4.14. This extrapolation

process was performed to ensure that there are monthly values for each PFT for each land grid cell in the data to allow for use with alternate vegetation distributions. The extrapolation process was done as a series of expanding search boxes for grid cells where the PFT did not exist in the current day PFT mapping. Where the search box crossed the equator months were offset by six months to account for Austral and Boreal seasons.

4.9 ICESAT Satellite Canopy Height

CLM5 Satellite Phenology (SP) configuration also prescribes canopy height top and bottom for all PFTs. For Tree PFTs the canopy height is derived from the ICESAT canopy height mapping of Simard et al. (2011) using Equations 4.9, 4.10 and 4.11. with PFT parameters from Table 4.4. For other PFTs and CFTs Canopy Height remains the same as those originally derived for CLM3 as listed in Oleson et al. (2004) and listed in Table 4.4.

$$Potential\ Canopy\ Top = \sum_{i=Tree\ PFTs} PFT\ Tree\ Frac_i \times PFT\ Top_i \qquad (4.9)$$

$$PFT\ Canopy\ Top_i = \frac{ICESAT\ Canopy\ Top}{Potential\ Canopy\ Top} \times PFT\ Top_i \qquad (4.10)$$

$$PFT\ Canopy\ Bottom_i = \frac{PFT\ Canopy\ Top_i}{PFT\ Top_i} \times PFT\ Bottom_i \qquad (4.11)$$

The tree canopy height conversion from the grid cell level ICESAT value to individual PFT canopy height top and bottom values follows a similar relationship to the PFT LAI and SAI calculations in the previous section. The initial PFT canopy height top is calculated using Equations 4.9 and 4.10. Here a tree potential canopy top height for the grid cell is calculated using the tree PFT fractions multiplied by the prescribed PFT canopy height top listed in Table 4.4. The relative canopy height observed by the ICESAT satellite is then used to scale each of tree PFT canopy height top values by the difference in potential and observed values for the grid cell. This same scaling is applied for the PFT canopy height bottom using the PFT values listed in Table 4.4.

Table 4.7: ICESAT CLM5 Global and IPCC Regional Plant Functional Type (PFT) Canopy Height in meters.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
ICESAT	10.5	7.8	7.2	11.0	12.5	14.7	18.1	2.1	10.4	23.8	9.0	0.0
NdlEvgTemp	18.3	15.5	20.2	19.8	17.9	17.5	20.6	20.5	17.9	18.6	20.4	17.1
NdlEvgBorl	15.9	8.5	19.1	18.3	17.5	15.7	9.8	0.0	13.8	21.0	18.1	8.5
NdlDecBorl	14.2	7.0	15.4	16.8	14.5	11.4	8.2	0.0	8.8	0.0	14.0	7.0
BrdEvgTrop	24.9	23.9	21.4	22.9	0.0	0.0	26.2	28.7	21.3	26.5	23.4	28.3
BrdEvgTemp	24.4	22.0	24.7	26.5	25.4	22.2	24.1	24.3	22.7	27.0	29.3	25.0
BrdDecTrop	12.5	12.7	10.6	13.5	0.0	0.0	12.6	14.6	13.9	12.2	12.2	19.8
BrdDecTemp	20.7	14.6	17.9	22.9	21.5	20.4	16.8	24.9	21.9	19.6	22.8	19.7
BrdDecBorl	19.3	10.5	23.2	21.1	20.8	19.7	14.2	0.0	16.7	23.8	21.5	0.0
ShrEvgTemp	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
ShrDecTemp	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
ShrDecBorl	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
GrsC3Arc	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
GrsC3	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
GrsC4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Crop	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

(a) ICESAT Canopy Height (m)
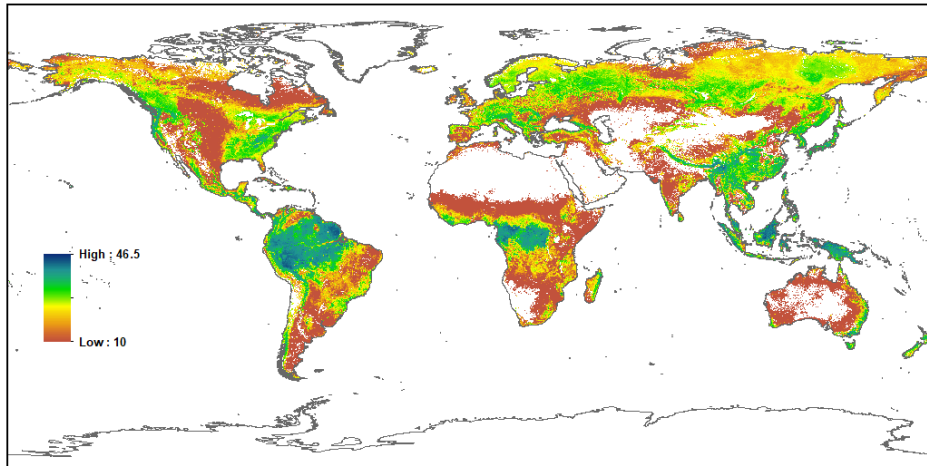
(b) CLM5 Broadleaf Evergreen Tree Canopy Top (m)
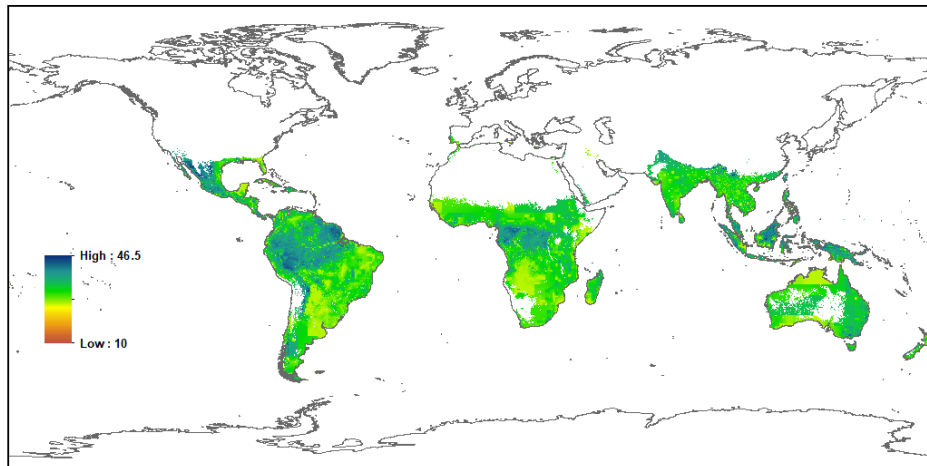
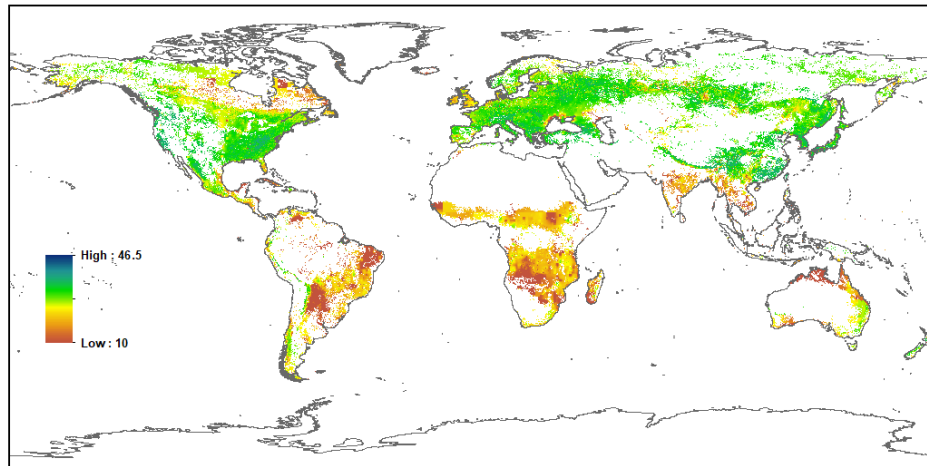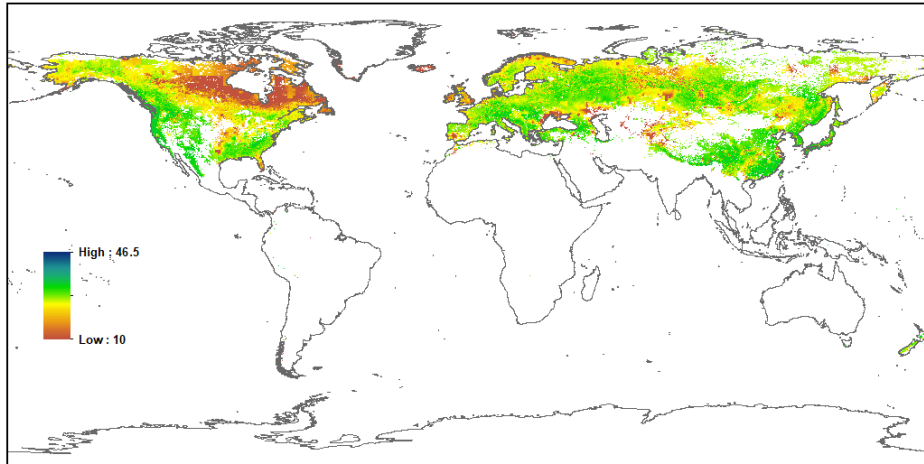**(c) CLM5 Broadleaf Deciduous Tree Canopy Top (m)**

Figure 4.15: Global Current Day CLM5 ICESAT Canopy Height Data for: (a) All Vegetation Combined; (b) Broadleaf Evergreen Trees; and (c) Broadleaf Deciduous Trees.

(a) CLM5 Needleleaf Evergreen Tree Canopy Top (m)
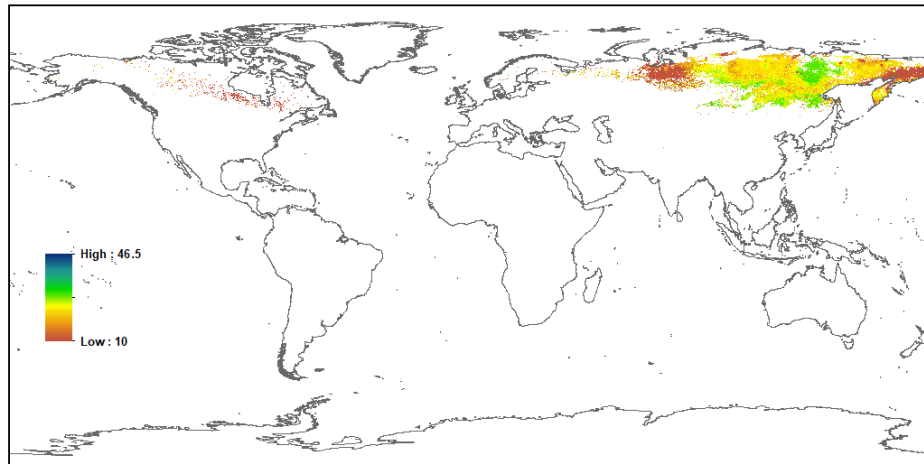

(b) CLM5 Needleleaf Deciduous Tree Canopy Top (m)

Figure 4.16: Global Current Day CLM5 ICESAT Canopy Height Data for: (a) Needleleaf Evergreen Trees; and (b) Needleleaf Deciduous Trees.

The ICESAT grid cell canopy height along with the results from the CLM5 PFT canopy height mapping for Broadleaf Evergreen Tree, and Broadleaf Deciduous Tree PFTs are shown in Figure 4.15. The CLM5 PFT canopy height mapping for Needleleaf Evergreen Tree and Needleleaf Deciduous Tree PFTs are shown in Figure 4.16. Like the PFT LAI and SAI mapping, the CLM5 PFT canopy height top and bottom input data files have the values for each PFT extrapolated globally to provide values for where the PFT does not exist in the current day PFT mapping.

The ICESAT canopy height mapping shows the highest values are in the tropical forests of the Amazon, Congo, and Southeast Asia with values over 30 meters tall. The grid cell canopy height rapidly drops off in the transition to savannas reflecting both the shorter trees and the lower tree density of these ecosystems. There also are higher canopy height values associated with the temperate and boreal forests, which again fall off as they transition to savannas or tundra.

The Broadleaf Evergreen Tree PFT canopy height mapping shows that in tropical forests the CLM5 PFT values match well with ICESAT grid cell height. For the savannas however the Broadleaf Evergreen Tree heights while lower are not the main driver of the lower ICESAT grid cell height. The Broadleaf Deciduous Tree canopy height mapping does show there are lower CLM5 PFT tree heights in savannas, which combines with the lower tree density in these areas to explain this large decrease in ICESAT height.

The Broadleaf Deciduous Tree and Needleleaf Evergreen Tree canopy height mapping combine to show the consistency between the CLM5 PFT canopy height values and the ICESAT mapping in temperate forests with both around 20 meters, and in boreal forests with both around 15 meters. The Needleleaf Evergreen Tree and Needleleaf Deciduous Tree canopy height mapping values also show how canopy heights rapidly decrease north of the boreal forests with heights dropping below 10 meters into the tundra.

The average ICESAT and CLM5 PFT canopy heights are calculated globally and for the IPCC regions in Table 4.7. The global average vegetated ICESAT canopy height is 10.5 meters, with Southeast Asia having the tallest canopy height at 23.8 meters and the Middle East having the lowest at 2.1 meters. The global average CLM5 tree PFT canopy height varied widely. Broadleaf Evergreen Tropical and Temperate Trees having the highest canopies at around 25 meters. Broadleaf Deciduous Tropical Trees were much shorter at 12.5 meters, while Broadleaf Deciduous Temperate and Boreal Trees were around 20 meters. Needleleaf Evergreen Temperate Trees were marginally taller than Needleleaf Evergreen Boreal Trees at 18.3 meters and 15.9 meters respectively. Needleleaf Deciduous Boreal Trees were again shorter at 14.2 meters.

4.10 MODIS Albedo and Soil Color

CLM5 calculates the instantaneous soil surface reflective properties ($\alpha_{soi,\Lambda}$) using Equation 4.12 with new global soil color maps and the instantaneous soil moisture from the top soil layer (θ_1). The reflectance properties are calculated for the visible and near infrared spectrums (Λ), along with the soil color properties from Table 4.8. The soil reflectance combines with the new PFT mapping, instantaneous PFT LAI and SAI values, leaf and stem optical properties, snow reflective properties, and the downwelling solar radiation in the direct and diffuse beams for the visible and near infrared spectrums, to calculate the instantaneous surface albedo with associated radiation absorption for all soil and vegetated surface elements. The surface radiation calculation in CLM5 uses the two-stream radiation model as fully described in the CLM5 Technote (Lawrence et al. 2018).

$$\alpha_{soi,\Lambda} = min\left(\left(\alpha_{sat,\Lambda} + max(0.11 - 0.4\,\theta_1, 0.0)\right), \alpha_{dry,\Lambda}\right) \qquad (4.12)$$

The new CLM5 soil color mapping is generated using monthly climatology data from the MODIS MCD43A2 and MCD43A3 version 5 eight-day visible and near infrared albedo, snow cover and quality data for the period 2001 to 2015 using the methods of Lawrence and Chase (2008). The MODIS monthly albedo and snow data was generated from a global 8-day climatology in the same manner as the MODIS LAI data using the 15 years of the raw MCD43A2 and MCD43A3 data at the 1 km resolution of the source data. The 8-day climatology was generated using all high-quality data for that day of the year for the 15 years. Any grid cells that had no high-quality data for a particular day were temporally filled using an inverse time weighted interpolation from the previous and next climatology values for the grid cell. The 8-day climatology values were then

averaged for each month for the grid cell to produce a monthly climatology of albedo and snow at the 1 km resolution.

Table 4.8: CLM5 Soil Color classes are from 1 – 20 with reflective properties for dry and saturated soils in the visible and near infrared spectrums as listed in the CLM5 Technote (Lawrence et al. 2018) and Lawrence and Chase (2008).

Color	α sat vis	α dry vis	α sat nir	α dry nir	Color	α sat vis	α dry vis	α sat nir	α dry nir
1	0.25	0.36	0.50	0.61	11	0.13	0.24	0.26	0.37
2	0.23	0.34	0.46	0.57	12	0.12	0.23	0.24	0.35
3	0.21	0.32	0.42	0.53	13	0.11	0.22	0.22	0.33
4	0.20	0.31	0.40	0.51	14	0.10	0.20	0.20	0.31
5	0.19	0.30	0.38	0.49	15	0.09	0.18	0.18	0.29
6	0.18	0.29	0.36	0.48	16	0.08	0.16	0.16	0.27
7	0.17	0.28	0.34	0.45	17	0.07	0.14	0.14	0.25
8	0.16	0.27	0.32	0.43	18	0.06	0.12	0.12	0.23
9	0.15	0.26	0.30	0.41	19	0.05	0.10	0.10	0.21
10	0.14	0.25	0.28	0.39	20	0.04	0.08	0.08	0.16

The new CLM5 soil colors were calculated by fitting the visible and near infrared soil reflectance values for each month, for each grid cell at the 1 km resolution, to find values that would reproduce the same monthly snow free surface albedo in CLM5 as were found with the MODIS monthly data. The soil reflectance values were generated with climatological monthly soil moisture and the two-stream radiation model run offline with monthly downwelling solar fluxes taken from a CMIP5 Historical CLM 4.5 simulation interpolated to the 1 km resolution of the satellite data. The fitted monthly soil colors for each grid cell were then averaged to give the best annual average soil color for a grid cell to be used for all months.

The MODIS July climatology blacksky (direct beam) shortwave (combined visible and near infrared) albedo and snow cover maps are shown in Figures 4.17a and 4.17b. The maps show that the highest albedo values occurred over Antarctica and Greenland with values in these areas over 0.75. These areas also have high snow cover for all months including July. The remaining land area is almost completely snow free for July allowing for at least one month of the year to have a snow free albedo to calculate soil color. For the snow free land, albedo values are generally much higher for sparsely vegetated area than the more densely vegetated forests.

The Sahara and Arabian Peninsula both have the highest snow free albedo with values greater than 0.40 in some areas. Other sparsely vegetated areas including inland Australia, the Tibetan plateau, Western Asia, and the deserts of the Namib, Mongolia, and the American Southwest generally have higher albedo with values over 0.20. The more densely vegetated areas have much lower albedo. The tropical forests of the Amazon, Congo, and Southeast Asia have albedo values around 0.11, with the Boreal forests having even lower albedo with values down to 0.08. In the transition between deserts and forests, the grasslands, croplands, shrublands and savannas generally have intermediate albedo with values in the range 0.12 to 0.20.

The new global CLM5 soil color map is shown in Figure 4.17c. The maps show that for areas of continual snow cover, including Antarctica and Greenland, CLM5 prescribes the default soil color class of 10, as the actual soil reflectivity cannot be detected by MODIS. For the remaining areas that have snow free months the areas with the highest albedo values result in the lowest soil colors, while the areas with the lowest albedo values have the highest soil colors, in line with the soil reflective properties of Table 4.8.

(a) MODIS Albedo BSA Shortwave Climatology July   2001 - 2015
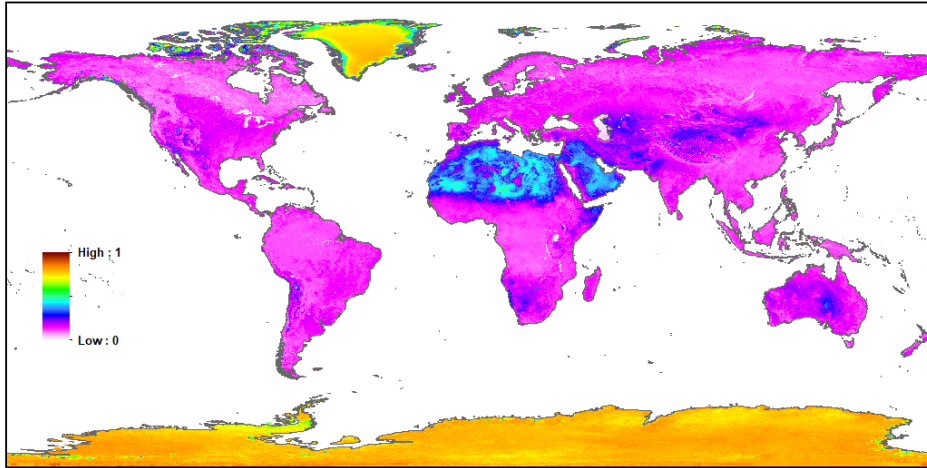
(b) MODIS Albedo Snow Cover Climatology July   2001 - 2015
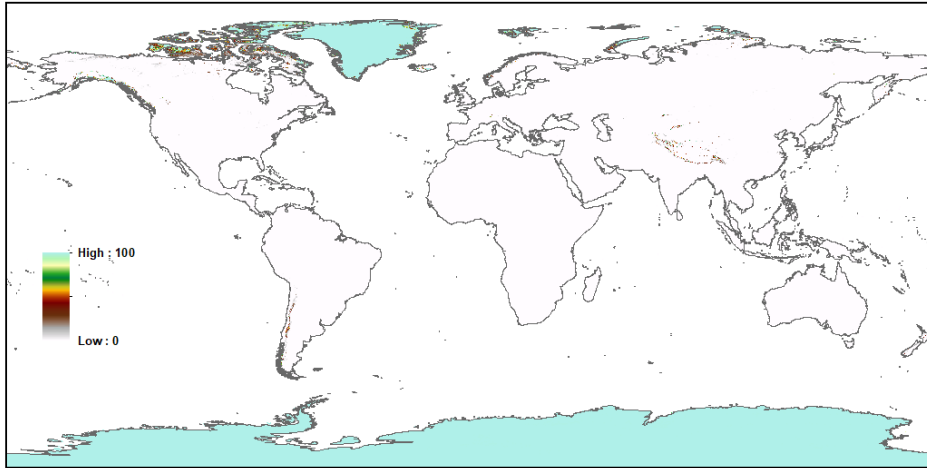
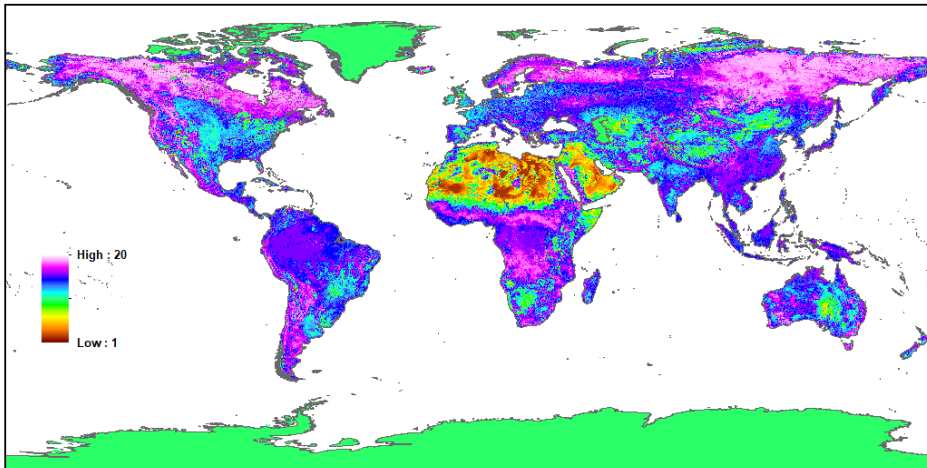**(c)** CLM5 MODIS Albedo Soil Color Class

Figure 4.17: Global Current Day CLM5 MODIS Albedo and Soil Color for: (a) July Blacksky Shortwave Albedo; (b) July Snow Cover; and (c) CLM5 Soil Color Class.

Table 4.9: MODIS CLM5 Global and IPCC Regional Average Shortwave Blacksky Monthly Climatology Albedo as Reflected Fraction, Snow as Percentage Cover, and CLM5 Soil Color. MODIS data is generated as the monthly climatology from 2001 – 2015 from 8 day 1 km MCD43B2 and MCD43B3 version 5 data.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Albedo Jan	0.30	0.22	0.16	0.26	0.42	0.22	0.13	0.27	0.36	0.12	0.18	0.79
Albedo Jul	0.22	0.22	0.16	0.16	0.14	0.15	0.13	0.27	0.19	0.12	0.16	0.78
Snow Jan	31.1	0.0	0.6	21.9	76.5	29.0	0.1	2.4	56.9	0.0	6.1	99.9
Snow Jul	10.4	0.0	0.1	0.4	0.5	1.1	0.2	0.0	12.5	0.0	0.8	99.8
Soil Color	14	12	14	14	16	14	15	10	15	15	14	10

For the very bright deserts of the Sahara and the Arabian Peninsula soil color as low as 1 are assigned. In other sparsely vegetated areas, the soil colors assigned are higher, ranging from 5 to 12, with lower reflective properties as shown in Table 4.8. The low albedo forested areas resulted in the highest soil colors ranging from 16 to 20, with the lowest associated soil reflectance properties. The remaining intermediately vegetated areas typically ranged in soil colors from 12 to 16. There are also cases where sparsely vegetated areas have low albedo values due to dark soils or underlying rocks. These areas include Western Australia and dark areas of the Sahara. In these areas CLM5 assigns higher soil colors to match the observed albedo.

4.11 Fire Model Population, GDP, Peatland, and Ag Burning Month

The CLM5 Fire Model has new representations of fire that include triggers and suppression of wildfire from natural and anthropogenic sources, as well as new representations for agricultural, deforestation, and peat fires (Li et al. 2012a, b, Li et al. 2013a). To support this functionality CLM5 prescribes global maps of annual population density, gross domestic production, peatland area fraction, and peak month of agricultural burning, to represent the human influence on fire from ignition and suppression (Figure 4.16 and Figure 4.17).

The global population density at 0.5x0.5 degrees resolution was developed annually for the 1850-2100 time period from decadal population density data for 1850–1980 from the Database of the Global Environment version 3.1 (HYDEv3.1) (Klein Goldewijk 2011) and population density data for 1990, 1995, 2000, and 2005 from the Gridded Population of the World version 3 dataset (GPWv3) (CIESIN, 2005). Additional CMIP6 Shared Socio-Economic Projection (SSP) global population maps for are used for the period 2015 – 2100. The global maps of population in 1850, 2014, and for 2100 under SSP3 are shown in Figure 4.16. The SSP3 population map represents the highest population growth and increase in density of the five CMIP6 baseline scenarios.

The 1850 global population map shows very low population density across most of the world with the exceptions the Ganges River plain, areas of China, and limited areas in Europe. The 2014 population map show the rapid historical increase in population density across India and China with localized urban increases elsewhere. The SSP3 2100 population scenario map shows the rapid projected increases in population density across India, Africa and Central America, but decreases in China and Europe. The change in average global population density as well as in the IPCC regions is shown in Table 4.10 for the historical period as well as all five CMIP6 baseline SSP scenarios.
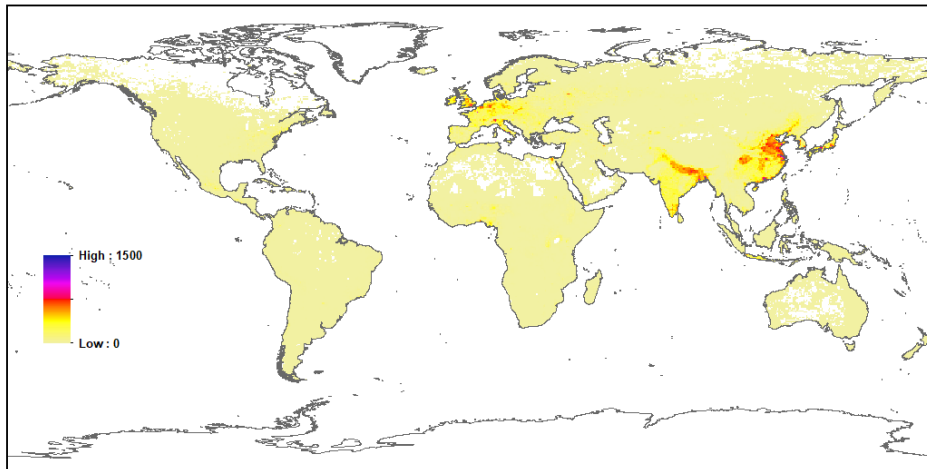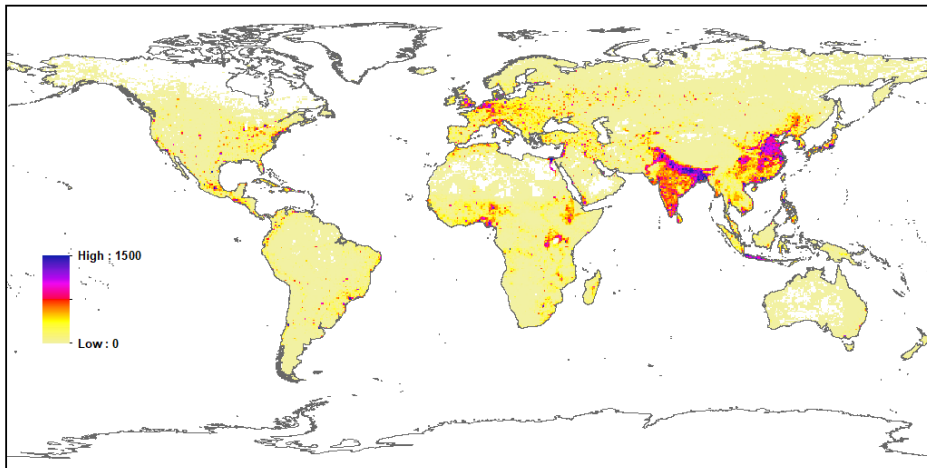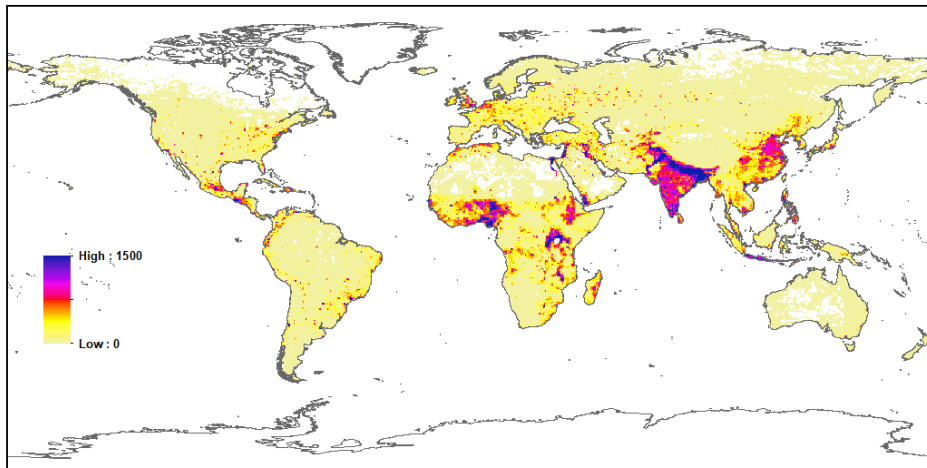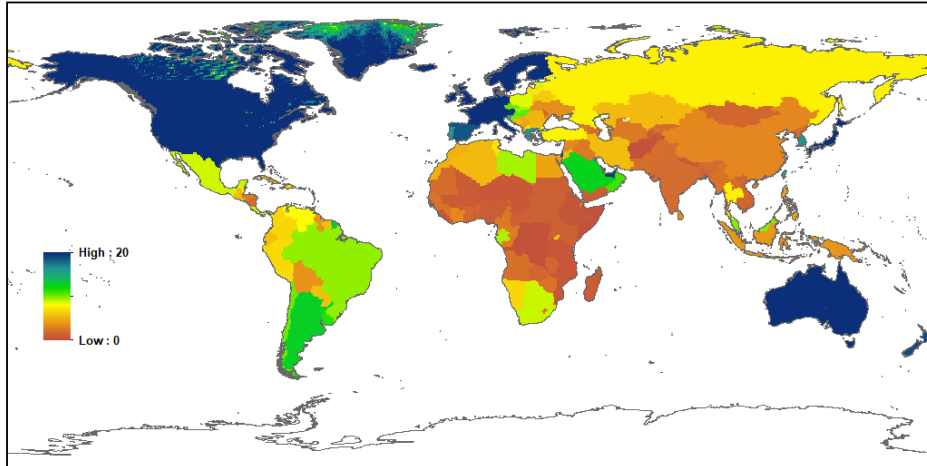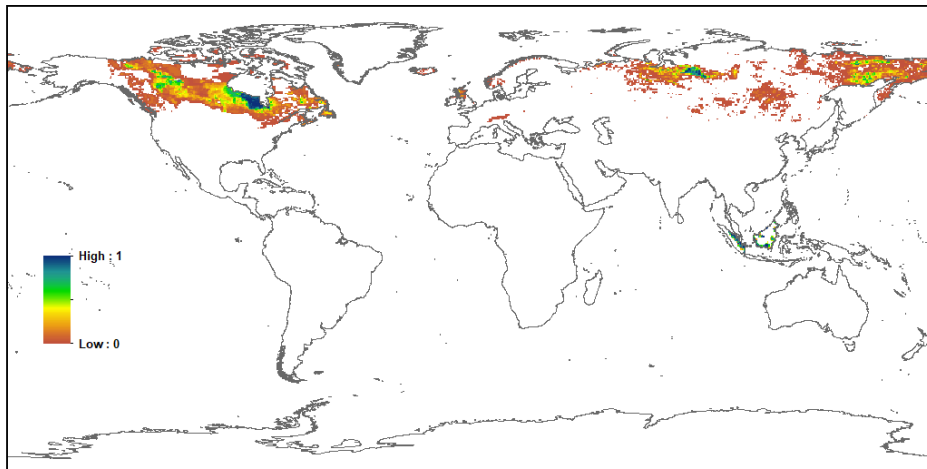
Figure 4.18: Global CLM5 Fire Model Population Data for: (a) 1850; (b) 2014; and (c) 2100 for the CMIP6 SSP3 Scenario.

(a) SRES GDP in US $1000s / Capita   2000

High : 20
Low : 0

(b) Peat Fraction from various sources

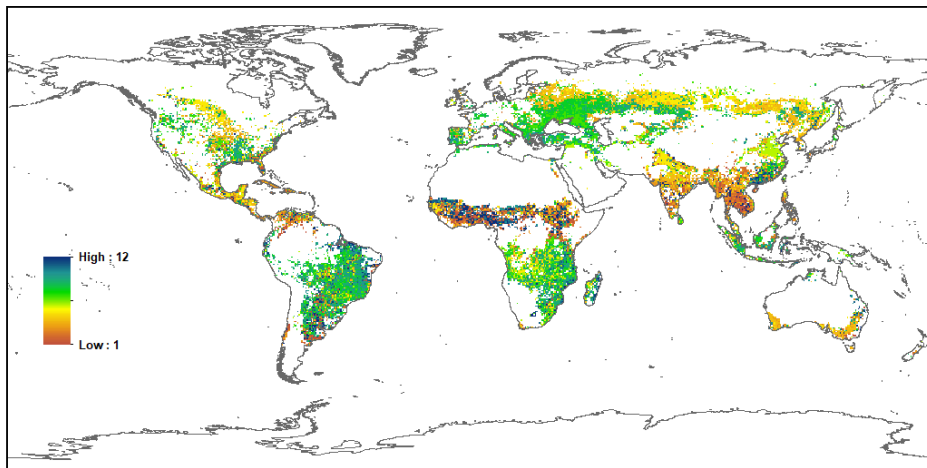

High : 1
Low : 0

(c) Agricultural Burning Month

High : 12
Low : 1

Figure 4.19: Global Current Day CLM5 Fire Model Surface Data for: (a) SRES 1995 Gross Domestic Product (GDP); (b) Peatland Fraction; and (c) Peak Month for Agricultural Burning.

Table 4.10: CLM5 Global and IPCC Regional Fire Model Properties. Average Population Density variables in People/km^2 are for 1850, 2014 and 2100 for all five CMIP6 SSPs. Other average variables are SRES Gross Domestic Product (GDP) in constant 1995 US $1000/Capita, Peat Area in millions of km^2, and Agriculture Burning Month.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Pop 1850	8.5	3.1	3.3	39.6	2.6	35.4	1.6	3.3	1.2	8.8	55.3	0.0
Pop 2014	48.3	37.2	16.3	129.9	11.5	102.2	29.5	44.0	16.9	117.1	336.1	0.1
Pop SSP1	47.2	65.8	13.7	62.8	7.7	98.1	23.5	58.5	24.6	92.8	311.8	0.1
Pop SSP2	60.8	88.9	13.6	74.2	10.3	100.0	32.5	80.1	24.3	121.4	445.8	0.1
Pop SSP3	83.9	128.4	8.0	97.4	13.8	72.0	52.3	124.0	13.9	174.3	728.8	0.0
Pop SSP4	62.2	119.3	10.7	54.1	7.6	80.8	27.5	111.2	19.2	114.8	409.1	0.0
Pop SSP5	49.8	64.0	20.0	63.7	7.8	120.2	22.0	61.6	37.8	90.7	305.4	0.1
GDP	7.5	1.0	24.2	0.9	2.6	17.5	4.3	4.5	26.2	1.4	0.4	0.0
Peat Area	1.99	0.00	0.00	0.00	0.48	0.01	0.00	0.00	1.12	0.38	0.00	0.00
Ag Burn Mon	6.6	6.9	5.2	6.0	6.3	7.4	7.3	7.4	6.5	5.3	5.0	0.0

The CLM5 Fire model Gross Domestic Production (GDP) per capita in 2000 is prescribed at 0.5x0.5 degrees resolution from van Vuuren et al. (2006). This data set is the base-year GDP data for IPCC-SRES and derived from country-level World Bank's World Development Indicators (WDI) measured in constant 1995 US dollars (World Bank, 2004) and the UN Statistics Database (UNSTAT, 2005). The global GDP map is shown in Figure 4.19a. The maps shows the high GDP per capita in the United States, Canada, Europe, Japan, Australia and New Zealand. Medium level GDP is shown in Brazil, Argentina, Mexico, South Africa, Libya, Saudi Arabia and Eastern Europe. Russia and other South American countries have low GDP, with remainder of Africa and Asia having very low GDP.

The average global and IPCC region GDP values are calculated in Table 4.10. These values show the same distribution of GDP as the global map. The table shows that North America has the highest GDP at $26.2k per capita, followed by the Asia Pacific Developed countries at $24.2k per capita, and Europe at $17.5k per capita. The lowest values of GDP were Southern Asia at $0.4k per capita, Eastern Asia at $0.9k per capita, and Africa at $1.0k per capita.

The CLM5 Fire model peatland area fraction was prescribed at 0.5x0.5 degrees resolution. The global peatland data was generated from data in Indonesia and Malaysian Borneo from Olson et al. (2001), data in Canada from Tarnocai et al. (2011), and bog, fen and mire data in boreal regions (north of 45N) outside Canada provided by the Global Lakes and Wetlands Database (GLWD) of Lehner and Döll (2004). The current day global peatland distribution is shown in Figure 4.19b. The maps shows the widespread areas of peatlands in North America, northern Russia and eastern Siberia. There are additional areas of peatlands in Indonesia on the islands of Sumatra and Borneo.

The global and IPCC region areas of peatland are calculated in Table 4.10. The global total area of peatland is just under 2.0 million km^2, with the regional totals showing more than half of that is in North America with 1.1 million km^2 of peatland. The other regions with substantial areas of peatland are Eurasia with 0.5 million km^2, and Southeast Asia with 0.4 million km^2.

The final surface data component of the CLM5 Fire model is the climatological peak month for agricultural waste burning. This is prescribed globally at 0.5x0.5 degrees resolution taken from the data of van der Werf et al. (2010). The global distribution of agricultural peak burning month is shown in Figure 4.19c. The map shows strong regional timing for agricultural burning with

distinct fire regimes. The global and IPCC region average agricultural peak burning months are calculated in Table 4.10. The subregional nature of the agricultural fire regimes are so mixed that these averages all fall in the middle of the year providing little guidance on the actual global or regional burning months.

From the global map the subregional agricultural fire regimes can be determined. In eastern South America and southern Africa the agricultural peak burning is in June through August corresponding with Austral winter. In Australia the agricultural peak burning is in April through May during the post-harvest season of Austral autumn. In northern areas of Africa, India and into Thailand the agricultural peak burning occurs December through February during the dry season. In the southern and western United States, southern Europe, Western Asia, southern Russia, and China the agricultural peak burning occurs from September through November in the post-harvest season of Boreal autumn. Further north in Canada and Russia peak burning occurs in March through April in Boreal spring. These distinct fire regimes are a key determinant of the agricultural fires in the CLM5 Fire model.

4.12 Volatile Organic Compounds

The MEGAN 2.1 Volatile Organic Compounds (VOCs) model implemented in CLM5 uses emission factors to calculate the production of speciated monoterpenes, sesquiterpenes, oxygenated VOCs as well as isoprene. A flexible scheme has been implemented in the CLM5 to specify a subset of these emissions. The CLM5 implementation allows for additional flexibility in grouping chemical compounds to form the lumped species frequently used in atmospheric chemistry.

To support the spatial and plant specific properties of VOC emissions, CLM5 has isoprene emission factors globally mapped for specific PFT groups at the 0.5x0.5 degrees resolution. These global maps are taken from the data of Guenther et al. (2012). The global maps of isoprene emission factors are shown for Broadleaf Tree, Fineleaf Evergreen Tree and Fineleaf Deciduous Tree PFT groups are shown in Figure 4.20, and for Shrub, Grass and Crop PFT groups in Figure 4.21.

For the Broadleaf Trees the highest emission factors are found in Europe into Russia, northern China, Australia and northern areas of North America. For Fineleaf Evergreen Trees however only the northern areas of North America are elevated. For Shrubs Australia is elevated as well as a small area through equatorial Africa, and the Canadian Prairies into northern Alberta and British Columbia. While the global maps show spatial variation in isoprene emission factors within each PFT group, the differences between PFT groups are three orders of magnitude larger.

Table 4.11: CLM5 Global and IPCC Regional Volatile Organic Compounds (VOCs) Properties through Emission Factors for Broadleaf Tree, Fineleaf Evergreen Tree, Fineleaf Deciduous Tree, Shrub, Grass and Crop PFT groups in μg isoprene/m^2/hr.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Broadleaf EF	10071	7234	14704	10204	12591	18644	11386	2477	13627	11634	8488	0
Fine Evg EF	1804	1943	2837	1435	1224	912	2595	1547	2082	2768	3087	1
Fine Dec EF	511	811	593	291	112	174	1254	51	48	1526	883	0
Shrub EF	8992	9297	2188	9775	7482	7549	9409	8631	9131	1113	1081	6
Grass EF	463	331	184	581	855	542	413	329	684	282	439	0
Crop EF	52	31	34	99	21	12	81	13	13	370	189	0

(a) Broadleaf Tree Emission Factor (µg isoprene m-2 h-1)
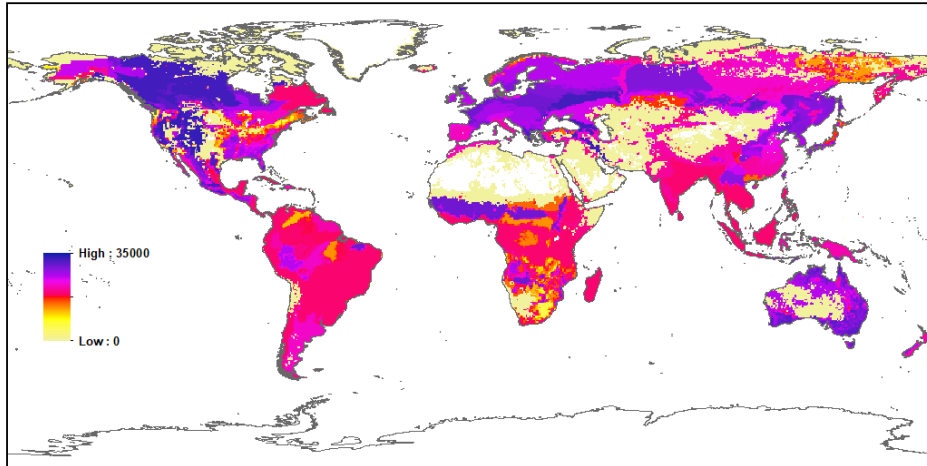
(b) Fineleaf Evergreen Tree Emission Factor (µg isoprene m-2 h-1)
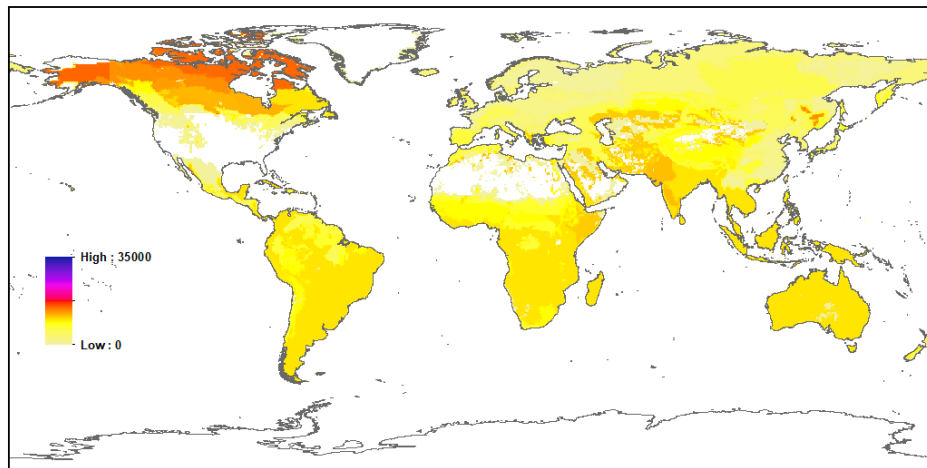
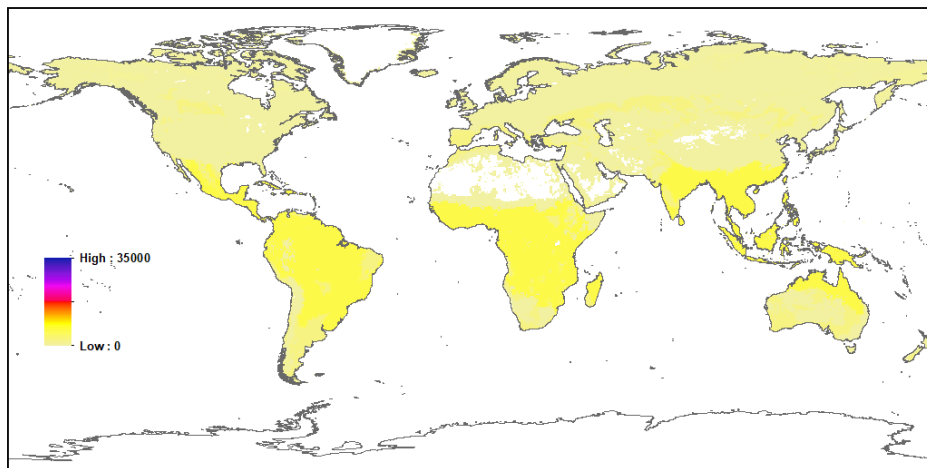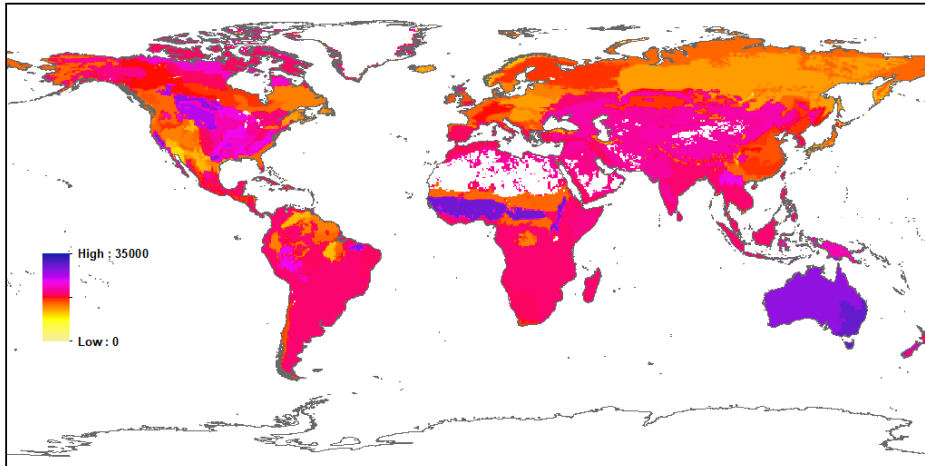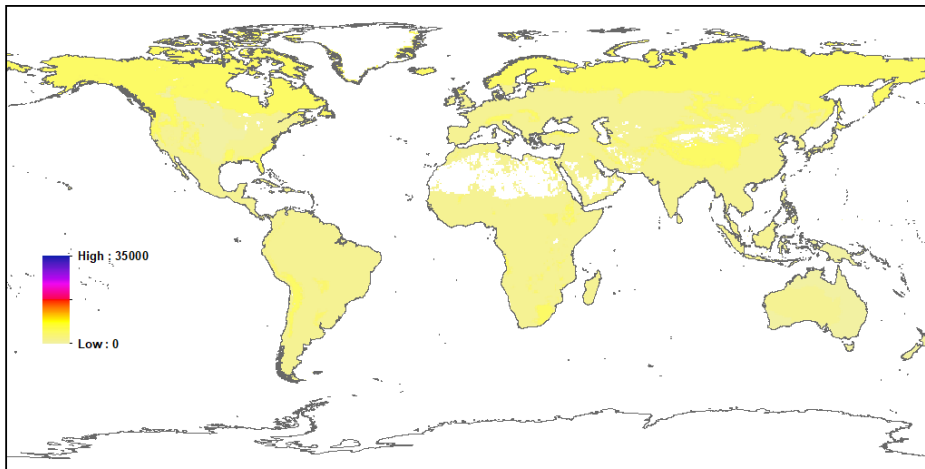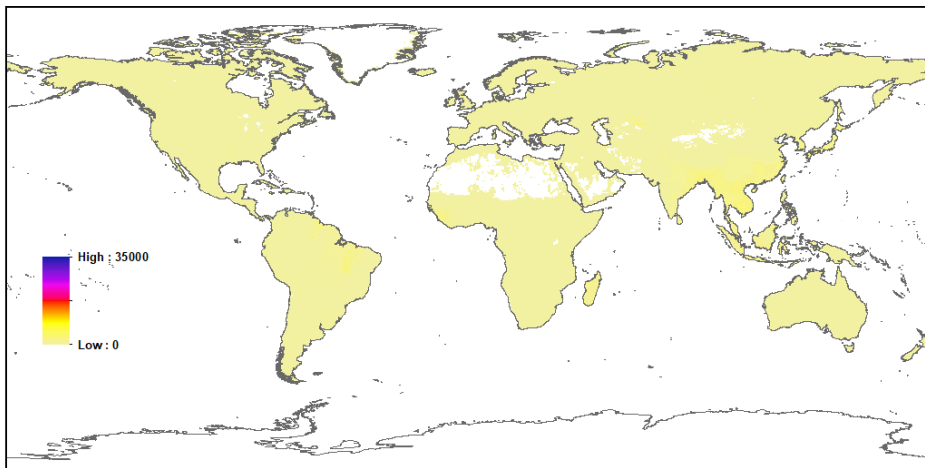(c) Fineleaf Deciduous Tree Emission Factor (µg isoprene m-2 h-1)

Figure 4.20: Global Current Day CLM5 Volatile Organic Compound Model PFT Group Isoprene Emission Data for: (a) Broadleaf Trees; (b) Fineleaf Evergreen Trees; and (c) Fineleaf Deciduous Trees.

(a) Shrub Emission Factor (μg isoprene m-2 h-1)

(b) Grass Emission Factor (μg isoprene m-2 h-1)

(c) Crop Emission Factor (μg isoprene m-2 h-1)

Figure 4.21: Global Current Day CLM5 Volatile Organic Compound Model PFT Group Isoprene Emission Data for: (a) Shrubs; (b) Grasses; and (c) Crops.

The global and IPCC regional analyses of average isoprene emission factors by PFT group are calculated in Table 4.11. The average global Broadleaf Tree emission factor of 10071 µg isoprene/m^2/hr is five times higher than the Fineleaf Evergreen Tree emission factor of 1804, it is 20 times the Fineleaf Deciduous Tree emission factor of 511, and it is 200 times the Crop emission factor of 52. Shrubs also have a high average global emission factor at 8992 which is nearly 20 times the emission factor of Grass at 463.

The spatial variation shown in the global mapping of emission factors for each of the PFT groups is also reflected in the regional emission factor analysis. The Broadleaf Tree emission factors range from 18644 in Europe down to 2477 in the Middle East. The Fineleaf Evergree Tree emission factors range from 3087 in Southern Asia down to 912 in Europe. The Fineleaf Deciduous Tree emission factors range from 1526 in Southeast Asia down to 48 in North America. For the non-Tree PFTs, Shrubs range from 9775 in East Asia down to 1081 in Southern Asia, while Grass range from 855 in Eurasia down to 184 in Asia Pacific Developed countries, and Crop range from 370 in Southeast Asia down to 13 in North America.

4.13 Methane

The representation of methane production in CLM5 follows previous versions of the CLM Methane model as documented in Riley et al. (2011). The Methane model represents the CH4 mass balance, CH4 production, ebullition, aerenchyma transport, CH4 Oxidation, reactive transport solution, including boundary conditions, numerical solution, water table interface, seasonal inundation effects, and impact of seasonal inundation on CH4 production. One of the key determinants of CH4 production is the inundation fraction of the CLM5 land units.

The Methane model uses a simplified dynamic representation of spatial inundation based on Prigent et al. (2007). This method uses a multi-satellite approach to estimate the global monthly inundated fraction (Fi) on a 0.25x0.25 degrees resolution grid from 1993 - 2000. A simple scaling factor is used to mimic the impact of seasonal inundation on CH4 production based on surface runoff. The three inundation factors used by the CLM5 Methane model are represented globally through maps of the Maximum Inundation Factor (F0), the seasonal Surface Runoff Factor (P3), and the Water Table Depth Factor (ZWT0).

Table 4.12: CLM5 Global and IPCC Regional Methane Model Properties for Maximum Inundation Fraction (F0), Surface Runoff Factor (1000s) (P3), and Water Table Factor (ZWT0).

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Max Inun - F0	0.11	0.11	0.17	0.12	0.08	0.15	0.14	0.12	0.14	0.11	0.16	0.01
Surf Run - P3	98.1	91.2	223.2	115.4	17.9	46.9	58.5	967.3	23.5	3.5	227.0	0.0
Water - ZWT0	12.4	8.7	11.5	19.4	7.1	14.7	15.7	12.0	12.5	28.4	41.0	0.0

The three inundation factors are mapped globally from the methods of Prigent et al. (2007) in Figure 4.20. The Maximum Inundation Factor (F0) map shows there are four main areas with high inundation factors in the northern areas of North America, India, China, and northern Russia. The Surface Runoff Factor (P3) is highly localized with only a few areas globally having higher values. The Water Table Depth Factor (ZWT0) has a more widespread representation with localized very high values in flood plains of major river systems. The major areas with high values are India, China, and northern North America.

(a) Maximum Inundation Fraction (F0)
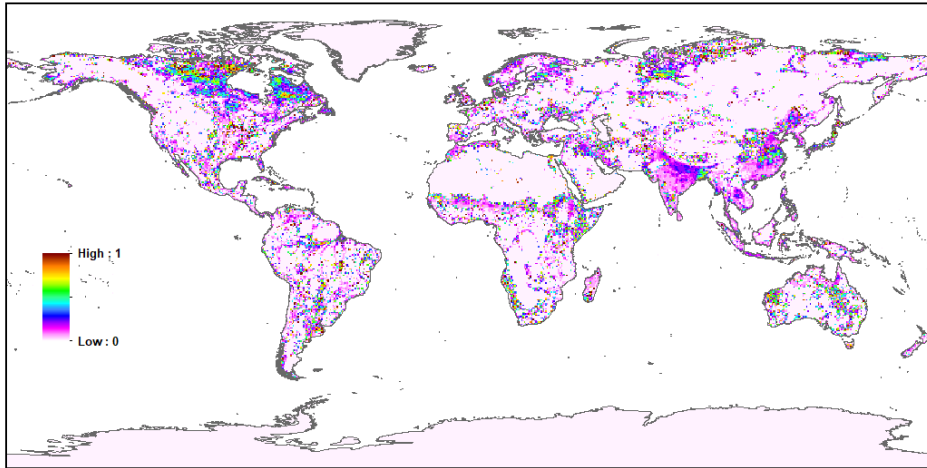
(b) Surface Runoff Factor (P3) for Inundation Calculation
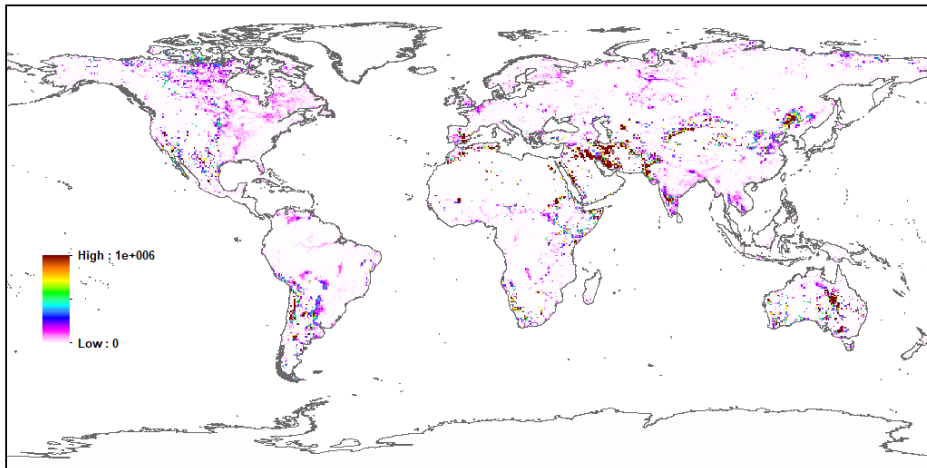
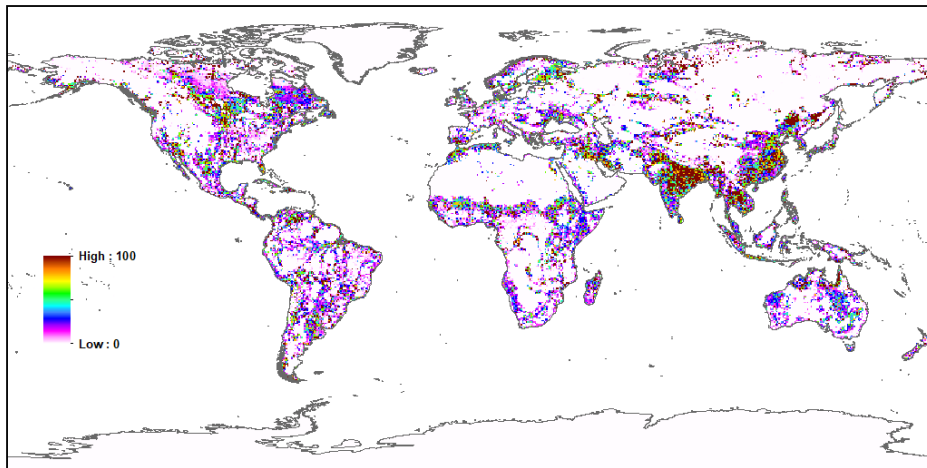(c) Water Table Depth Factor (ZWT0) for Inundation Calculation

Figure 4.22: Global Current Day CLM5 Methane Model Surface Data for: (a) Maximum Inundation Fraction (F0); (b) Surface Runoff Factor (P3); and (c) Water Table Depth Factor (ZWT0).

The global and IPCC regional analyses for the average inundation factors are calculated in Table 4.12. The Maximum Inundation Fraction (F0) has an average global value of 0.11. This ranges from 0.17 in the Asia Pacific Developed countries to 0.08 in Eurasia. The Surface Runoff Factor (P3) is a scalar that is multiplied by the seasonal runoff in mm/second. The higher the surface runoff factor the more that the seasonal runoff contributes to the inundation calculation in the model. The highest P3 values are in the Middle East at 967.3k and lowest values are in Southeast Asia at 3.5k.

The Water Table Depth Factor (ZWT0) mediates the influence of the model calculated water table depth on inundation fraction, following the model description in Riley et al. (2011). The larger the ZWT0 value the less sensitive the inundation fraction is to water table depth. Southern Asia has the largest ZWT0 value at 41.0 with Eurasia having the lowest value at 7.1.

4.14 IPCC Assessment Report 6 Global Regions

While not an input variable to CLM5, the IPCC Assessment Report 6, 10 global regions have been used as a consistent regional analysis framework throughout this technote. The 10 global regions have been compiled from the country listings provided in the report. The 10 global regions are shown in Figure 4.23, with the region abbreviations and long names shown in Table 4.13.
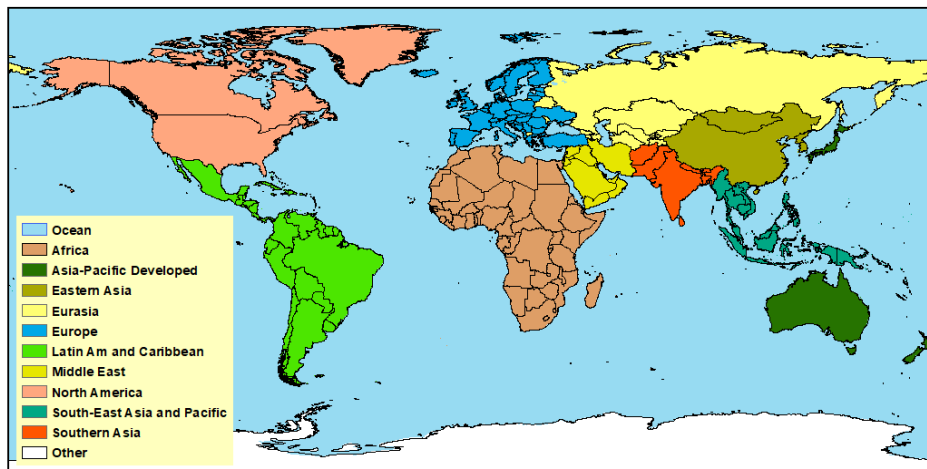
Figure 4.23: Global IPCC 10 Regions from Assessment Report 6 for regional analysis.

Table 4.13: IPCC Assessment Report 6, 10 Region Definitions Abbreviations and Long Names.

1. AFR	2. APD	3. EAS	4. ERA	5. EUR	6. LAC	7. MEA	8. NAM	9. SEA	10. SAS
Africa	Asia Pacific Develop.	Eastern Asia	Eurasia	Europe	Latin America & Carib.	Middle East	North America	South East Asia	Southern Asia

CHAPTER 5.

CMIP6 – LUMIP – LUH2 LAND USE

5.1 CMIP6 – LUMIP Overview

The Coupled Model Intercomparison Project (CMIP), under the auspices of the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM), has become one of the foundational elements of climate science, coordinating the design and dissemination of global climate model simulations for the past, current and future climate periods (Eyring et al. 2016). The CMIP Phase 6 (CMIP6) project includes 34 climate models from 27 climate and Earth system modeling institutions that have performed both the Historical (DECK) simulations from 1850 to 2014, and the future Shared Socioeconomic Pathways (SSPs) simulations from 2015 to 2100 (Tebaldi et al. 2021). NCAR has supplied two climate models for these simulations through the Community Earth System Model (CESM2), with the default (CAM6) and the Whole Atmosphere (WACCM6) configurations.

The CMIP6 climate and Earth system simulations are coordinated across modeling centers through shared consistent prescriptions of climate forcing conditions for each simulation period. The time series data includes prescription of greenhouse gases and aerosols including CO_2, CH_4, N_2O, Sulfur and Black Carbon (Eyring et al. 2016), as well as land use and land cover change (LULCC) from the Land Use Model Intercomparison Project (LUMIP) (Lawrence et al. 2017). In both the CAM6 and WACCM6 configurations, the NCAR CESM2 land surface is represented by CLM5 with the LULCC prescribed from time series data as detailed in the next four chapters.

The CMIP6 LUMIP project is responsible for coordinating the shared land use time series to be used by all modeling institutions for the simulations of the Historical (DECK), the Integrated Assessment Model (IAM) generated SSPs of the Scenario Model Intercomparison Project (ScenarioMIP) (O'Neill et al. 2016), and the Transient (mid) Holocene simulations of the Paleoclimate Modeling Intercomparison Project (PMIP4) (Kageyama et al. 2018). To support the shared development and distribution of the Paleo, Historical and SSP time series, LUMIP through the Land Use Harmonization 2 (LUH2) project (Hurtt et al. 2020), has generated continuous harmonized annual global time series data at 0.25 degrees for historical reconstructions for 850 to 2015, and projected land use from ScenarioMIP Integrated Assessment Models (IAMs) for five Baseline scenarios and three Mitigation scenarios for 2015 to 2100.

This chapter provides a detailed description of the CMIP6 – LUMIP Historical LULCC time series from 850 to 2014, developed as part of the LUH2 project as detailed in Lawrence et al. (2017) and Hurtt et al. (2020). The chapter also provides a high-level overview of all of the CMIP6 LUH2 time series representations for the PMIP4 Transient Holocene from 6000bce to 849ce, the Last Millennium from 850 to 1849, the DECK Historical period for 1850 to 2014, and the ScenarioMIP Baseline SSP scenarios for 2015 to 2100. The CLM5 representation of land use and land cover change only contains a subset of the full LUH2 time series data, so only the CLM5 relevant data

is included in this Technote. For a full description of the unrepresented components of the LUH2 time series see Hurtt et al. (2020).

Chapter 8 provides a detailed description of the CLM5 Land Data Tool, including data preparation for the tool using data described in previous chapters, and the CLM5 Mksurf surface data generation tool required to build the final input files for CLM5. Chapter 6 provides a detailed description of the CLM5 Historical LULCC time series that results from the LUH2 Historical time series data combined in the CLM5 Land Data Tool. Chapter 7 provides a higher-level analysis of all of the CMIP6 CLM5 LULCC time series data from 6000 BCE through to the current day with five baseline SSP scenarios to 2100.

Table 5.1 LUH2 Annual Land Use States.

Variable Name	Full Name	GLM2 Historical Data Sources
primf	Primary Forest	HYDE 3.2, combined with Miami LU biomass, with Landsat changes for 2000 – 2012
primn	Primary Non Forest	HYDE 3.2 combined with Miami LU biomass, with Landsat changes for 2000 – 2012
secdf	Secondary Forest	HYDE 3.2 combined with Miami LU biomass, with Landsat changes for 2000 – 2012
secdn	Secondary Non Forest	HYDE 3.2 combined with Miami LU biomass, with Landsat changes for 2000 – 2012
c3ann	Crop C3 Annual	HYDE 3.2 combined with Monfreida for 2000 and rotation rates from Sahajpal for 2011 – 2014
c4ann	Crop C4 Annual	HYDE 3.2 combined with Monfreida for 2000 and rotation rates from Sahajpal for 2011 – 2014
c3per	Crop C3 Perennial	HYDE 3.2 combined with Monfreida for 2000 and rotation rates from Sahajpal for 2011 – 2014
c4per	Crop C4 Perennial	HYDE 3.2 combined with Monfreida for 2000 and rotation rates from Sahajpal for 2011 – 2014
c3nfx	Crop C3 Nitrogen Fixing	HYDE 3.2 combined with Monfreida for 2000 and rotation rates from Sahajpal for 2011 – 2014
pastr	Pasture	HYDE 3.2 directly
range	Rangeland	HYDE 3.2 directly
urban	Urban	HYDE 3.2 directly

5.2 The Land Use Harmonization 2 (LUH2) Project

The Land Use Harmonization 2 (LUH2) project was developed to generate a new harmonized set of land use scenarios that smoothly connect historical reconstructions of land use with eight future projections in the format required for ESMs. The data produced estimates the fractional land use patterns, underlying annual land use transitions, and key agricultural, forestry and biofuel management for the time period 850 – 2100 at the 0.25^O x 0.25^O resolution. The LUH2 project follows the LUH1 project that provided a considerably smaller set of data for 1500 – 2100 at the

0.5 x 0.5 resolution for the ESM simulations of CMIP5 (Hurtt et al. 2011). The LUH1 data set formed the basis of land use and land cover change in CLM4 and CLM4.5 (Lawrence et al. 2012, 2018).

The LUH2 historical reconstruction from 850 to 2015 is predominantly based on the History of the Global Environment database (HYDE 3.2), with land use derived from historical population densities as detailed in Klein Goldewijk et al. (2017). For the modern period the mapping is steered towards satellite information using the land cover data of the UNFAO agricultural land use (FAO 2020a), the FAO Global Forest Resources Assessment for 2000 (FAO 2000), and Landsat forest change detection from Hansen et al. (2013), as described in Hurtt et al. (2020). Going further back into the past, the mapping is steered towards spatially explicit maps of climate, soil, slope, and proximity to rivers and lakes. Agricultural management including nitrogen fertilizer and irrigation, as well as shifting cultivation and wood harvest, are compiled from multiple sources for different components of the historical time period as detailed later in this chapter.

Table 5.2 LUH2 Annual Land Use States and Transitions Matrix. Three possible cases: Excluded (Excl), Wood Harvest (Wood Harv) changes State, and Transition (Trans) changes State.

	primf	primn	secdf	secdn	c3ann	c4ann	c3per	c4per	c3nfx	pastr	range	urban
primf	Excl	Excl	Wood Harv	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans
primn	Excl	Excl	Trans	Wood Harv	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans
secdf	Excl	Excl	Wood Harv	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans
secdn	Excl	Excl	Trans	Wood Harv	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans
c3ann	Excl	Excl	Trans	Trans	Excl	Trans	Trans	Trans	Trans	Trans	Trans	Trans
c4ann	Excl	Excl	Trans	Trans	Trans	Excl	Trans	Trans	Trans	Trans	Trans	Trans
c3per	Excl	Excl	Trans	Trans	Trans	Trans	Excl	Trans	Trans	Trans	Trans	Trans
c4per	Excl	Excl	Trans	Trans	Trans	Trans	Trans	Excl	Trans	Trans	Trans	Trans
c3nfx	Excl	Excl	Trans	Trans	Trans	Trans	Trans	Trans	Excl	Trans	Trans	Trans
pastr	Excl	Excl	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Excl	Trans	Trans
range	Excl	Excl	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Excl	Trans
urban	Excl	Excl	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Trans	Excl

The LUH2 timeseries data is generated using the Global Land Use Model 2 (GLM2) which is a development from the GLM1 model used with LUH1. The GLM2 model simulates annual sub grid land use states for 12 land use types shown in Table 5.1, along with a matrix of land use transitions from one year to the next shown in Table 5.2. A range of annual grid level forestry, agricultural

and biofuel management actions are prescribed for the land use states, with the CLM5 relevant management actions shown in Table 5.3. GLM2 solves the land surface state and use for every grid cell at each annual time step constrained by a range of inputs including land use maps, crop type and rotation rates, shifting cultivation rates, agricultural management, wood harvest, forest transitions, and potential biomass and biomass recovery rates (Hurtt et al. 2020).

Table 5.3 LUH2 Annual Land Use Management Actions used for CLM5 data.

Variable Name	Full Name	Data Notes
irrig_c3ann	Crop C3 Annual Irrigation	Irrigation as a fraction of C3 Annual Crop area
irrig_c4ann	Crop C4 Annual Irrigation	Irrigation as a fraction of C4 Annual Crop area
irrig_c3per	Crop C3 Perennial Irrigation	Irrigation as a fraction of C3 Perennial Crop
irrig_c4per	Crop C4 Perennial Irrigation	Irrigation as a fraction of C4 Perennial Crop
irrig_c3nfx	Crop C3 Nitrogen Fixing Irrigation	Irrigation as a fraction of C3 N Fixing Crop
fertl_c3ann	Crop C3 Annual Fertilizer	Grid cell average nitrogen fertilizer for C3 Annual Crops in kgN/ha
fertl_c4ann	Crop C4 Annual Fertilizer	Grid cell average nitrogen fertilizer for C4 Annual Crops in kgN/ha
fertl_c3per	Crop C3 Perennial Fertilizer	Grid cell average nitrogen fertilizer for C3 Perennial Crops in kgN/ha
fertl_c4per	Crop C4 Perennial Fertilizer	Grid cell average nitrogen fertilizer for C4 Perennial Crops in kgN/ha
fertl_c3nfx	Crop C3 Nitrogen Fixing Fertilizer	Grid cell average nitrogen fertilizer for C3 Nitrogen Fixing Crops in kgN/ha
primf_harv	Primary Forest Wood Harvest	Both harvest area as fraction of grid cell and biomass in kgC per grid cell.
primn_harv	Primary Non-Forest Harvest	Both harvest area as fraction of grid cell and biomass in kgC per grid cell.
secmf_harv	Secondary Mature Forest Harv	Both harvest area as fraction of grid cell and biomass in kgC per grid cell.
secyf_harv	Secondary Young Forest Harv	Both harvest area as fraction of grid cell and biomass in kgC per grid cell.
secnf_harv	Secondary Non-Forest Harvest	Both harvest area as fraction of grid cell and biomass in kgC per grid cell.

Given the extremely large amount of data produced in LUH2 only the historical CLM5 relevant land use states, shifting cultivation, agricultural management, and wood harvest are investigated in detail for this technote. The remaining key LUH2 timeseries data are described at the end of the chapter at a higher level providing a contextual investigation of changes in land surface state and management from the mid Holocene (6000bce to 850ce), through the last millennium (850 –

1850), into the historical period (1850 – 2014) and over the five future baseline SSPs (2015 – 2100). The LUH2 descriptions are provided to show the land use timeseries data prior to the generation of CLM5 land use and land cover data described in Chapters 6, 7 and 8.

5.3 Historical Land Use States Timeseries

The LUH2 historical land use states for 2014 and 850 are calculated globally and for the 10 IPCC regions for each land use state, as well as for the grouped land use states of forests, non-forests, grazing and crops in Table 5.4. The current day 2014 distributions of grouped land use states of forests, non-forests and crops are mapped in Figure 5.1, with the individual current day land use distributions mapped in Figures 5.4, 5.5, 5.6 and 5.7. The individual land use states are plotted for 2014 and 850 in Figure 5.2 with differences between 2014 and 850, and the annual timeseries plotted in Figure 5.3.

The LUH2 global area of all forests in 2014 is 36.8 million km^2, which is comprised of 22.1 million km^2 of Primary Forest and 14.7 million km^2 of Secondary Forest. This compares to the 27.7 million km^2 of forests found in the MODIS land cover mapping for the same period as reported in Table 2.2, and the 40.9 million km^2 reported in the FAO Global Forest Resources Assessment for 2000 (FAO 2000). The largest areas of LUH2 forests were found in Eurasia at 8.9 million km^2, North America at 8.5 million km^2, and Latin America at 8.0 million km^2, with smaller areas of forests in Africa and Southeast Asia. The largest areas of 2014 Primary Forest were in Latin America at 6.8 million km^2, and North America at 5.8 million km^2.

The LUH2 global area of forests in 850 was 47.3 million km^2, which made up a corresponding loss of 10.5 million km^2 over the period to 2014. The LUH2 850 forest area was almost completely Primary Forest with a small area of Secondary Forest at 0.4 million km^2 associated with shifting cultivation at the beginning of the period. The biggest losses in LUH2 forest area were in Latin America at -3.2 million km^2, East Asia at -2.1 million km^2, North America at -1.6 million km^2, Africa at -1.1 million km^2, and Southeast Asia at -1.0 million km^2. The timeseries plot of Figure 5.3b shows there was little forest loss until 1800 with a slow transition from Primary Forest to Secondary Forest from 850 to 1850. After 1800 however, the transition from Primary Forest to Secondary Forest accelerates with an increasing loss of total forest associated with the increase in other land use classes.

The current day 2014 LUH2 map of total forests in Figure 5.1a, shows five main forest areas globally. These correspond with the Amazon, the Congo, Southeast Asia, Boreal North America, and Boreal Eurasia. The current day 2014 LUH2 Primary Forest and Secondary Forest mapping shown in Figures 5.4a and b, identifies distinct areas for both classes of forest. The Primary Forest is found within the central Amazon, Congo, Borneo into Papua New Guinea, and high latitudes of the Boreal Forest in North America and Eurasia. The Secondary Forest by contrast are found in the Temperate Forests of North America and China, as well as the lower latitudes of the Boreal Forest through Scandinavia and Eurasia.

The biggest differences in LUH2 forest area compared to the MODIS land cover mapping were in North America and Eurasia where LUH2 had 4.1 million km^2 and 3.3 million km^2 more current day forest respectively than MODIS. These differences primarily represent the LUH2 mapping of forests into high latitude tundra shrub and grasslands in the MODIS mapping. Other areas with higher LUH2 forest area were in Africa and East Asia, which in LUH2 were 0.9 million km^2 and 0.7 million km^2 higher respectively. These differences represent LUH2 mapping of forests into areas mapped as woody savannas in MODIS. Both of these types of differences in forest area

mapping reflect differences in the definitions of forest tree cover, as well as differences in the methodological approaches of current day remote sensing of forest cover in MODIS, compared to the historical modeling of forest areas with GLM2 in LUH2 constrained by the land cover data of the UNFAO agricultural data base (FAO 2020a), the FAO Global Forest Resources Assessment for 2000 (FAO 2000), climate, soils, slope and human land use.

Table 5.4 Annual Global and Regional LUH2 Land Use States Area for 850 and 2014 in Millions of km^2.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
forest	36.8	3.4	0.6	2.4	8.9	1.5	8.0	0.0	8.5	3.1	0.6	0.0
non-forest	42.8	14.6	3.7	2.3	6.4	1.9	4.5	2.8	4.7	0.4	1.8	0.0
grazing	32.6	8.9	3.4	5.1	3.6	0.9	5.6	2.0	2.7	0.2	0.5	0.0
crop	15.6	2.8	0.5	1.3	1.6	1.8	2.0	0.3	2.0	1.1	2.1	0.0
primf	22.1	2.5	0.2	0.8	3.8	0.2	6.8	0.0	5.8	2.0	0.1	0.0
primn	27.6	6.3	3.2	1.8	6.0	0.8	3.7	0.3	4.2	0.3	1.0	0.0
secdf	14.7	0.8	0.4	1.7	5.1	1.3	1.2	0.0	2.7	1.1	0.5	0.0
secdn	15.3	8.3	0.5	0.5	0.5	1.0	0.8	2.5	0.4	0.1	0.8	0.0
c3ann	8.6	0.9	0.5	0.8	1.5	1.3	0.6	0.2	1.1	0.5	1.3	0.0
c4ann	3.0	1.1	0.0	0.2	0.1	0.2	0.5	0.0	0.4	0.1	0.3	0.0
c3per	1.4	0.4	0.0	0.1	0.0	0.1	0.3	0.0	0.0	0.4	0.1	0.0
c4per	0.3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
c3nfx	2.4	0.4	0.0	0.2	0.0	0.1	0.5	0.0	0.6	0.1	0.4	0.0
pastr	7.9	3.0	0.1	1.1	0.6	0.7	1.5	0.0	0.6	0.1	0.2	0.0
range	24.8	5.8	3.3	4.0	3.0	0.2	4.0	2.0	2.1	0.0	0.3	0.0
urban	0.5	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0
Year 850												
forest	47.3	4.4	0.8	4.5	9.6	1.9	11.2	0.0	10.1	4.0	0.9	0.0
non-forest	76.1	23.4	7.4	5.9	10.5	3.5	8.8	4.8	7.9	0.6	3.4	0.0
grazing	3.3	1.7	0.0	0.5	0.4	0.4	0.0	0.2	0.0	0.0	0.1	0.0
crop	1.7	0.2	0.0	0.2	0.1	0.4	0.1	0.1	0.0	0.0	0.6	0.0
primf	46.9	4.3	0.8	4.5	9.6	1.8	11.1	0.0	10.1	4.0	0.8	0.0
primn	75.4	23.0	7.4	5.9	10.5	3.5	8.8	4.7	7.9	0.6	3.3	0.0
secdf	0.4	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.0
secdn	0.7	0.4	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.0
c3ann	1.0	0.1	0.0	0.1	0.0	0.3	0.0	0.1	0.0	0.0	0.4	0.0
c4ann	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
c3per	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
c4per	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
c3nfx	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
pastr	0.4	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
range	2.9	1.6	0.0	0.5	0.4	0.2	0.0	0.2	0.0	0.0	0.1	0.0
urban	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

The LUH2 current day global area of all non-forest (not including crops, pastures, rangelands or urban) in 2014 is 42.8 million km^2. The current day non-forest area comprises of 27.6 million km^2 of Primary Non Forest and 15.3 million km^2 of Secondary Non Forest. The largest areas of non-forest were in Africa with 14.6 million km^2, Eurasia with 6.4 million km^2, and North America with 4.7 million km^2. Other areas with large amounts of non-forest were Latin America, Asia Pacific and the Middle East. The composition of Primary Non Forest was less than half of the total non-forest area for Africa compared with up to 95% for Eurasia, North America and other regions.

(a) LUH2 All Forest (Primary + Secondary) 2014   (%)
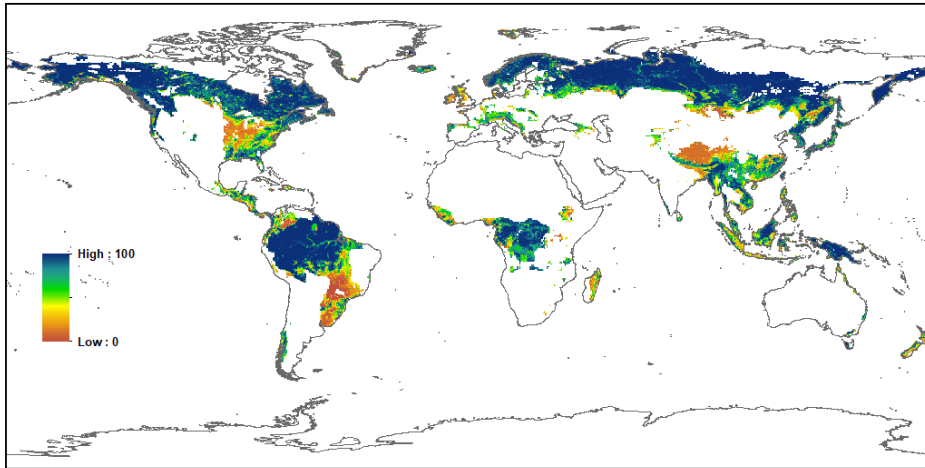
(b) LUH2 All Non Forest (Primary + Secondary) 2014   (%)
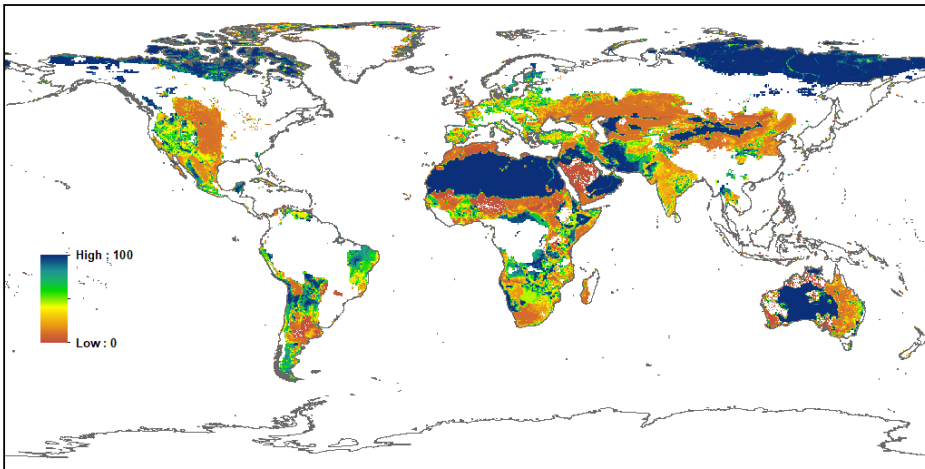
(c) LUH2 All Crop 2014   (%)
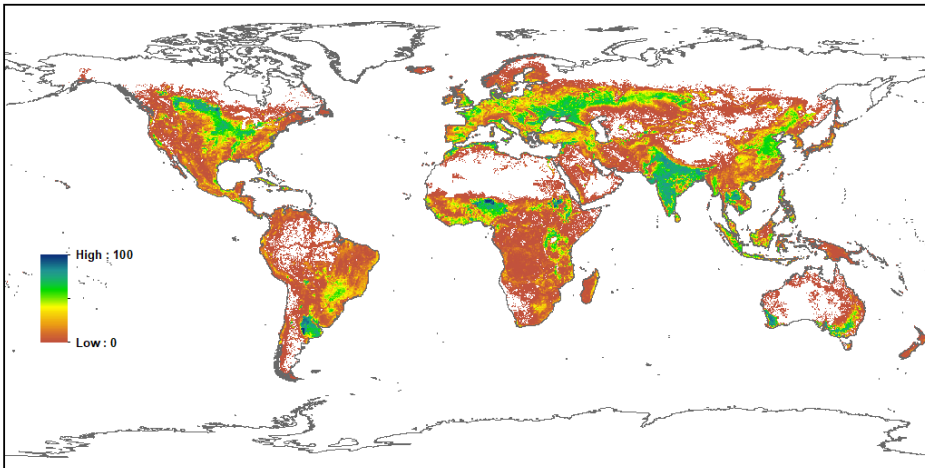
Figure 5.1: Current Day 2014 LUH2 Aggregated Land Use States for: (a) All Forests; (b) All Non Forests; and (c) All Crop.
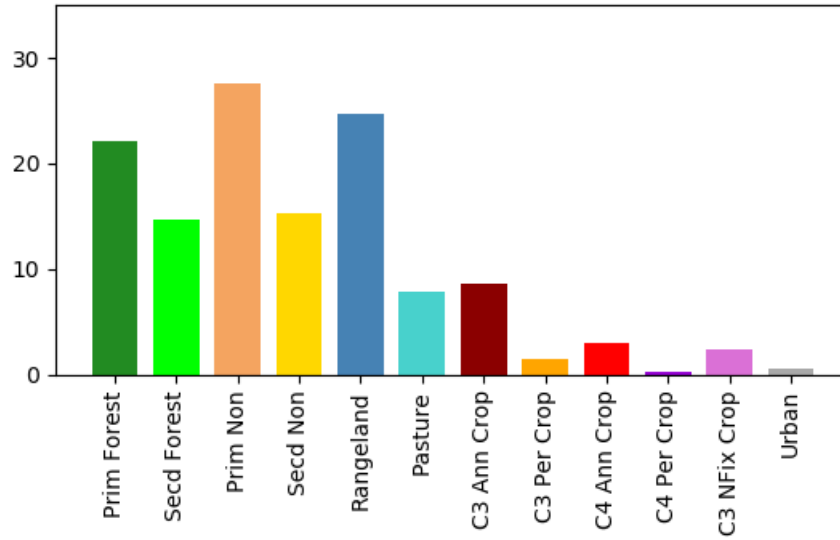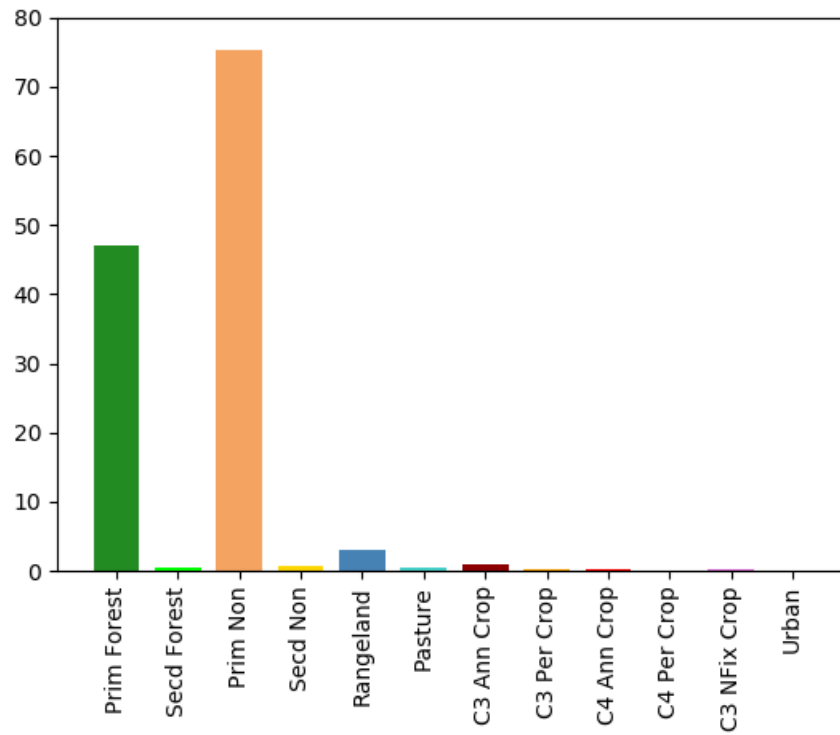
Figure 5.2: LUH2 Global Land Use States for: (a) Current Day 2014; and (b) Historical 850.
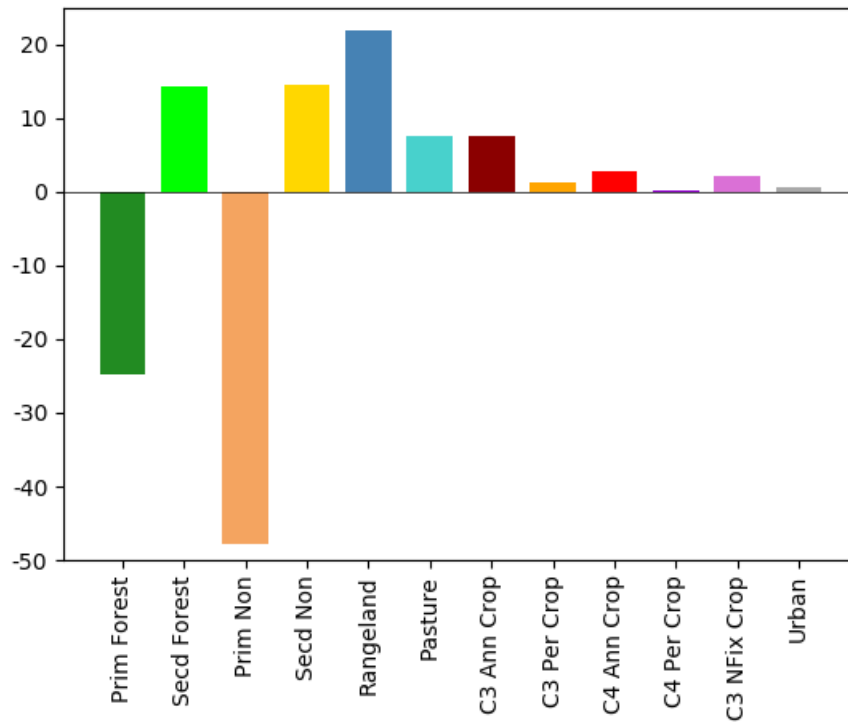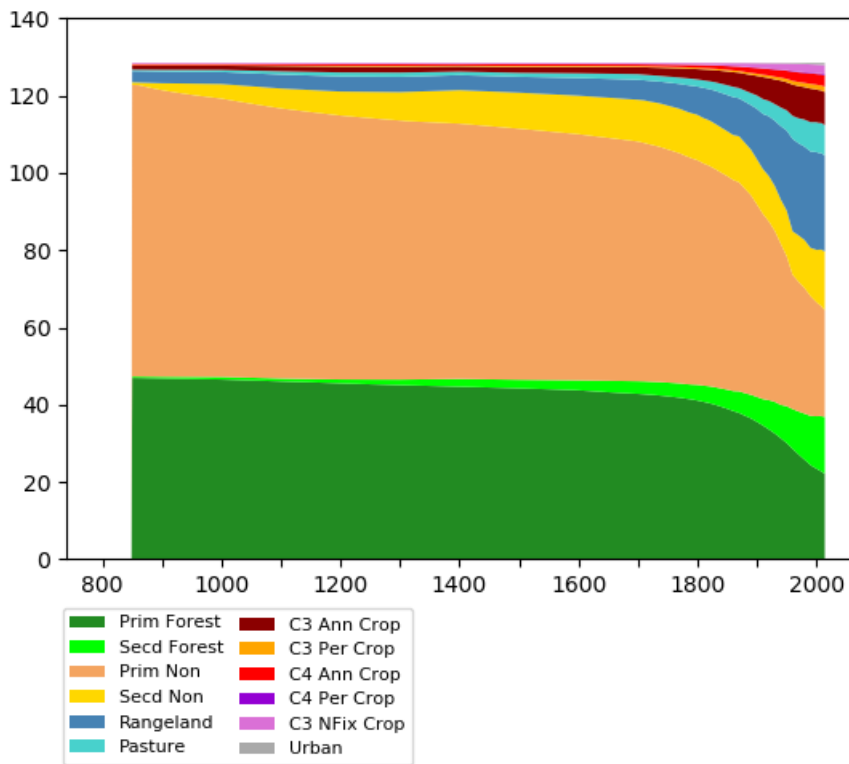
Figure 5.3: LUH2 Change in Global Land Use States for: (a) Difference between Current Day 2014 and Historical; and (b) Annual values from 850 to 2014.

The LUH2 global area of all non-forest in 850 was 76.1 million km^2, which represents a loss of -33.3 million km^2 over the period. Like the forest area, the LUH2 850 non-forest area was almost completely Primary Non Forest with a small area of Secondary Non Forest at 0.7 million km^2 associated with shifting cultivation at the beginning of the period. The largest losses in non-forest area were in Africa with -8.8 million km^2, Latin America with -4.3 million km^2, and Eurasia with -4.1 million km^2. The timeseries plot of Figure 5.3b shows there was little non-forest loss until 1700 with a slow transition from Primary Non Forest to Secondary Non Forest. After 1700 however the transition from Primary Non Forest to Secondary Non Forest accelerates. After 1700 there is also a rapid loss of total non-forest associated with the large expansion of Rangeland and Pasture land use over the period.

The current day 2014 LUH2 map of total non-forests in Figure 5.1b shows five main forest areas globally. These correspond with the Sahara, the Arabian Peninsula into Central Asia, Central Australia, the high Arctic in North America, and northern Siberia. The current day 2014 LUH2 Primary Non Forest and Secondary Non Forest mapping shown in Figures 5.4c and 5.5a identifies distinct areas for both classes of non-forest. The Primary Non Forest is found in the high Arctic, Central Asia and Central Australia. The Secondary Non Forest by contrast are found in North Africa into Western Asia.

The next largest global current day LUH2 land use state is grazing area with 32.6 million km^2 in 2014. The LUH2 grazing area is a combination of the current day LUH2 Rangeland area of 24.8 million km^2 and the LUH2 Pasture area of 7.9 million km^2. The largest areas of current day grazing are Africa with 8.9 million km^2, Latin America with 5.6 million km^2 and East Asia with 5.1 million km^2. Other large areas of current day grazing are in Eurasia, Asia Pacific, and North America. Globally Rangeland accounted for 75% of grazing land with Pasture making up the remaining 25%. In wetter regions Pasture made up the majority of grazing with Pasture accounting for 86% of grazing in Southeast Asia and 79% in Europe. In more arid regions Rangeland made up the majority of grazing, with Rangeland accounting for 98% of grazing in the Middle East and in Asia Pacific.

The LUH2 global area of grazing in 850 was 3.3 million km^2 representing a tenfold increase of 29.3 million km^2 over the period to 2014. Globally in 850 Rangeland accounted for 89% of all grazing land with Pasture only accounting for the remaining 11%. Over half of the grazing area in 850 was in Africa with 1.7 million km^2, with the regions of East Asia, Europe, Eurasia and the Middle East accounting for another 1.5 million km^2 combined. The timeseries plot of Figure 5.3b shows there was little increase in grazing area from 850 up to 1700, with slow expansions in Rangeland and Pasture over that period. After 1700 however there are rapid expansions of both Rangeland and Pasture land use up to 2014.

The current day 2014 LUH2 maps of Rangeland and Pasture in Figures 5.5b and c show the widespread extent of current day grazing. The two maps show the distinct climate division of Rangeland in more arid areas and Pasture in more moist areas. Current day Rangeland covers six main areas starting in western North America, southern South America, southern Africa, more arid areas of northern Africa into the Middle East, Central Asia and inland Australia. Pasture areas by contrast are found predominantly on the wetter boundaries of the Rangeland areas and through the eastern North America, Europe and China. While there are areas where both Pasture and Rangeland coexist, the majority of their distributions are mutually exclusive reflecting the strong role climate plays in the GLM2 and HYDE 3.2 methods for allocating both of these land use states.

(a) LUH2 Primary Forest 2014   (%)
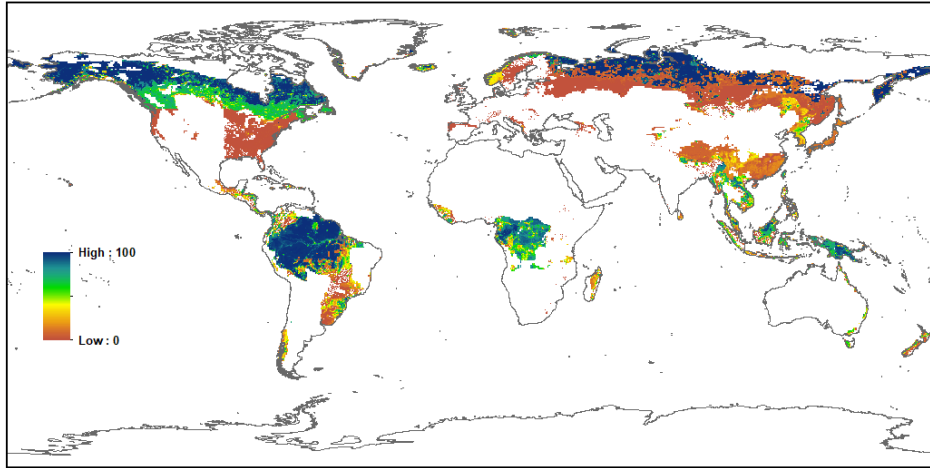
(b) LUH2 Secondary Forest 2014   (%)
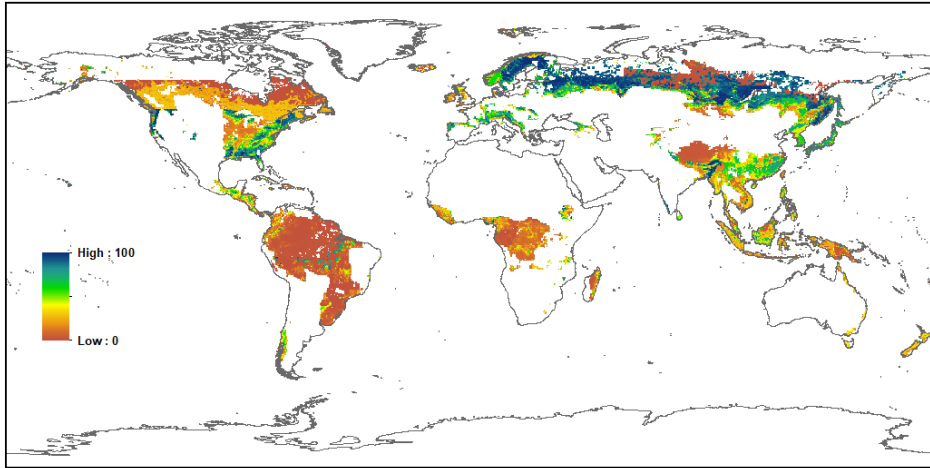
(c) LUH2 Primary Non Forest 2014   (%)
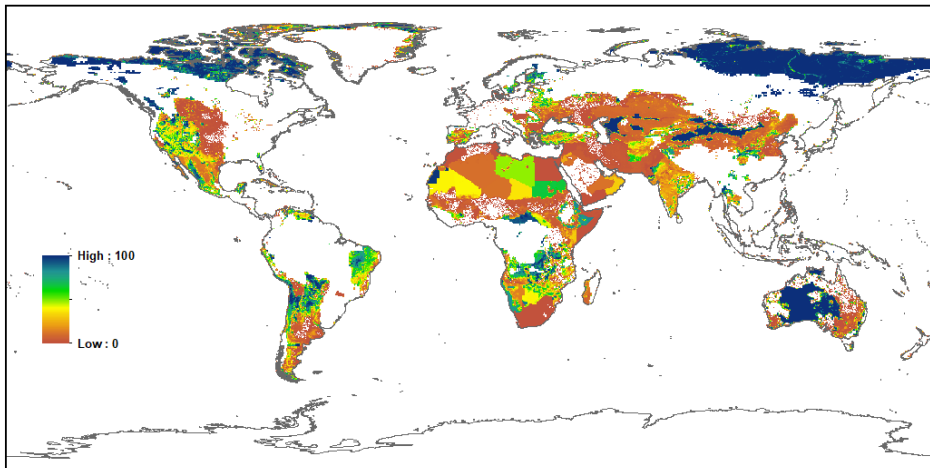
Figure 5.4: Current Day 2014 LUH2 Land Use States for: (a) Primary Forest; (b) Secondary Forest; and (c) Primary Non Forest.
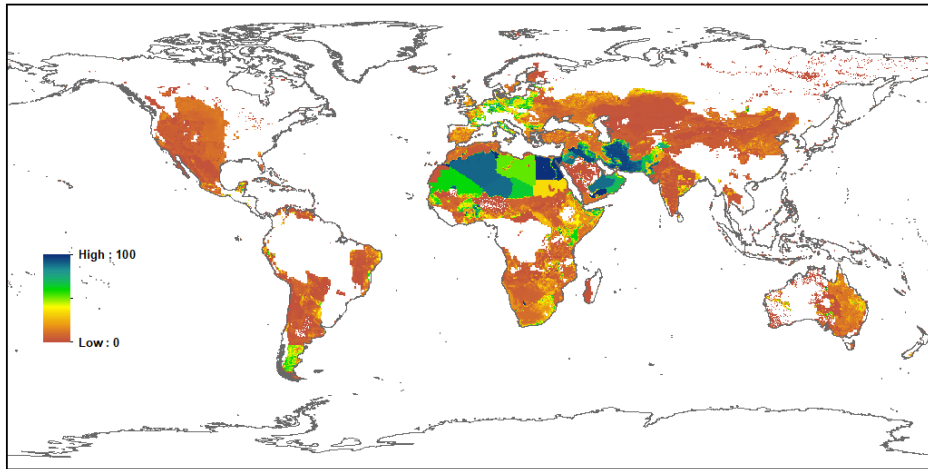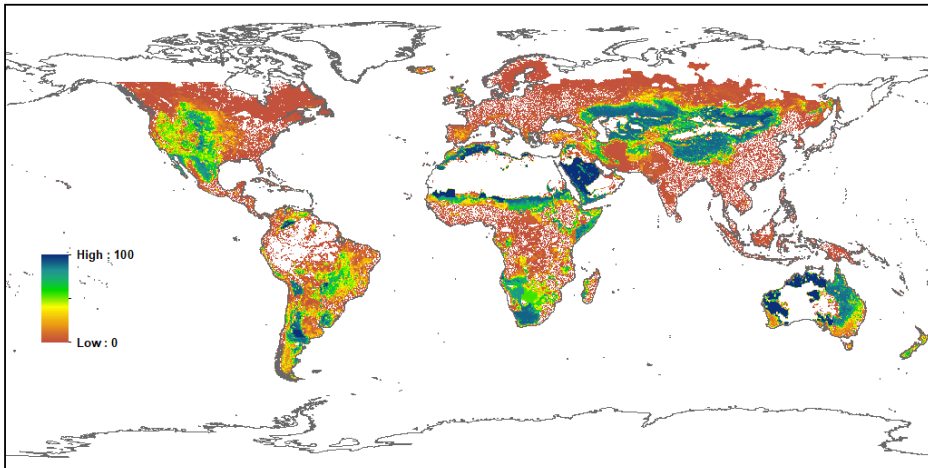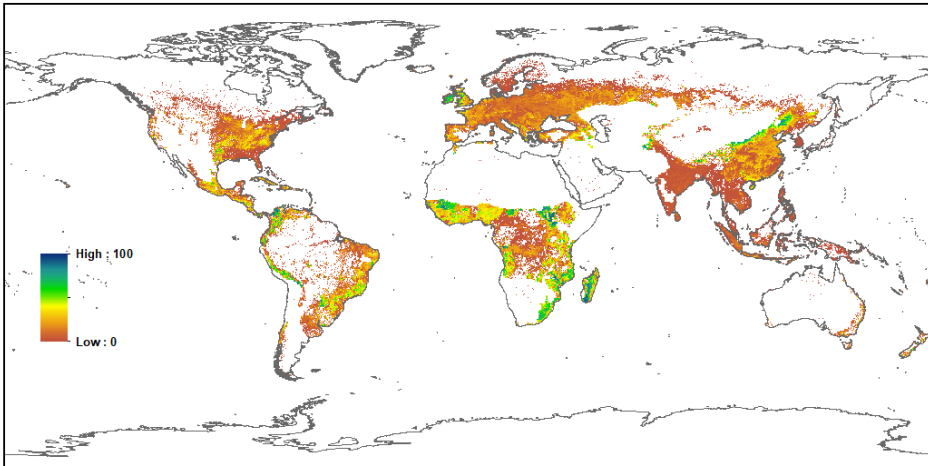
Figure 5.5: Current Day 2014 LUH2 Land Use States for: (a) Secondary Non Forest; (b) Rangeland; and (c) Pasture.

(a) LUH2 C3 Annual Crop 2014 (%)
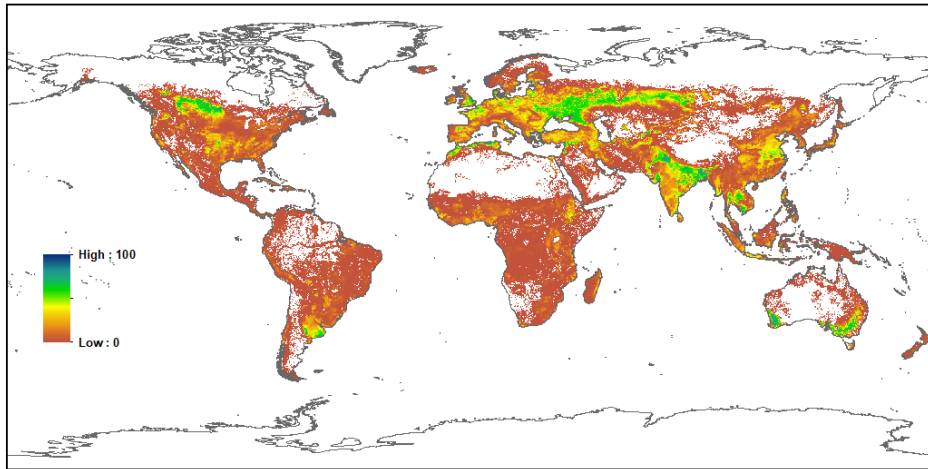
(b) LUH2 C4 Annual Crop 2014 (%)
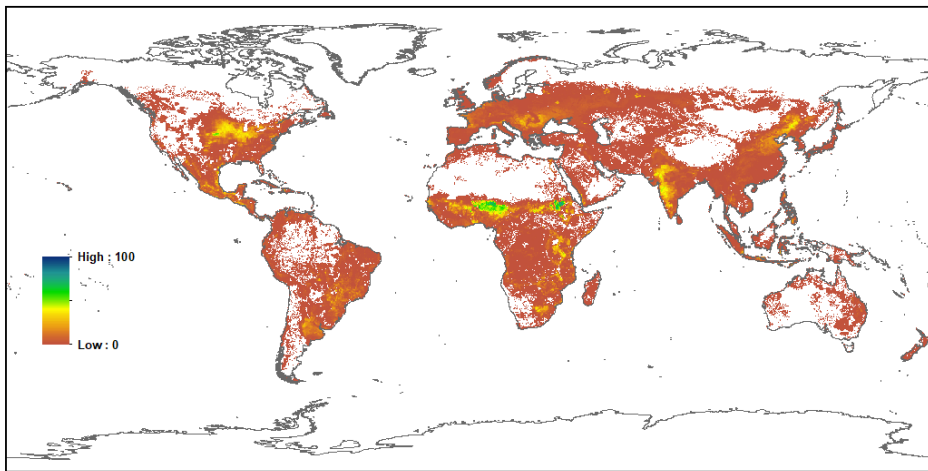
(c) LUH2 C3 Perennial Crop 2014 (%)
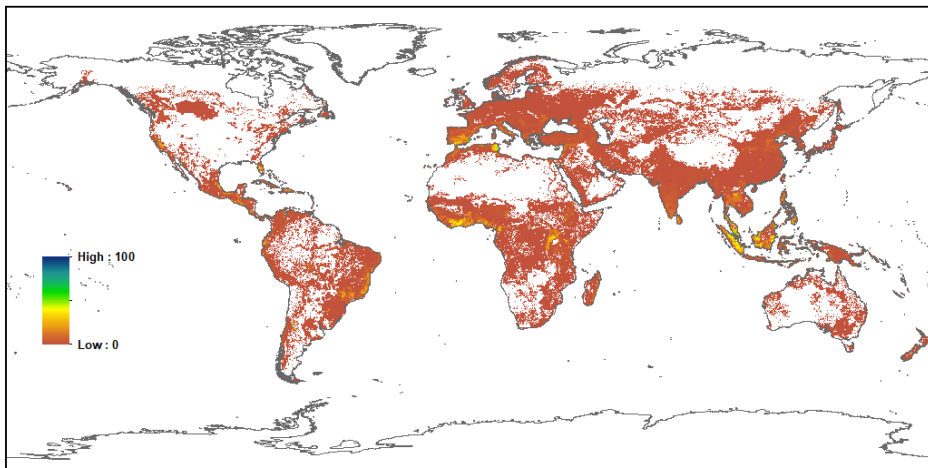
Figure 5.6: Current Day 2014 LUH2 Land Use States for: (a) C3 Annual Crop; (b) C4 Annual Crop; and (c) C3 Perennial Crop.

(a) LUH2 C4 Perennial Crop 2014   (%)
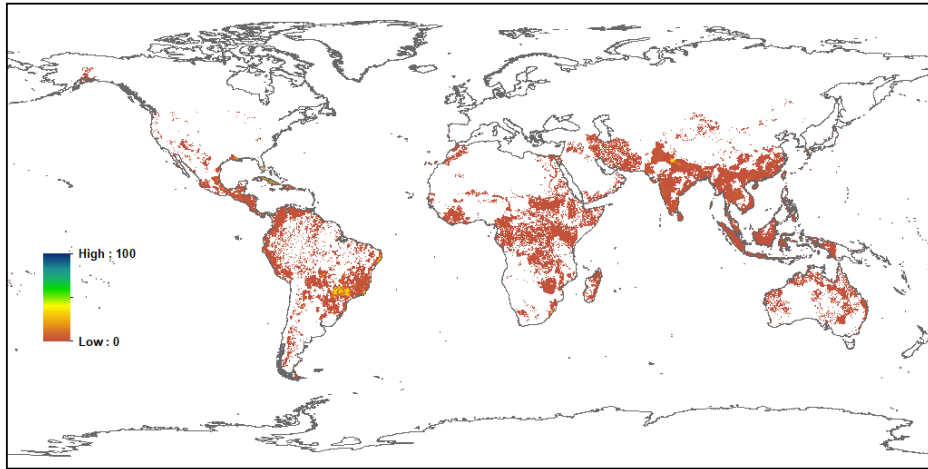
(b) LUH2 C3 N Fixing Crop 2014   (%)
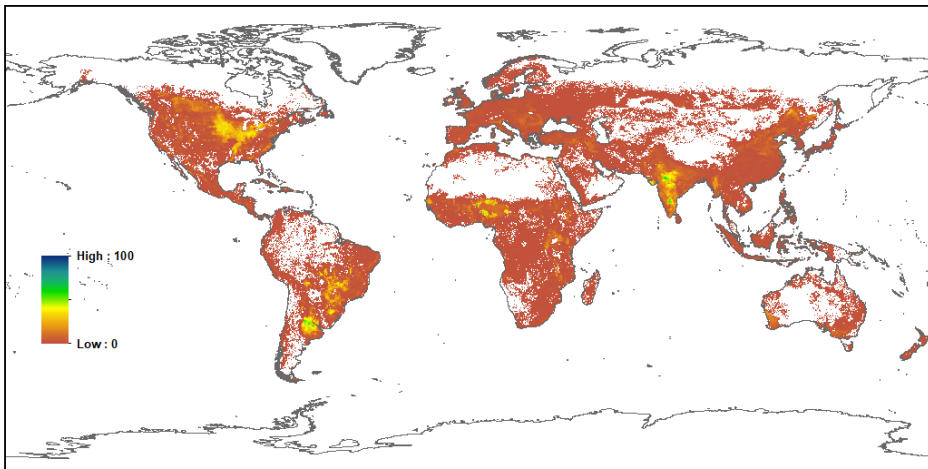
(c) LUH2 Urban 2014   (%)
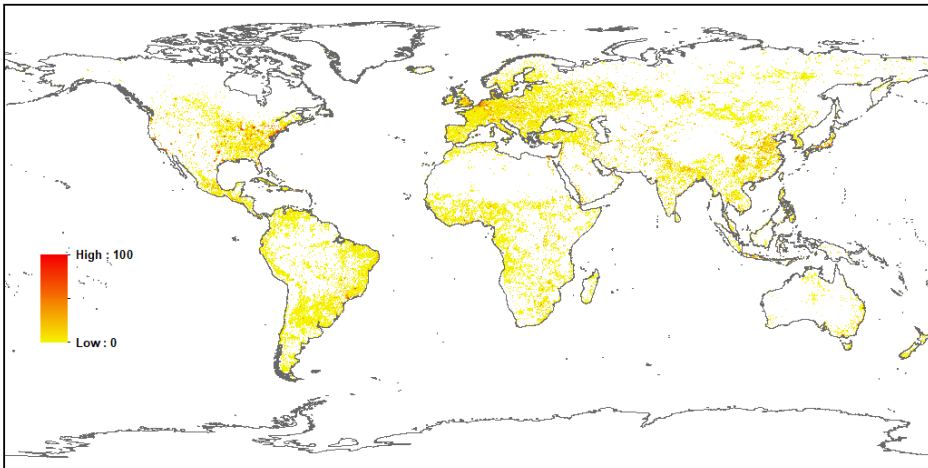
Figure 5.7: Current Day 2014 LUH2 Land Use States for: (a) C4 Perennial Crop; (b) C3 Nitrogen Fixing Crop; and (c) Urban.

The current day LUH2 global area of all crops in 2014 is 15.6 million km^2, which is comprised of 8.6 million km^2 of C3 Annual Crop, 3.0 million km^2 of C4 Annual Crop, 1.4 million km^2 of C3 Perennial Crop, 0.3 million km^2 of C4 Perennial Crop, and 2.4 million km^2 of C3 Nitrogen Fixing Crop. The total crop area compares to the 14.4 million km^2 of global crops for the year 2016 in the UN FAOSTAT database and the 13.6 million km^2 for the year 2000 in the MIRCA data set as reported in Table 3.3. This also compares to the 12.7 million km^2 of global crops for the year 2005 from the FAO/EarthStat data used to generate current day CLM5 PFTs and CFTs reported in Table 2.5. and Table 3.4. The largest areas of 2014 LUH2 crops were found in Africa at 2.8 million km^2, Southern Asia at 2.1 million km^2, North America at 2.0 million km^2, Latin America at 2.0 million km^2, Europe with 1.8 million km^2, Eurasia with 1.6 million km^2, and East Asia with 1.3 million km^2. Other regions with less 2014 LUH2 crop area were found in Southeast Asia with 1.1 million km^2, Asia Pacific with 0.5 million km^2, and the Middle East with 0.3 million km^2.

The LUH2 global area of all crops in 850 was 1.7 million km^2, which represents a corresponding increase in crops of 13.9 million km^2 over the period to 2014. The LUH2 850 crop area was predominantly made up of C3 Annual Crop at 1.0 million km^2, with smaller contributions from C4 Annual Crop at 0.3 million km^2, C3 Perennial Crop at 0.2 million km^2 and C3 Nitrogen Fixing Crop at 0.2 km^2. Nearly two thirds of the LUH2 crop area in 850 was in the areas of Southern Asia with 0.6 million km^2 and Europe with 0.4 million km^2. The regions of Africa, East Asia, Latin America and the Middle East accounted for another 0.6 million km^2 combined. The timeseries plot of Figure 5.3b shows there was little increase in crop area from 850 up to 1700, with slow expansions in all crop types over that period. After 1700 however there is a rapid expansion in C3 Annual Crop which continues up to 2014. The large expansion of C4 Annual Crop and C3 Nitrogen Fixing Crop however does not start until after 1800.

The current day 2014 LUH2 map of all crops in Figure 5.1c shows nine main crop areas globally. These correspond with the areas of the Mid-West into the Great Plains of North America, southern Brazil and northern Argentina, Europe into southern Russia, Nigeria and the Rift Valley of Africa, Pakistan, India and Bangladesh, China, Southeast Asia, and southern Australia. These are the same current day crop areas as found in the EarthStat/FAOSTAT crop map used to generate current day CLM5 PFTs and CFTs shown in Figure 2.18c, and as the crop areas of the MODIS land cover map shown in Figure 2.1a. Of the other two data sets he spatial distribution of the LUH2 current day crops more closely resembles the MODIS land cover mapping. The EarthStat/FAOSTAT mapping by contrast is more concentrated in particular regions, with large areas without any crops. LUH2 current day crops are more dispersed and reflective of the remotely sensed data.

The current day 2014 LUH2 individual crop type mapping shown in Figures 5.6 and 5.7, demonstrates that the LUH2 data corresponds well with the individual crop mapping of the EarthStat/FAOSTAT data of Chapter 3. The effectiveness of the mapping between the LUH2 crop types and the EarthStat/FAOSTAT also can be seen in the data preparation of Chapter 8, and with the generation of the CLM5 raw PFTs and CFTs from the LUH2 data with the CLM5 Land Use Data Tool in Chapter 6. Both of these components show that with the correct LUH2 crop type supporting data the current day EarthStat/FAOSTAT individual crop distributions can effectively be recreated from the LUH2 data for current day.

The current day 2014 LUH2 C3 Annual Crop mapping shown in Figure 5.6a, shows that the LUH2 data corresponds well with the current day EarthStat/FAOSTAT distributions of Wheat (Figure 3.3c), Rice (Figure 3.4c), Cotton (Figure 3.5a), Barley (Figure 3.6a), Cassava (Figure 3.6b), Fodder Grass (Figure 3.8a), Potatoes (Figure 3.9c), Rapeseed (Figure 3.10b), Rye (Figure 3.10c), Sugarbeet (Figure 3.11b), and Sunflower (Figure 3.11c). The 2014 LUH2 C4 Annual Crop

mapping shown in Figure 5.6b, corresponds well with the current day EarthStat/FAOSTAT mapping of Temperate Corn (Figure 3.3a), Tropical Corn (Figure 3.3b), Millet (Figure 3.9a), and Sorghum (Figure 3.11a).

The current day 2014 LUH2 C3 Perennial Crop mapping shown in Figure 5.6c, shows the LUH2 data corresponds well with the combined widely distributed EarthStat/FAOSTAT mapping of Citrus (Figure 3.6c), Cocoa (Figure 3.7a), Coffee (Figure 3.7b), Datepalm (Figure 3.7c), Grapes (Figure 3.8b), and Oilpalm (Figure 3.9b). The current day LUH2 C4 Perennial Crop mapping shown in Figure 5.7a, corresponds only with the current day EarthStat/FAOSTAT Sugarcane mapping (Figure 3.5b). There is good agreement for the major areas of both maps however the LUH2 mapping has very small amounts of C4 Perennial Crop at large distances from the Sugarcane distribution. This will be addressed further in the land use data preparation of Chapter 8. The current day LUH2 C3 Nitrogen Fixing Crop mapping shown in Figure 5.7b, again corresponds well with the EarthStat/FAOSTAT current day mapping of Temperate Soybean (Figure 3.4a), Tropical Soybean (Figure 3.4b), Groundnuts (Figure 3.8c), and Pulses (Figure 3.10a).

The final component of the LUH2 states data is the Urban mapping. The current day LUH2 global area of Urban, shown in Table 5.4 for 2014, is 0.5 million km^2. This is substantially lower than the 0.8 million km^2 of the CLM5 Urban combined mapping of Tall Building District, High Density and Medium Density urban classes from Jackson et al. (2010), shown in Table 4.1. It is however, the same as the 0.5 million km^2 from the MODIS Urban land cover data, shown in Table 2.5. The current day LUH2 2014 Urban land use state map of Figure 5.7c shows the LUH2 mapping has more widely distributed urban area compared to the Jackson et al. (2010) CLM Urban mapping of Figure 4.3a, and the MODIS derived Urban mapping of Figure 2.13c. The mapping and the regional analysis show that LUH2 had much lower Urban areas in India and China than found in the CLM Urban mapping. The LUH2 land use state is much smaller than all other states beside C4 Perennial Crop. Historically the LUH2 Urban area is only 3,900 km^2 in 850, making the contribution to the global LUH2 land use timeseries very small for the entire historical period as shown in Figure 5.3.

5.4 Historical Shifting Cultivation

The LUH2 land use timeseries considers Shifting Cultivation to be a specific land use involving the sequence of clearing, agricultural use for one to several years, and then subsequent abandonment to forest or other natural vegetation for a period of regeneration which may take up to several decades to renew the fertility of the land (Hurtt et al. 2020). The annual Shifting Cultivation areas and rates are generated in LUH2 from the expert-opinion based rates of Heinimann et al. (2017), consistent with the other annual land use states and transitions of the GLM2 model. The Shifting Cultivation land use is represented in the LUH2 data as the difference between the transitions between crop land use states, and the Secondary Forest and Secondary Non Forest states that do not result in changes to the absolute area of those states.

For CLM5 land data generation process, the area of land involved in Shifting Cultivation for both Forest and Non Forest are explicitly calculated for both land use states using the residual transitions once the changes in states are accounted for. The global and regional areas of LUH2 Shifting Cultivation are calculated for 850 and 2014 in Table 5.5. The annual time series of Shifting Cultivation for both states and for all land is shown in Figure 5.8. The current day 2014 distributions of Shifting Cultivation are shown globally for both states and for all land in Figure 5.9.

Table 5.5 Annual Global and Regional LUH2 Shifting Cultivation Area for 850 and 2014 in Millions of km^2.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
All Areas	0.28	0.11	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.08	0.03	0.00
Forest	0.17	0.04	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.07	0.02	0.00
Non Forest	0.11	0.07	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.01	0.00
Year 850												
All Areas	0.07	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.03	0.00
Forest	0.04	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00
Non Forest	0.03	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00

The global current day 2014 LUH2 Shifting Cultivation area was 0.28 million km^2, exactly matching the current day estimate of Heinimann et al. (2017). Of this area, Secondary Forest Shifting Cultivation accounted for 0.17 million km^2, with Secondary Non Forest Shifting Cultivation accounting for the other 0.11 million km^2. The largest current day regions of Shifting Cultivation were in Africa with 0.11 million km^2, Southeast Asia with 0.08 km^2, Latin America with 0.06 km^2, and Southern Asia with 0.03 km^2. The Non Forest Shifting Cultivation dominated in Africa accounting for nearly 70% while Forest Shifting Cultivation dominated in Southeast Asia, Latin America and Southern Asia. The dominance of the tropical regions reflects the description of shifting cultivation from Heinimann et al. (2017), locating the practice in central and tropical South America, tropical Africa, and tropical Southeast Asia.
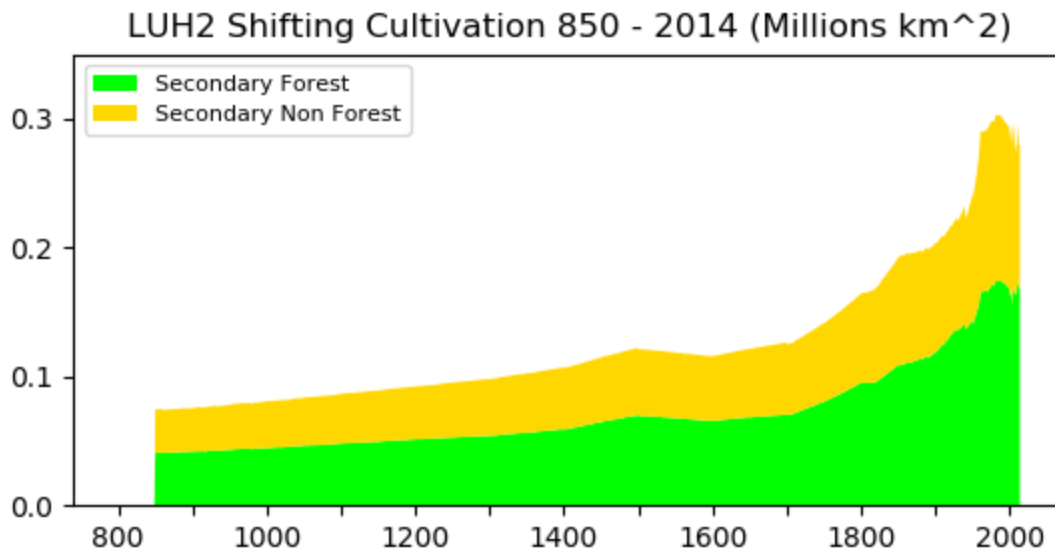
Figure 5.8: LUH2 Change in Global Shifting Cultivation for Secondary Forest and Secondary Non Forest Land Use States as annual values from 850 to 2014.

(a) LUH2 All Shifting Cultivation 2014   **(%)**
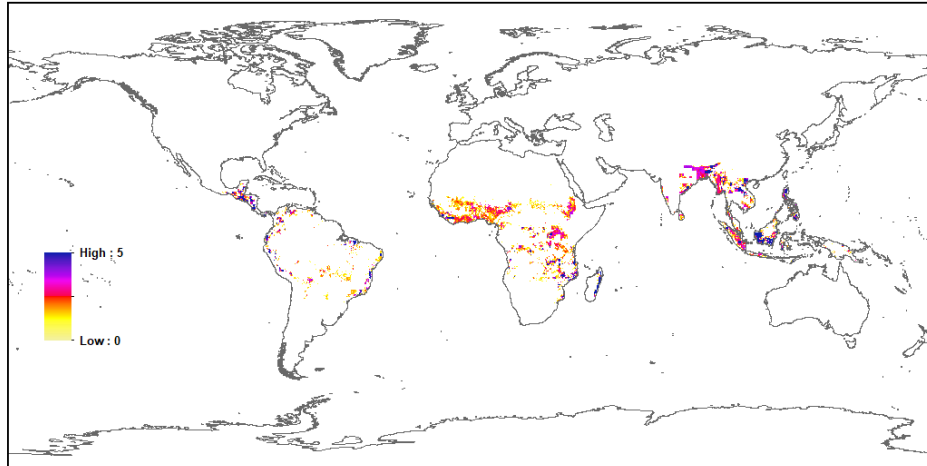
(b) LUH2 Forest Shifting Cultivation 2014   **(%)**
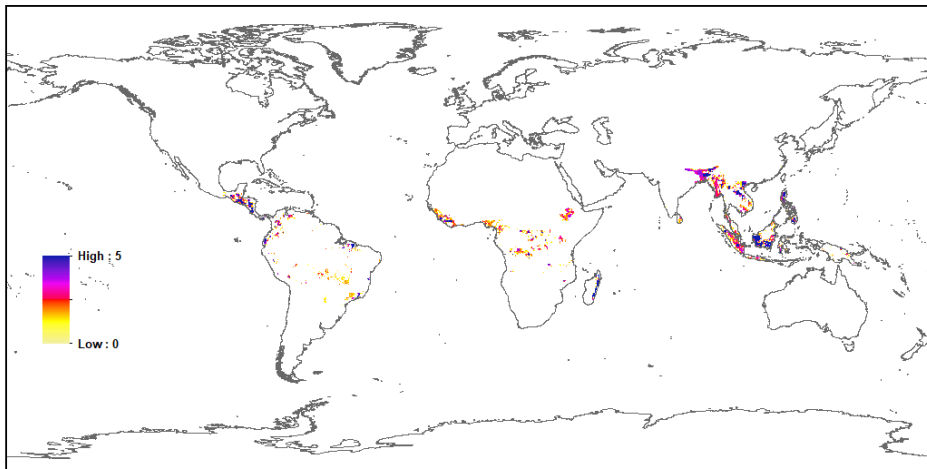
(c) LUH2 Non Forest Shifting Cultivation 2014   **(%)**
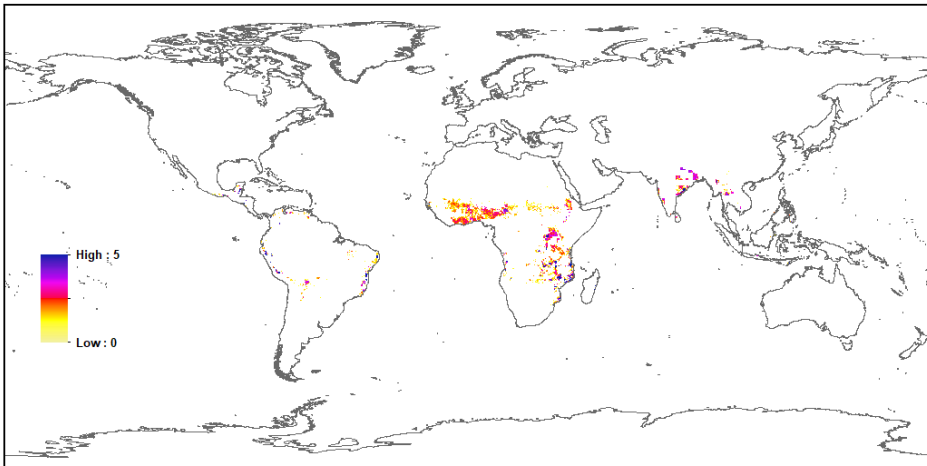
Figure 5.9: Current Day 2014 LUH2 Shifting Cultivation for: (a) All Land Use States; (b) Forest Land Use States; and (c) Non Forest Land Use States.

On the original LUH2 land use data the 850 Shifting Cultivation is set to 0.0 and slowly ramps up to the 0.07 million km^2 over the first 10 years of the timeseries. For generating CLM5 land surface data however, shifting cultivation continues back through time over the Transient Holocene to 6000bce. To enable this the LUH2 transitions, and Secondary Forest and Secondary Non Forest states for the first 10 years of the historical time series and all of the Transient Holocene period are weighted by the HYDE 3.2 population data at the grid cell level to generate new time series data for these years as discussed at the end of the chapter. Because of this change the 850 Shifting Cultivation reported here is closer to the original LUH2 861 value than the 850 value.

The CLM5 LUH2 850 Shifting Cultivation area is largest in Southern Asia with 0.03 million km^2 which is close to the current day 2014 value. The 850 Shifting Cultivation of other regions however is much lower than the 2014 values. The biggest difference in 850 Shifting Cultivation is in Africa with only 0.02 million km^2 representing a 0.09 million km^2 increase in Shifting Cultivation over the period. Other big differences in 850 Shifting Cultivation were found in Southeast Asia with 0.01 million km^2, and in Latin America with 0.02 million km^2, representing increases in Shifting Cultivation of 0.07 million km^2 and 0.04 million km^2 respectively.

The current day 2014 LUH2 All Shifting Cultivation mapping of Figure 5.9a, shows how the shifting cultivation is restricted to three main areas. These areas reflect the regional analysis of Table 5.5, and the description of shifting cultivation made in Heinimann et al. (2017). The three key areas are Central America through tropical South America, tropical Africa, and India through Bangladesh into Southeast Asia. The current day 2014 LUH2 Forest Shifting Cultivation map of Figure 5.9b shows these areas are concentrated in Central America, the Congo, Bangladesh and Southeast Asia, while the Non Forest Shifting Cultivation map of Figure 5.9c shows that these areas are found predominantly in the woody savannas of Africa and India.

5.5 Historical Irrigation and Fertilizer

The LUH2 land use time series includes historical crop management activities through information on irrigation, flooded agriculture, and industrial nitrogen fertilizer application. The irrigation data and the area of flooded rice are prescribed from HYDE 3.2, which is proportionally allocated to the LUH2 crop types. The C3 Annual Crop which associated with rice is the only flooded crop considered. Industrial fertilizer application for 1961 – 2011 was based on Zhang et al. (2015), which is originally derived from FAOSTAT country level fertilizer data (FAO 2020b). The industrial nitrogen fertilizer application from 1915 to 1960 was hindcast using the 1961 distributions scaled by the global annual fertilizer data from Smil (2001). For the 2012 to 2015 period, the same process was applied using the 2011 distribution scaled by the global annual fertilizer data from the International Fertilizer Association (IFA 2014). The scaling rates in both cases are applied equally to all crops in all regions relative to the base country level crop specific nitrogen fertilizer rates of the reference year.

The global and regional areas of all LUH2 irrigation along with individual crop irrigation areas are calculated for 2014 and 850 in Table 5.6. The individual 2014 crop irrigation areas are plotted in Figure 5.10a, with the transient annual crop irrigation areas shown in Figure 5.10b for 850 to 2014. The total global area of current day LUH2 irrigation is 2.66 million km^2 in 2014, which represents 17% of all LUH2 crops in 2014. This compares to the FAOSTAT value of 2.51 million km^2 for the year 2000, the FAOSTAT value of 2.80 million km^2 for the year 2016, and the MIRCA 2000 value of 3.18 million km^2 as shown in Table 3.3. The 850 LUH2 irrigation area was 0.04 million km^2, which was less than 2% of the current day value. The time series plot of annual

irrigation in Figure 5.10b shows that irrigation remained very low from 850 until 1700 where it increased slowly to 1800 before increasing rapidly after 1800 to current day.

Regionally the largest areas of current day 2014 LUH2 irrigation areas were found in Southern Asia with 0.89 million km^2 representing 42% of crops, and in Eastern Asia with 0.61 million km^2 representing 48% of crops. Smaller current day LUH2 irrigation areas were found in North America with 0.22 million km^2 representing only 11% of crops, Southeast Asia with 0.20 million km^2 representing 19% of crops, and in Latin America with 0.18 million km^2 representing only 9% of crops. The smallest area of LUH2 irrigation was in Asia Pacific at 0.04 million km^2, while the lowest irrigation rate was in Africa at 4% of crops.

By crop type the global current day LUH2 C3 Annual Crop irrigation was the largest area with 1.66 million km^2, with C4 Annual Crop irrigation area at 0.38 million km^2 and C3 Nitrogen Fixing Crop irrigation at 0.33 million km^2. The current day LUH2 irrigation map in Figure 5.1 shows the global distribution of irrigation in 2014 with large concentrations of irrigated crop in northern India and Pakistan along the Indus and Ganges rivers, and in northeastern China. Compared to the current day CLM5 irrigation mapping in Figure 3.5c however, there is less LUH2 irrigation across Europe into southern Russia and across North America.

Table 5.6 Annual Global and Regional LUH2 Irrigation Area in Millions of km^2 for 850 and 2014.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
All Irrigation	2.66	0.12	0.04	0.61	0.13	0.15	0.18	0.14	0.22	0.20	0.89	0.00
Irrig c3ann	1.66	0.06	0.03	0.40	0.11	0.09	0.05	0.09	0.10	0.13	0.60	0.00
Irrig c4ann	0.38	0.03	0.00	0.10	0.01	0.02	0.05	0.01	0.05	0.02	0.10	0.00
Irrig c3per	0.22	0.01	0.00	0.04	0.01	0.03	0.04	0.01	0.01	0.04	0.03	0.00
Irrig c4per	0.07	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.03	0.00
Irrig c3nfx	0.33	0.01	0.00	0.06	0.00	0.01	0.03	0.01	0.06	0.01	0.13	0.00
Year 850												
All Irrigation	0.04	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Irrig c3ann	0.03	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Irrig c4ann	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Irrig c3per	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Irrig c4per	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Irrig c3nfx	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The global and regional amounts of all LUH2 nitrogen fertilizer along with individual crop fertilizer amounts are calculated for 2014 in Table 5.7. The global 850 LUH2 fertilizer amount was 0.0 TgN for all crops, which is why these values are not listed. The individual 2014 LUH2 crop level fertilizer amounts are plotted in Figure 5.10a. The total global current day LUH2 nitrogen fertilizer is 108.7 TgN in 2014, representing an average fertilizer application rate of 69.7 kgN/ha. This compares to the FAOSTAT global value of 100 TgN for the year 2012 as reported by Zhang (2015).

The global individual crop type average fertilizer rates show strong differences, with C4 Perennial Crop having the highest rate of 120.8 kgN/ha, followed by C4 Annual Crop and C3 Annual Crop rates of 78.2 and 77.8 kgN/ha respectively. C3 Nitrogen Fixing Crop had the lowest rates at 23.6 kgN/ha. The time series plot of annual nitrogen fertilizer in Figure 5.10c shows that LUH2 industrial nitrogen fertilizer only started in 1915 with a slow ramp up to 1950, after which fertilizer application rapidly increases globally to current day values in 2014.
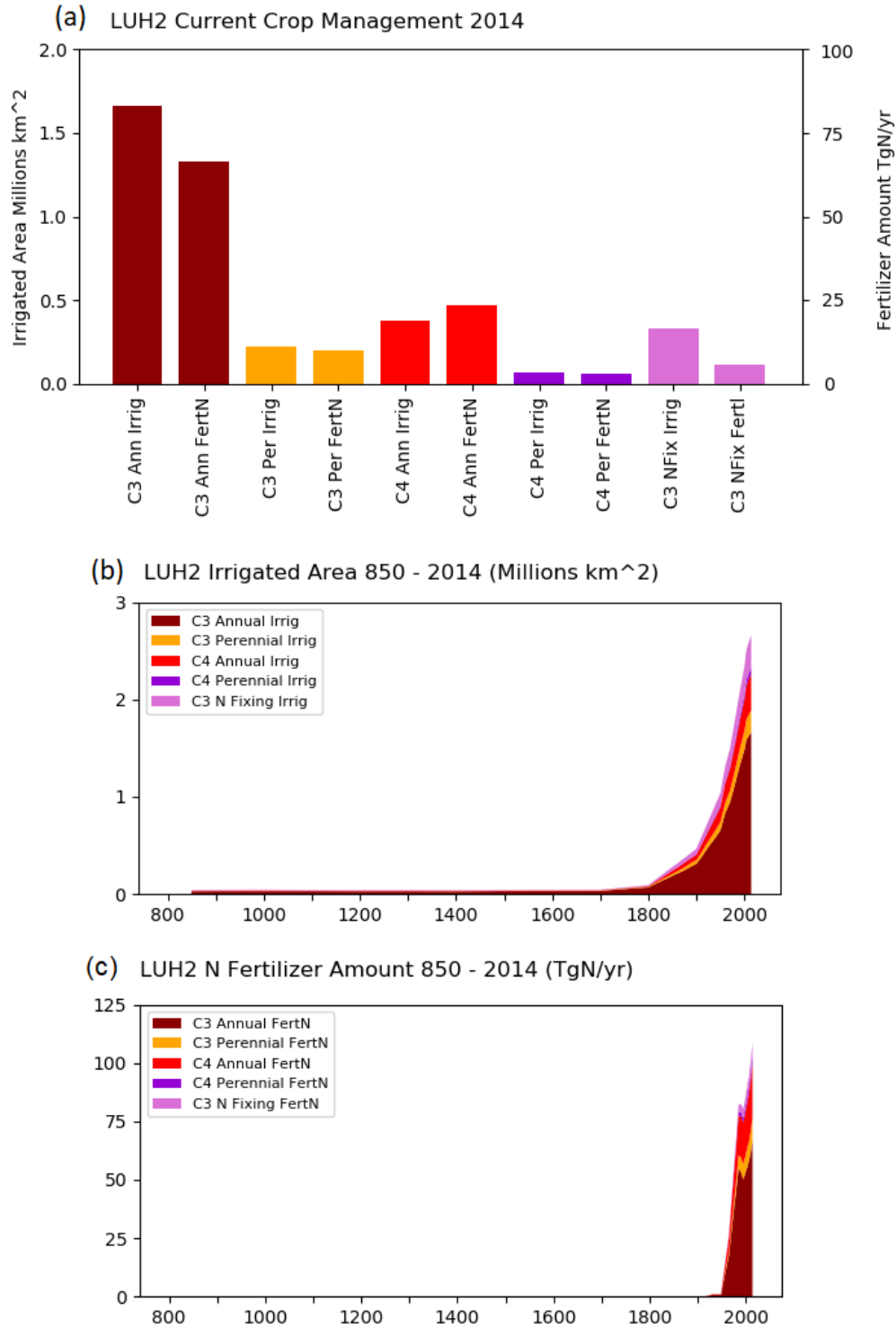
Figure 5.10: LUH2 Global Agricultural Management for Crop Level Irrigation and Nitrogen Fertilizer for: (a) Current Day 2014; (b) Annual values of Irrigation from 850 to 2014; and (c) Annual values of Nitrogen Fertilizer from 850 to 2014.

(a) LUH2 All Irrigated Crops 2014 (%)
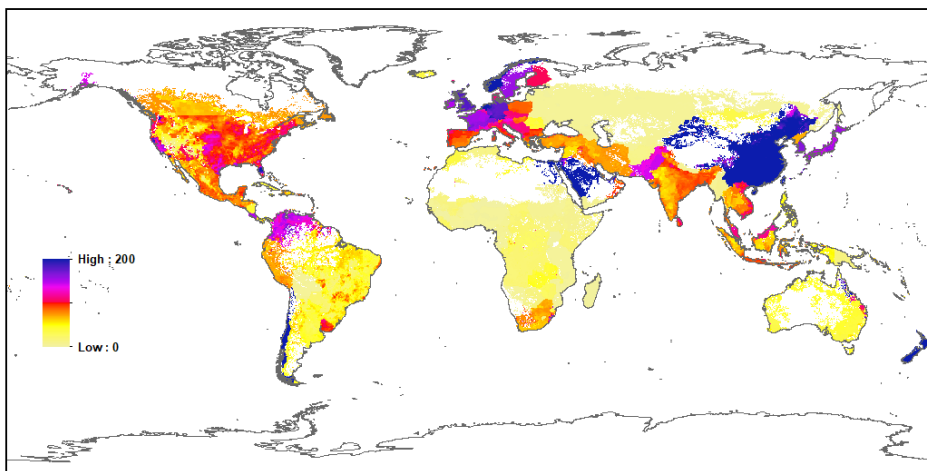
High : 100
Low : 0
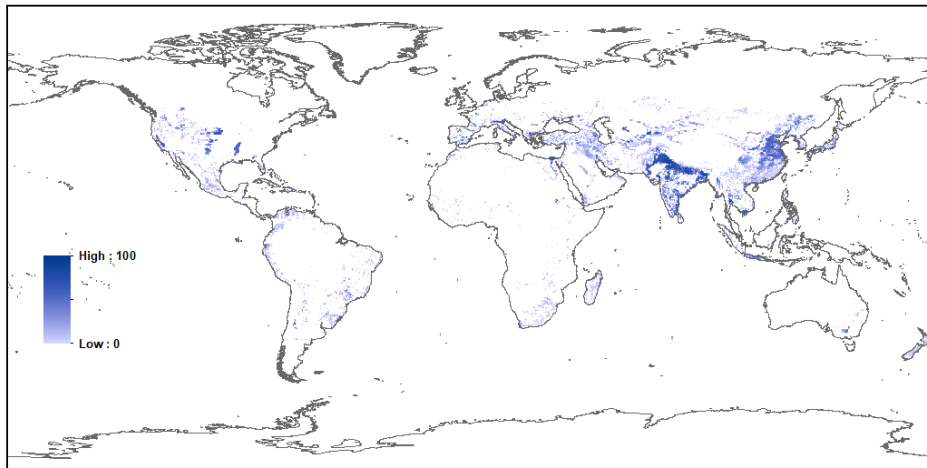
(b) LUH2 All Fertilizer 2014 kgN/ha

High : 200
Low : 0

Figure 5.11: Current Day 2014 LUH2 Agricultural Management for: (a) Irrigated Crops as percentage of grid cell; and (b) Average Crop Nitrogen Fertilizer in kgN/ha.

Regionally the largest amount of current day LUH2 fertilizer was in East Asia with 27.6 TgN, at an average rate of 219.8 kgN/ha. This rate is over two times the rate of any other region. The next largest amount of fertilizer was in North America with 18.2 TgN, at an average rate of 89.1 kgN/ha. This is followed by Southern Asia with 17.5 TgN at an average rate of 82.6 kgN/ha, and Europe with 16.1 TgN and average rate of 90.2 kgN/ha. The Middle East, Latin America, and Asia Pacific regions all had similar fertilizer rates between 52.3 and 86.8 kgN/ha, but due to the smaller crop areas of these regions they had smaller total fertilizer use. Africa and Eurasia both had much lower average fertilizer rates at 15.6 and 5.8 kgN/ha respectively.

The global map of current day average LUH2 nitrogen fertilizer application in Figure 5.11b shows similar differences to the regional analysis, but with differences in fertilizer rates within the regions following the different crop distributions within each region. The highest fertilizer application rates, around 200 kgN/ha, are found across China, Japan, the Middle East and areas in Europe. Lower fertilizer rates around 100 kgN/ha are shown through North America, other areas of Europe, areas

of South America, Western Asia into Pakistan, India and Southeast Asia, and in South Africa. The remaining crop areas have low fertilizer applications below 50 kgN/ha.

Table 5.7 Annual Global and Regional LUH2 Fertilizer Amount in TgN and Average Fertilizer Application Rate in kgN/ha for 2014.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
All Fertilizer	108.7	4.4	2.7	27.6	1.0	16.1	10.8	3.0	18.2	7.5	17.5	0.0
Fert c3ann	66.6	2.0	2.3	18.1	0.9	11.8	3.4	2.0	9.7	3.9	12.7	0.0
Fert c4ann	23.6	1.4	0.1	5.2	0.0	2.9	4.1	0.4	7.1	0.8	1.3	0.0
Fert c3per	9.9	0.4	0.2	2.6	0.1	0.9	1.6	0.4	0.6	2.3	0.6	0.0
Fert c4per	3.1	0.2	0.1	0.3	0.0	0.0	1.5	0.0	0.1	0.3	0.7	0.0
Fert c3nfx	5.6	0.4	0.1	1.5	0.0	0.4	0.3	0.1	0.6	0.2	2.1	0.0
Year 2014												
All Fert Rate	69.7	15.6	52.3	219.8	5.8	90.2	54.2	86.8	89.1	68.4	82.6	91.3
Rate c3ann	77.8	22.3	49.9	234.7	5.8	90.4	57.6	82.9	92.5	72.9	100.2	103.5
Rate c4ann	78.2	12.8	102.2	214.5	3.3	140.0	83.0	165.3	182.7	99.7	41.8	183.1
Rate c3per	69.9	12.5	299.2	326.3	21.7	62.9	61.7	106.4	211.5	62.2	68.4	61.4
Rate c4per	120.8	90.8	212.9	365.9	206.7	140.9	100.9	210.9	165.7	106.0	148.7	107.2
Rate c3nfx	23.6	8.7	10.9	93.1	2.4	34.5	5.6	36.1	10.5	21.3	54.1	44.4

5.6 Historical Wood Harvest

The historical wood harvest in LUH2 is based on national statistics for 199 countries based on data from HYDE 3.2 of Klein Goldewijk et al. (2017), combined with the country level FAO national wood harvest volume data for 1961 to 2015 from FAO (2020c), adjusted to biomass carbon amounts using the methods of Houghton and Hackler (2000). For years from 1920 to 1960 harvest rates are hindcast using HYDE 3.2 population data combined with the 1920 national wood harvest totals from Zon and Sparhawk (1923) and early 1960s country level data from the FAO wood harvest database. Prior to 1920 LUH2 provides three alternative wood harvest pathways. For this technote we only investigate the baseline historical wood harvest pathway. The baseline pathway uses nationally scaled values of the wood harvest data of Kaplan et al. (2017) from 850 to 1800, with the annual wood harvest from 1800 to 1920 time interpolated between the Kaplan et al. (2017) 1800 wood harvest values and the Zon and Sparhawk (1923) 1920 wood harvest values.

The global and regional areas of LUH2 wood harvest along with the individual wood harvest class areas are calculated for 2014 and 850 in Table 5.8. The amount of LUH2 wood harvest biomass along with the individual wood harvest class biomass amounts are calculated for 2014 and 850 in Table 5.9. The total global area of current day LUH2 wood harvest is 0.85 million km^2/yr, with an associated 1.38 PgC/yr in biomass extracted for 2014. The global individual 2014 wood harvest class area and biomass values are plotted against each other in Figure 5.12a. The plot shows that the greatest area of 2014 wood harvest was in the Secondary Non Forest class at 0.59 million km^2/yr, but with a relatively small amount of biomass extracted at 0.16 PgC/yr. Conversely, the largest amount of current day biomass was in Primary Forest 0.64 PgC/yr, but with a relatively small wood harvest area of 0.07 million km^2/yr. The plot also shows that Secondary Mature Forest and Secondary Young Forest classes both had small global areas of 0.03 million km^2/yr each but made up a large amount of the biomass wood harvest at 0.25 PgC/yr each. The differences in area versus biomass reflects the differences in wood density in the different forest and non-forest wood harvest classes.

The total global 850 LUH2 wood harvest area was 0.04 million km^2, which was less than 5% of the current day area representing a 20 fold increase in wood harvest area over the historical period. The total global 850 LUH2 wood harvest biomass was 0.02 PgC which is less than 2% of the current day and reflecting a nearly 70 fold increase in biomass extraction. The global transient 850 to 2014 annual wood harvest areas values are shown in Figure 5.12b, with the transient wood harvest biomass shown in Figure 5.12c.

Table 5.8 Annual Global and Regional LUH2 Wood Harvest Area in Millions of km^2 for 850 and 2014.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
Harvest area	0.85	0.58	0.00	0.02	0.03	0.08	0.01	0.01	0.04	0.01	0.08	0.00
primf area	0.07	0.01	0.00	0.01	0.02	0.00	0.01	0.00	0.01	0.01	0.00	0.00
primn area	0.11	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
secmf area	0.03	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
secyf area	0.03	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
secnf area	0.59	0.49	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.00	0.04	0.00
Year 850												
Harvest area	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
primf area	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
primn area	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
secmf area	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
secyf area	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
secnf area	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.9 Annual Global and Regional LUH2 Wood Harvest Biomass Amount in PgC for 850 and 2014.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
Harv amount	1.38	0.28	0.03	0.14	0.09	0.16	0.19	0.00	0.18	0.13	0.18	0.00
primf amount	0.64	0.07	0.02	0.05	0.09	0.00	0.18	0.00	0.09	0.12	0.01	0.00
primn amount	0.07	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
secmf amount	0.25	0.00	0.01	0.09	0.00	0.05	0.01	0.00	0.09	0.00	0.00	0.00
secyf amount	0.25	0.06	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.01	0.12	0.00
secnf amount	0.16	0.11	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.02	0.00
Year 850												
Harv amount	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
primf amount	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
primn amount	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
secmf amount	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
secyf amount	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
secnf amount	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
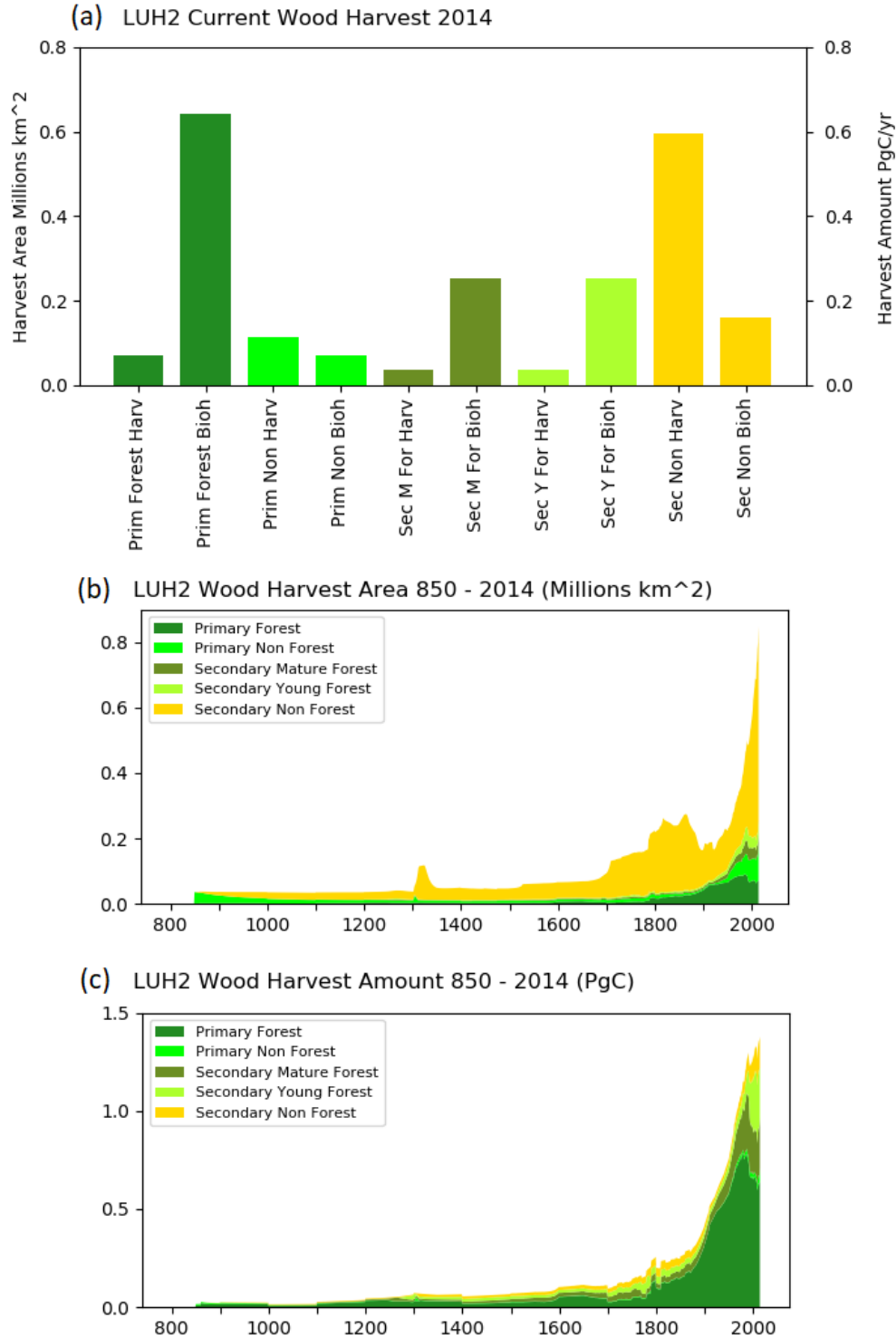
Figure 5.12: LUH2 Global Forest Management for Wood Harvest Area and Wood Harvest Biomass Amount for: (a) Current Day 2014; (b) Annual values of Wood Harvest Area from 850 to 2014; and (c) Annual values of Wood Harvest Biomass Amount from 850 to 2014.

(a) LUH2 All Wood Harvest 2014   Millions kgC
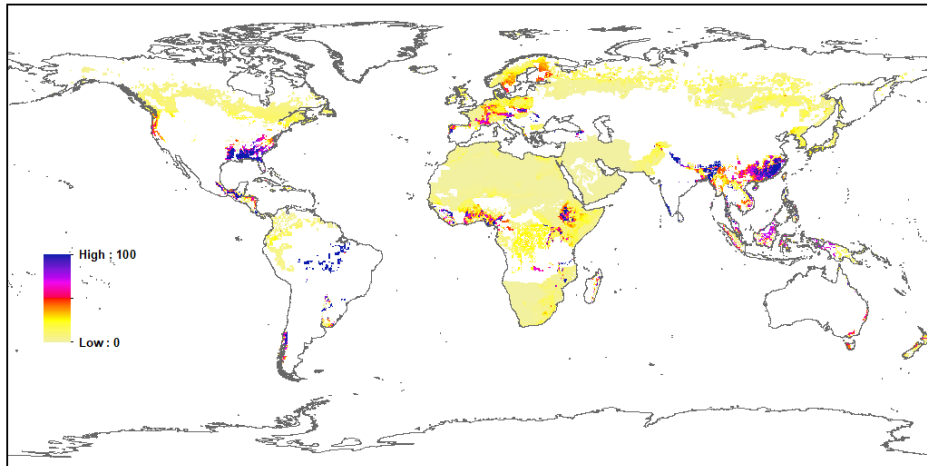
(b) LUH2 Primary Forest Wood Harvest 2014   Millions kgC
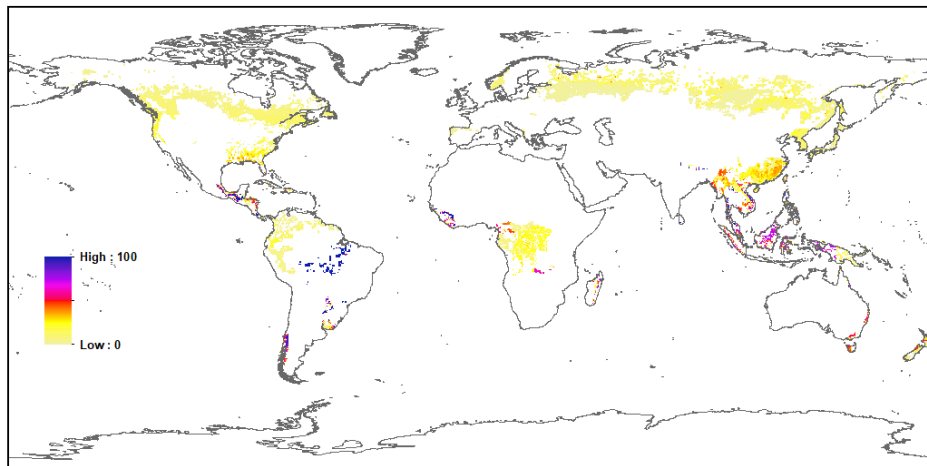
(c) LUH2 Primary Non Forest Wood Harvest 2014   Millions kgC
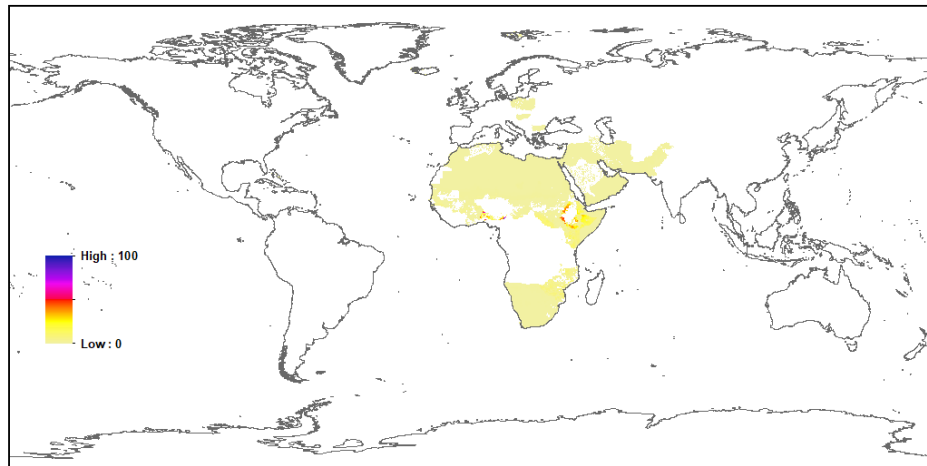
Figure 5.13: Current Day 2014 LUH2 Wood Harvest Biomass Amount for: (a) All Land Use States; (b) Primary Forest; and (c) Primary Non Forest.

(a) LUH2 Secondary Mature Forest Wood Harvest 2014   Millions kgC
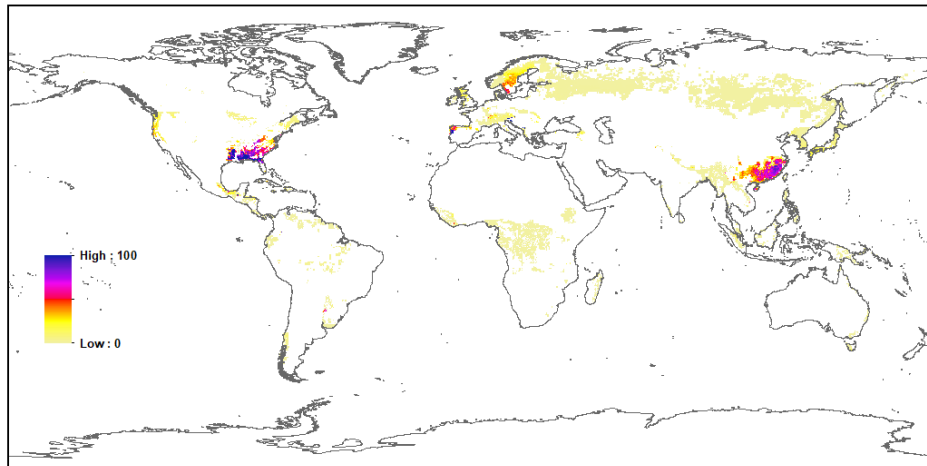
(b) LUH2 Secondary Young Forest Wood Harvest 2014   Millions kgC
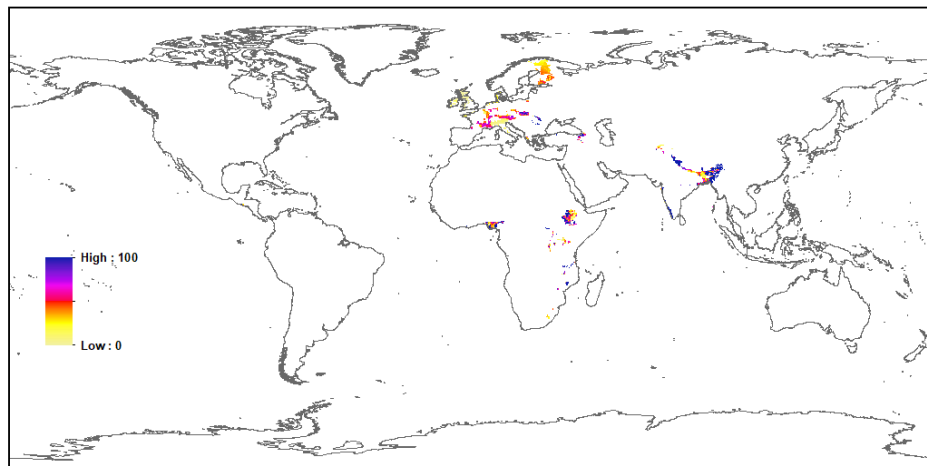
(c) LUH2 Secondary Non Forest Wood Harvest 2014   Millions kgC
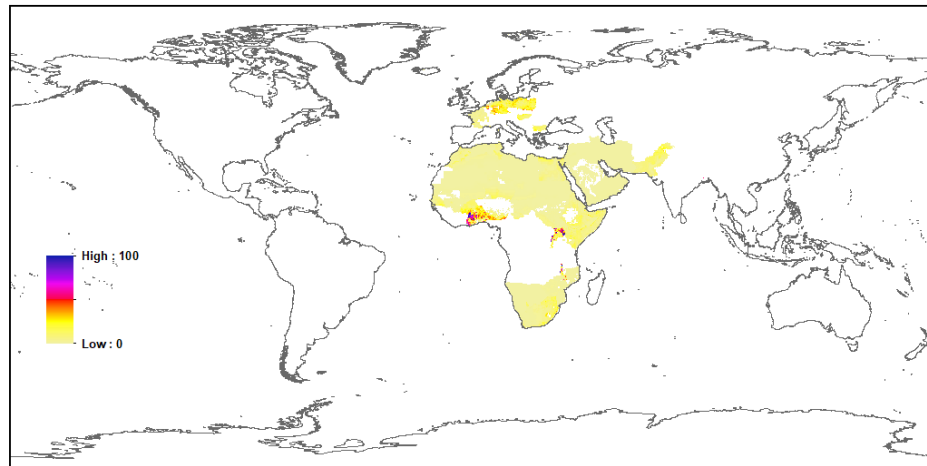
Figure 5.14: Current Day 2014 LUH2 Wood Harvest Biomass Amount for: (a) Mature Secondary Forest; (b) Young Secondary Forest; and (c) Secondary Non Forest.

The time series plot of annual wood harvest area shows that wood harvest area remained low from 850 until 1700 and was predominantly Primary Non Forest and Secondary Non Forest classes for that period. After 1700 the non-forest classes grow rapidly to 1850 before decreasing through 1950 before a large increase to current day. The Primary Forest area of wood harvest is very low until the late 1700s after which there is a steady increase through to the end of the 1970s with a small drop to a constant level until the current day. The Secondary Mature Forest and Secondary Young Forest wood harvest area are very small prior to 1900, with equal sized increases from 1900 to current day levels.

The time series plot of annual LUH2 wood harvest biomass shows that the total amount of carbon removed slowly ramped up from 850 until 1700. From 1700 to 1900 there was a slow ramp up of biomass extraction that changes a much larger ramp up from 1900 to current day. Prior to 1700 the biomass carbon comes from both forest and non-forest sources in similar amounts. After 1700 Primary Forest wood harvest becomes the primary source of biomass until 1900. After 1900 Secondary Mature Forest and Secondary Young Forest wood harvest become increasing important, while Primary Forest continues to be the major source of wood harvest biomass until the current day.

Regionally, Africa has the largest area of current day wood harvest at 0.58 million km^2, which makes up almost 70% of the global area of wood harvest. Africa also has the largest amount of wood harvest biomass at 0.28 PgC/yr. The African wood harvest area is predominantly Secondary Non Forest and Primary Non Forest with low associated biomass amounts. The remaining much smaller areas of African Primary Forest and Secondary Young Forest account for an equal amount of biomass harvest. The second largest regional current day wood harvest biomass amount is Latin America at 0.19 PgC/yr but on a much smaller area of 0.01 million km^2, which is almost all Primary Forest wood harvest.

The third largest current day wood harvest biomass amount is Southern Asia at 0.18 PgC/yr with a wood harvest area of 0.08 million km^2. The wood harvest area of Southern Asia is made up of large areas of non-forest and much smaller areas of forest. The fourth larges wood harvest biomass amount is North America with 0.18 PgC/yr, but with a wood harvest area of only 0.04 million km^2. The North America wood harvest was split between Primary Forest and Secondary Mature Forest with a small contribution from Secondary Young Forest. Other regions with substantial wood harvest biomass amounts were Europe with 0.16 PgC/yr, East Asia with 0.14 PgC/yr, Southeast Asia with 0.13 PgC/yr, and Eurasia with 0.09 PgC/yr.

The global map of current day LUH2 wood harvest biomass in Figure 5.13a reflects the regional analysis, with six main areas with high wood harvest and large expanses of low biomass extraction. The individual wood harvest biomass maps of Figures 5.13 and 5.14 help explain what is contributing to each of these areas. The first area of current day high wood harvest biomass is the southeast of the USA which is a combination of Primary Forest and Secondary Mature Forest wood harvest. The next area is southeastern Amazon which is all Primary Forest wood harvest. The next area is central Europe into Scandinavia, which is a combination of Secondary Mature Forest, Secondary Young Forest and Secondary Non Forest.

The next obvious area is central Africa which is a combination of Secondary Young Forest and Secondary Non Forest with smaller contributions from the other classes. The area after that is northern India into China, which is predominantly Secondary Young Forest wood harvest in India and Secondary Mature Forest wood harvest in China. The last major area is of current day wood harvest is Southeast Asia which is predominantly Primary Forest. Other widespread low intensity

wood harvest occurs across the Boreal Forest and over the savannas of Africa and the Middle East.

5.7 All CMIP6 LUH2 Time Series

Beyond the CMIP6 Historical period of 1850 to 2014, additional climate modeling simulations have been developed for the Transient (mid) Holocene from 6000bce to 850ce, and the Last Millennium from 850 to 1850 under the PMIP4 project. The Last Millennium simulations overlap with the LUH2 historical reconstruction and so are directly used for CLM5 land use data for that period. The Transient (mid) Holocene however covers the period before the LUH2 data. To address this for CLM5, an annual land use time series is generated using the first 10 years of the LUH2 historical time series (850 – 859 CE) combined with the HYDE 3.2 population data for the same period. The population data for the Holocene prior to the LUH2 data is then used to scale land use for each grid based on the population of that year relative to the beginning of the LUH2 time period. The population and land use data are highly consistent with LUH2 given that the HYDE 3.2 time series is a primary component of the historical reconstruction for this time (Hurtt et al. 2020).

Continuing from the end of the CMIP6 Historical period there are a range of future SSPs generated by ScenarioMIP as evolutions of the global and regional energy system consistent with the SSP narratives and their associated challenges for mitigation and adaptation. The SSPs depict vastly different energy futures, featuring a wide range of possible energy demand developments and energy supply structures. These differences emerge due to a combination of assumptions with respect to the main drivers of the energy system, including technological change, economic growth, population, emergence of new energy services, energy intensity of services, and assumptions with respect to costs and availability of future fossil fuel resources and their alternatives (Riahi et al. 2017) and (O'Neill et al. 2016).

As part of the scenario generation process the IAMs produce time series of land use for each of the SSPs as described in Popp et al. (2017). The SSP land use output from each of the IAMs have different starting points depending on the underlying data assumptions of the models. To prevent these differences impacting the continuity and consistency of the individual SSPs relative to each other and the historical reconstruction, a process of harmonization was performed for each SSP against the historical time series for the years 2015 and continuing through the time series to 2100 as described in (Hurtt et al. 2020). The harmonization process aims to preserve the land use changes dictated by the IAMs without introducing other inconsistencies from the underlying modeling process.

The SSP1-2.6 and SSP2-4.5 time series both have reforestation and afforestation components generated by the IAMs that result in additional forest areas that cannot be supported by the GLM2 model. To address this the LUH2 time series data includes the additional annual variable of Added Tree Cover to represent the forest gain not captured by the GLM2 model. For the CLM5 land data generation process the annual Added Tree Cover components are added to Secondary Forest so that they can represent the reforestation and afforestation through increases in the CLM5 tree PFT percentages. The added Secondary Forest is taken from the available annual Primary Non Forest and Secondary Non Forest land use states.

The change in time series values for each CMIP6 period have been calculated globally, with the change in LUH2 land use states shown in Table 5.10, the change in shifting cultivation shown in Table 5.11, the change in crop management in Table 5.12, and the change in wood harvest area

and biomass amount shown in Table 5.13. All of the global annual LUH2 time series values for the combined Forest, Non Forest, and Crop states are plotted in Figure 5.15. All of the global annual LUH2 time series values for Pasture, Rangeland and Urban states are plotted in Figure 5.16. All of the global annual LUH2 time series values for Shifting Cultivation area, Irrigated Crop area, and Nitrogen Fertilizer amount are plotted in Figure 5.17. Finally, all of the global annual LUH2 time series values for Wood Harvest area and biomass amount are plotted in Figure 5.18. Descriptions of each of the time series data are compiled from the tables and plots for the remainder of this chapter as a reference for their use as the basis of the transient CLM5 land use data generated in Chapters 6, 7 and 8.

Table 5.10: Change in LUH2 States over All CMIP6 Time Series in Millions of km^2.

	Trans Hol	Last Mil	Historical	SSP1	SSP2	SSP3	SSP4	SSP5
forest	-0.90	-3.43	-7.06	7.18	2.54	-3.41	-1.41	-0.94
non forest	-3.91	-9.22	-24.07	-3.25	-2.52	-4.27	-4.53	0.14
grazing	3.13	8.62	20.73	-6.33	-4.45	1.61	2.73	-1.91
crop	1.68	4.00	9.91	1.95	3.93	5.64	2.70	2.07
primf	-1.30	-7.96	-16.85	-5.57	-7.68	-8.29	-8.81	-7.93
primn	-4.61	-20.13	-27.72	-5.53	-6.20	-7.08	-7.47	-4.87
secdf	0.39	4.53	9.78	12.75	10.22	4.89	7.40	6.99
secdn	0.69	10.91	3.65	2.28	3.69	2.81	2.94	5.01
c3ann	1.01	2.53	5.00	-0.74	1.72	1.81	-0.31	0.36
c4ann	0.27	0.69	2.06	-0.46	0.97	2.12	0.44	1.15
c3per	0.16	0.27	0.98	1.30	0.52	0.68	0.33	0.12
c4per	0.02	0.03	0.20	2.14	0.08	0.08	2.25	0.06
c3nfx	0.22	0.48	1.67	-0.28	0.64	0.95	-0.01	0.38
pastr	0.38	2.15	5.36	-3.61	-1.77	0.94	1.74	-0.89
range	2.75	6.47	15.37	-2.72	-2.68	0.68	0.99	-1.02
urban	0.00	0.03	0.50	0.45	0.50	0.43	0.51	0.65

Transient Holocene 6000bce to 850ce: The almost 7000 years before the historical LUH2 data is characterized by much lower global and regional populations in the HYDE 3.2 data, with substantially lower areas of cropping and grazing. At the beginning of the period LUH2 Forest areas are near their natural extent of 48.24 million km^2 and Non Forest areas are at 80.02 million km^2. Over the period grazing increases by 3.13 million km^2 through increases in Pasture of 0.38 million km^2 and Rangeland of 2.75 million km^2. Cropping increases by 1.68 million km^2 mostly through increases in the C3 Annual Crop, but other crop also increased from their near zero values. The increases in land use over the period resulted in a loss of Forest area of -0.90 million km^2, leaving a Forest area of 47.33 million km^2 in 850. The increase in land use also created a loss of Non Forest area of -3.91 million km^2, resulting in Non Forest area of 76.11 million km^2 in 850. Shifting Cultivation increased over the period from zero to the 0.07 million km^2 found at the beginning of the LUH2 Historical data. Irrigation increased over the period to 0.04 million km^2 although this was less than 3% of all crops. Wood harvest increased in area from zero to 0.03 million km^2/yr with biomass extraction increasing to 0.01 PgC/yr.

Table 5.11: Change in LUH2 Shifting Cultivation over All CMIP6 Time Series in Millions of km^2.

	Trans Hol	Last Mil	Historical	SSP1	SSP2	SSP3	SSP4	SSP5
all	0.07	0.12	0.09	-0.12	-0.13	-0.15	-0.20	-0.16
forest	0.04	0.07	0.06	-0.10	-0.09	-0.10	-0.12	-0.12
non forest	0.03	0.05	0.03	-0.03	-0.04	-0.05	-0.08	-0.04

Last Millennium 850 to 1850: The 1000 years of the Last Millennium had a substantial rise in global LUH2 grazing area of 8.62 million km^2, resulting in a grazing area of 11.75 million km^2 in 1850. The period also had a large increase in crop area of 4.00 million km^2, resulting in a crop area of 5.69 million km^2 in 1850. The increase in grazing came through an increase in Pasture of 2.15 million km^2 and Rangeland of 6.47 million km^2. The increase in crop area came primarily through an increase in C3 Annual Crop but there were increases in other crops as well. The increases in land use caused a global loss of Forest area of -3.43 million km^2, resulting in a Forest area of 43.90 million km^2 in 1850. There also was a loss of Non Forest area of -9.22 million km^2, resulting in a Non Forest area of 66.89 km^2 in 1850. Shifting Cultivation increased over the period by 0.12 million km^2, to finish at 0.19 million km^2 in 1850. Irrigation increased by 0.24 million km^2 over the period, resulting in an irrigated area of 0.28 million km^2 in 1850. This was still only 3% of all crops in 1850. Global wood harvest increased over the period by 0.21 million km^2/yr and 0.24 PgC/yr, resulting in a wood harvest rate of 0.24 million km^2/yr and 0.26 PgC/yr in 1850.

Table 5.12: Change in LUH2 Crop Management over All CMIP6 Time Series. Irrigation in Millions of km^2 and Fertilizer in TgN/yr.

	Trans Hol	Last Mil	Historical	SSP1	SSP2	SSP3	SSP4	SSP5
irrigation	0.04	0.24	2.39	0.16	-0.09	1.38	0.02	0.68
fertilizer	0.00	0.00	108.73	60.89	94.33	59.44	31.92	-10.31
irrig_c3ann	0.03	0.16	1.48	-0.10	-0.06	0.56	-0.17	0.25
irrig_c4ann	0.00	0.03	0.34	-0.01	0.00	0.48	-0.03	0.26
irrig_c3per	0.00	0.02	0.21	0.12	-0.02	0.10	-0.01	0.05
irrig_c4per	0.00	0.00	0.06	0.14	-0.01	0.01	0.27	0.00
irrig_c3nfx	0.01	0.03	0.30	0.00	-0.01	0.22	-0.03	0.13
fertl_c3ann	0.00	0.00	66.58	10.71	44.66	32.02	2.86	-3.93
fertl_c4ann	0.00	0.00	23.56	1.33	37.03	13.53	3.53	-2.45
fertl_c3per	0.00	0.00	9.88	19.11	1.68	7.22	3.25	-2.64
fertl_c4per	0.00	0.00	3.12	25.78	6.44	2.34	15.25	-1.02
fertl_c3nfx	0.00	0.00	5.59	3.96	4.52	4.34	7.03	-0.28

Historical 1850 to 2014: The 165 years of the Historical period saw a rapid increase in land use with an increase in global grazing area of 20.73 million km^2 and an increase of crop area of 9.91 million km^2. The increase in grazing resulted in a global grazing area of 32.47 million km^2 in 2014, while the increase in cropping resulted in a crop area 15.59 million km^2. The increase in land use caused a loss of Forest area of -7.06 million km^2 over the period, resulting in a Forest area of 36.84 million km^2 in 2014. The increase in land use also caused a loss of Non Forest area of -24.07 million km^2, resulting in a Non Forest area 42.82 million km^2 in 2014. Global Shifting Cultivation increased over the period by 0.12 million km^2, resulting in an area of Shifting Cultivation of 0.28 million km^2 in 2014.

Crop management also increased rapidly over the Historical period with nitrogen fertilizer application increasing from zero over the last 60 years of the Historical period to a global rate of 108.73 TgN/yr. This represented an average fertilizer application rate of 69.47 kgN/ha in 2014. Irrigation also increased rapidly adding 2.39 million km^2 over the period, resulting in 2.66 million km^2 of irrigated crop area in 2014. This also increased the fraction of irrigated crops to 11% of all crops in 2014. Wood harvest biomass increased by 1.12 PgC/yr over the Historical period to have rate of 1.38 PgC/yr in 2014. The wood harvest area changes were more complex as the area of Forest wood harvest increased continuously through the period, while the area of Non Forest decreased over the first 100 years before increasing rapidly over the last 65 years. Total wood

harvest area ultimately increased by 0.60 million km^2/yr, resulting in a final value of 0.85 million km^2/yr in 2014.

SSP1-2.6 Sustainability (Taking the Green Road) 2015 to 2100: This scenario was modeled for the CMIP6 ScenarioMIP using the IMAGE 3.0 IAM (Stehfest et al. 2014). In this future scenario the world shifts gradually, but pervasively, toward a more sustainable path, emphasizing more inclusive development that respects perceived environmental boundaries. Land use is strongly regulated, e.g. tropical deforestation rates are strongly reduced. Crop yields are rapidly increasing in low- and medium-income regions, leading to a faster catching-up with high income countries. Healthy diets with low animal-calorie shares and low waste prevail. In an open, globalized economy, food is traded internationally. In SSP1-2.6, international cooperation for climate change mitigation starts early (after 2020). All land use emissions are priced at the level of carbon prices in the energy sector.

In response to the SSP1-2.6 narrative the LUH2 data has the largest scenario increase in Forest area of 7.18 million km^2, resulting in a Forest area of 44.02 million km^2 in 2100. Most of the increase in Forest area comes from reforestation and afforestation associated with the Added Tree Cover variable. The dotted line of SSP1-2.6 in the Forest area plot of Figure 5.15b shows that without the Added Tree Cover variable the change in Forest area would have been close to zero. The increase in Forest is accompanied by the largest decrease in grazing of -6.33 million km^2, which results in a grazing area of only 26.14 million km^2 by 2100. There is also the smallest scenario increase in Crop area of 1.95 million km^2, which results in a Crop area of 17.54 million km^2 in 2100. To allow for the increases in both Forest and Crop there is a decrease in Non Forest area of -3.25 million km^2, which resulted in a Non Forest area of 39.57 million km^2 in 2100.

Table 5.13: Change in LUH2 Wood Harvest over All CMIP6 Time Series. Harvest Area in Millions of km^2/yr and Biomass Harvest Amount in PgC/yr.

	Trans Hol	Last Mil	Historical	SSP1	SSP2	SSP3	SSP4	SSP5
harvest area	0.03	0.21	0.60	-0.22	0.66	2.35	2.65	1.99
harv amount	0.01	0.24	1.12	-0.48	0.13	0.16	0.49	0.31
primf_harv	0.00	0.02	0.05	-0.03	0.01	-0.02	0.00	0.00
primn_harv	0.03	-0.03	0.11	-0.07	-0.07	-0.05	-0.07	-0.06
secmf_harv	0.00	0.00	0.03	0.00	0.03	0.02	0.05	0.03
secyf_harv	0.00	0.00	0.03	0.00	0.03	0.02	0.05	0.03
secnf_harv	0.00	0.21	0.39	-0.12	0.67	2.39	2.61	1.99
primf_bioh	0.01	0.14	0.49	-0.34	-0.12	-0.25	-0.13	-0.10
primn_bioh	0.01	0.00	0.07	-0.04	-0.04	-0.02	-0.03	-0.03
secmf_bioh	0.00	0.04	0.21	0.00	0.22	0.12	0.46	0.27
secyf_bioh	0.00	0.03	0.22	-0.05	-0.01	0.09	0.06	-0.01
secnf_bioh	0.00	0.04	0.12	-0.05	0.07	0.22	0.13	0.18

The Shifting Cultivation area of SSP1-2.6 decreased by -0.24 million km^2, resulting in a Shifting Cultivation area of 0.31 million km^2 in 2100. All SSPs had similar decreases in Shifting Cultivation. The crop management also increased over the period with a total global fertilizer amount of 169.62 TgN/yr, and average application rate of 96.70 kgN/ha. This is the second highest fertilizer usage of the scenarios. The global irrigated area increased by 0.16 million km^2, resulting in an irrigated crop area of 2.82 million km^2 in 2100. This was the middle of the scenario changes in irrigation leaving 9% of all crops irrigated in 2100. SSP1-2.6 was the only scenario where wood harvest was reduced. The wood harvest area was reduced by -0.22 million km^2/yr, resulting in a wood

harvest area of 0.62 million km^2/yr in 2100. The wood harvest biomass amount was also reduced by -0.48 PgC/yr, resulting in a wood harvest amount of 0.90 PgC/yr in 2100.

SSP2-4.5 Middle of the Road 2015 to 2100: This scenario was modeled for the CMIP6 ScenarioMIP using the MESSAGE IAM (Riahi et al. 2007; Fricko et al 2017). In this future scenario the world follows a path in which social, economic, and technological trends do not shift markedly from historical patterns. Land use change is incompletely regulated, i.e. tropical deforestation continues, although at slowly declining rates over time. Rates of crop yield increase decline slowly over time, but low-income regions catch up to a certain extent. Caloric consumption and animal calorie shares converge slowly towards high levels. International trade remains to large extent regionalized. In SSP2-4.5, international cooperation for climate change mitigation is delayed due to a transition phase to a uniform carbon price until 2040. In this transition phase, emissions from agricultural production are priced at the level of energy sector emissions, while avoided deforestation and afforestation are not incentivized before 2030.

Following the SSP2-4.5 narrative the LUH2 data has the second largest scenario increase in Forest area of 2.54 million km^2, resulting in a Forest area of 28.02 million km^2 in 2100. Like SSP1-2.6, the increase in Forest area comes from the reforestation and afforestation associated with the Added Tree Cover variable, with the dotted line for SSP2-4.5 without the Added Tree Cover in Figure 5.15b having close to zero change in Forest area. The increase in Forest is accompanied by the second biggest decrease in global grazing area of -4.45 million km^2, resulting in a grazing area of 28.02 million km^2 in 2100. There is also the second largest of the scenarios increase in Crop area of 3.93 million km^2, resulting in a Crop area of 19.52 million km^2 in 2100. The increases in Forest and Crop area require a decrease in Non Forest area of -2.52 million km^2, which results in a Non Forest area of 40.30 million km^2 in 2100.

The SSP2-4.5 Shifting Cultivation area decreased by -0.26 million km^2, resulting in a Shifting Cultivation area of 0.28 million km^2 in 2100. The scenario had the largest fertilizer application rate increase over the period with a total global fertilizer amount of 203.06 TgN/yr, at an average application rate of 104.01 kgN/ha in 2100. The global irrigated area decreased by -0.09 million km^2, resulting in an irrigated crop area of 2.57 million km^2 in 2100. This was the only scenario to reduce irrigation leaving only 8% of crops irrigated in 2100. SSP2-4.5 had an increase in wood harvest area of 0.66 million km^2/yr, resulting in a wood harvest area of 1.51 million km^2/yr in 2100. The wood harvest biomass amount also increased by 0.13 PgC/yr, resulting in a wood harvest amount of 1.50 PgC/yr in 2100. This was the second lowest wood harvest scenario with only SSP1-2.6 being lower.

SSP3-7.0 Regional Rivalry (A Rocky Road) 2015 to 2100: This scenario was modeled for the CMIP6 ScenarioMIP using the AIM IAM (Fujimori et al. 2017). In this future scenario resurgent nationalism, concerns about competitiveness and security, and regional conflicts push countries to increasingly focus on domestic or, at most, regional issues, including food and energy security. Land use change is hardly regulated. Rates of crop yield increase decline strongly over time, especially due to very limited transfer of new agricultural technologies to developing countries. Unhealthy diets with high animal shares and high food waste prevail. A regionalized world leads to reduced trade flows for agricultural goods. In SSP3-7.0, forest mitigation activities and abatement of agricultural GHG emissions are limited due to major implementation barriers such as low institutional capacites in developing countries. In addition, they are delayed as a consequence of low international cooperation. In 2020, high income countries start the transition to a unifom carbon price until 2040, whereas low-income countries start in 2030 and converge until 2050.
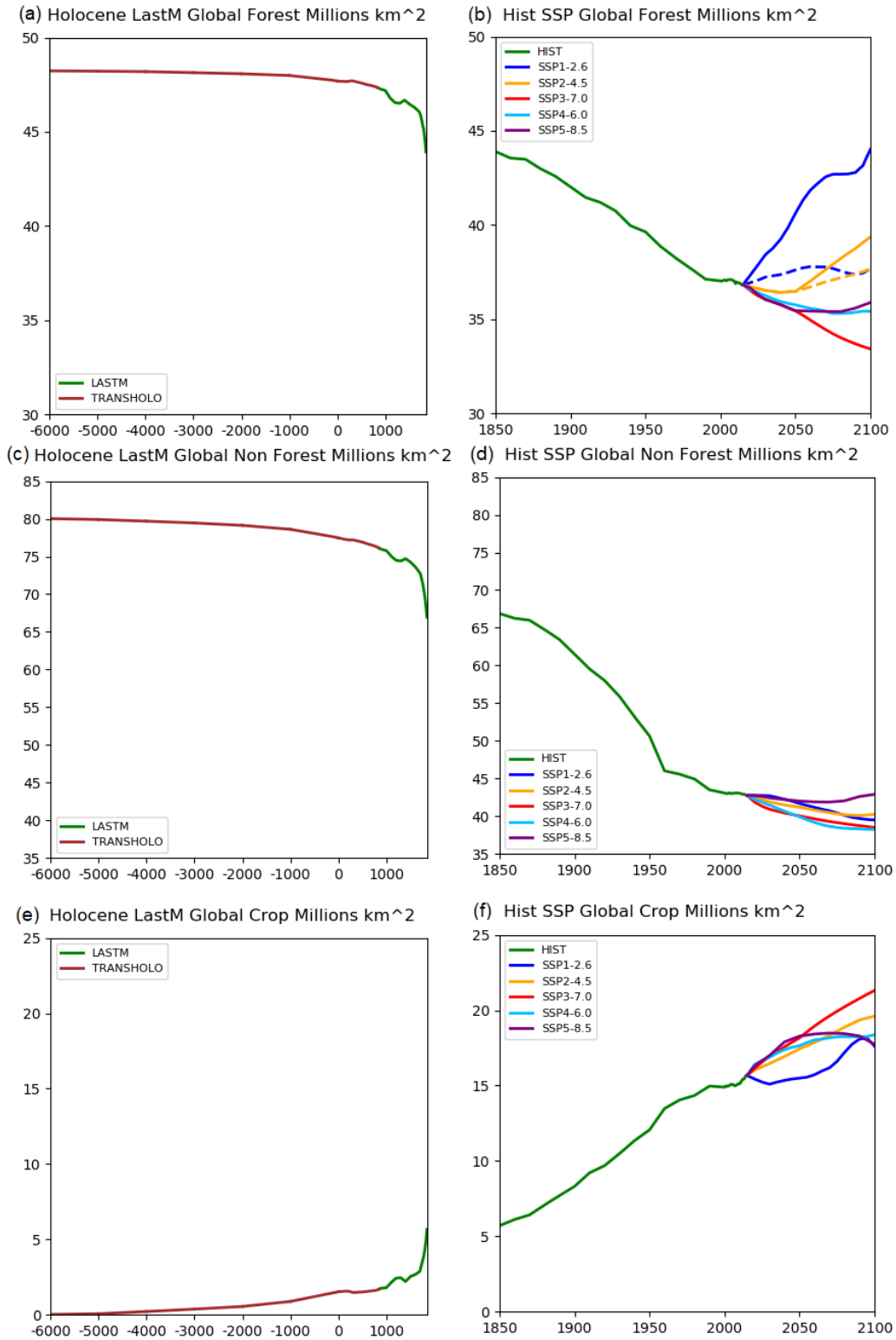
Figure 5.15: All CMIP6 LUH2 Timeseries Global Annual Land Use States for: (a and b) All Forests; (c and d) All Non Forests; and (e and f) All Crops.
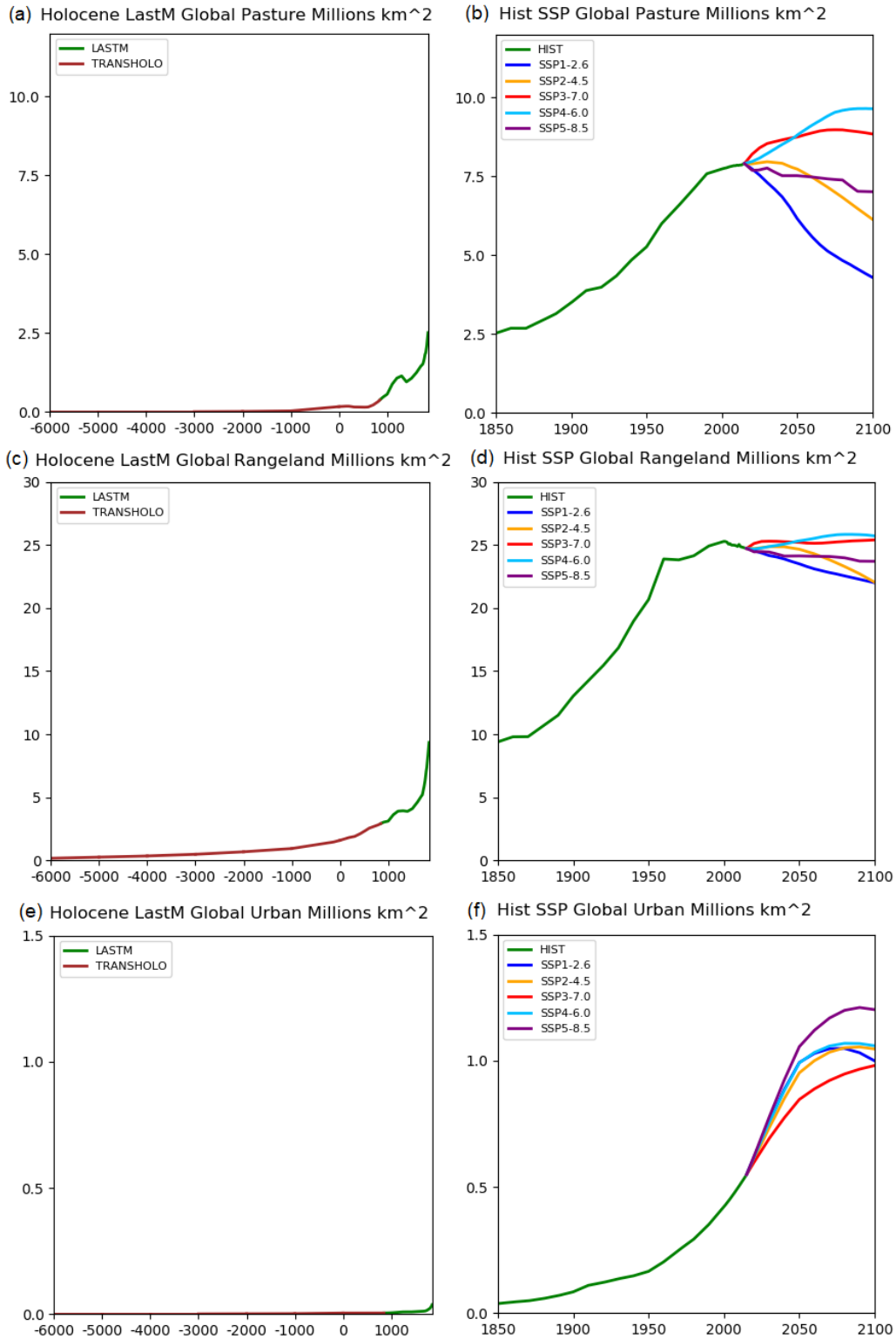
Figure 5.16: All CMIP6 LUH2 Timeseries Global Annual Land Use States for: (a and b) Pasture; (c and d) Rangeland; and (e and f) Urban.
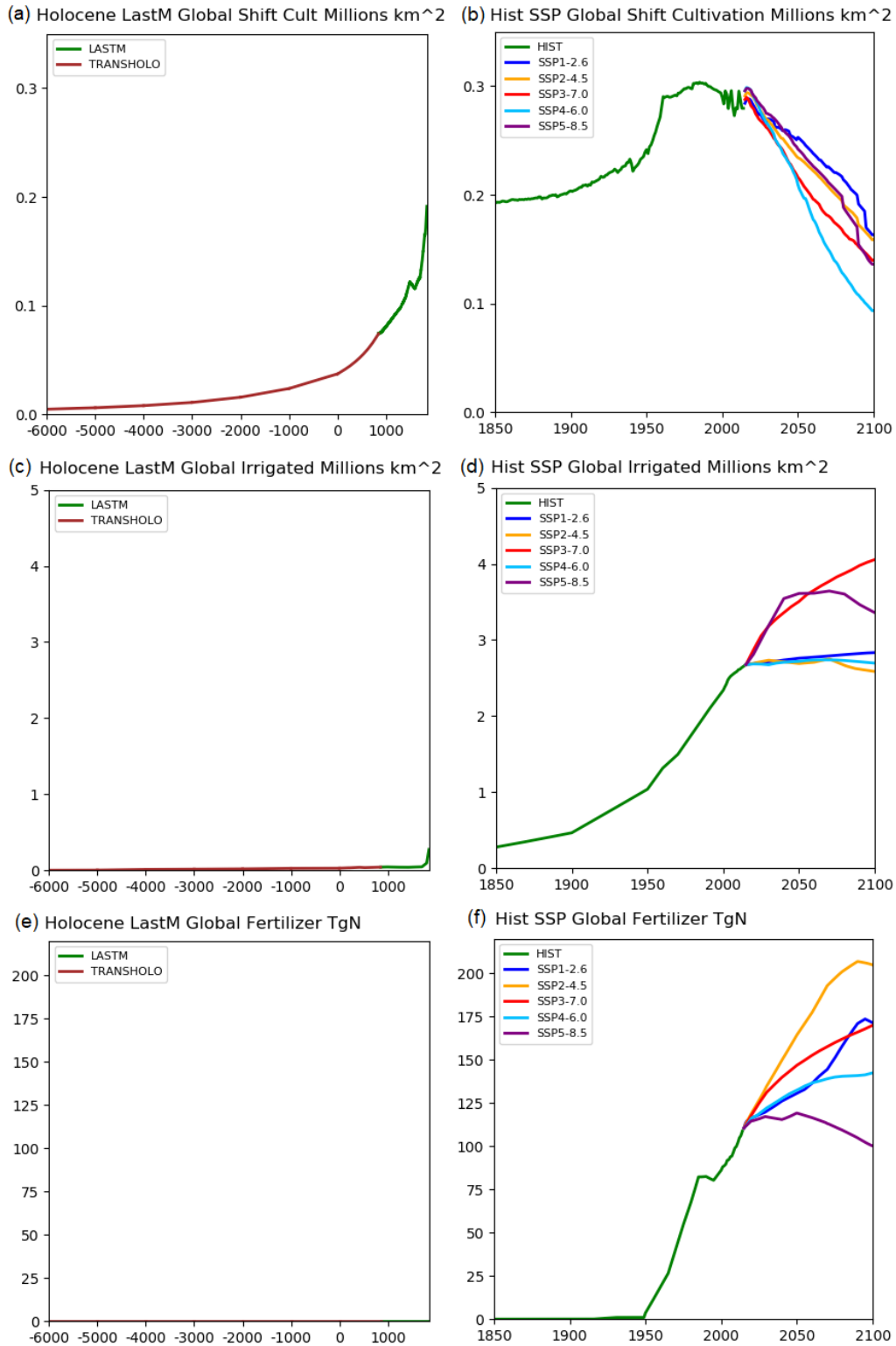
Figure 5.17: All CMIP6 LUH2 Timeseries Global Annual Agricultural Management for: (a and b) Shifting Cultivation; (c and d) Irrigated Crop Area; and (e and f) Nitrogen Fertilizer Use.
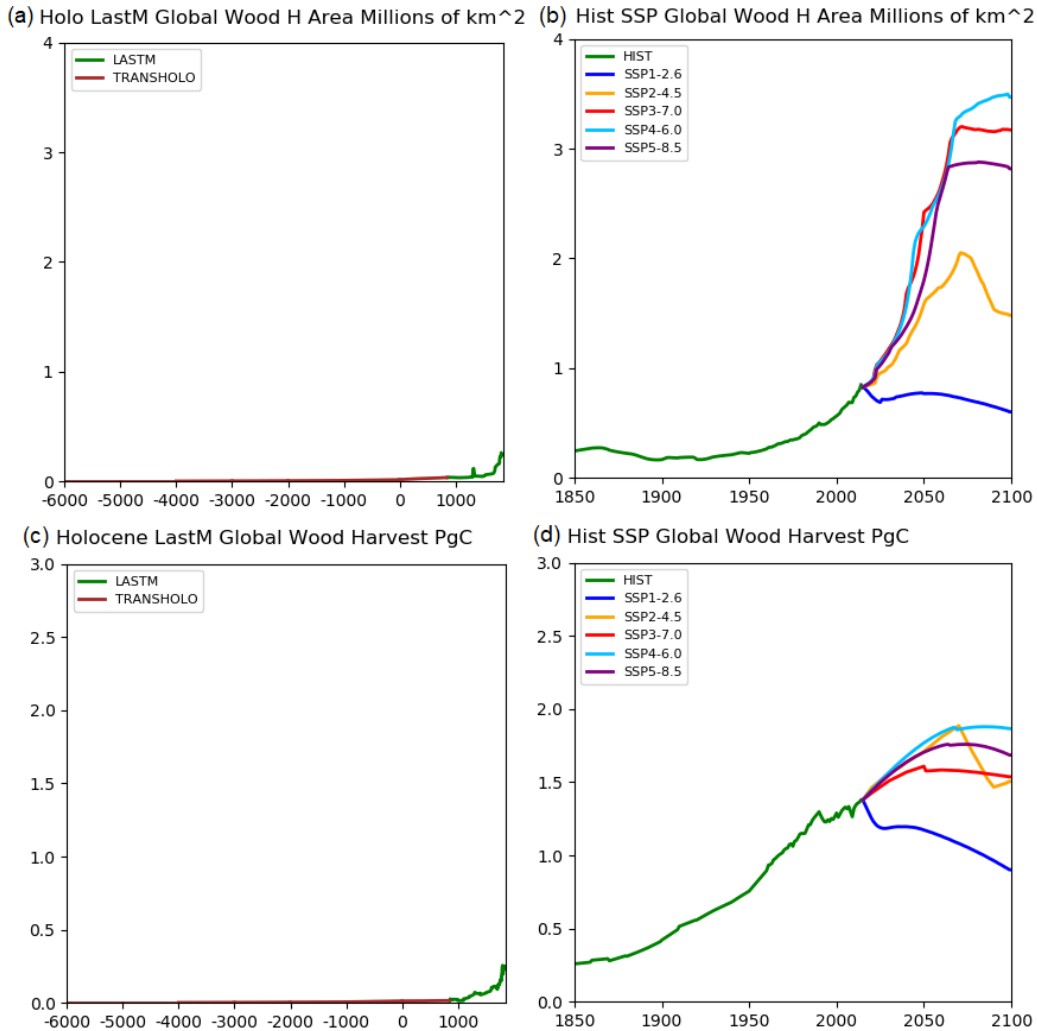
Figure 5.18: All CMIP6 LUH2 Timeseries Global Annual Forest Management for: (a and b) Wood Harvest Area; and (c and d) Wood Harvest Biomass Amount.

Due to the lack of land use regulation and declining crop yields the SSP3-7.0 LUH2 data has the largest scenario increase in Crop area with an addition of 5.64 million km^2, resulting in a Crop area of 21.23 million km^2 in 2100. The scenario also had the second largest increase in grazing of 1.61 million km^2, resulting in a grazing area of 34.09 million km^2 in 2100. The large increases in Crop and grazing area are accompanied by the largest scenario decrease in Forest area of -3.41 million km^2, resulting in a Forest area of only 33.43 million km^2 in 2100. The scenario also had the second largest decrease in Non Forest area of -4.27 million km^2, resulting in a Non Forest area of 38.54 million km^2 in 2100.

The SSP3-7.0 Shifting Cultivation area decreased by -0.30 million km^2, resulting in a Shifting Cultivation area of 0.25 million km^2 in 2100. The scenario had an increase in total global fertilizer amount to 168.18 TgN/yr, at an average application rate of 79.23 kgN/ha in 2100. The scenario had the largest increase in irrigation area of 1.38 million km^2, resulting in an irrigated crop area of 4.04 million km^2 in 2100. The large increase in irrigated area was offset by the large increase in total crop area so that only 10% of crops were irrigated in 2100. SSP3-7.0 had the second largest

increase in wood harvest area of 2.35 million km^2/yr, resulting in a wood harvest area of 3.20 million km^2/yr in 2100. The wood harvest biomass amount also increased by the third largest amount of 0.13 PgC/yr, resulting in a wood harvest amount of 1.54 PgC/yr in 2100.

SSP4-6.0 Inequality (A Road Divided) 2015 to 2100: This scenario was modeled for the CMIP6 ScenarioMIP using the GCAM IAM (Wise et al. 2014). In this future scenario highly unequal investments in human capital, combined with increasing disparities in economic opportunity and political power, lead to increasing inequalities and stratification both across and within countries. Land use change is strongly regulated in high income countries, but tropical deforestation still occurs in poor countries. High income countries achieve high crop yield increases, while low-income countries remain relatively unproductive in agriculture. Caloric consumption and animal calorie shares converge towards medium levels. Food trade is globalized, but access to markets is limited in poor countries, increasing vulnerability for non-connected population groups. In SSP4-6.0, international cooperation for climate change mitigation starts early (after 2020). But emissions from agricultural and land use are incompletely priced, with limited incentives for avoided deforestation and afforestation before 2030.

The ineffective land use emission pricing results in the SSP4-6.0 LUH2 data having the largest scenario increase in grazing area of 2.73 million km^2, with a global grazing area of 35.20 million km^2 in 2100. The ineffective land use controls also lead to the third largest increase in Crop area of 2.70 million km^2, resulting in a Crop area of 35.20 million km^2 in 2100. The increases in grazing and Crop area require the largest decrease in Non Forest area of -4.53 million km^2 and the second largest decrease in Forest area of -1.41 million km^2. These decreases leave a Non Forest area of 38.29 million km^2 and a Forest area of 35.43 in million km^2 2100.

The SSP4-6.0 has the largest decrease in Shifting Cultivation area of -0.40 million km^2, resulting in the lowest Shifting Cultivation area of 0.14 million km^2 in 2100. The scenario has the second lowest fertilizer increase with a total global fertilizer amount to 140.65 TgN/yr, at an average application rate of 76.90 kgN/ha in 2100. The scenario also has a very small increase in irrigation area of only 0.02 million km^2, leaving irrigation nearly unchanged at 2.68 million km^2 in 2100. The lack of increase in irrigated area results in only 8% of crops being irrigated in 2100. The scenario also has the largest increase in wood harvest. The wood harvest area increases by 2.65 million km^2/yr, resulting in a wood harvest area of 3.49 million km^2/yr in 2100. The wood harvest biomass amount also increases by 0.49 PgC/yr, resulting in a wood harvest amount of 1.87 PgC/yr in 2100.

SSP5-8.5 Fossil-Fueled Development (Taking the Highway) 2015 to 2100: This scenario was modeled for the CMIP6 ScenarioMIP using the REMIND-MagPIE IAM (Kriegler et al. 2017). In this future scenario driven by the economic success of industrialized and emerging economies, this world places increasing faith in competitive markets, innovation and participatory societies to produce rapid technological progress and development of human capital as the path to sustainable development. Land use change is incompletely regulated, i.e. tropical deforestation continues, although at slowly declining rates over time. Crop yields are rapidly increasing. Unhealthy diets with high animal shares and high waste prevail. Barriers to international trade are strongly reduced, and strong globalization leads to high levels of international trade. In SSP5-8.5, all land use emissions are priced at the level of carbon prices in the energy sector. But in contrast to SSP1-2.6, international cooperation for climate change mitigation is delayed due to a transition phase to a uniform carbon price until 2040.

The slowing rates of tropical deforestation and the rising crop yields result in the SSP5-8.5 LUH2 data having the second smallest increase in Crop area of 2.70 million km^2, with a Crop area of 17.66 million km^2 in 2100. The land use changes and crop yield rises of the scenario also result

in a reduction of grazing area of -1.91 million km^2, with a global grazing area of 35.89 million km^2 in 2100. The increase in Crop area and decrease in grazing area had mixed results on the other land use classes with a relatively small decrease in Forest area of -0.94 million km^2 and an even smaller increase in Non Forest area of 0.14 million km^2. The net result of these changes was a Forest area of 35.89 million km^2 and a Non Forest area of 42.95 million km^2 in 2100.

The SSP5-8.5 Shifting Cultivation decreased by -0.32 million km^2, resulting in a Shifting Cultivation area of 0.23 million km^2 in 2100. The scenario also had the only decrease in fertilizer use with total global fertilizer amount reduced to 98.42 TgN/yr and an average application rate of 55.74 kgN/ha in 2100. For the decrease in fertilizer application rate to be consistent with the scenario narrative of increasing crop yields the REMIND-MagPIE IAM needed to achieve these crop yields through other technology change assumptions not included in the LUH2 data. The global irrigation area of the scenario increased by 0.68 million km^2, resulting in an irrigated crop area of 3.34 million km^2 in 2100. The increase in irrigated area combined with the increase in total crop area to maintain the current day 11% of crops being irrigated in 2100. SSP5-8.5 had the third largest increase in wood harvest area of 1.99 million km^2/yr, resulting in a wood harvest area of 2.84 million km^2/yr in 2100. The wood harvest biomass amount also increased by the second largest amount of 0.49 PgC/yr, resulting in a wood harvest amount of 2.84 PgC/yr in 2100.

CHAPTER 6.

CLM5 HISTORICAL TIME SERIES

6.1 CLM5 Historical Time Series Overview

Annual CLM5 Land Unit, Plant Functional Type (PFT), Crop Functional Type (CFT), PFT Shifting Cultivation, CFT Irrigation and Fertilizer, and PFT Wood Harvest were generated using the CLM5 Land Use Data Tool as described in Chapter 8. The CLM5 Land Use Data Tool combines the LUH2 annual land cover and land use data described in Chapter 5, with the current day distributions of Land Units and PFTs described in Chapters 2. The tool also combines the current day 1961 – 2015 annual distributions of CFTs developed from EarthSTAT and FAOSTAT data in Chapter 3. The MksurfData tool described in Chapter 8, combines the annual PFT and CFT distributions along with other current day data described in Chapter 4 to produce final CLM5 surface data and land use time series data for each LUH2 time series at a specified model resolution and domain extent.

This chapter describes in detail the CLM5 Historical period from 850 – 2014ce for the large quantity of data generated for each year by the CLM5 Land Use Data Tool as described in Chapter 8. The analyses are performed globally and regionally at the 0.25 degree resolution of the LUH2 data as well as at the finite volume 0.9x1.25 degree resolution of the default CESM model. The Historical period corresponds with the detailed description of the LUH2 land use data for the same period in Chapter 5. Given the large number of other time series listed in Chapter 5, higher level descriptions are provided for all the other land use time series data of the CMIP6 LUMIP and PaleoMIP projects in Chapter 7. These other time series cover annual data for 6000 bce to 2014, and for then five baseline Shared Socio-economic Pathways (SSPs) for 2015 to 2100.

6.2 CLM5 Land Unit and Plant Functional Type (PFT) Historical Data

The global area for each of the annual Land Unit and PFT distributions generated by the CLM5 Land Use Data Tool is shown for 850 and 2014 in Figure 6.1. The difference in PFT and Crop Land Unit area, as well as the annual PFT and Crop component areas over the 850 to 2014 time period are shown in Figure 6.2. The global and IPCC regional area for each Land Unit and PFT for 850 and 2014 are listed in Table 6.1. The surface data area for Glaciers, Lakes, and Urban are kept constant for all time series at their current day values as there is no capacity to have these values prescribed transiently in CLM5. The current day 2014 global raw maps for each of the CLM5 PFT and Crop Land Unit components are shown at 0.25 degrees resolution in Figures 6.3, 6.4, 6.5, 6.6, and 6.7.

The total global land area for the CLM5 land surface data with the finite volume 0.9x1.25 degree resolution domain file is 148.7 million km^2. This is consistent with the MODIS 1 km land cover data set at 148.1 million km^2. The global CLM5 Glacier Land Unit area for the land surface data was 15.6 million km^2, which was slightly lower than the CISM2 Glacier extent at 16.2 million km^2,

and higher than the MODIS Land Cover Glacier at 15.3 million km^2. The global CLM5 Lake Land Unit area was 2.7 million km^2, which was lower than both the GLWD data is at 2.9 million km^2, and the MODIS Land Cover Lake at 2.8 million km^2. The global CLM5 Urban Land Unit area was 0.8 million km^2, which was consistent with the Urban mapping of Jackson et al. (2010), and higher than the MODIS Land Cover Urban at 0.5 million km^2. The prescribed areas of Glacier, Lake and Urban Land Units are fixed for all time periods so real-world changes in these Land Units are not captured in the CLM5 land surface data.

Table 6.1 Annual Global and Regional CLM5 Land and Plant Functional Type (PFT) Area for 850 and 2014 in Millions of km^2, using the default CLM5 Domain File Land Mask.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Land	148.7	30.0	8.4	11.4	21.7	6.6	20.5	5.3	21.8	4.8	5.2	12.9
Glacier	15.6	0.0	0.0	0.1	0.1	0.0	0.0	0.0	2.0	0.0	0.0	12.8
Lake	2.7	0.3	0.0	0.1	0.6	0.1	0.2	0.1	1.1	0.0	0.0	0.0
Urban	0.8	0.1	0.0	0.2	0.0	0.1	0.1	0.0	0.0	0.1	0.2	0.0
Year 2014												
Tree	33.8	4.3	1.1	2.0	6.6	1.8	9.0	0.0	5.7	2.8	0.7	0.0
Shrub	9.5	1.1	2.0	0.0	3.3	0.1	0.9	0.1	1.7	0.0	0.1	0.0
Grass	44.5	10.6	2.8	5.0	7.6	2.6	7.0	0.5	6.6	0.9	1.2	0.0
Bare	27.0	10.9	1.9	2.9	1.8	0.1	1.3	4.2	2.6	0.0	1.0	0.1
NdlEvgTemp	4.8	0.0	0.2	0.8	1.1	0.8	0.0	0.0	1.7	0.0	0.1	0.0
NdlEvgBorl	6.0	0.0	0.0	0.3	2.5	0.4	0.0	0.0	2.8	0.0	0.0	0.0
NdlDecBorl	2.1	0.0	0.0	0.1	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BrdEvgTrop	14.0	3.2	0.2	0.0	0.0	0.0	7.7	0.0	0.0	2.6	0.3	0.0
BrdEvgTemp	1.2	0.1	0.4	0.1	0.0	0.0	0.4	0.0	0.0	0.1	0.1	0.0
BrdDecTrop	1.9	1.0	0.1	0.0	0.0	0.0	0.6	0.0	0.0	0.1	0.1	0.0
BrdDecTemp	2.8	0.0	0.2	0.4	0.4	0.5	0.2	0.0	0.9	0.0	0.1	0.0
BrdDecBorl	1.0	0.0	0.0	0.2	0.5	0.0	0.1	0.0	0.2	0.0	0.0	0.0
ShrEvgTemp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ShrDecTemp	4.5	1.1	2.0	0.0	0.1	0.0	0.8	0.1	0.3	0.0	0.1	0.0
ShrDecBorl	4.9	0.0	0.0	0.0	3.2	0.1	0.1	0.0	1.5	0.0	0.0	0.0
GrsC3Arc	8.3	0.0	0.0	1.6	3.3	0.3	0.3	0.0	2.7	0.0	0.1	0.0
GrsC3	16.3	1.6	0.8	2.6	4.1	2.1	1.7	0.4	2.9	0.1	0.3	0.0
GrsC4	19.9	9.0	2.1	0.8	0.2	0.1	4.9	0.0	1.0	0.9	0.8	0.0
Crop	14.8	2.7	0.5	1.2	1.6	1.7	1.9	0.3	2.0	1.0	2.0	0.0
Year 850												
Tree	44.5	6.6	1.4	3.4	7.2	2.5	12.0	0.0	6.9	3.5	1.2	0.0
Shrub	9.8	1.2	2.1	0.0	3.4	0.2	1.0	0.2	1.8	0.0	0.1	0.0
Grass	46.1	10.5	3.1	4.6	8.5	3.1	5.7	0.6	7.3	1.2	1.9	0.0
Bare	27.7	11.1	1.9	3.0	1.9	0.2	1.4	4.4	2.7	0.0	1.2	0.1
NdlEvgTemp	6.4	0.0	0.2	1.5	1.5	1.1	0.1	0.0	2.0	0.0	0.1	0.0
NdlEvgBorl	6.3	0.0	0.0	0.5	2.6	0.4	0.0	0.0	2.8	0.0	0.0	0.0
NdlDecBorl	2.1	0.0	0.0	0.1	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BrdEvgTrop	18.6	4.7	0.3	0.1	0.0	0.0	9.8	0.0	0.0	3.2	0.6	0.0
BrdEvgTemp	2.1	0.1	0.5	0.2	0.0	0.0	0.9	0.0	0.0	0.1	0.2	0.0
BrdDecTrop	2.9	1.7	0.1	0.0	0.0	0.0	0.9	0.0	0.0	0.1	0.2	0.0
BrdDecTemp	4.6	0.0	0.3	0.7	0.6	0.9	0.2	0.0	1.8	0.0	0.1	0.0
BrdDecBorl	1.3	0.0	0.0	0.4	0.5	0.0	0.1	0.0	0.2	0.0	0.0	0.0
ShrEvgTemp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ShrDecTemp	4.9	1.2	2.1	0.0	0.1	0.1	0.9	0.1	0.3	0.0	0.1	0.0
ShrDecBorl	4.9	0.0	0.0	0.0	3.2	0.1	0.1	0.0	1.5	0.0	0.0	0.0
GrsC3Arc	8.0	0.0	0.0	1.4	3.3	0.3	0.2	0.0	2.7	0.0	0.1	0.0
GrsC3	18.2	1.5	1.0	2.4	4.9	2.6	1.5	0.5	3.4	0.0	0.3	0.0
GrsC4	19.9	9.0	2.1	0.7	0.3	0.2	3.9	0.0	1.1	1.1	1.5	0.0
Crop	1.6	0.2	0.0	0.2	0.1	0.4	0.1	0.1	0.0	0.0	0.6	0.0
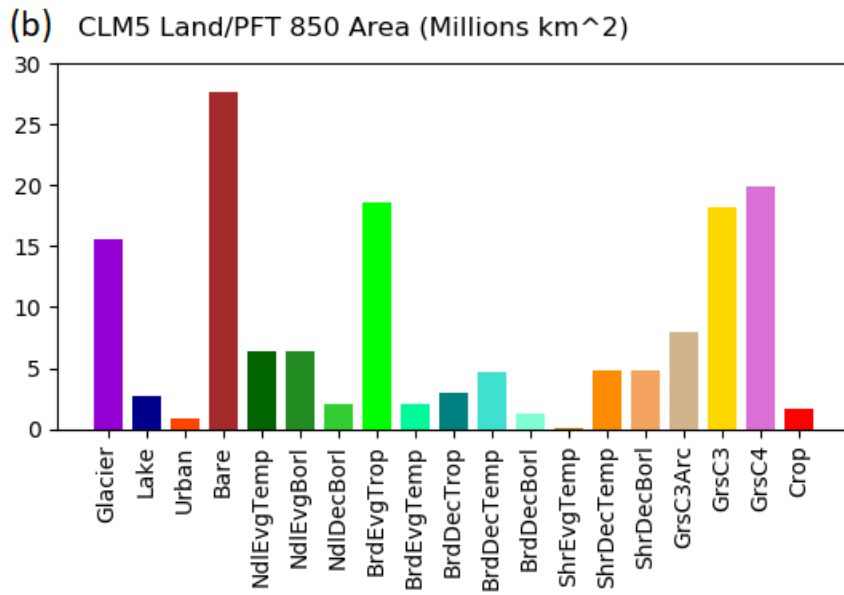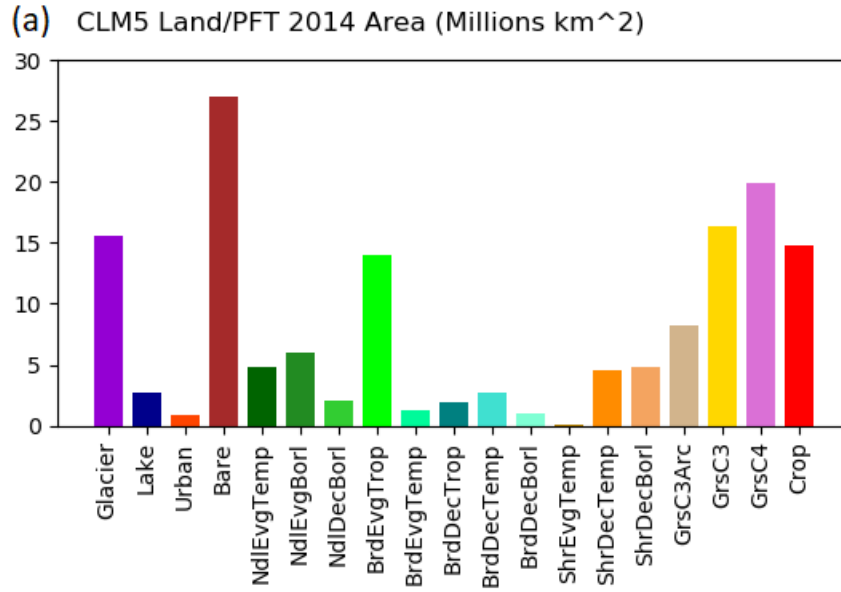
Figure 6.1: CLM5 Global Land Use and Plant Functional Type Area for: (a) Current Day 2014; and (b) Historical 850
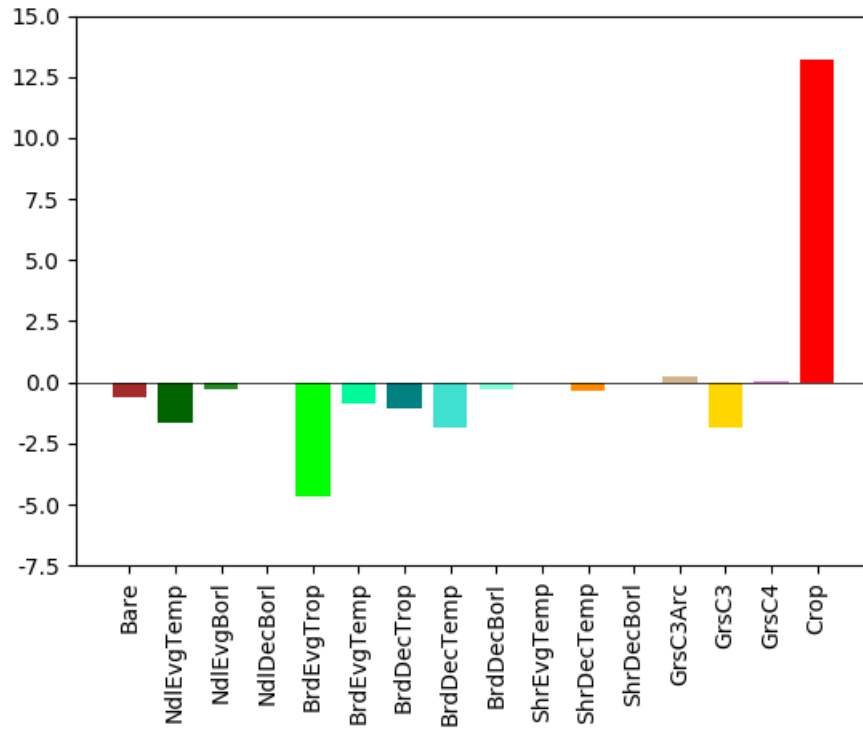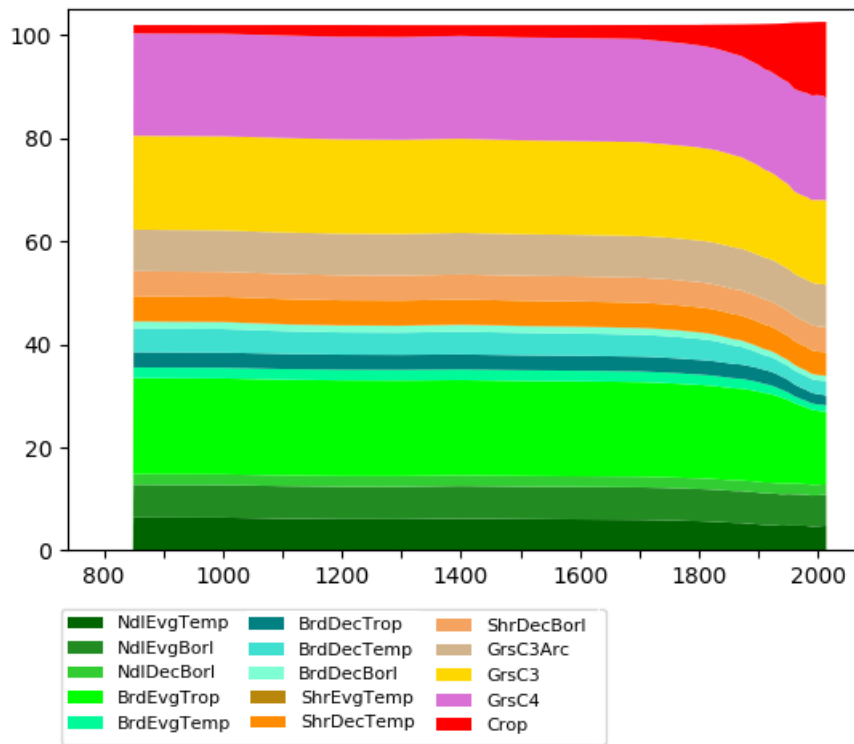
Figure 6.2: CLM5 Change in Global Land Use and Plant Functional Type Area for: (a) Difference between Current Day 2014 and 850; and (b) Annual values from 850 to 2014.

(a) CLM5 Bare Soil 2014 Raw 0.25 Degrees   (%)
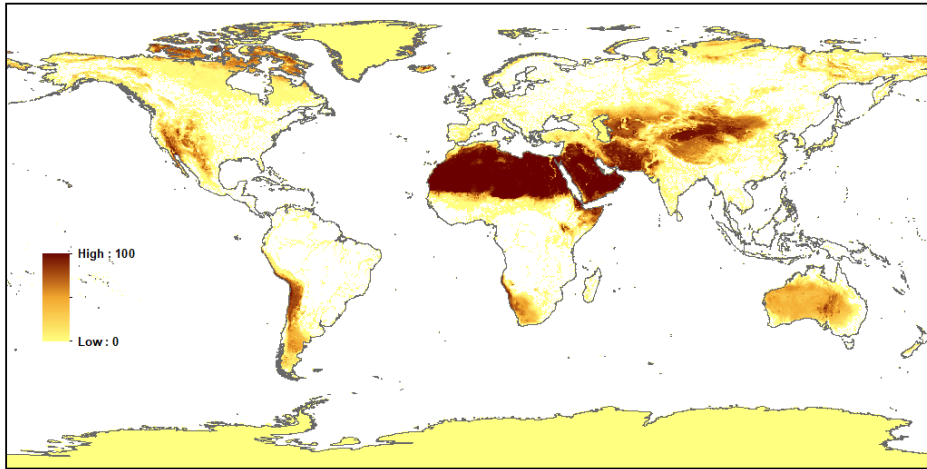
(b) CLM5 Needleleaf Evergreen Temperate 2014 Raw 0.25 Degrees   (%)
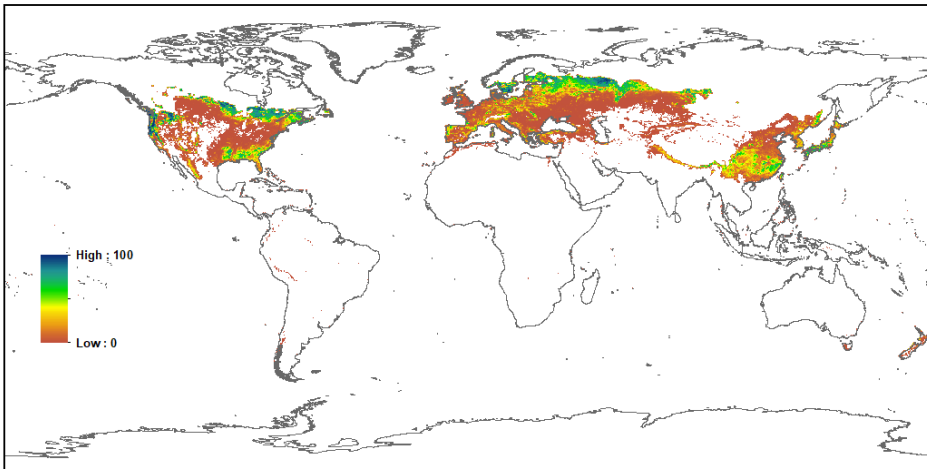
(c) CLM5 Needleleaf Evergreen Boreal 2014 Raw 0.25 Degrees   (%)
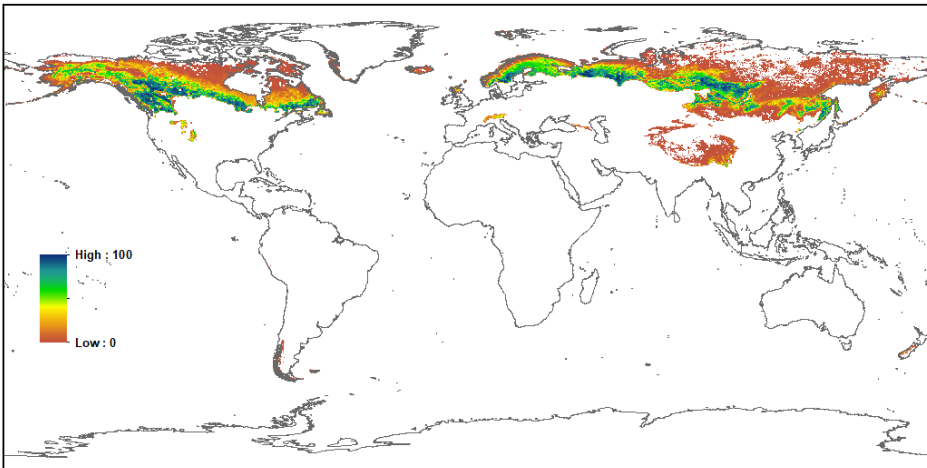
Figure 6.3: Global Current Day CLM5 PFT Mapping 2014: (a) Percent Bare Ground; (b) Percent Needleleaf Evergreen Temperate Tree; and (c) Percent Needleleaf Evergreen Boreal Tree.

(a) CLM5 Needleleaf Deciduous Boreal 2014 Raw 0.25 Degrees   (%)
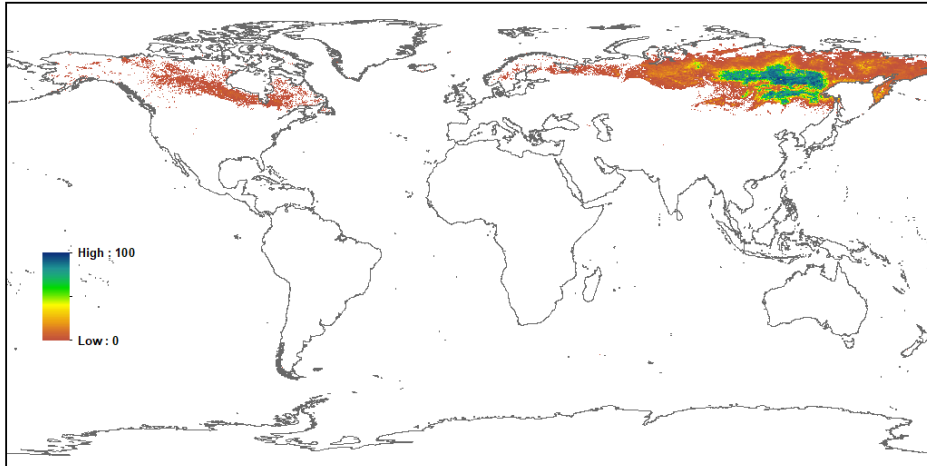
(b) CLM5 Broadleaf Evergreen Tropical 2014 Raw 0.25 Degrees   (%)
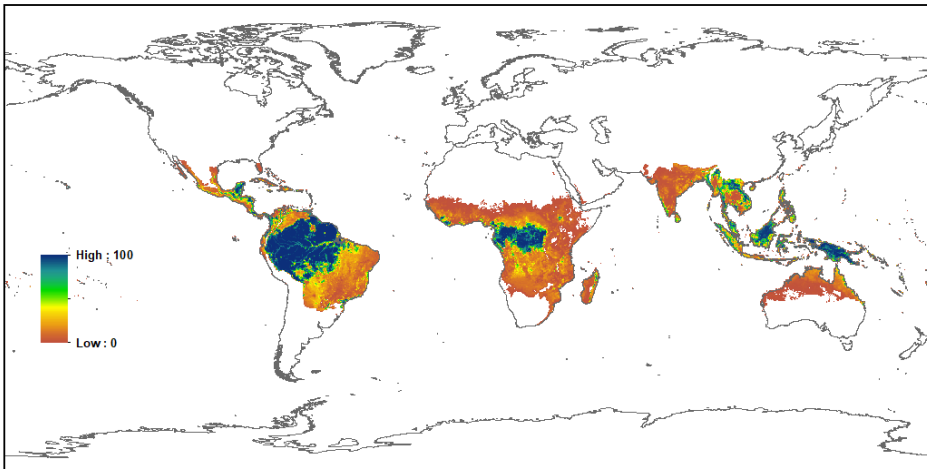
(c) CLM5 Broadleaf Evergreen Temperate 2014 Raw 0.25 Degrees   (%)
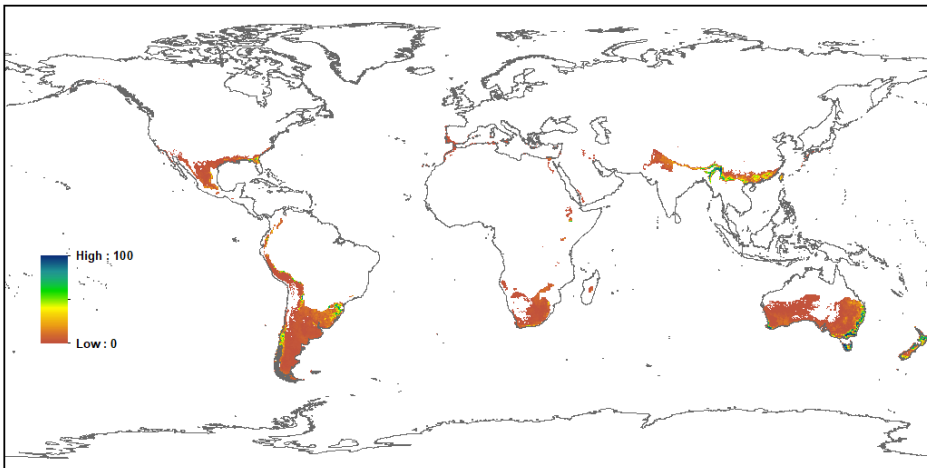
Figure 6.4: Global Current Day CLM5 PFT Mapping 2014: (a) Percent Needleleaf Deciduous Boreal Tree; (b) Percent Broadleaf Evergreen Tropical Tree; and (c) Percent Broadleaf Evergreen Temperate Tree.

Figure 6.5: Global Current Day CLM5 PFT Mapping 2014: (a) Percent Broadleaf Deciduous Tropical Tree; (b) Percent Broadleaf Deciduous Temperate Tree; and (c) Percent Broadleaf Deciduous Boreal Tree.

(a) CLM5 Shrub Deciduous Temperate 2014 Raw 0.25 Degrees   (%)
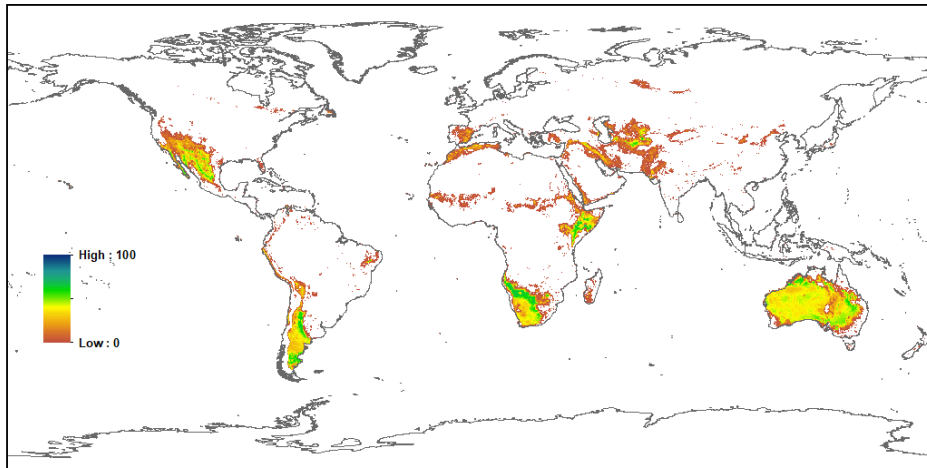
(b) CLM5 Shrub Deciduous Boreal 2014 Raw 0.25 Degrees   (%)
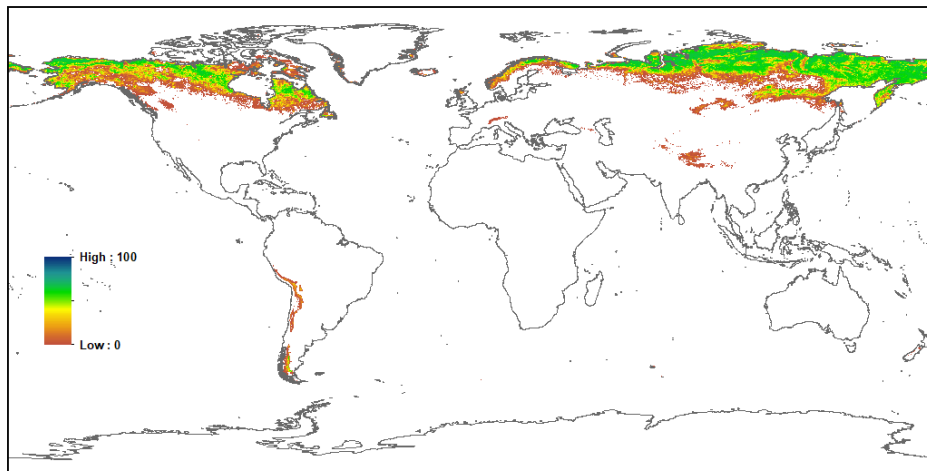
(c) CLM5 Grass C3 Arctic 2014 Raw 0.25 Degrees   (%)
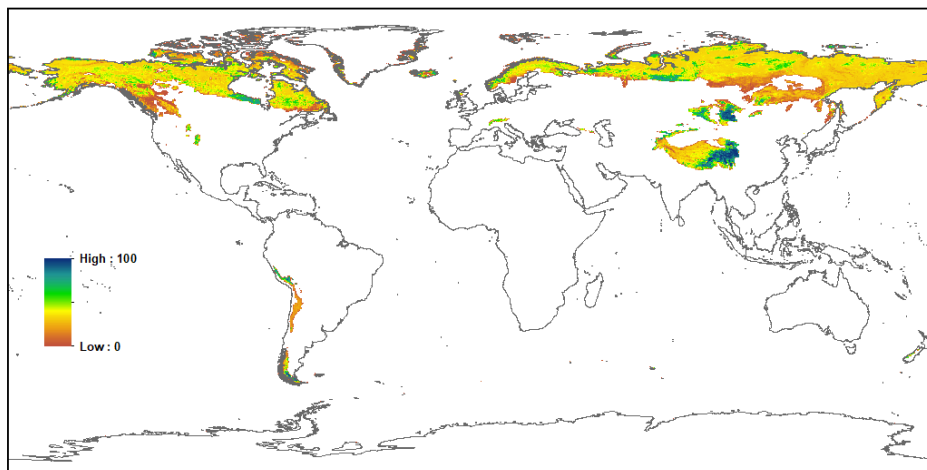
Figure 6.6: Global Current Day CLM5 PFT Mapping 2014: (a) Percent Deciduous Temperate Shrub; (b) Percent Deciduous Boreal Shrub; and (c) Percent C3 Arctic Grass.

(a) CLM5 Grass C3 2014 Raw 0.25 Degrees (%)
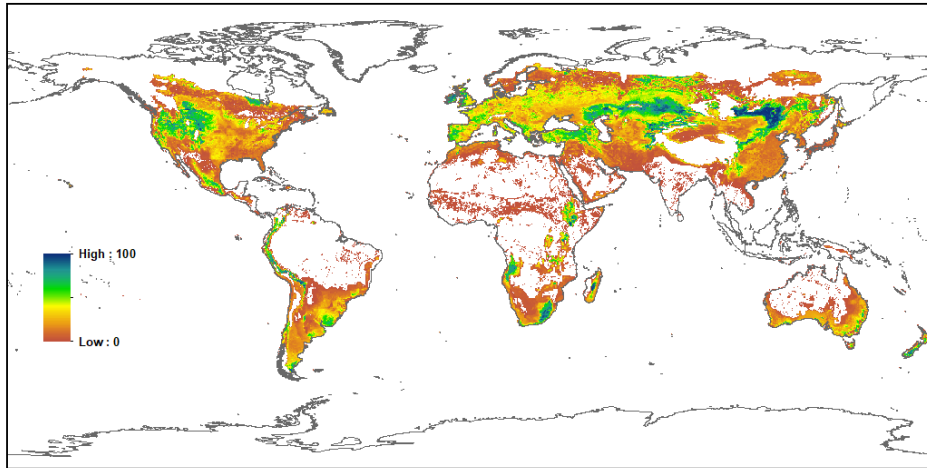
(b) CLM5 Grass C4 2014 Raw 0.25 Degrees (%)
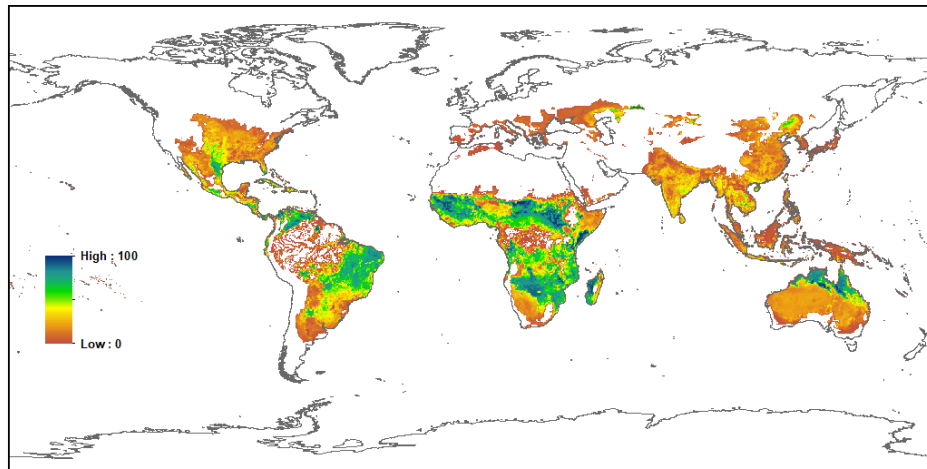
(c) CLM5 Crop 2014 Raw 0.25 Degrees (%)
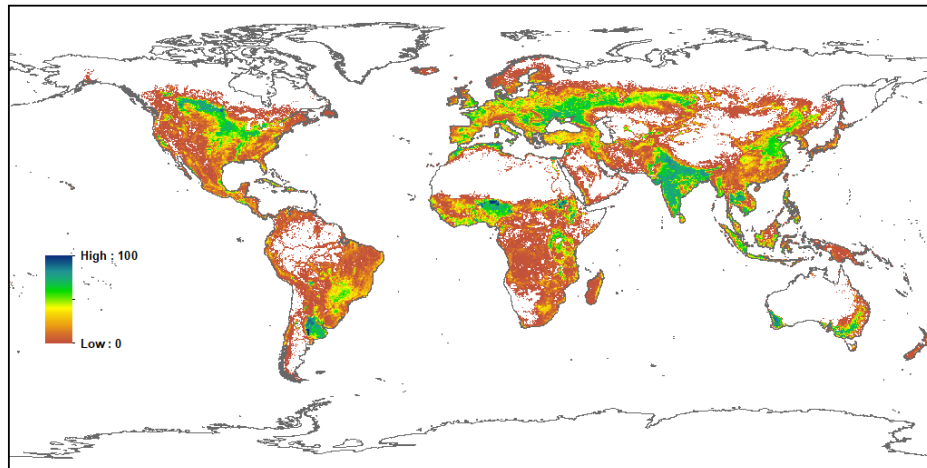
Figure 6.7: Global Current Day CLM5 PFT Mapping 2014: (a) Percent C3 Grass; (b) Percent C4 Grass; and (c) Percent Crop.

6.3 CLM5 Active Crop Functional Type (CFT) Historical Data

Table 6.2 Annual Global and Regional CLM5 Active Crop Functional Type (CFT) Area for 850 and 2014 in Millions of km^2, using the default CLM5 Domain File Land Mask.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
Temp Corn	0.93	0.01	0.00	0.20	0.10	0.19	0.07	0.01	0.34	0.00	0.02	0.00
Trop Corn	0.96	0.42	0.00	0.02	0.00	0.00	0.37	0.00	0.00	0.06	0.08	0.00
Wheat	3.50	0.20	0.26	0.29	0.76	0.59	0.20	0.17	0.63	0.00	0.43	0.00
Temp Soybean	0.62	0.00	0.00	0.07	0.02	0.02	0.13	0.00	0.37	0.00	0.00	0.00
Trop Soybean	0.37	0.01	0.00	0.01	0.00	0.00	0.27	0.00	0.00	0.01	0.07	0.00
Rice	1.59	0.14	0.03	0.26	0.01	0.01	0.13	0.01	0.02	0.44	0.54	0.00
Cotton	0.48	0.10	0.00	0.04	0.03	0.01	0.04	0.01	0.10	0.00	0.14	0.00
Sugarcane	0.23	0.01	0.00	0.01	0.00	0.00	0.13	0.00	0.00	0.03	0.04	0.00
Other Crop	6.63	1.93	0.21	0.34	0.73	0.88	0.61	0.16	0.60	0.43	0.79	0.00
All Crop	14.82	2.72	0.50	1.21	1.62	1.70	1.91	0.35	1.98	0.97	1.98	0.00
Year 850												
Temp Corn	0.06	0.00	0.00	0.01	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Trop Corn	0.08	0.02	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.02	0.00
Wheat	0.39	0.03	0.00	0.06	0.02	0.13	0.01	0.05	0.00	0.00	0.10	0.00
Temp Soybean	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trop Soybean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rice	0.28	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.01	0.20	0.00
Cotton	0.08	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.04	0.00
Sugarcane	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Other Crop	0.77	0.15	0.00	0.04	0.03	0.19	0.05	0.05	0.00	0.01	0.25	0.00
All Crop	1.61	0.20	0.01	0.18	0.06	0.36	0.11	0.10	0.01	0.03	0.58	0.00

(a) CLM5 CFT 2014 Area (Millions km^2)
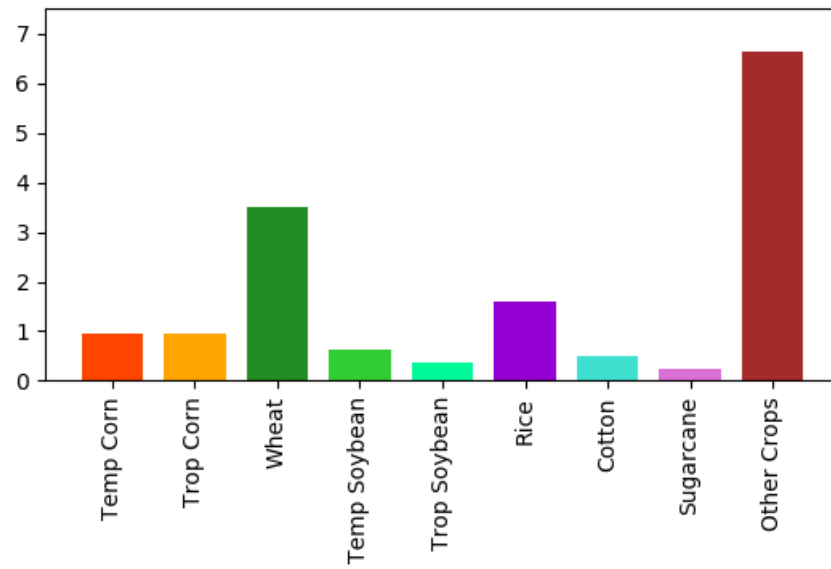
(b) CLM5 CFT 850 Area (Millions km^2)
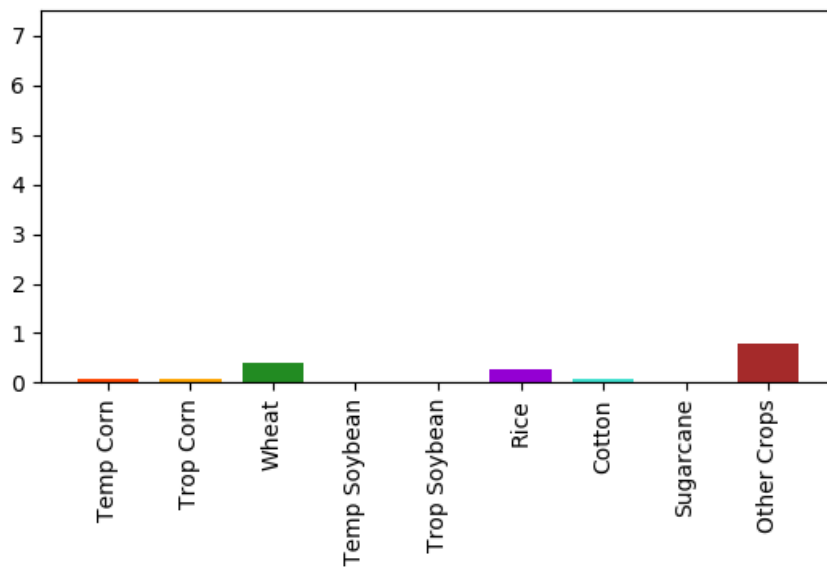
Figure 6.8: CLM5 Global Active and Other Crop Functional Type Area for: (a) Current Day 2014; and (b) Historical 850

(a) CLM5 Change in CFT 2014 - 850 (Millions km^2)
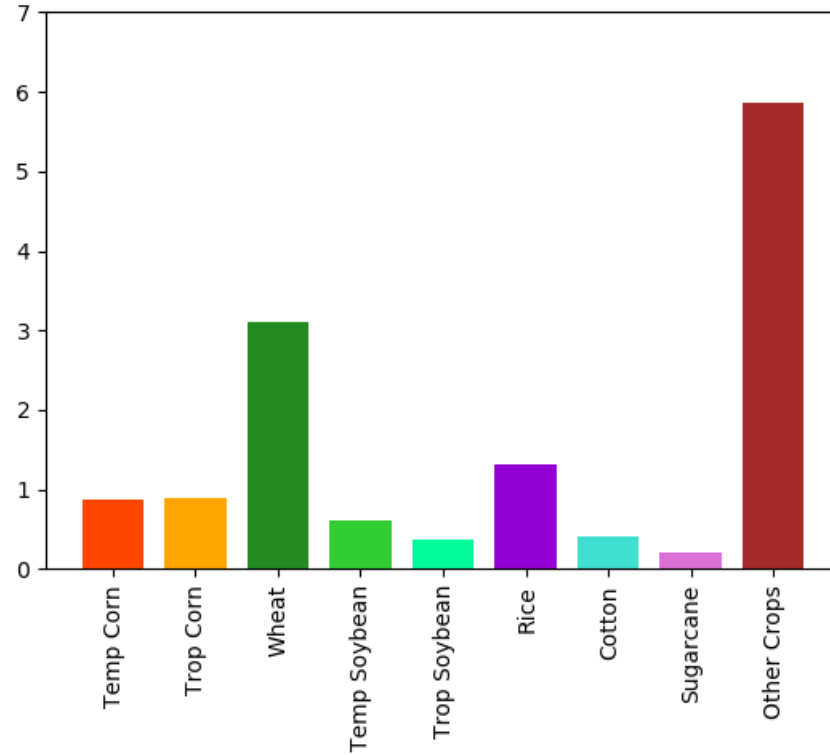
(b) CLM5 CFT Area 850 - 2014 (Millions km^2)
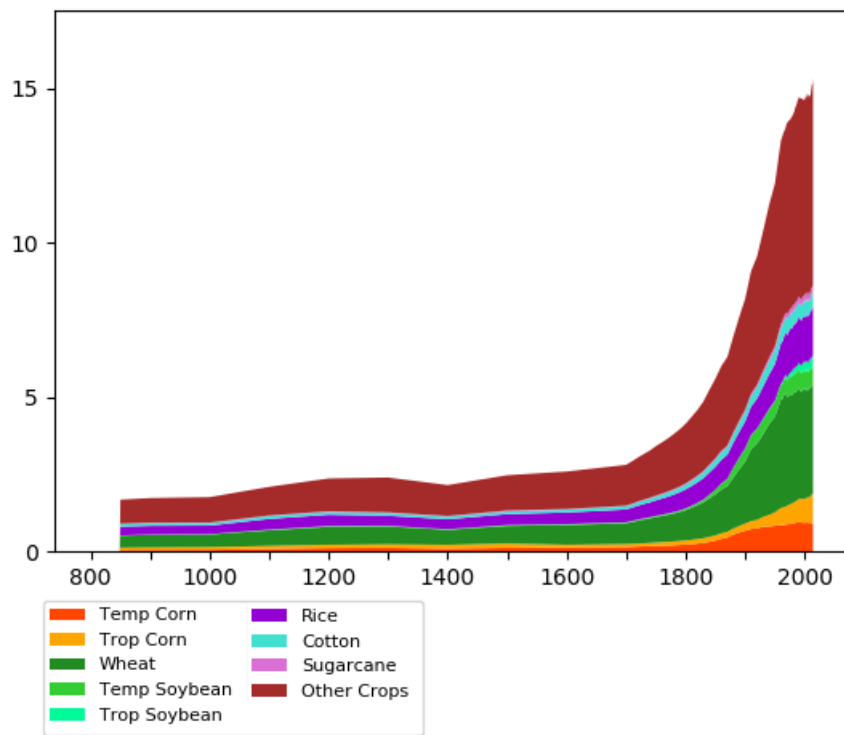
Figure 6.9: CLM5 Change in Global Active and Other Crop Functional Type Area for: (a) Difference between Current Day 2014 and 850; and (b) Annual values from 850 to 2014.

(a) CLM5 Temperate Corn 2014 Raw 0.25 Degrees   (%)
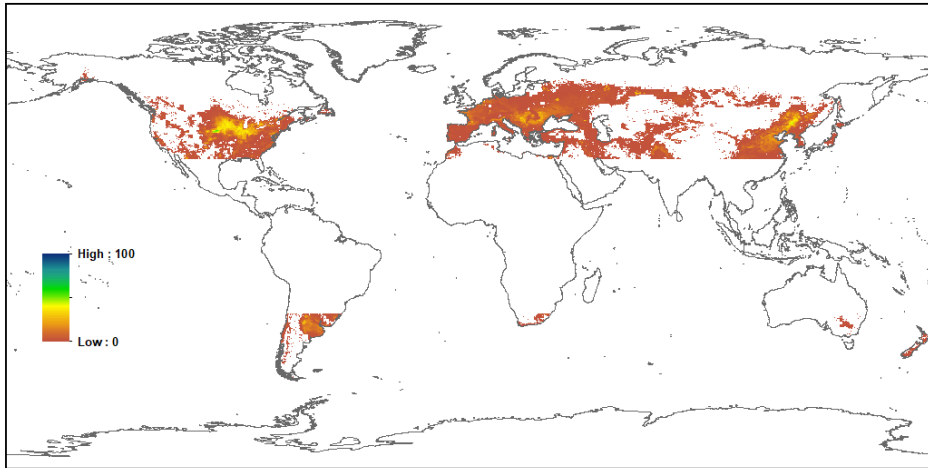
(b) CLM5 Tropical Corn 2014 Raw 0.25 Degrees   (%)
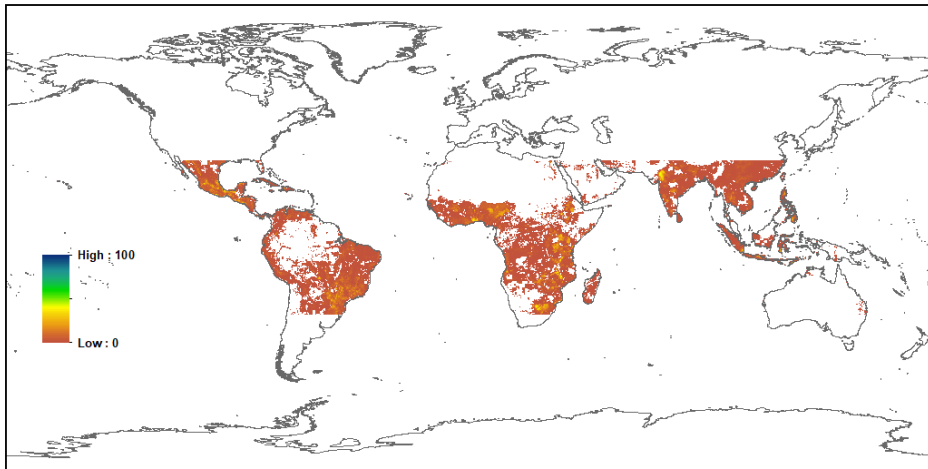
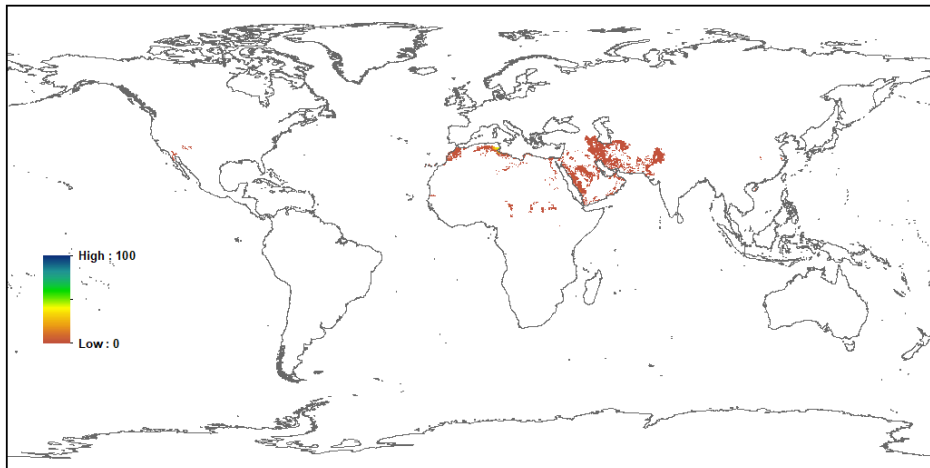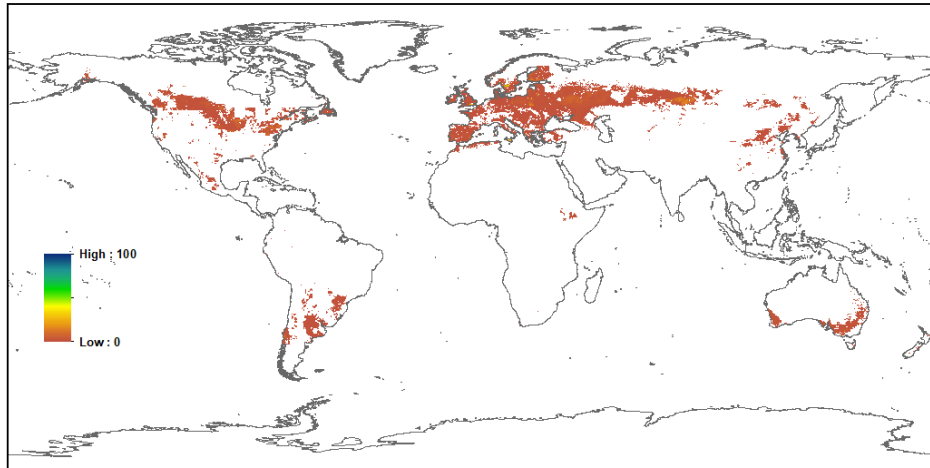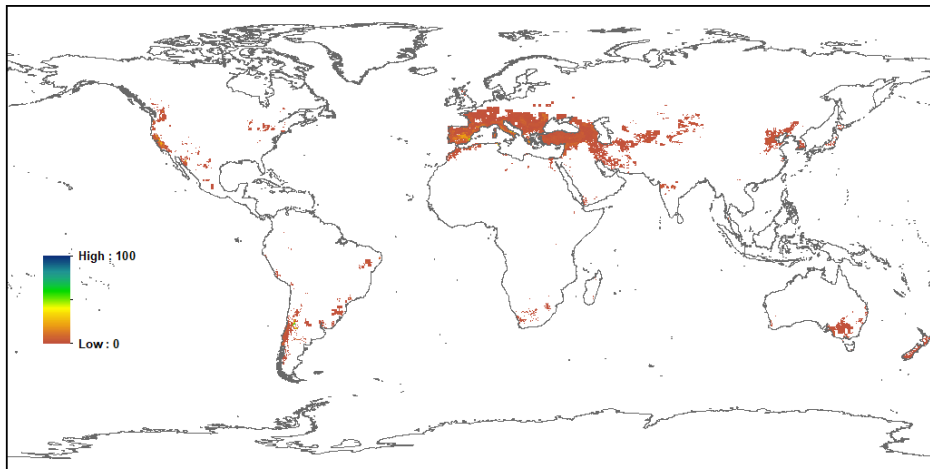**(c) CLM5 Wheat 2014 Raw 0.25 Degrees   (%)**
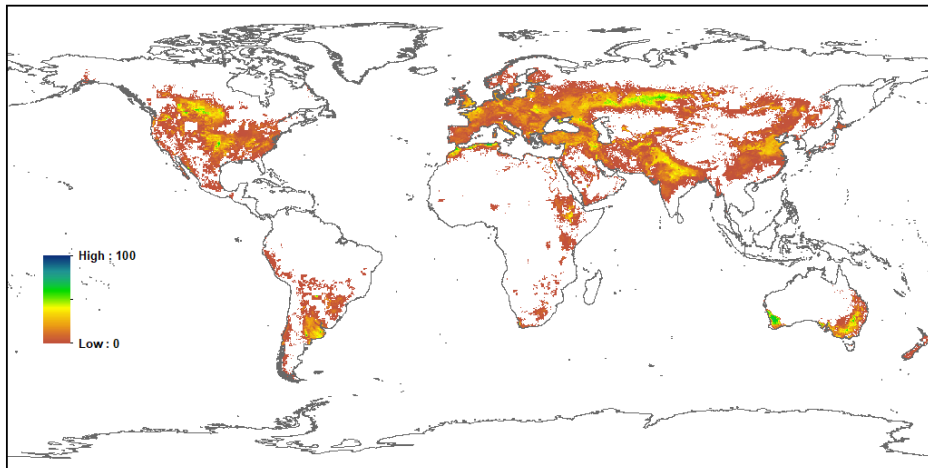
Figure 6.10: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Temperate Corn; (b) Percent Tropical Corn; and (c) Percent Wheat.

(a) CLM5 Temperate Soybean 2014 Raw 0.25 Degrees (%)
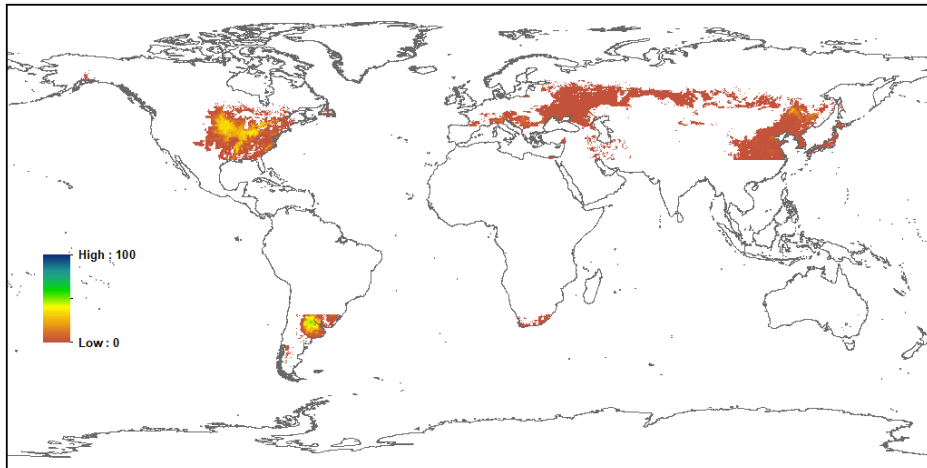
(b) CLM5 Tropical Soybean 2014 Raw 0.25 Degrees (%)
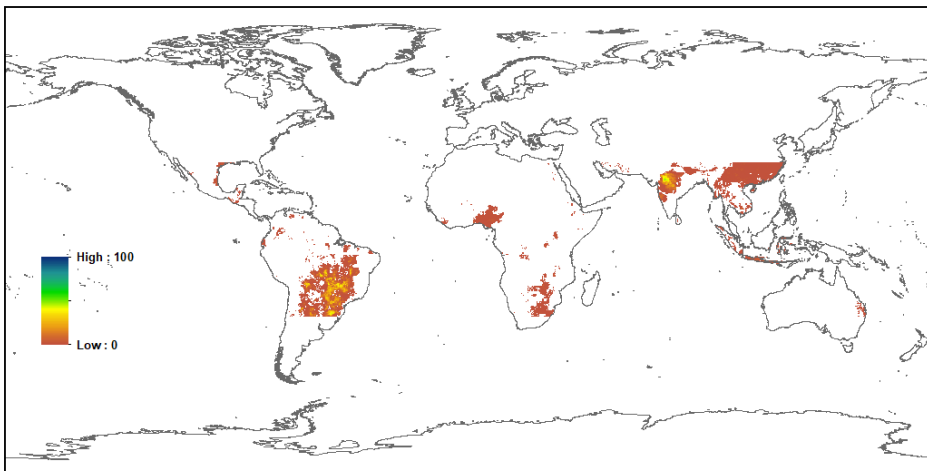
(c) CLM5 Rice 2014 Raw 0.25 Degrees (%)
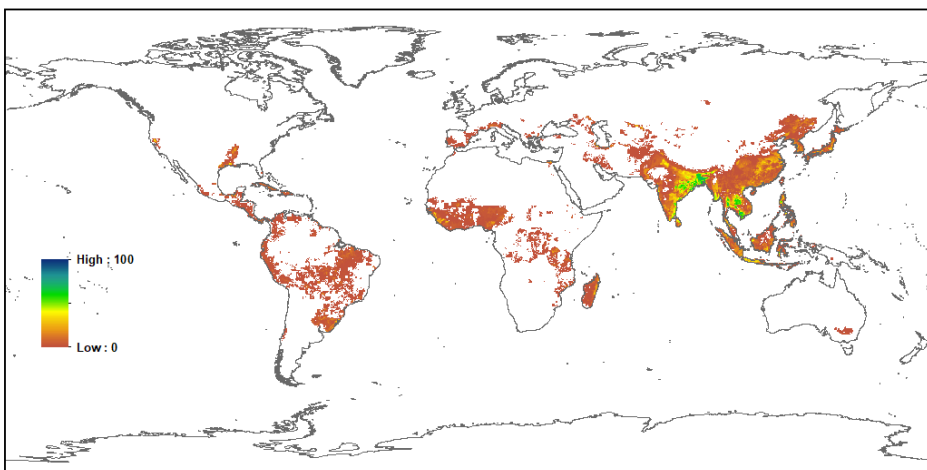
Figure 6.11: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Temperate Soybean; (b) Percent Tropical Soybean; and (c) Percent Rice.

(a) CLM5 Cotton 2014 Raw 0.25 Degrees   (%)
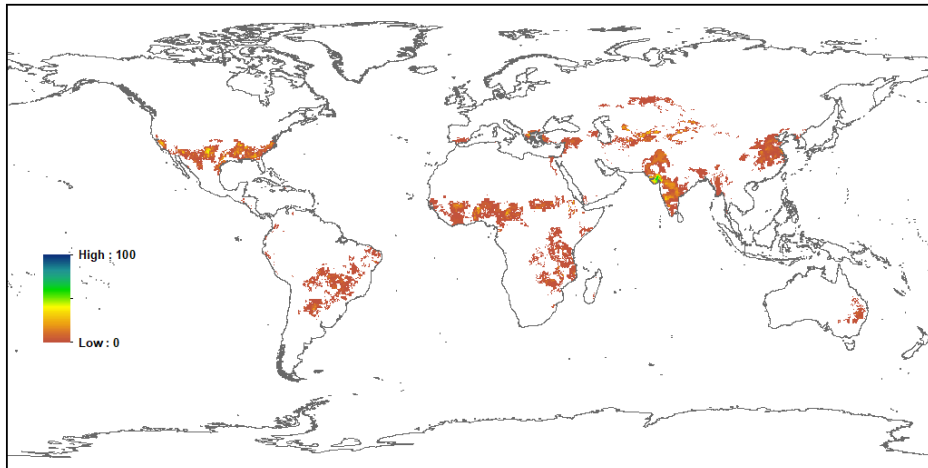
(b) CLM5 Sugarcane 2014 Raw 0.25 Degrees   (%)
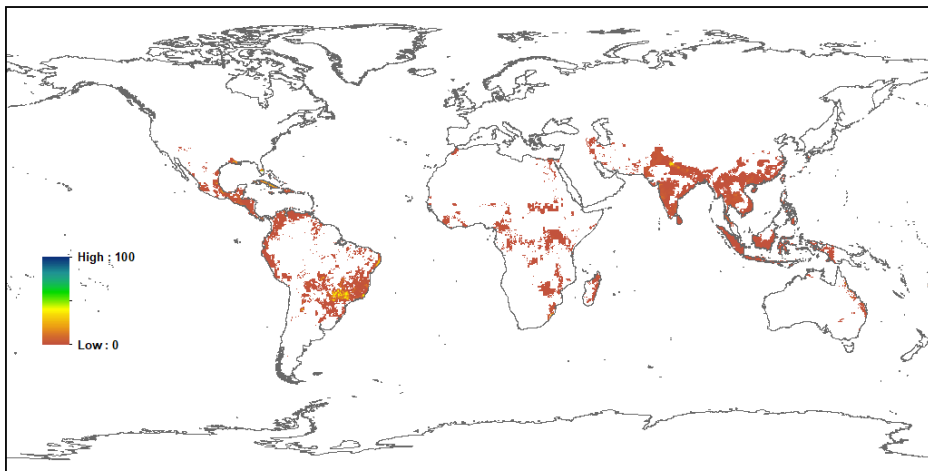
(c) CLM5 Other Crops 2014 Raw 0.25 Degrees   (%)
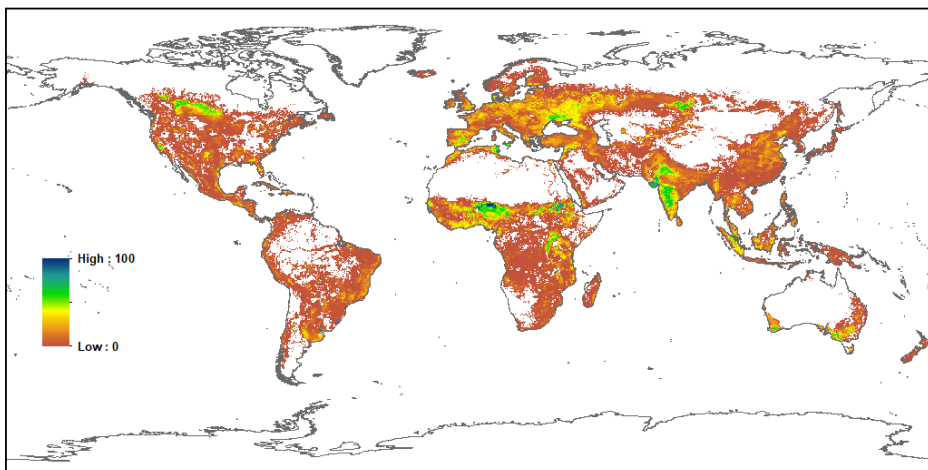
Figure 6.12: Global Current Day CLM5 CFT Mapping 2014: a) Percent Cotton; (b) Percent Sugarcane; and (c) Percent Other Crops.

6.4 CLM5 Other Crop Functional Type (CFT) Historical Data

Table 6.3 Annual Global and Regional CLM5 Unrepresented Crop Functional Type (CFT) Area for 850 and 2014 in Millions of km^2, using the default CLM5 Domain File Land Mask. Unrepresented CFTs are simulated using the functionally closest Active CFT parameters and phenology.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
Barley	0.90	0.09	0.07	0.01	0.30	0.26	0.02	0.05	0.08	0.00	0.01	0.00
Cassava	0.35	0.24	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.03	0.00	0.00
Citrus	0.26	0.05	0.00	0.06	0.00	0.03	0.04	0.01	0.01	0.01	0.04	0.00
Cocoa	0.15	0.07	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.04	0.00	0.00
Coffee	0.31	0.13	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.06	0.01	0.00
Datepalm	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Foddergrass	0.20	0.00	0.01	0.00	0.08	0.05	0.01	0.00	0.05	0.00	0.00	0.00
Grapes	0.17	0.01	0.00	0.01	0.02	0.10	0.01	0.01	0.01	0.00	0.00	0.00
Groundnuts	0.29	0.17	0.00	0.03	0.00	0.00	0.01	0.00	0.01	0.02	0.05	0.00
Millet	0.39	0.26	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.11	0.00
Oilpalm	0.27	0.04	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.21	0.00	0.00
Potatoes	0.26	0.02	0.00	0.05	0.06	0.06	0.02	0.00	0.03	0.00	0.02	0.00
Pulses	0.97	0.25	0.04	0.04	0.01	0.08	0.10	0.04	0.13	0.04	0.25	0.00
Rapeseed	0.43	0.00	0.06	0.06	0.04	0.09	0.00	0.00	0.13	0.00	0.06	0.00
Rye	0.09	0.00	0.00	0.00	0.05	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Sorghum	0.61	0.42	0.01	0.00	0.00	0.00	0.05	0.01	0.03	0.00	0.08	0.00
Sugarbeet	0.06	0.00	0.00	0.00	0.02	0.03	0.00	0.00	0.01	0.00	0.00	0.00
Sunflower	0.40	0.05	0.00	0.01	0.12	0.13	0.06	0.00	0.01	0.01	0.01	0.00
Year 850												
Barley	0.12	0.02	0.00	0.01	0.01	0.05	0.00	0.02	0.00	0.00	0.01	0.00
Cassava	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Citrus	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cocoa	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Coffee	0.03	0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
Datepalm	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Foddergrass	0.03	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Grapes	0.05	0.01	0.00	0.00	0.00	0.04	0.00	0.01	0.00	0.00	0.00	0.00
Groundnuts	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Millet	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
Oilpalm	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Potatoes	0.03	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Pulses	0.16	0.02	0.00	0.01	0.00	0.03	0.01	0.01	0.00	0.00	0.09	0.00
Rapeseed	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
Rye	0.02	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Sorghum	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
Sugarbeet	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Sunflower	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(a) CLM5 Other CFT 2014 Area (Millions km^2)
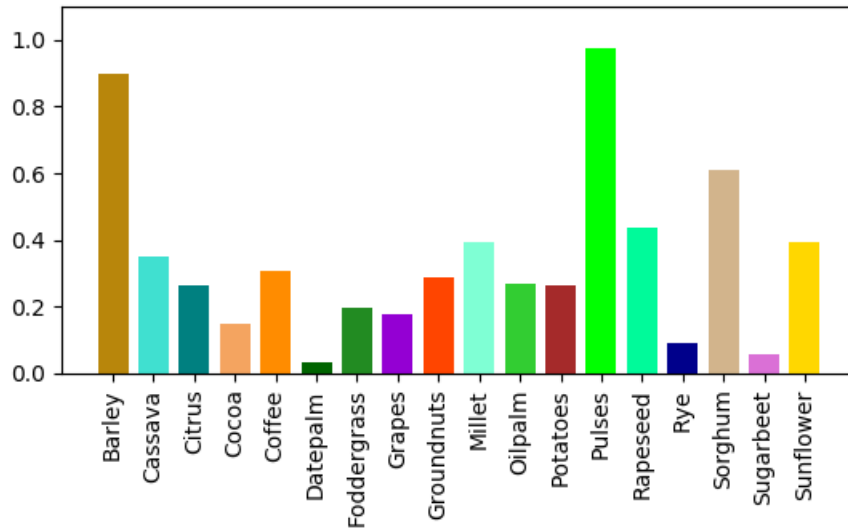
(b) CLM5 Other CFT 850 Area (Millions km^2)
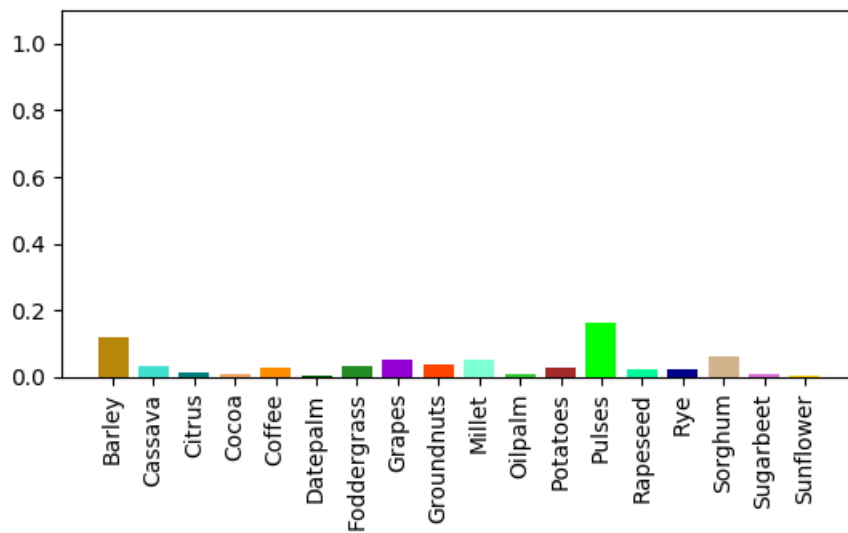
Figure 6.13: CLM5 Global Not Yet Represented Crop Functional Type Area (Remapped in CLM5) for: (a) Current Day 2014; and (b) Historical 850
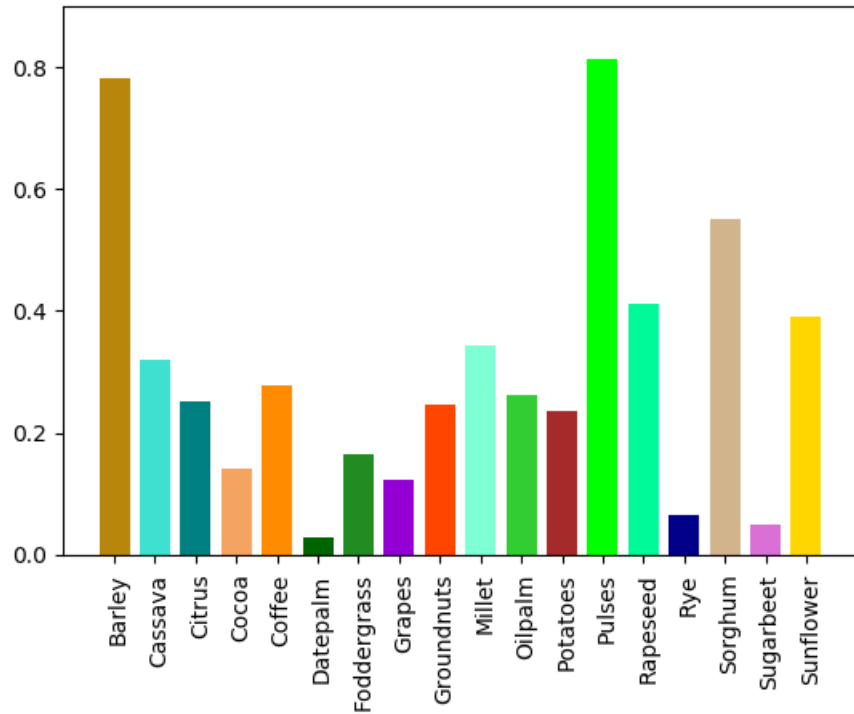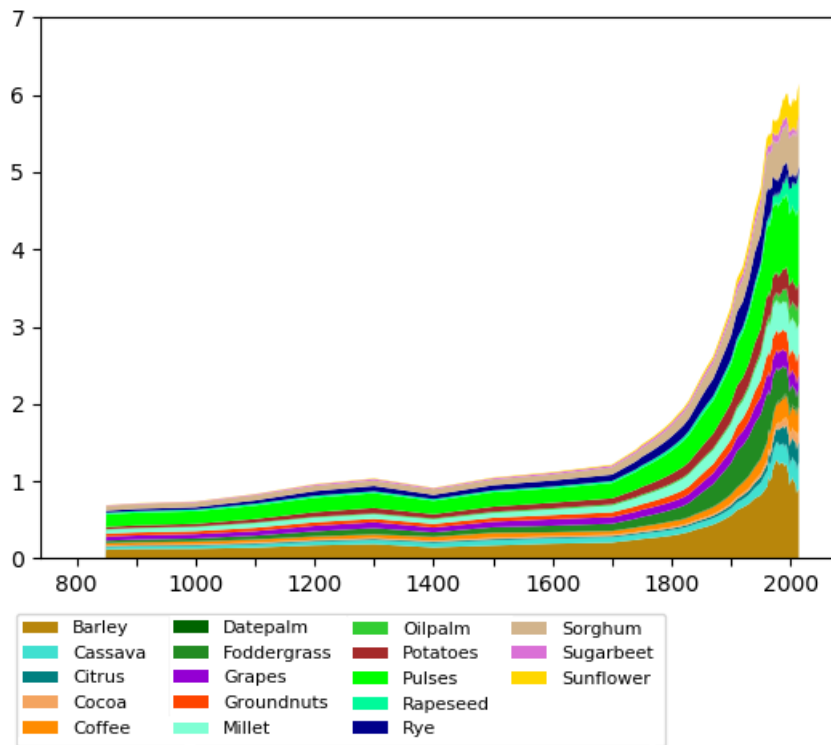
Figure 6.14: CLM5 Change in Not Yet Represented Crop Functional Type Area: (a) Difference between Current Day 2014 and 850; and (b) Annual values from 850 to 2014.

(a) CLM5 Barley 2014 Raw 0.25 Degrees   (%)
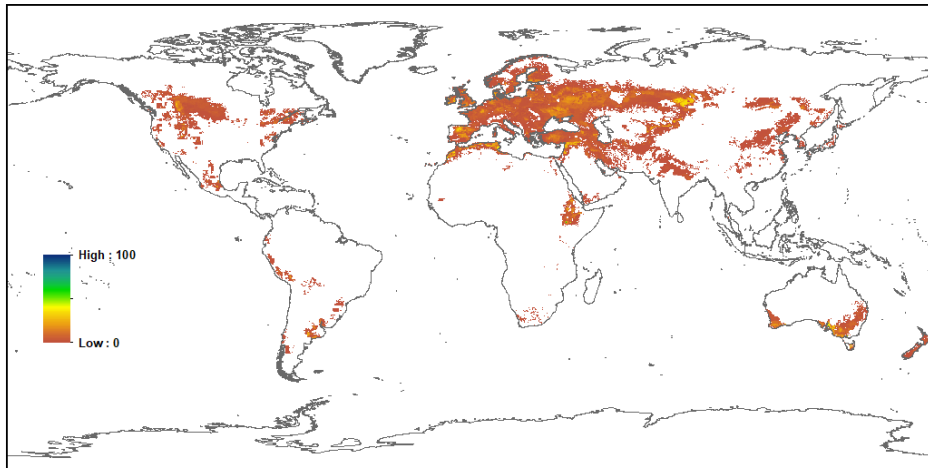
(b) CLM5 Cassava 2014 Raw 0.25 Degrees   (%)
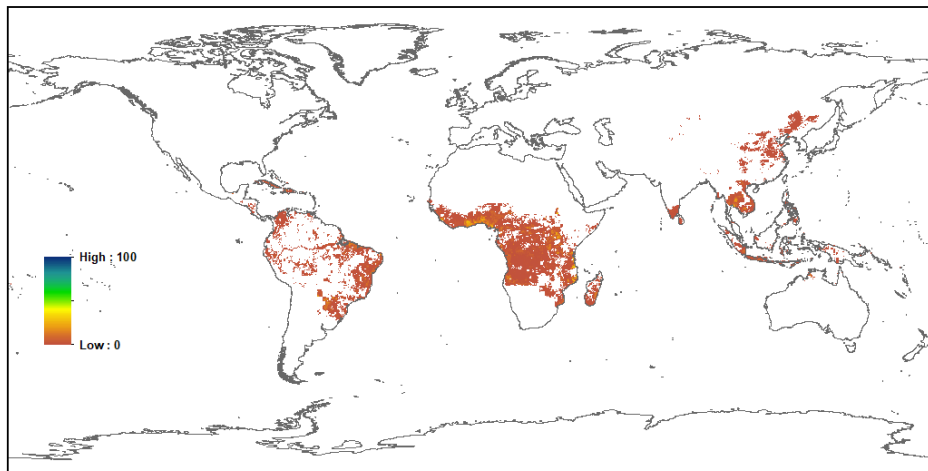
(c) CLM5 Citrus 2014 Raw 0.25 Degrees   (%)
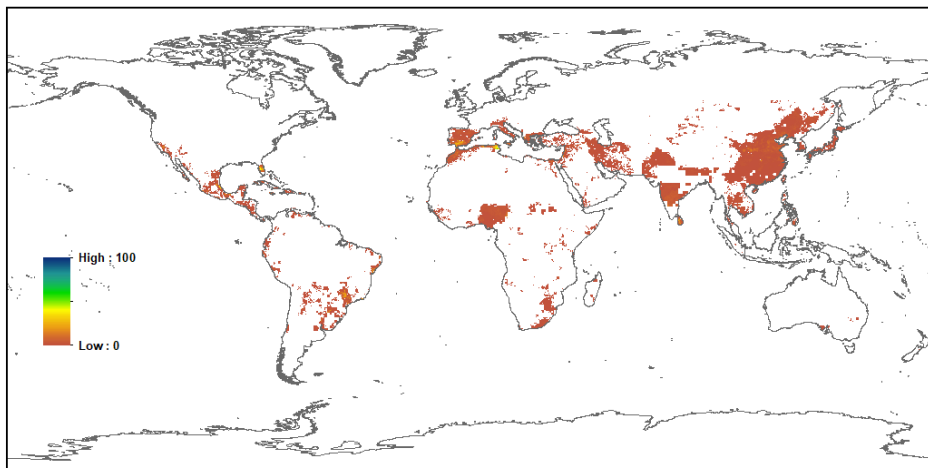
Figure 6.15: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Barley; (b) Percent Cassava; and (c) Percent Citrus.

(a) CLM5 Cocoa 2014 Raw 0.25 Degrees   (%)
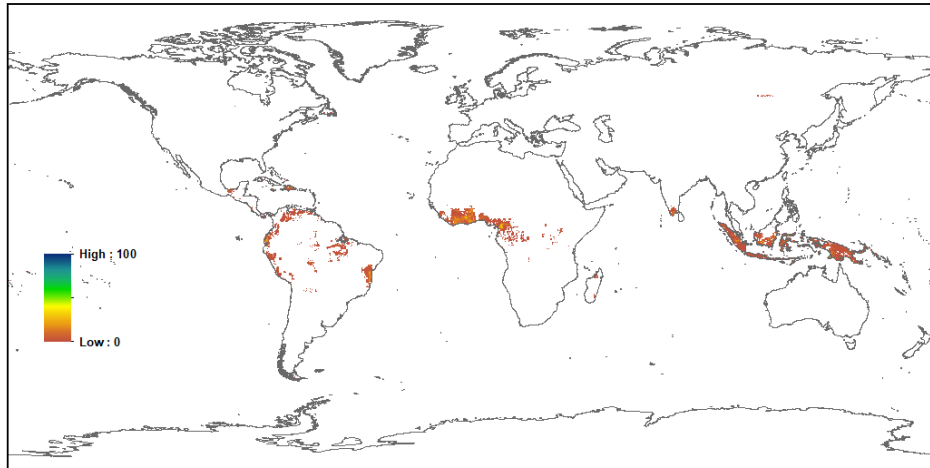
(b) CLM5 Coffee 2014 Raw 0.25 Degrees   (%)
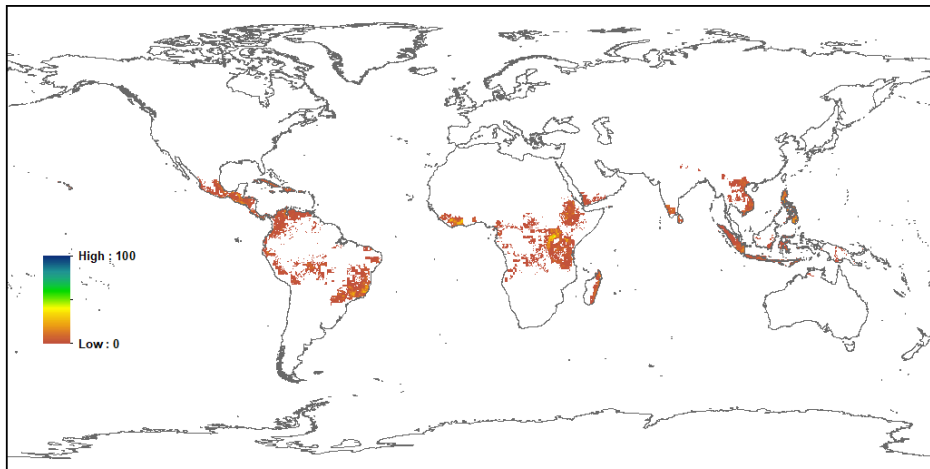
(c) CLM5 Datepalm 2014 Raw 0.25 Degrees   (%)

Figure 6.16: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Cocoa; (b) Percent Coffee; and (c) Percent Datepalm.

(a)   CLM5 Fodder Grass 2014 Raw 0.25 Degrees   (%)

(b)   CLM5 Grapes 2014 Raw 0.25 Degrees   (%)

(c)   CLM5 Groundnuts 2014 Raw 0.25 Degrees   (%)
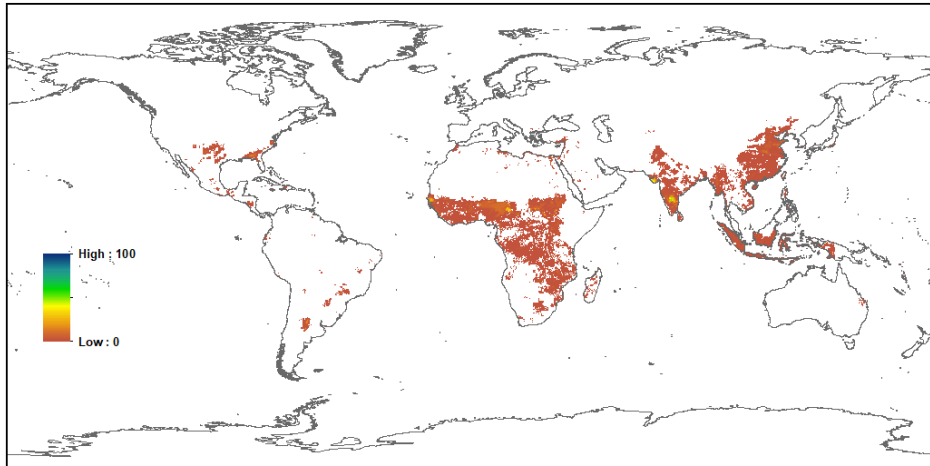
Figure 6.17: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Fodder Grass; (b) Percent Grapes; and (c) Percent Groundnuts.

(a) CLM5 Millet 2014 Raw 0.25 Degrees   (%)
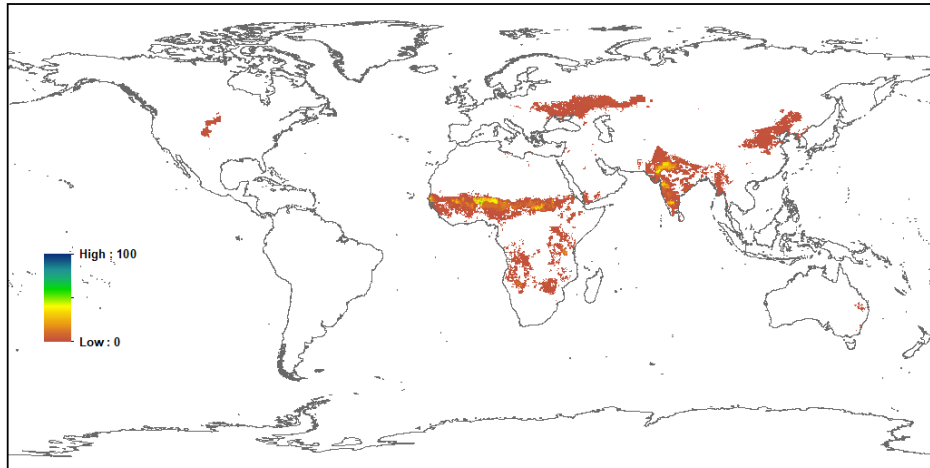
(b) CLM5 Oilpalm 2014 Raw 0.25 Degrees   (%)
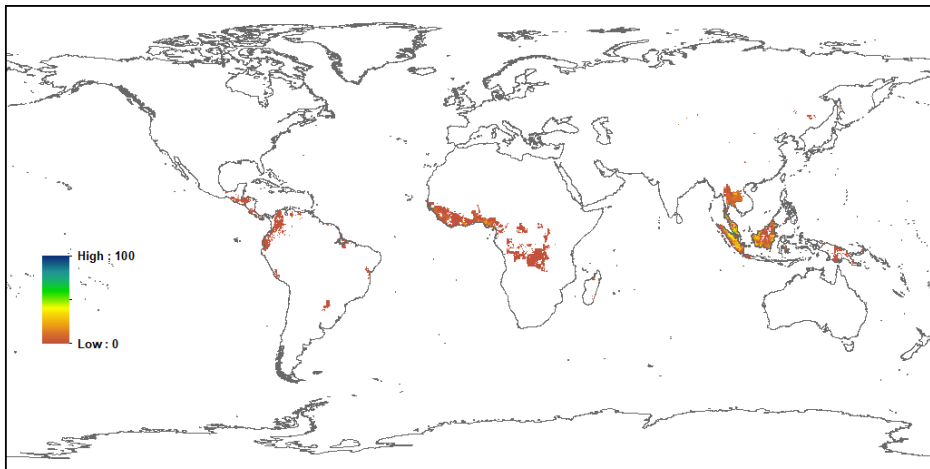
(c) CLM5 Potatoes 2014 Raw 0.25 Degrees   (%)
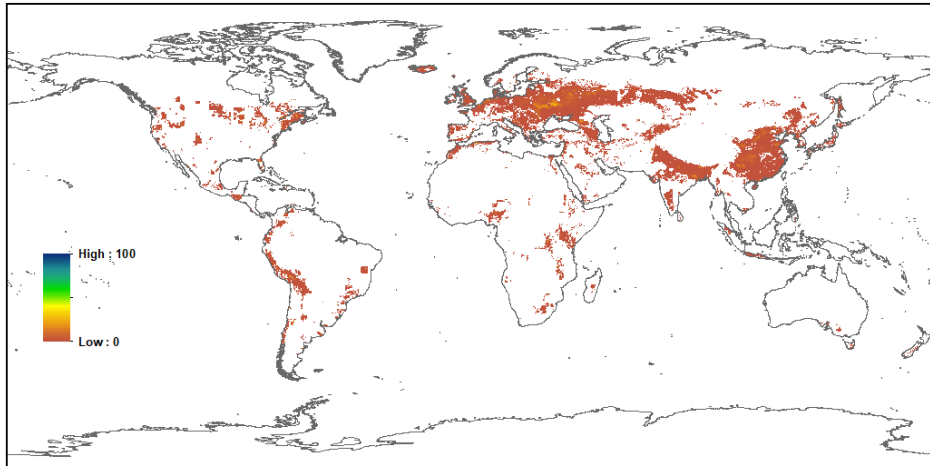
Figure 6.18: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Millet; (b) Percent Tropical Oil Palm; and (c) Percent Potatoes.

(a) CLM5 Pulses 2014 Raw 0.25 Degrees   (%)
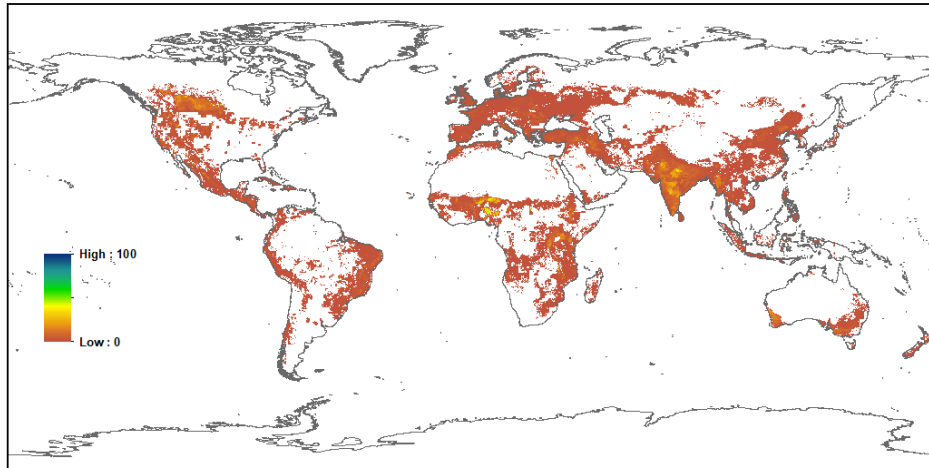
(b) CLM5 Rapeseed 2014 Raw 0.25 Degrees   (%)
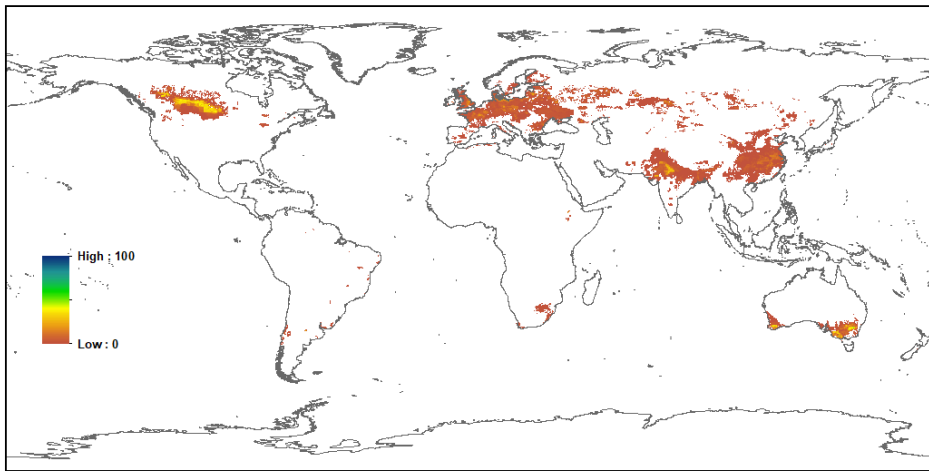
(c) CLM5 Rye 2014 Raw 0.25 Degrees   (%)
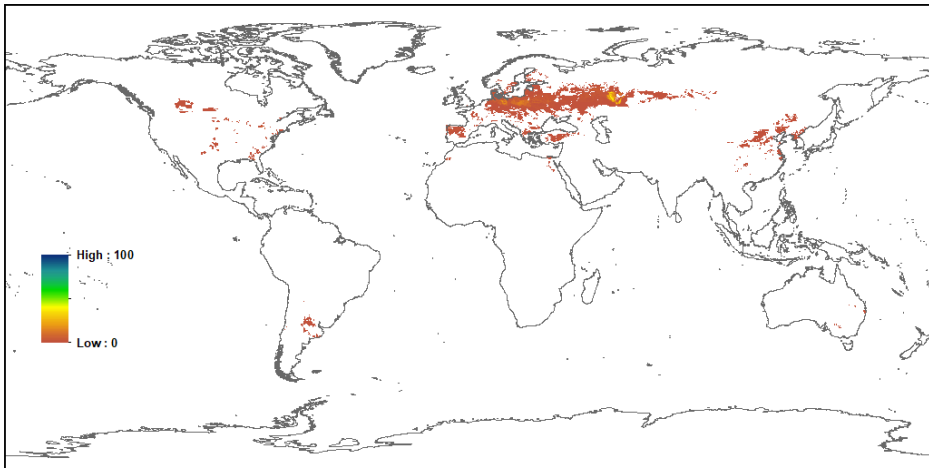
Figure 6.19: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Pulses; (b) Percent Rapeseed; and (c) Percent Rye.

(a) CLM5 Sorghum 2014 Raw 0.25 Degrees   (%)
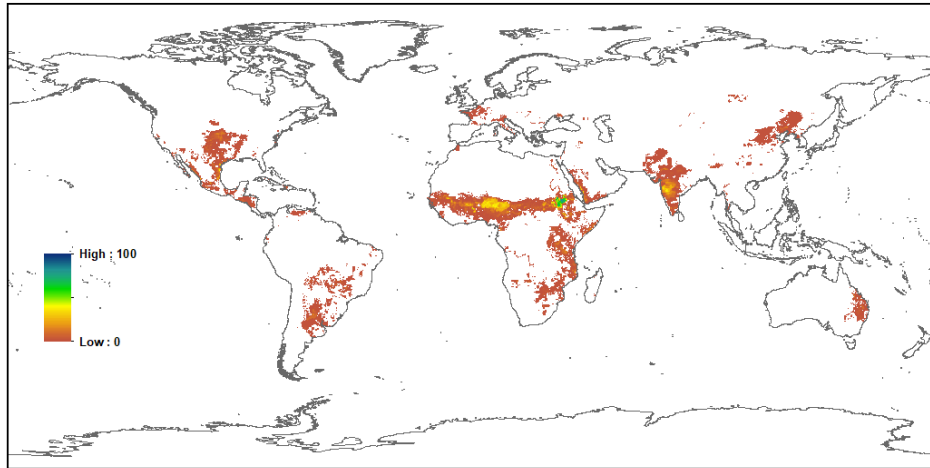
(b) CLM5 Sugarbeet 2014 Raw 0.25 Degrees   (%)
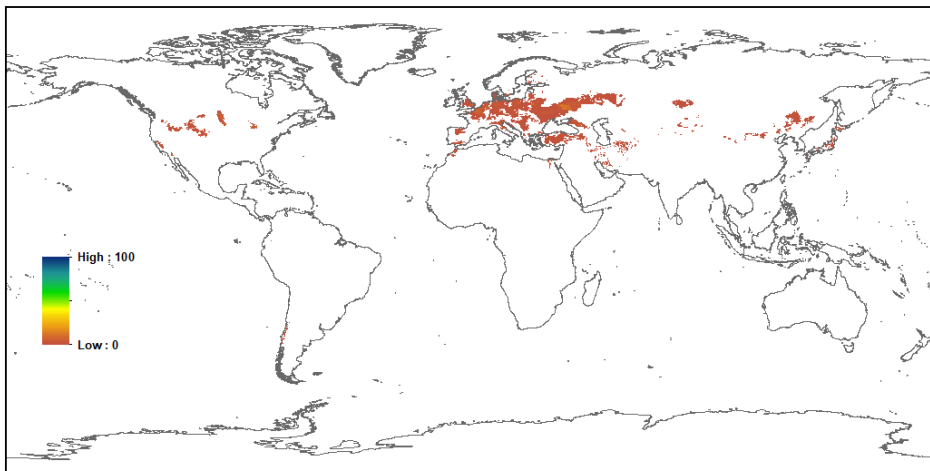
(c) CLM5 Sunflower 2014 Raw 0.25 Degrees   (%)
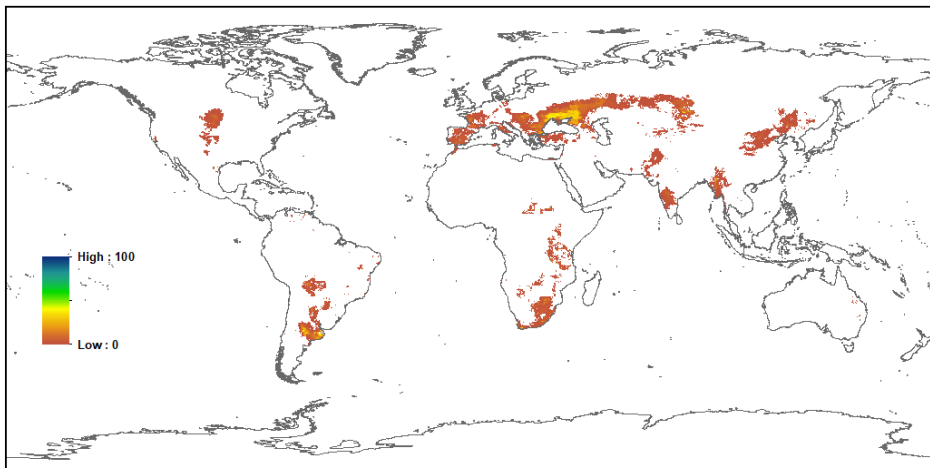
Figure 6.20: Global Current Day CLM5 CFT Mapping 2014: (a) Percent Sorghum; (b) Percent Sugarbeet; and (c) Percent Sunflower.

6.5 CLM5 Shifting Cultivation Historical Data

Table 6.4 Annual Global and Regional CLM5 Shifting Cultivation Area for 850 and 2014 in Millions of km^2, using the default CLM5 Domain File Land Mask.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
All PFTs	0.27	0.11	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.08	0.03	0.00
Tree	0.15	0.05	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.06	0.01	0.00
Shrub	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grass	0.12	0.07	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.01	0.00
Bare	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlEvgTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlEvgBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdEvgTrop	0.13	0.04	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.05	0.01	0.00
BrdEvgTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecTrop	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ShrEvgTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ShrDecTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ShrDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GrsC3Arc	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GrsC3	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GrsC4	0.10	0.05	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.01	0.00
Year 850												
All PFTs	0.05	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00
Tree	0.03	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Shrub	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grass	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Bare	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlEvgTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlEvgBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdEvgTrop	0.03	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
BrdEvgTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecTrop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ShrEvgTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ShrDecTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ShrDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GrsC3Arc	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GrsC3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GrsC4	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
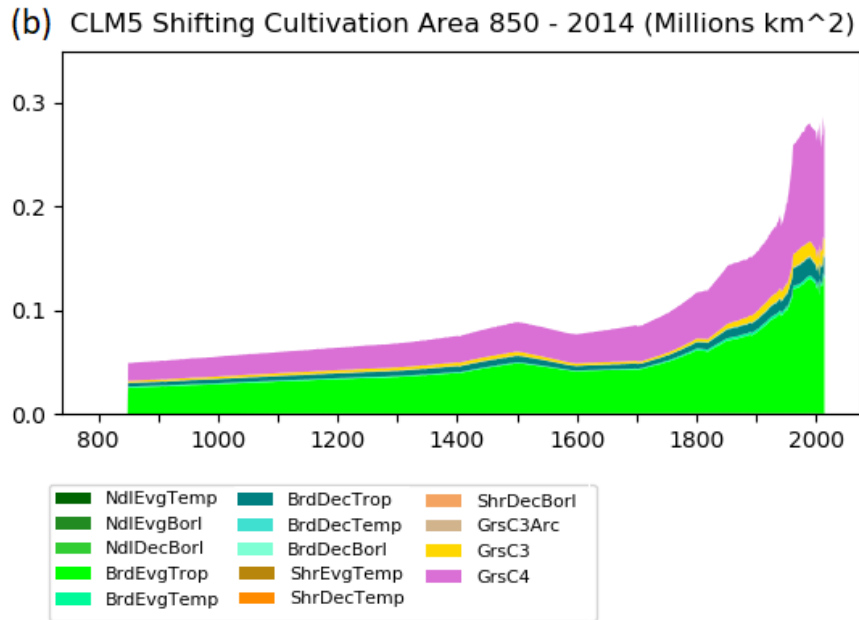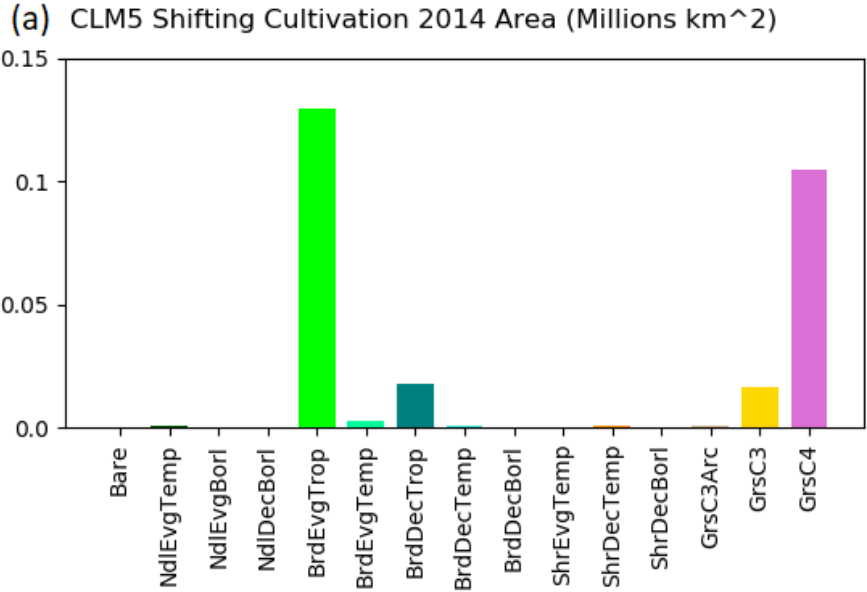
Figure 6.21: CLM5 Global Shifting Cultivation for Plant Functional Types for: (a) Current Day 2014; and (b) Annual values from 850 to 2014.

(a) CLM5 All PFTs Shifting Cultivation 2014 Raw 0.25 Degrees (%)
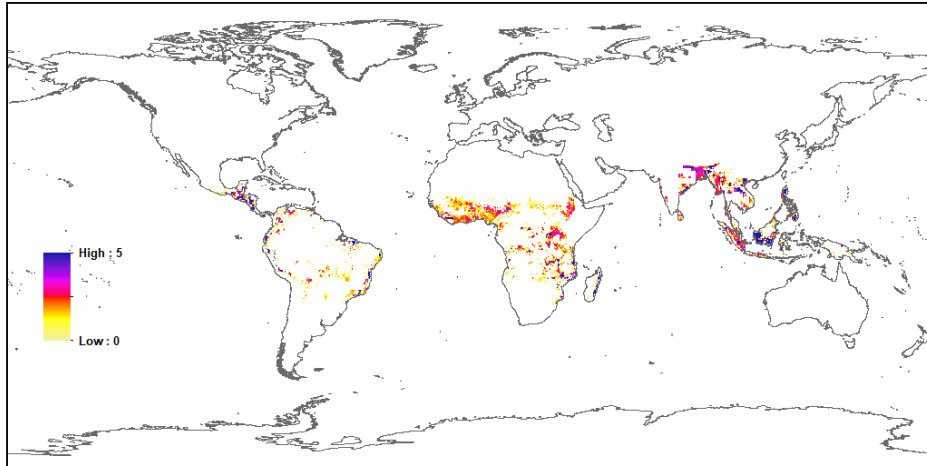
(b) CLM5 Tree PFTs Shifting Cultivation 2014 Raw 0.25 Degrees (%)
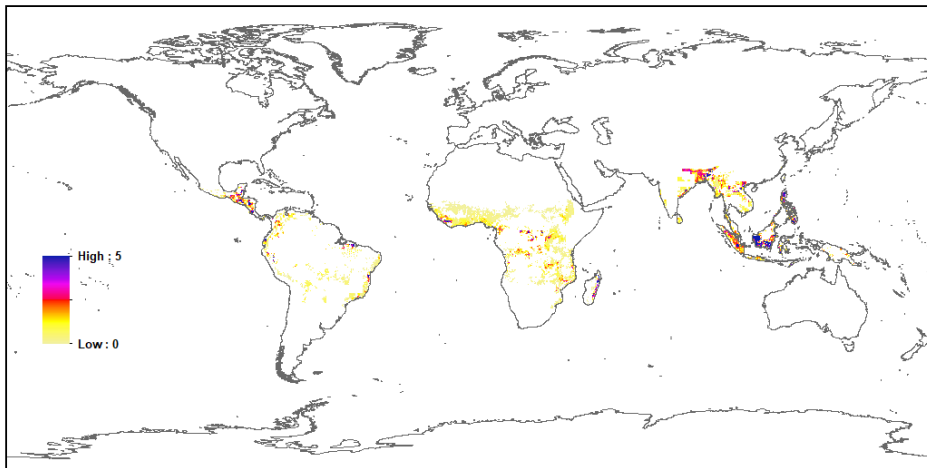
(c) CLM5 Grass PFTs Shifting Cultivation 2014 Raw 0.25 Degrees (%)
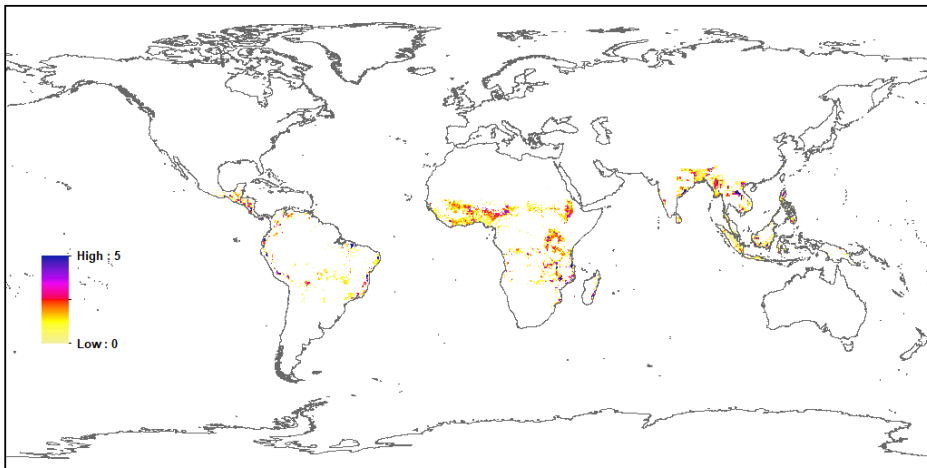
Figure 6.22: Global Current Day CLM5 Shifting Cultivation Mapping 2014:

6.6 CLM5 Irrigation and Fertilizer Historical Data

Table 6.5 Annual Global and Regional CLM5 Irrigated Crop Functional Type (CFT) Area for 850 and 2014 in Millions of km^2, using the default CLM5 Domain File Land Mask.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
All Irrigated	2.47	0.11	0.04	0.57	0.12	0.15	0.16	0.13	0.21	0.17	0.82	0.00
Temp Corn	0.17	0.00	0.00	0.09	0.01	0.02	0.00	0.01	0.04	0.00	0.01	0.00
Trop Corn	0.09	0.01	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.01	0.03	0.00
Wheat	0.61	0.03	0.01	0.14	0.05	0.05	0.01	0.07	0.04	0.00	0.22	0.00
Temp Soybean	0.06	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
Trop Soybean	0.04	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00
Rice	0.56	0.01	0.01	0.15	0.01	0.00	0.02	0.00	0.01	0.10	0.24	0.00
Cotton	0.15	0.00	0.00	0.03	0.02	0.00	0.00	0.00	0.03	0.00	0.06	0.00
Sugarcane	0.06	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.03	0.00
Other Crop	0.88	0.05	0.01	0.16	0.06	0.07	0.07	0.05	0.08	0.05	0.28	0.00
Year 850												
All Irrigated	0.04	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Temp Corn	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trop Corn	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wheat	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Temp Soybean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trop Soybean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rice	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cotton	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sugarcane	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other Crop	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.6 Annual Global and Regional CLM5 Nitrogen Fertilizer by Crop Functional Type (CFT) for 2014 in Amount in TgN/year and Rate in KgN/Ha, using the default CLM5 Domain File Land Mask.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Amount 2014												
All Fertilizer	93.1	3.9	4.1	23.7	2.0	12.6	8.9	3.0	16.7	5.8	12.9	0.0
Temp Corn	13.0	0.1	0.0	3.8	0.1	2.4	0.6	0.1	5.8	0.0	0.1	0.0
Trop Corn	4.8	0.8	0.0	0.5	0.0	0.0	2.7	0.0	0.1	0.6	0.3	0.0
Wheat	26.6	0.6	2.2	6.1	0.9	4.7	1.0	1.7	6.2	0.0	3.3	0.0
Temp Soybean	1.0	0.0	0.0	0.6	0.0	0.1	0.1	0.0	0.3	0.0	0.0	0.0
Trop Soybean	0.6	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.3	0.0
Rice	14.1	0.2	0.4	5.9	0.0	0.0	0.6	0.1	0.2	2.8	3.9	0.0
Cotton	3.3	0.2	0.1	0.8	0.0	0.1	0.2	0.0	0.8	0.0	1.2	0.0
Sugarcane	2.3	0.1	0.0	0.2	0.0	0.0	1.1	0.0	0.0	0.2	0.5	0.0
Other Crop	30.8	2.1	1.5	6.6	1.0	5.5	2.6	1.1	4.2	2.1	4.3	0.0
Rate 2014												
All Fertilizer	62.8	14.4	82.3	195.2	12.3	74.6	46.3	87.5	84.5	59.7	65.1	32.3
Temp Corn	139.1	132.8	206.6	186.6	5.9	125.6	89.6	125.8	169.1	0.0	58.5	0.0
Trop Corn	49.9	18.5	29.5	193.5	0.0	0.0	72.5	139.3	131.2	92.0	32.2	90.0
Wheat	76.1	31.3	83.6	208.3	11.9	78.4	53.5	100.0	97.5	67.3	78.6	0.0
Temp Soybean	16.0	24.1	33.0	81.3	7.8	21.8	4.1	4.4	8.3	0.0	3.6	0.0
Trop Soybean	15.6	6.3	2.0	84.8	0.0	0.0	4.8	2.6	5.7	24.5	45.5	0.0
Rice	88.8	14.4	158.8	221.5	22.2	65.3	46.8	75.9	70.5	64.5	72.5	0.0
Cotton	69.1	15.4	123.0	193.0	3.5	66.7	38.6	53.5	83.1	8.0	86.4	0.0
Sugarcane	99.9	66.7	122.5	317.3	71.5	71.1	84.4	120.9	108.6	91.6	125.9	0.0
Other Crop	46.4	11.0	70.0	192.8	13.4	62.2	42.5	72.0	69.7	48.7	54.8	31.6
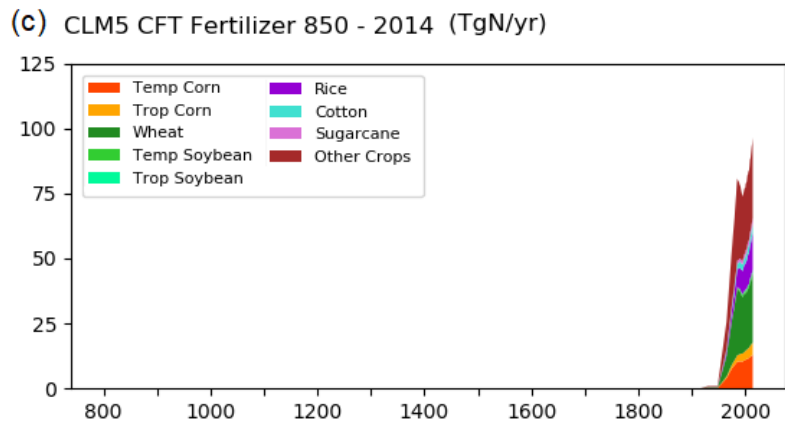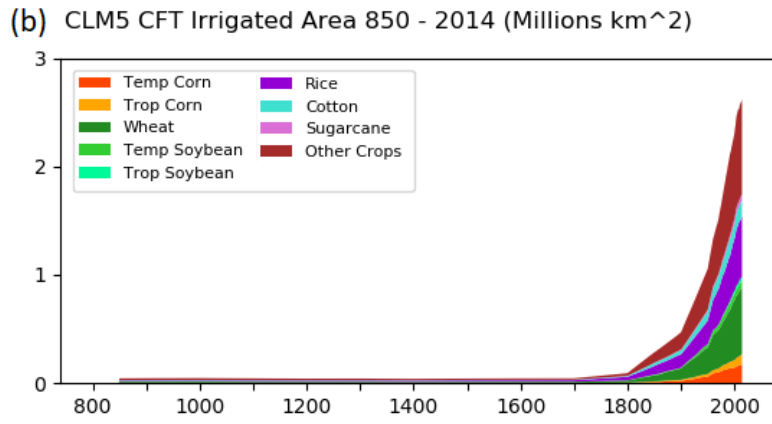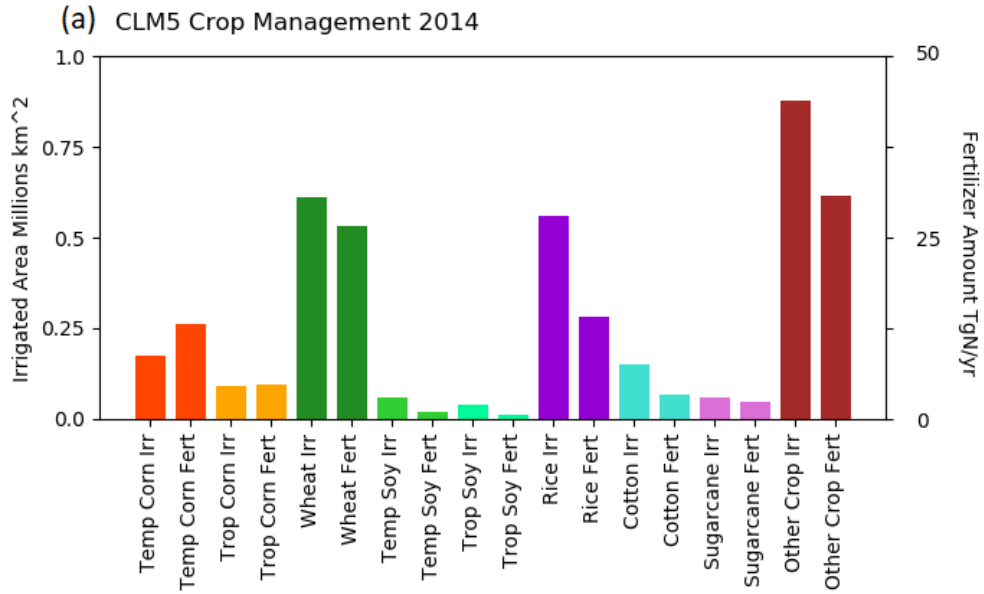
Figure 6.23: CLM5 Global Agricultural Management for Crop Functional Type Irrigation and Nitrogen Fertilizer for: (a) Current Day 2014; (b) Annual values of Irrigation from 850 to 2014; and (c) Annual values of Nitrogen Fertilizer from 850 to 2014.

(a) CLM5 All Irrigated Crops 2014 Raw 0.25 Degrees (%)
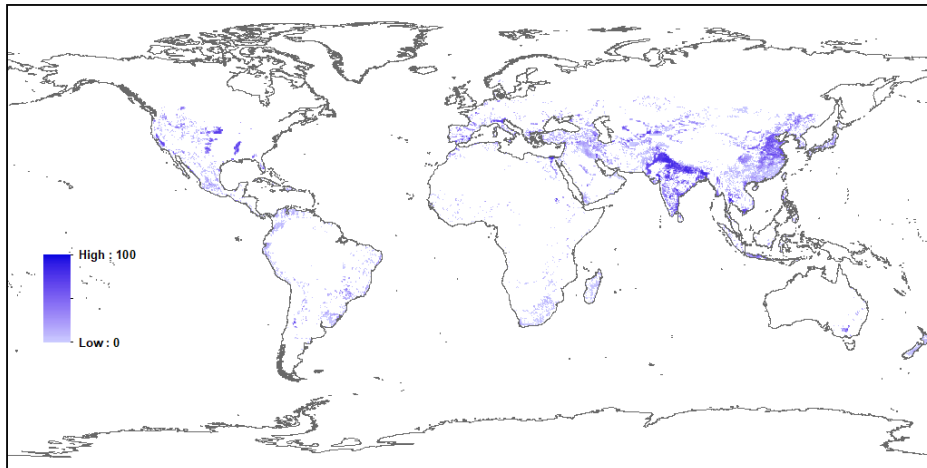
(b) CLM5 All Crop Fertilizer 2014 Raw 0.25 Degrees kgN/ha
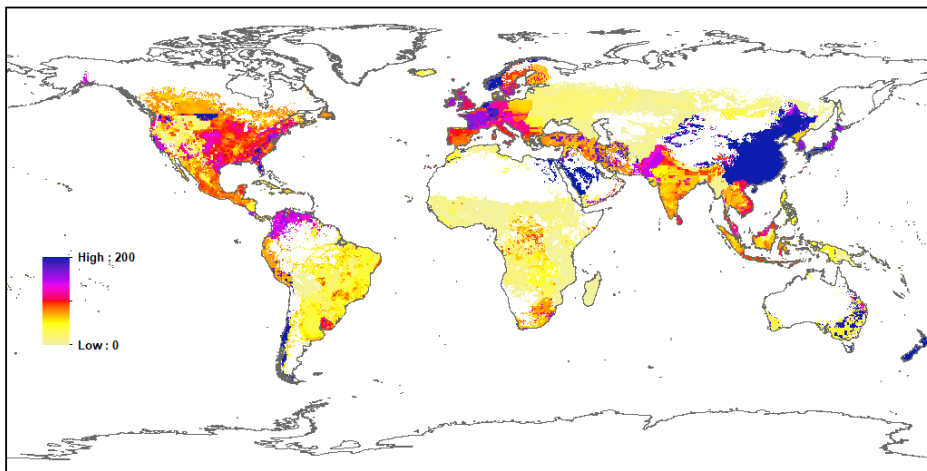
Figure 6.24: Global Current Day CLM5 Crop Management Mapping 2014 for: (a) Irrigated Crops as percentage of grid cell; and (b) Average Crop Nitrogen Fertilizer in kgN/ha.

7.6 CLM5 Wood Harvest Historical Data

Table 7.6 Annual Global and Regional CLM5 Wood Harvest for 850 and 2014 in PgC/yr, using the default CLM5 Domain File Land Mask.

	Glob	AFR	APD	EAS	ERA	EUR	LAC	MEA	NAM	SEA	SAS	OTH
Year 2014												
All Tree	1.49	0.29	0.03	0.15	0.10	0.19	0.20	0.00	0.19	0.15	0.20	0.00
NdlEvgTemp	0.33	0.00	0.01	0.08	0.01	0.08	0.00	0.00	0.10	0.00	0.04	0.00
NdlEvgBorl	0.13	0.00	0.00	0.01	0.05	0.04	0.00	0.00	0.03	0.00	0.00	0.00
NdlDecBorl	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdEvgTrop	0.58	0.20	0.00	0.01	0.00	0.00	0.17	0.00	0.00	0.13	0.08	0.00
BrdEvgTemp	0.12	0.01	0.01	0.02	0.00	0.00	0.02	0.00	0.01	0.01	0.03	0.00
BrdDecTrop	0.09	0.07	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00
BrdDecTemp	0.18	0.00	0.00	0.03	0.00	0.06	0.00	0.00	0.04	0.00	0.03	0.00
BrdDecBorl	0.03	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Year 850												
All Tree	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlEvgTemp	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlEvgBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NdlDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdEvgTrop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdEvgTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecTrop	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecTemp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BrdDecBorl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
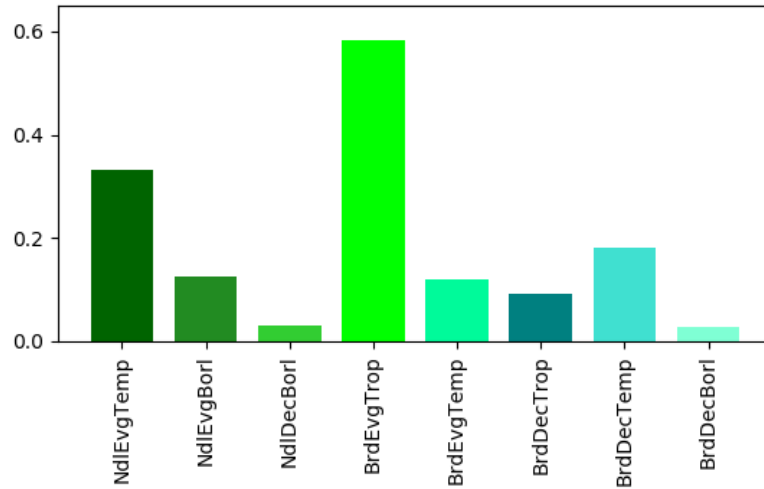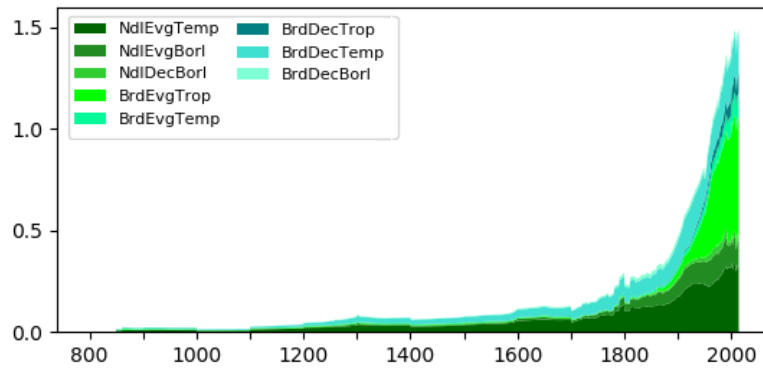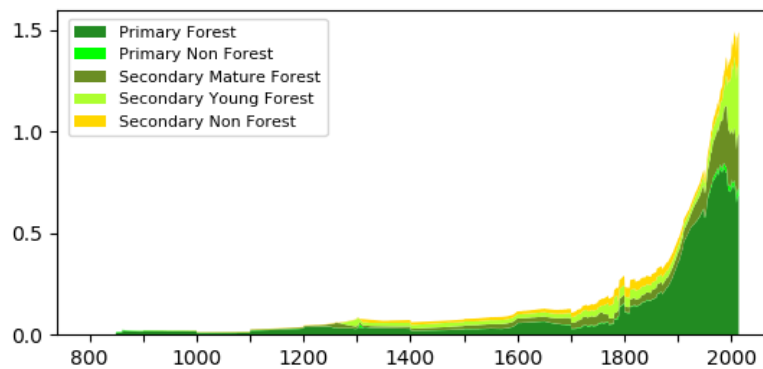
Figure 6.25: CLM5 Global Forest Management Wood Harvest Biomass Amount for: (a) Current Day 2014 for Tree Plant Functional Types; (b) Annual values of Wood Harvest Amount by for Tree Plant Functional Types from 850 to 2014; and (c) Annual values of Wood Harvest Biomass Amount from LUH2 Land Class for 850 to 2014.

(a) CLM5 All Trees Wood Harvest Amount 2014 — Millions kgC
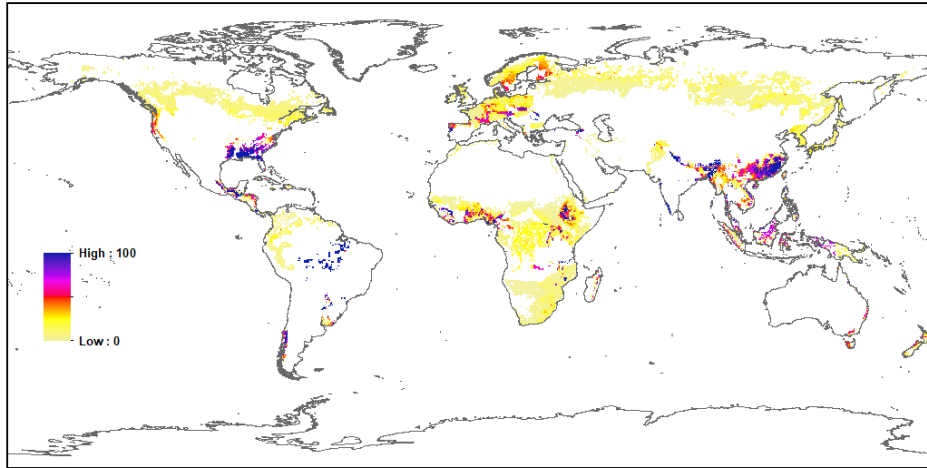
(b) CLM5 Needleleaf Wood Harvest Amount 2014 — Millions kgC
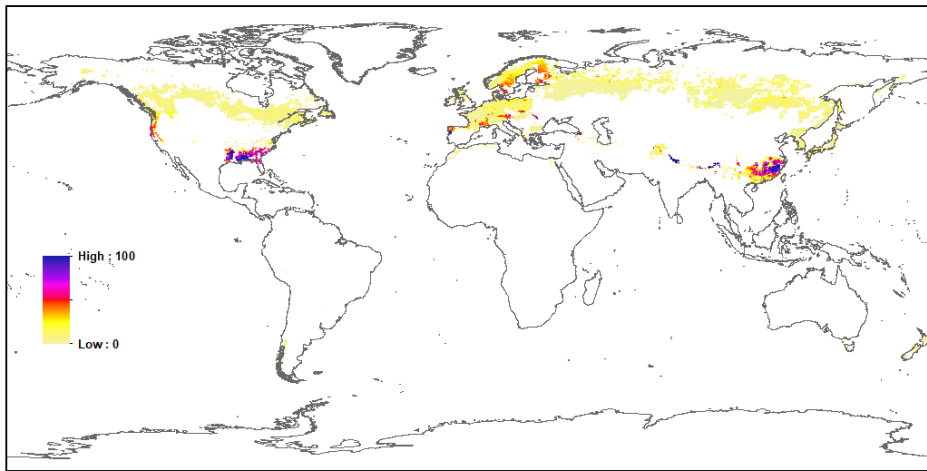
(c) CLM5 Broadleaf Wood Harvest Amount 2014 — Millions kgC
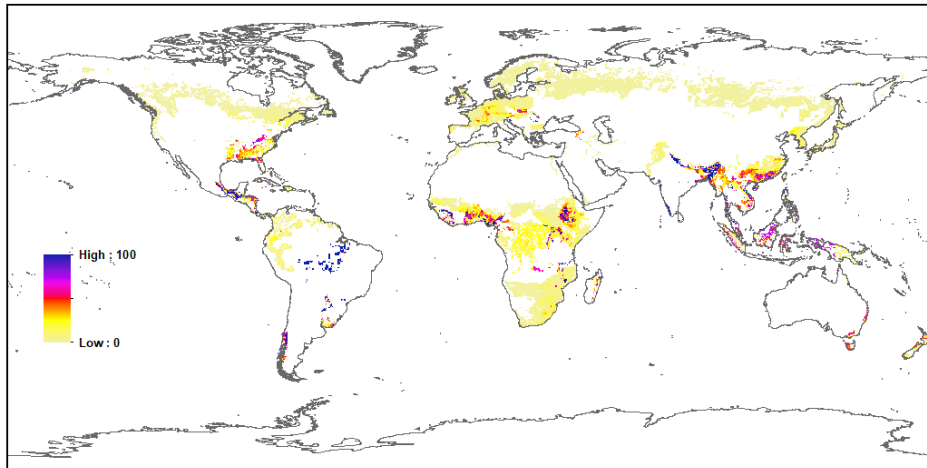
Figure 6.26: Global Current Day CLM5 Wood Harvest Biomass Amount Mapping 2014 for: (a) All Tree PFTs; (b) Needleleaf Tree PFTs; and (c) Broadleaf Tree PFTs.

CHAPTER 7.

CLM5 ALL CMIP6 TIME SERIES

7.1 CLM5 All CMIP6 Time Series Overview

CLM5 provides a consistent framework for Land Cover along with annual Land Use and Land Cover Change (LULCC) time series data from 6000 BCE through to the current day, with a range of baseline and mitigation Shared Socioeconomic Pathway (SSP) future scenarios to 2100. There are three additional extensions to 2300 without LULCC provided for key SSPs. The CLM5 LULCC prescription for each of these time series are developed from data in LUH2 format as described in Chapter 5 combined with CLM5 Land Data Tool and mksurf tools described in Chapter 6. This Chapter describes the nature and details of each of these periods in the CLM5 LULCC time series starting from prehuman disturbance.

The CLM5 current day climate land cover time series begins with the Earth in a pre-anthropogenic state (No-Anthro) for the year 6000 BCE. This is a reconstruction of the potential vegetation of the Earth prior to human disturbance under a current day climate. From the No-Anthro state, CLM5 provides annual LULCC data for the PMIP transient Holocene from 6000 BCE through 849 CE as described in *** PMIP Reference ***. Following the Transient Holocene, CLM5 provides annual LULCC data for the PMIP Last Millennium from 850 – 1849 CE as described in *** PMIP Reference ***. The CMIP6 Historical DECK period is prescribed in CLM5 through annual LULCC for 1850 – 2014 directly from the Historical LUH2 time series described in Lawrence et al. (2017) and Hurtt et al. (2020). The CLM5 LULCC framework provides time series data for the five baseline and three mitigation SSPs to 2100 as described in Hurtt et al. (2020). In addition, there are three SSP extensions to 2300 that have no LULCC. For simplicity and clarity, we only include the five baseline SSPs to 2100 in this Chapter.

The future SSP scenarios describe alternative path-dependent evolutions of the energy system consistent with their SSP narratives and their associated challenges for mitigation and adaptation. The SSPs depict vastly different energy futures, featuring a wide range of possible energy demand developments and energy supply structures. These differences emerge due to a combination of assumptions with respect to the main drivers of the energy system, including technological change, economic growth, emergence of new energy services, energy intensity of services, and assumptions with respect to costs and availability of future fossil fuel resources and their alternatives (Riahi et al. 2017) and (O'Neill et al. 2016). *** Add References ***

(a) Holocene LastM Global Tree Millions km^2
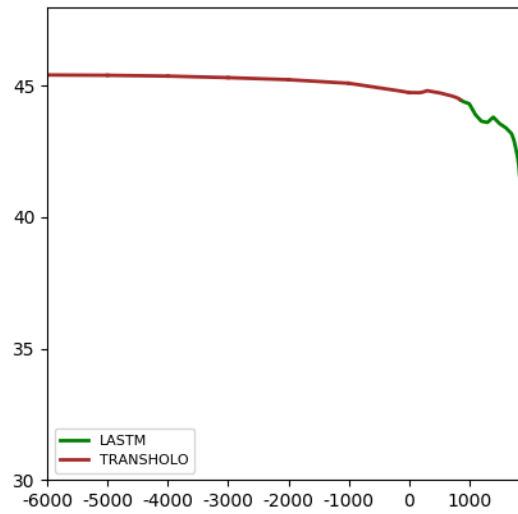
(b) Hist SSP Global Tree Millions km^2
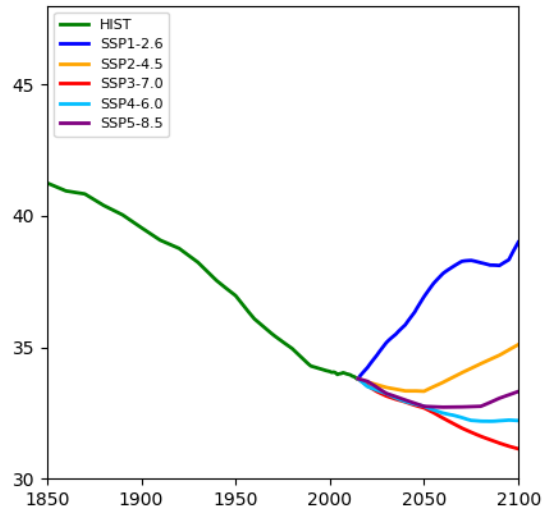
(c) Holocene LastM Global Grass Millions km^2
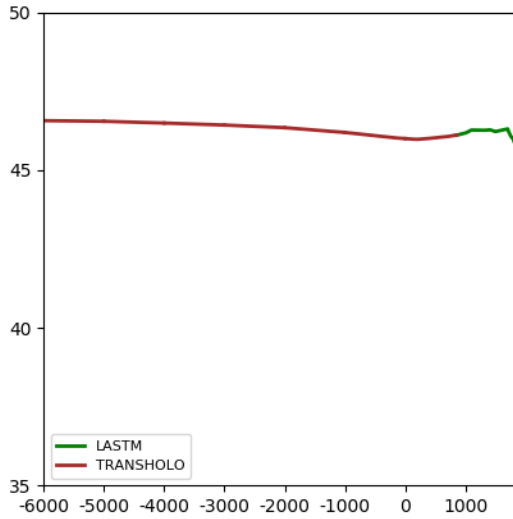
(d) Hist SSP Global Grass Millions km^2
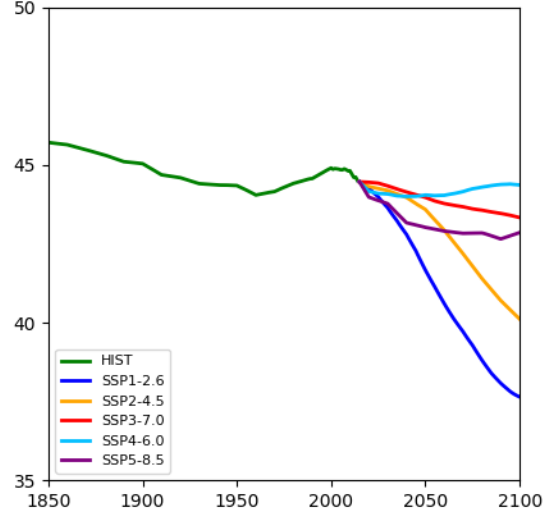
(e) Holocene LastM Global Shrub Millions km^2
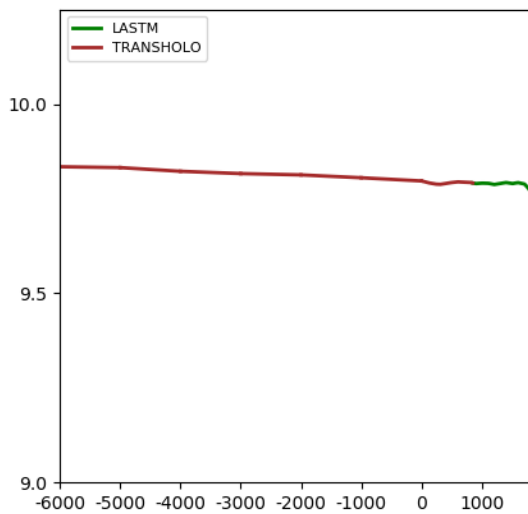
(f) Hist SSP Global Shrub Millions km^2
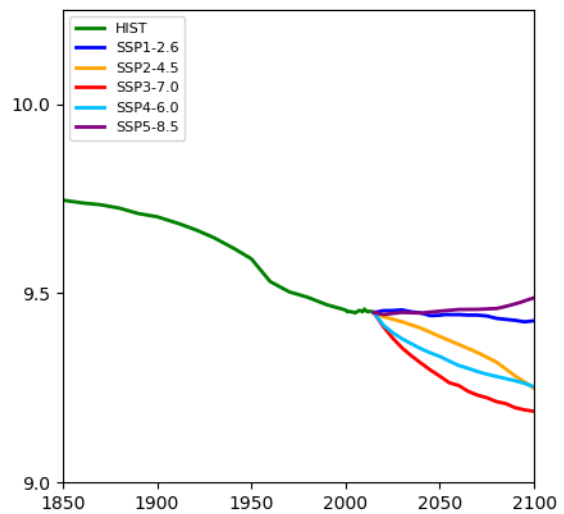
(a) Holocene LastM Global Crop Millions km^2
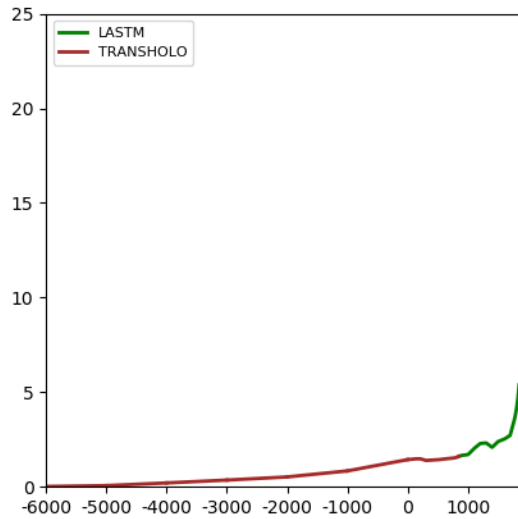
(b) Hist SSP Global Crop Millions km^2
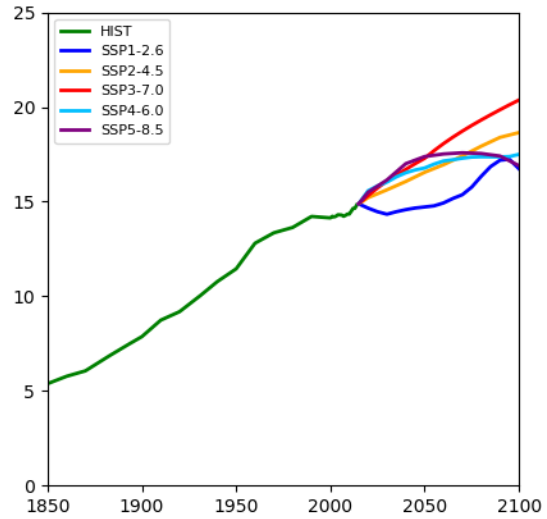
(c) Holocene LastM Global Irrig Crop Millions km^2
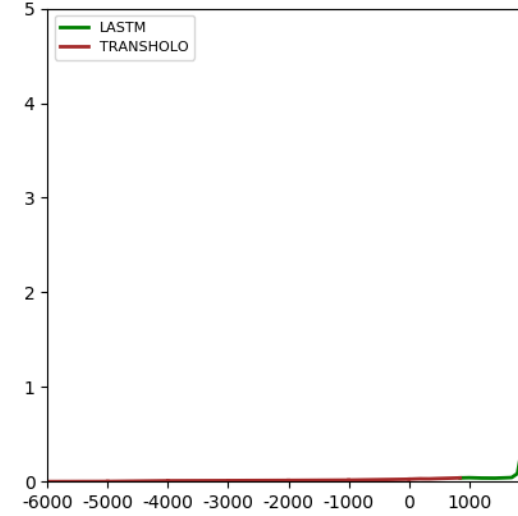
(d) Hist SSP Global Irrig Crop Millions km^2
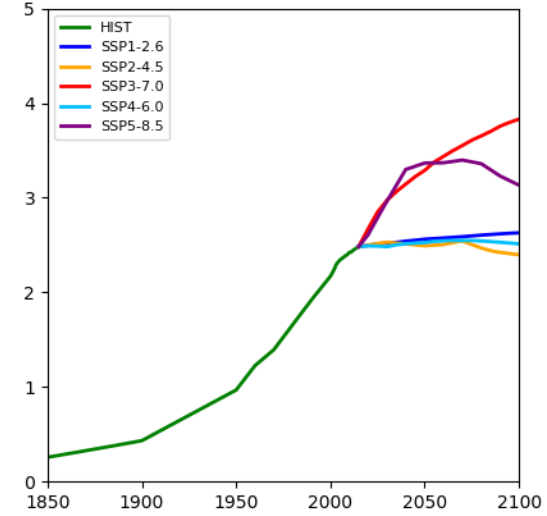
(e) Holocene LastM Global Fertilizer TgN
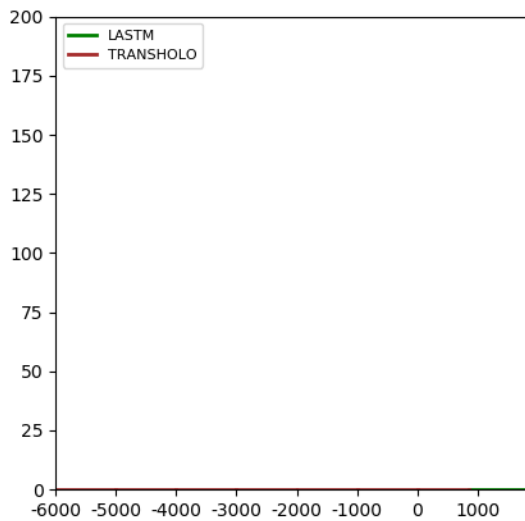
(f) Hist SSP Global Fertilizer TgN

(a) Holocene LastM Global Wood Harvest PgC
(b) Hist SSP Global Wood Harvest PgC
(c) Holocene LastM Shifting (Tree) Millions km^2
(d) Hist SSP Global Shifting (Tree) Millions km^2
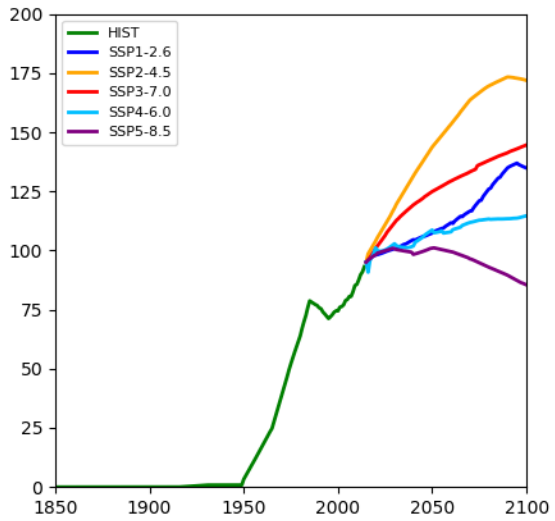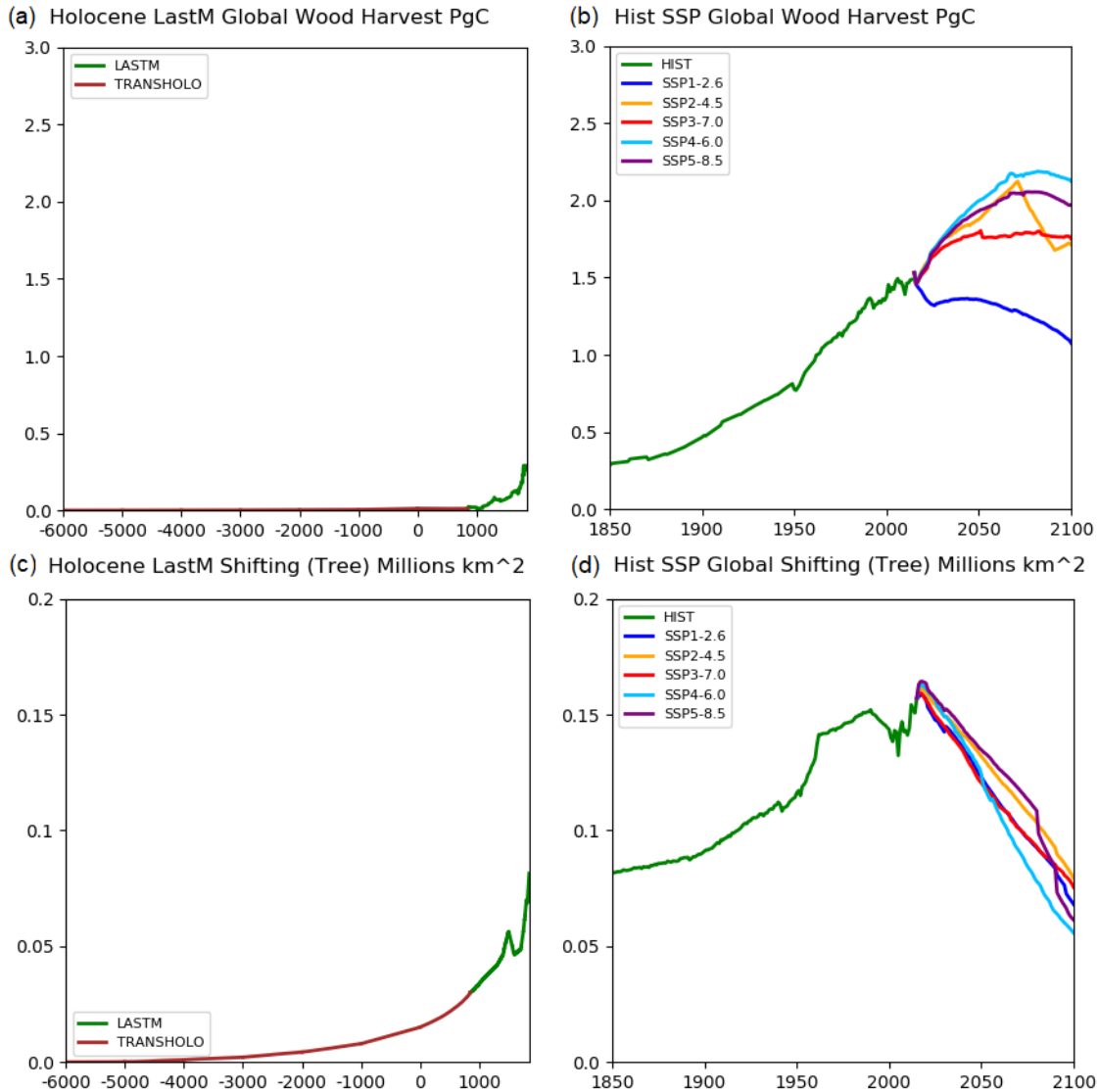
The detailed prescription of the CMIP6 SSP climate forcings is derived through Integrated Assessment Model (IAM) studies by various modeling groups as detailed by O'Neill et al. (2017) and Tebaldi et al. (2021) combined with secondary post processing of each forcing type. The pathways for the energy and land-use systems in the SSP scenarios translate into a wide range of green house gas and aerosol emissions and resultant concentrations. For CLM5 the LULCC associated with each SSP has been generated by land use models in the individual IAMs Calvin et al. (2018). This data has been harmonized and translated into the LUH2 format as described by Hurtt et al. (2020).

7.2 CLM5 No Anthro

The CLM5 No-Anthro Land Cover is generated from the CLM5 Land Data Tool with the first year (850 CE) of the LUH2 historical time series used as a template with the 850 Land Use removed and replaced by the most consistent LUH2 Primary Forest and Non-Forest land cover classes based on coexisting classes and nearest neighbors in the absence of coexistence. The vegetation

distributions and bioclimatic rules are taken from current day which is highly consistent with the climate of the Mid Holocene that 6000 BCE represents.

(a) CLM5 Bare Soil No Anthro Raw 0.25 Degrees   (%)
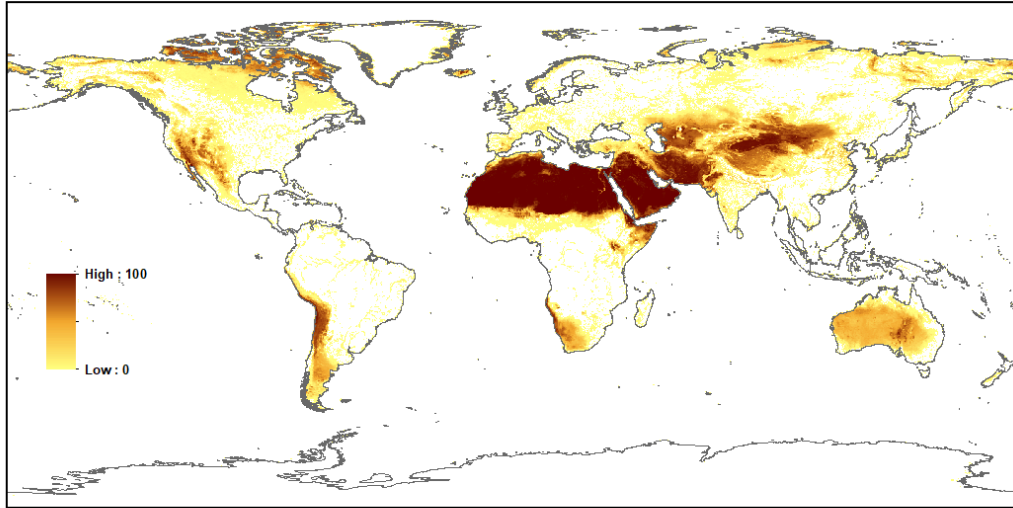
(b) CLM5 Needleleaf Evergreen Temperate No Anthro Raw 0.25 Deg   (%)
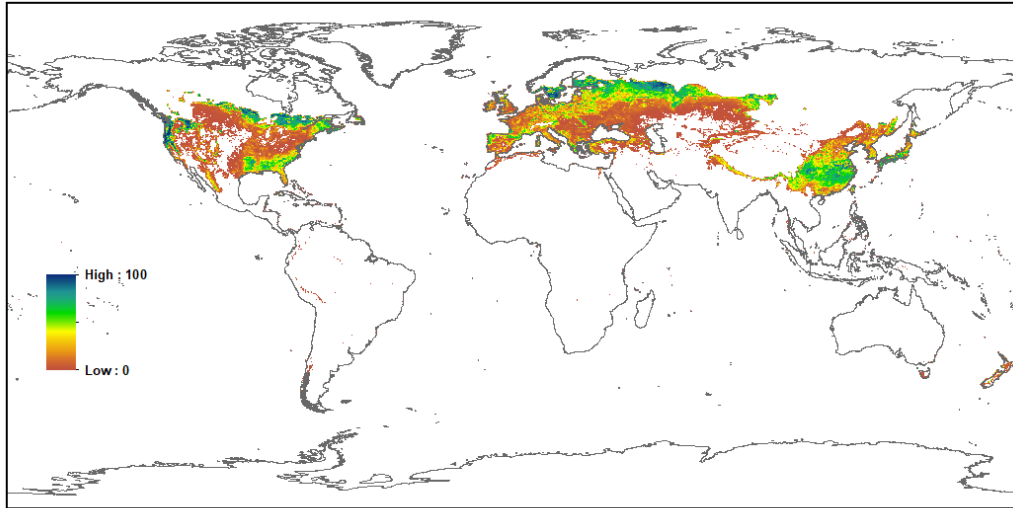
(c) CLM5 Needleleaf Evergreen Boreal No Anthro Raw 0.25 Degrees   (%)
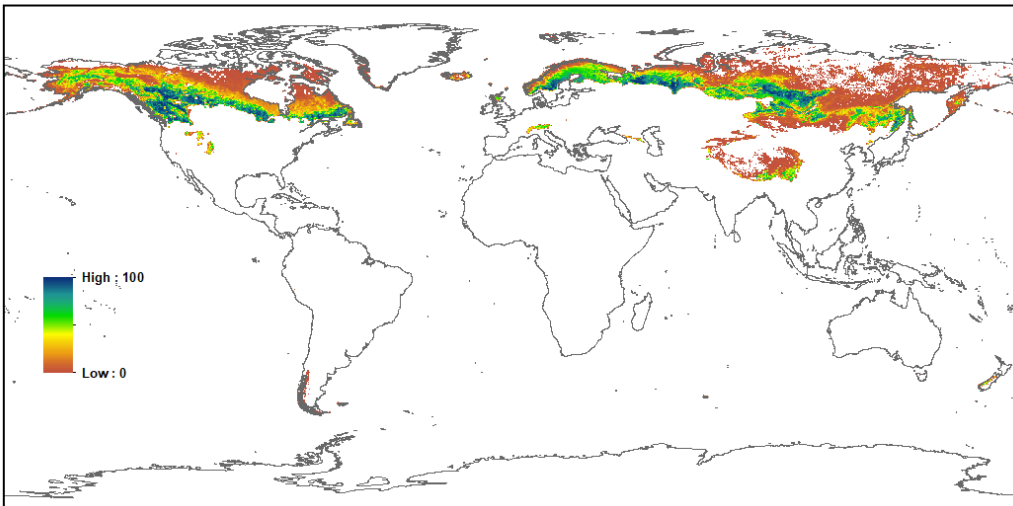
(a) CLM5 Needleleaf Deciduous Boreal No Anthro Raw 0.25 Degrees   (%)
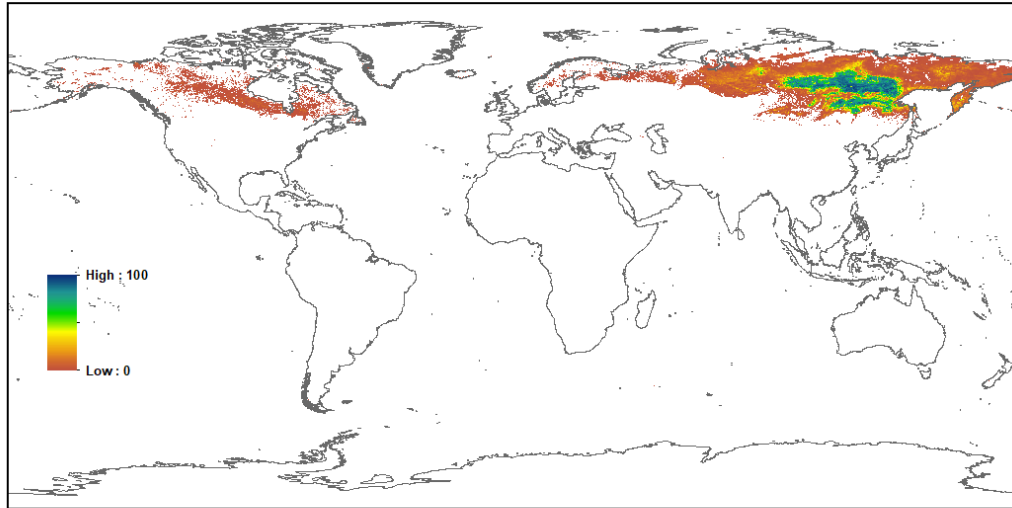
(b) CLM5 Broadleaf Evergreen Tropical No Anthro Raw 0.25 Degrees   (%)
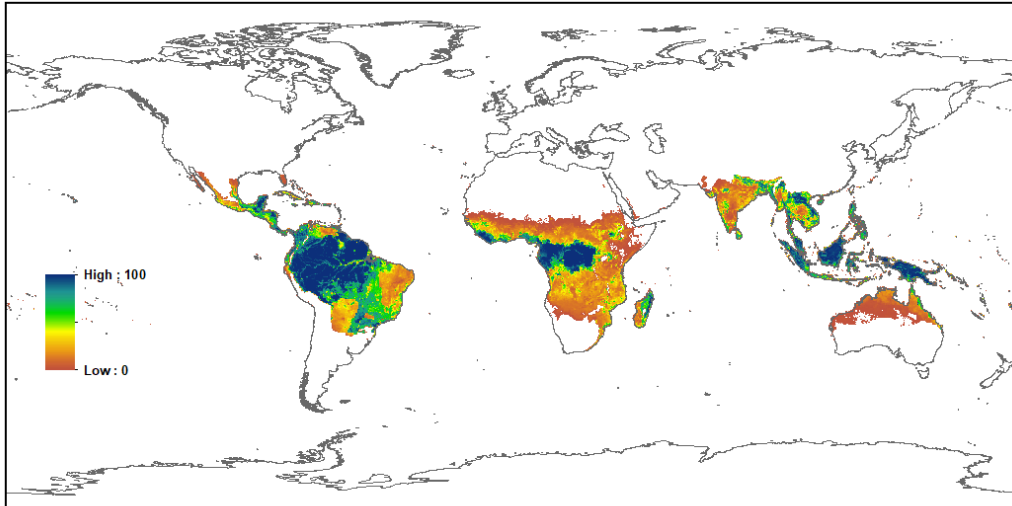
(c) CLM5 Broadleaf Evergreen Temperate No Anthro Raw 0.25 Degrees (%)
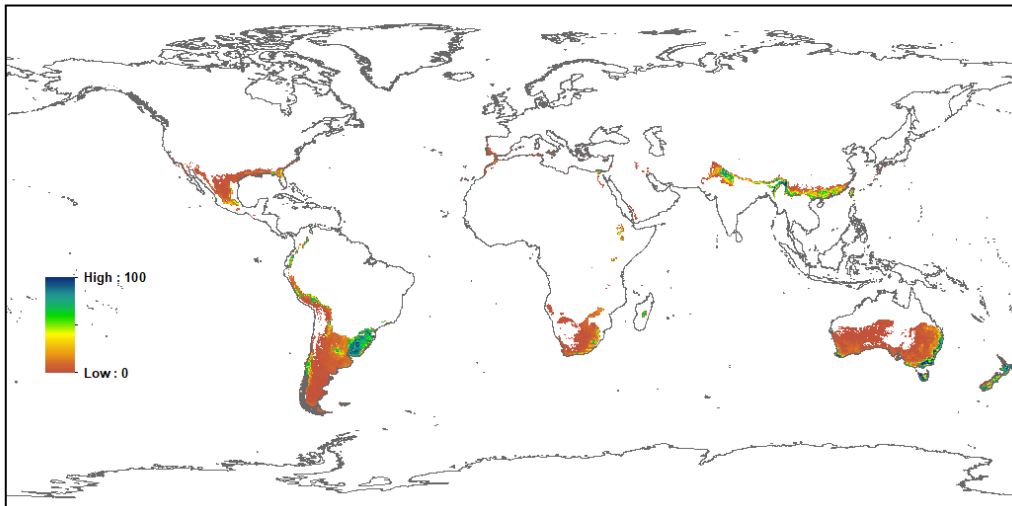
(a) CLM5 Broadleaf Deciduous Tropical No Anthro Raw 0.25 Degrees   (%)
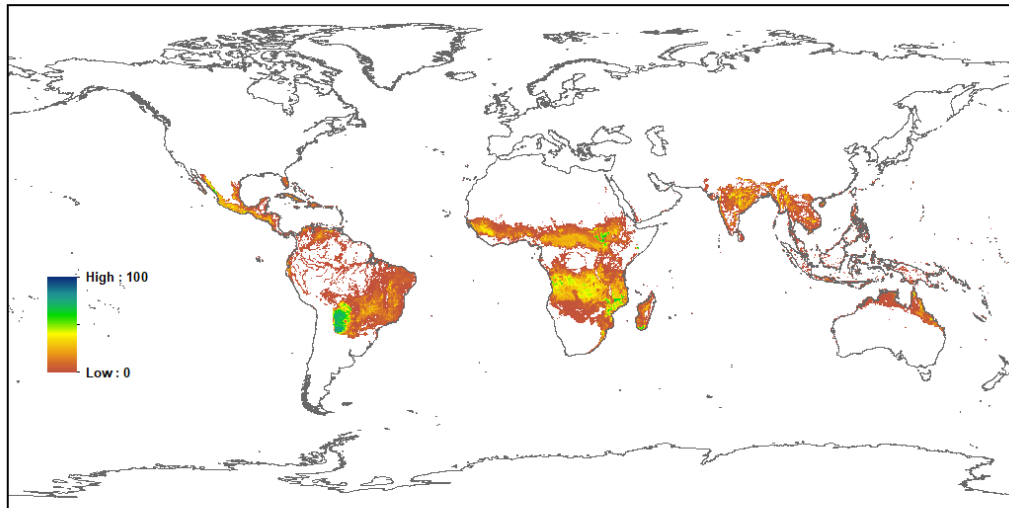
(b) Broadleaf Deciduous Temperate No Anthro Raw 0.25 Degrees   (%)
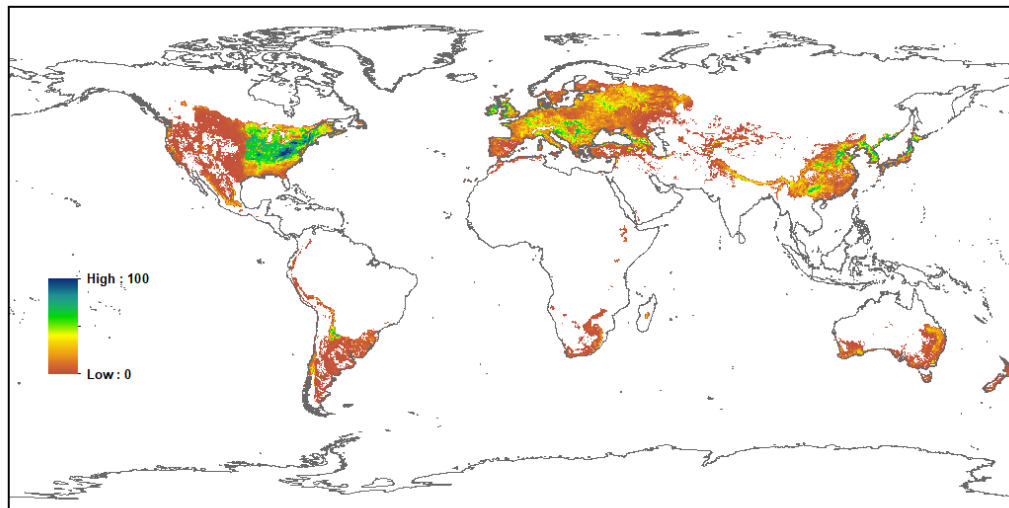
(c) CLM5 Broadleaf Deciduous Boreal No Anthro Raw 0.25 Degrees   (%)
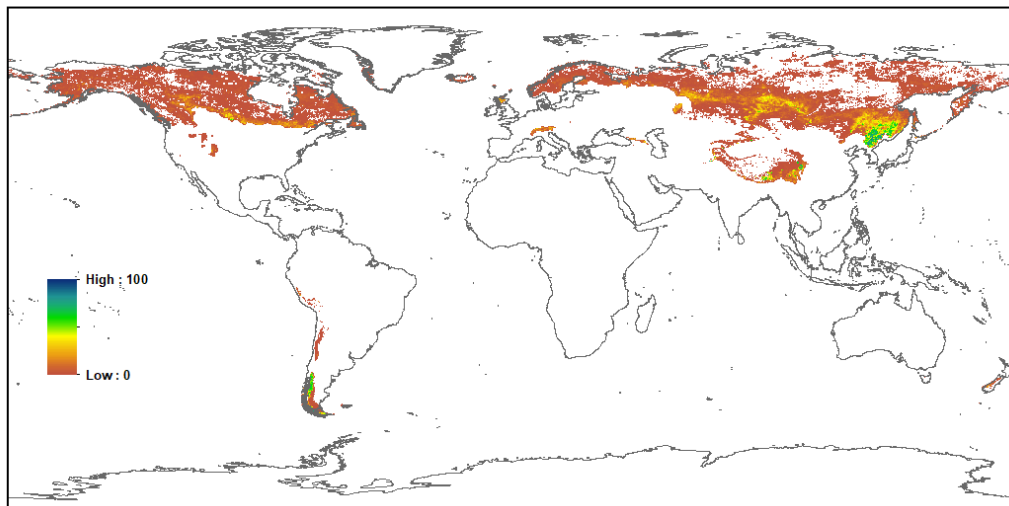
(a) CLM5 Shrub Deciduous Temperate No Anthro Raw 0.25 Degrees   (%)
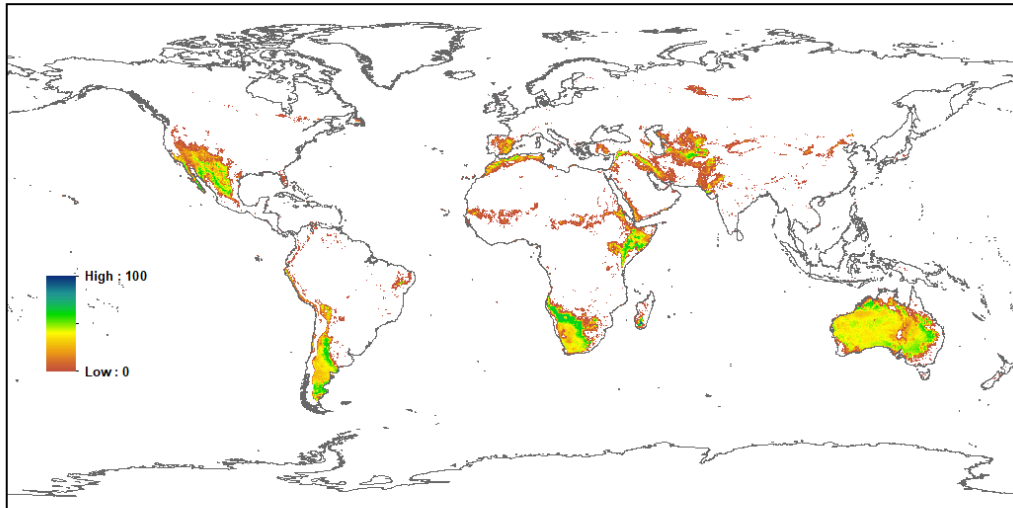
(b) CLM5 Shrub Deciduous Boreal No Anthro Raw 0.25 Degrees   (%)
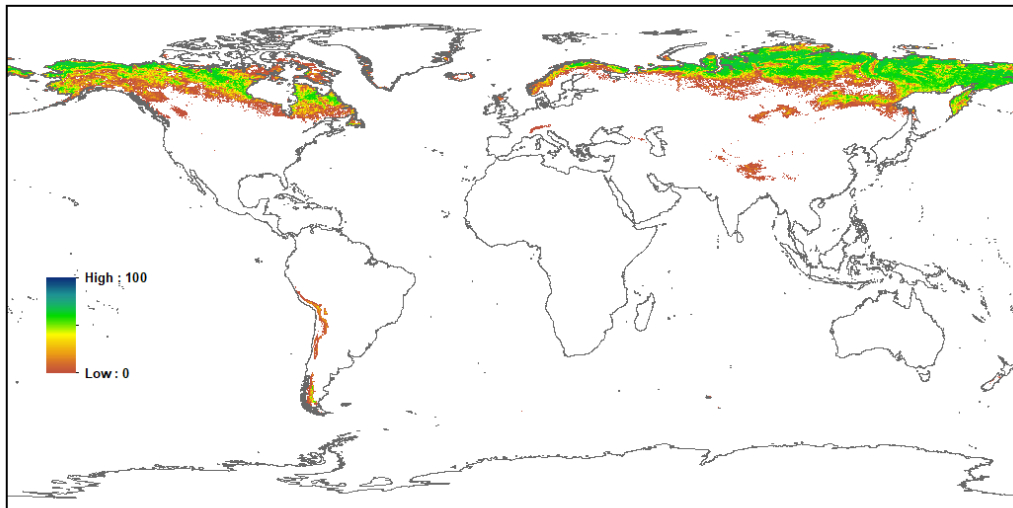
(c) CLM5 Grass C3 Arctic No Anthro Raw 0.25 Degrees   (%)
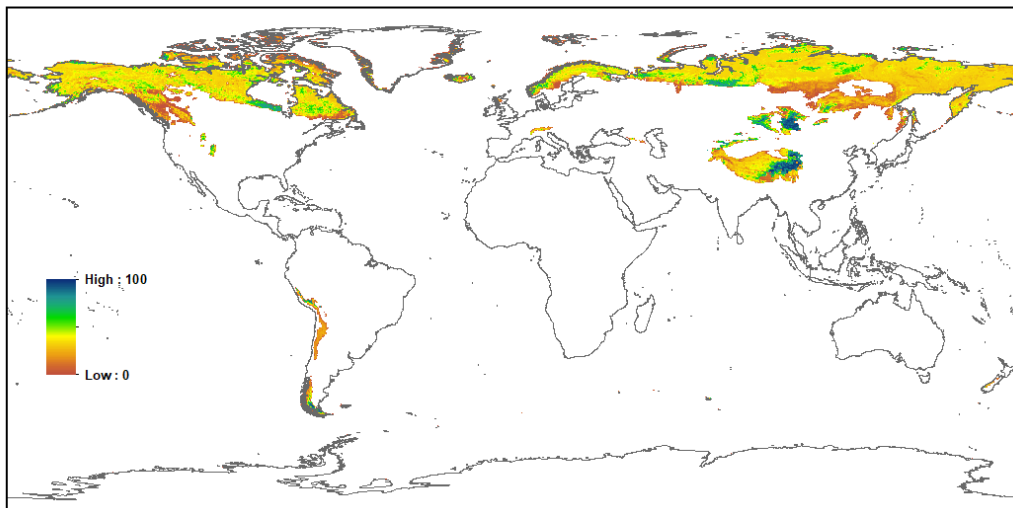
(a) CLM5 Grass C3 No Anthro Raw 0.25 Degrees   (%)
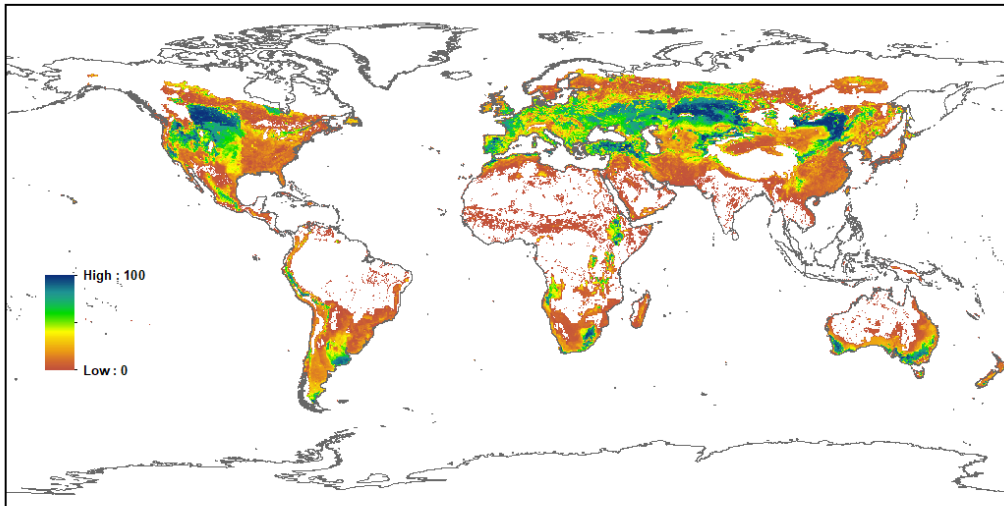
(b) CLM5 Grass C4 No Anthro Raw 0.25 Degrees   (%)
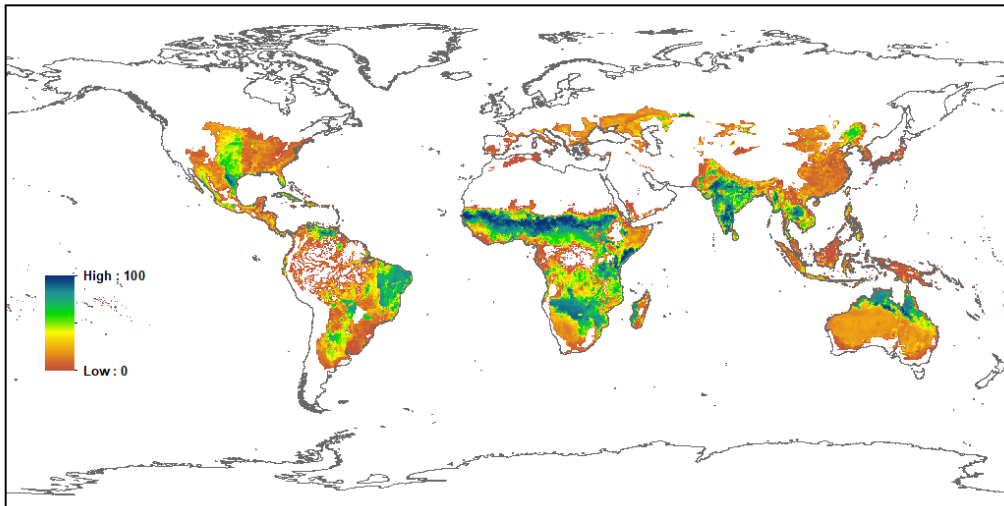
7.3 CLM5 Transient Holocene

The CLM5 Transient Holocene annual LULCC time series is generated from the CLM5 Land Data Tool with all years reconstructed for the period 6000 BCE – 850 CE with a version of the LUH2 time series generated for the PMIP transient Holocene project *** PMIP Reference ***. As no CMIP6 LUH2 data existed for this period from LUMIP the underlying LUH2 time series data is generated from first 10 years of the LUH2 historical time series (850 – 859 CE) combined with the HYDE 3.2 Holocene land use time series data from Klein Goldewijk et al. (2017). These data are highly consistent given that HYDE is a primary component of the historical reconstruction at this time (Hurtt et al. 2020).

(a) CLM5 Trans Holo Change in Tree PFTs 850ce - 6000bce   (%)
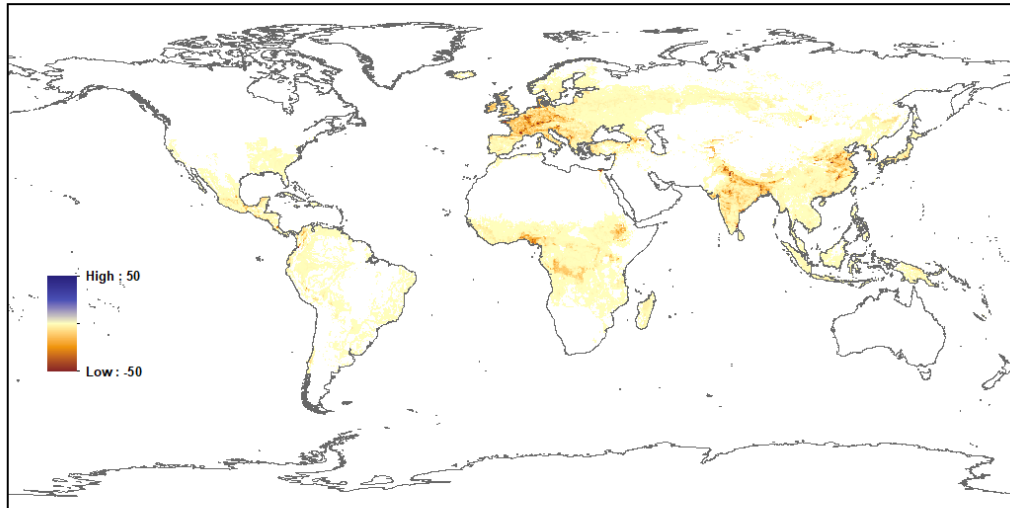
(b) CLM5 Trans Holo Change in Grass PFTs 850ce - 6000bce   (%)
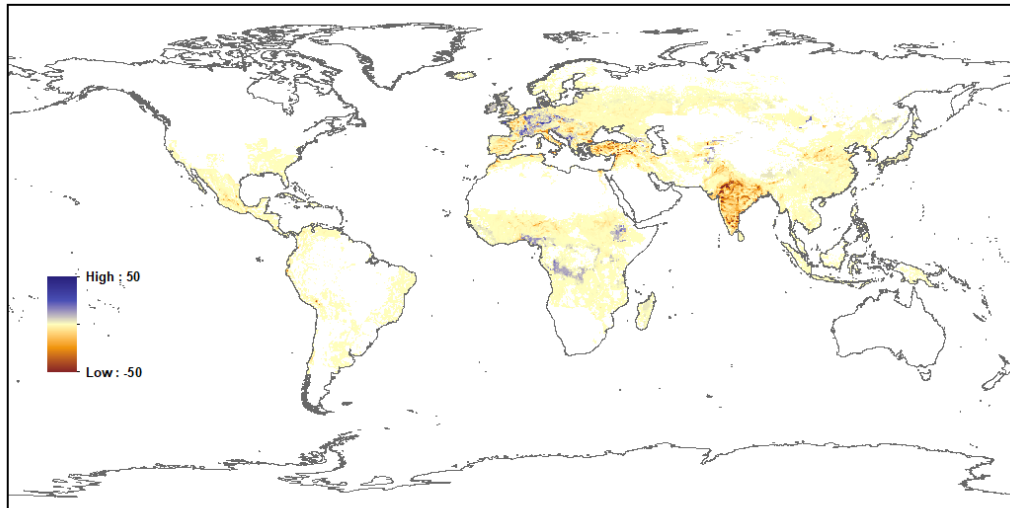
(c) CLM5 Trans Holo Change in Crop CFTs 850ce - 6000bce   (%)
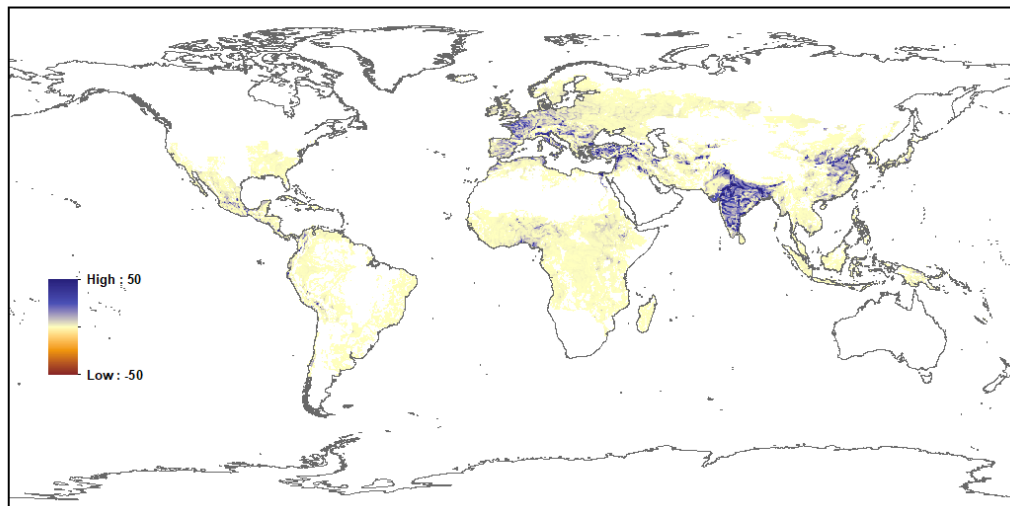
7.4 CLM5 Last Millennium

The CLM5 Last Millennium LULCC data is generated directly from the annual LUH2 LULCC time series data for 850 – 1850 period as described in Lawrence et al. (2017) and Hurtt et al. (2020). The LUH2 data is combined with the CLM5 Land Data Tool.
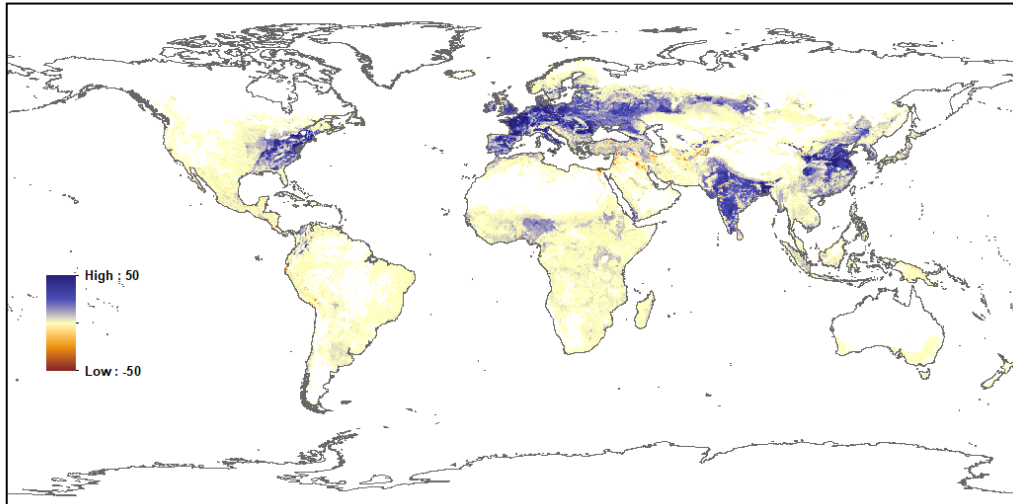
(a) CLM5 Last Millennium Change in Tree PFTs 1850 - 850 (%)
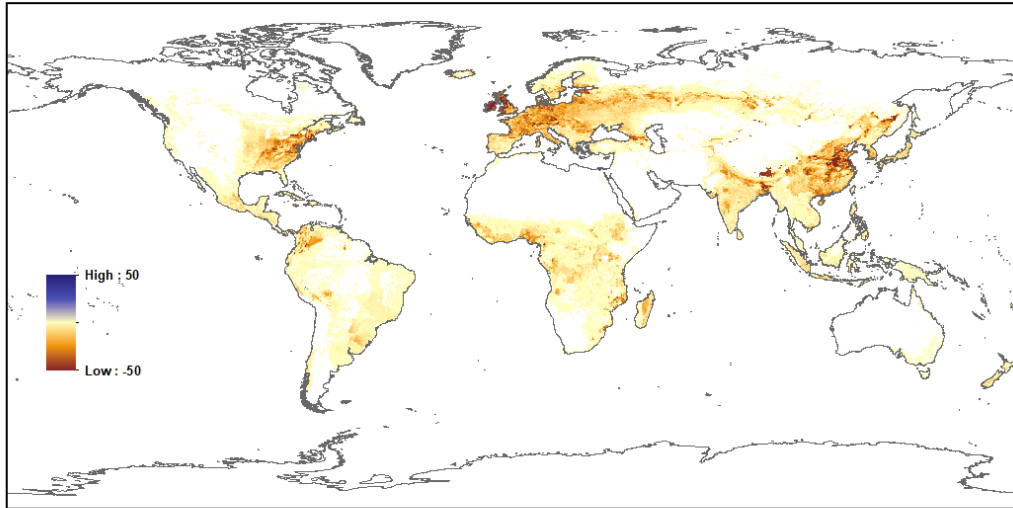
(b) CLM5 Last Millennium Change in Grass PFTs 1850 - 850 (%)
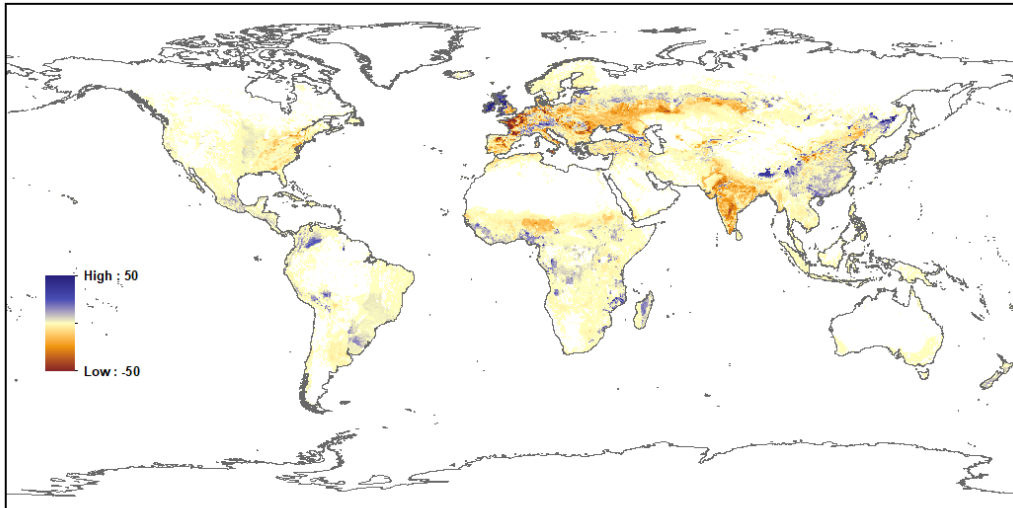
(c) CLM5 Last Millennium Change in Crop CFTs 1850 - 850 (%)

7.5 CLM5 Historical

The CLM5 Historical period is generated directly from the annual LUH LULCC time series data for 1850 – 2014 as described in Lawrence et al. (2017) and Hurtt et al. (2020). The Historical LUH2 time series data is combined with CLM5 Land Data Tool.

(a) CLM5 Historical Change in Tree PFTs 2014 - 1850   (%)
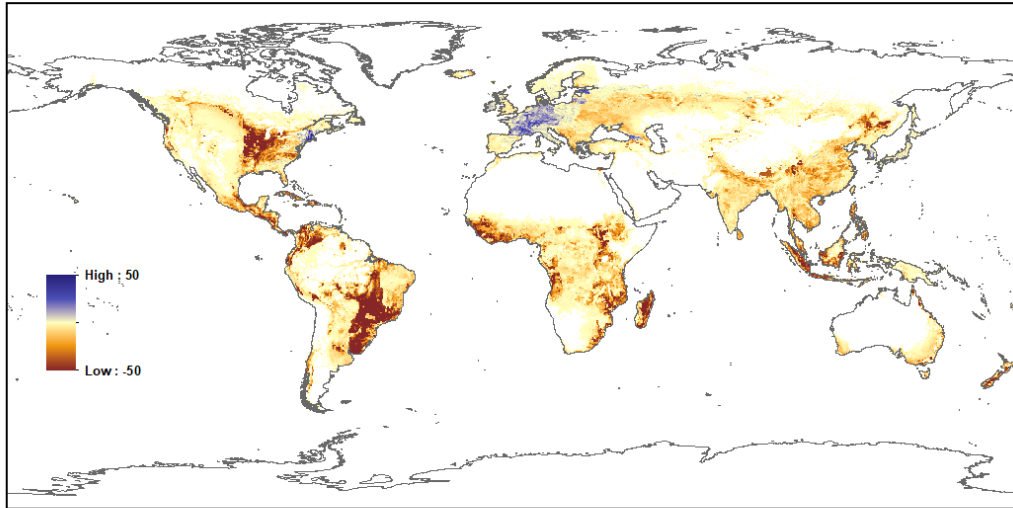
(b) CLM5 Historical Change in Grass PFTs 2014 - 1850   (%)
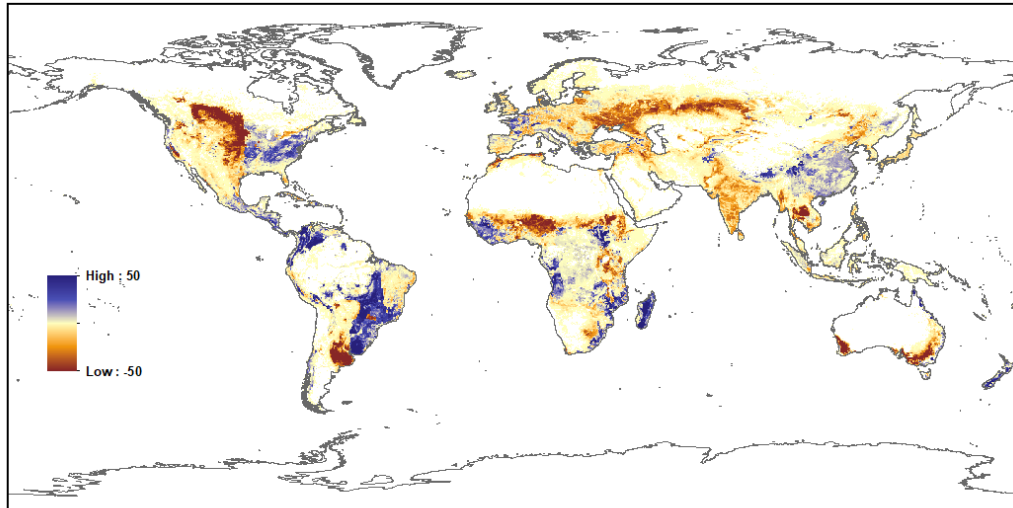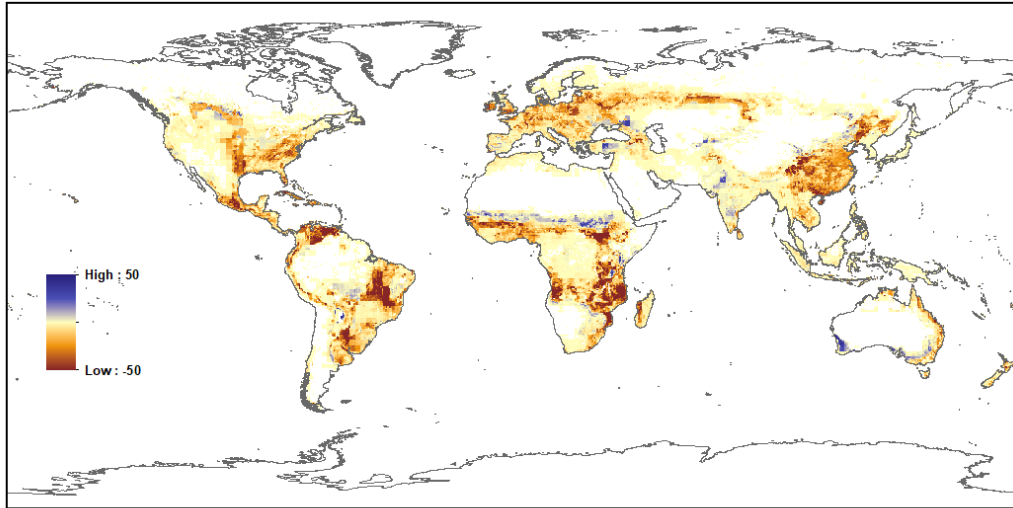

(c) CLM5 Historical Change in Crop CFTs 2014 - 1850   (%)
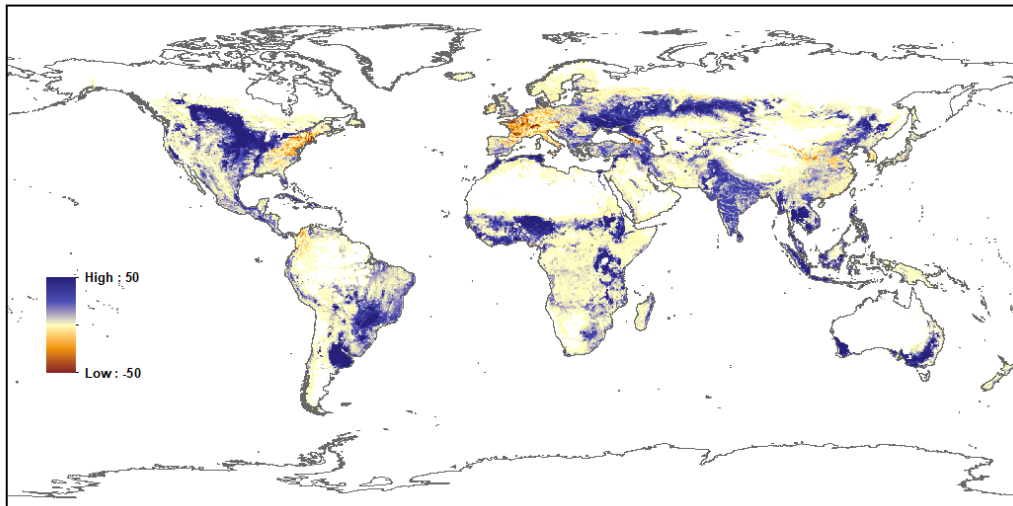
7.6 CLM5 SSP 1-2.6
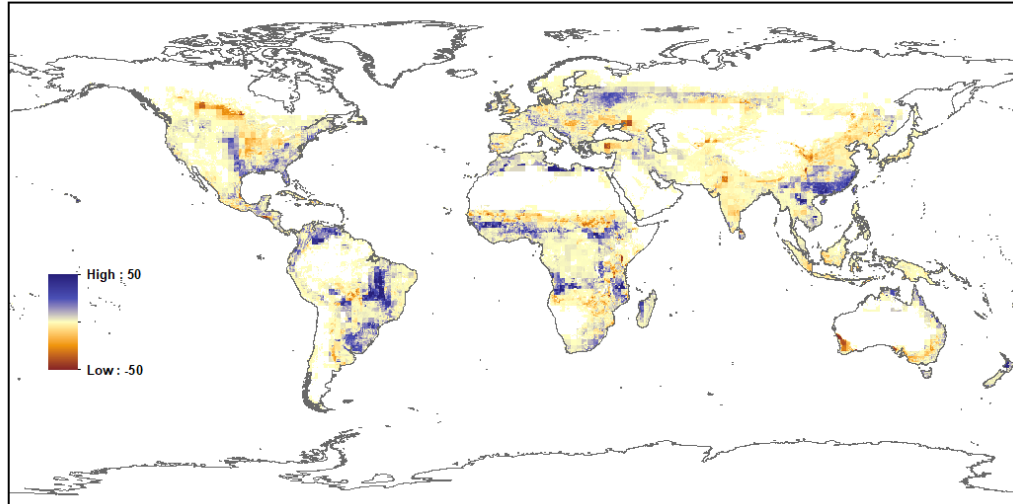
(a) CLM5 SSP1-2.6 Change in Tree PFTs 2100 - 2015   (%)
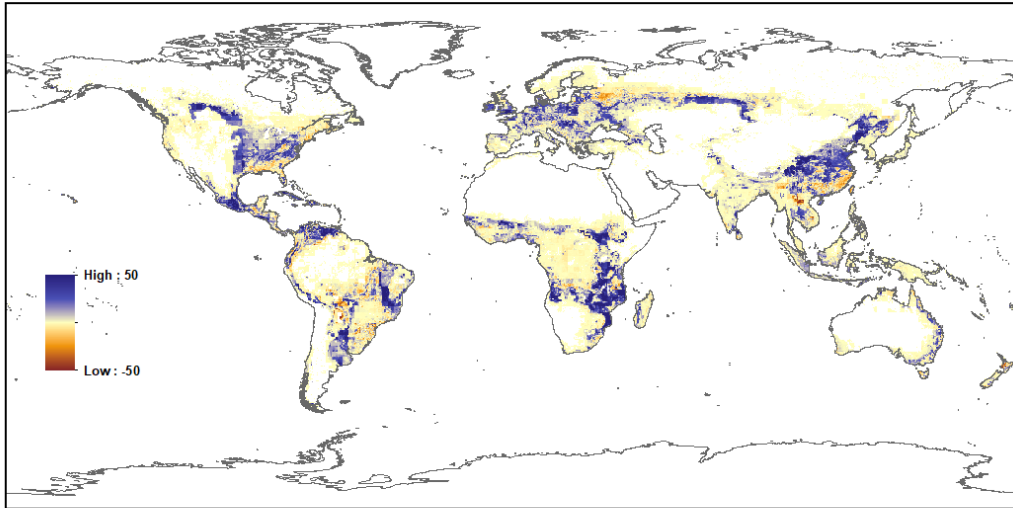
(b) CLM5 SSP1-2.6 Change in Grass PFTs 2100 - 2015   (%)
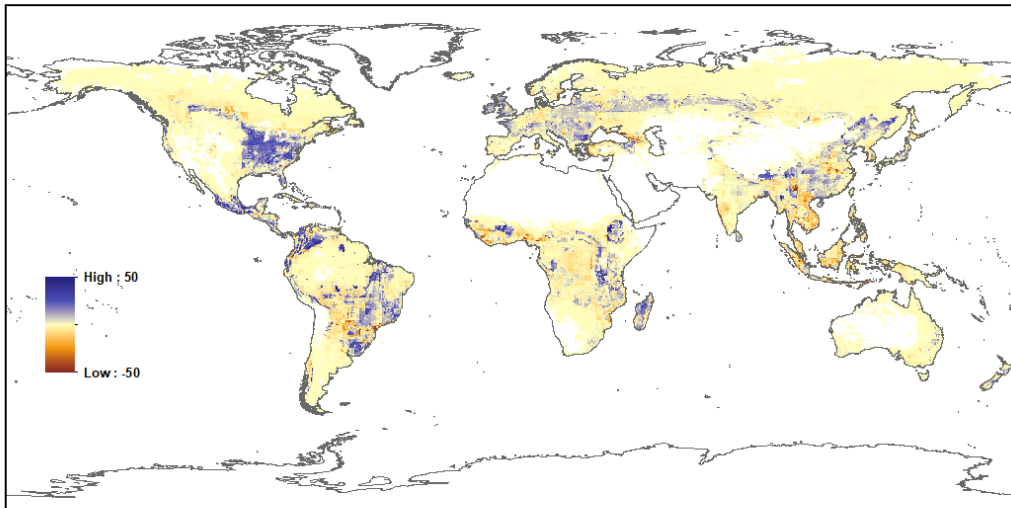

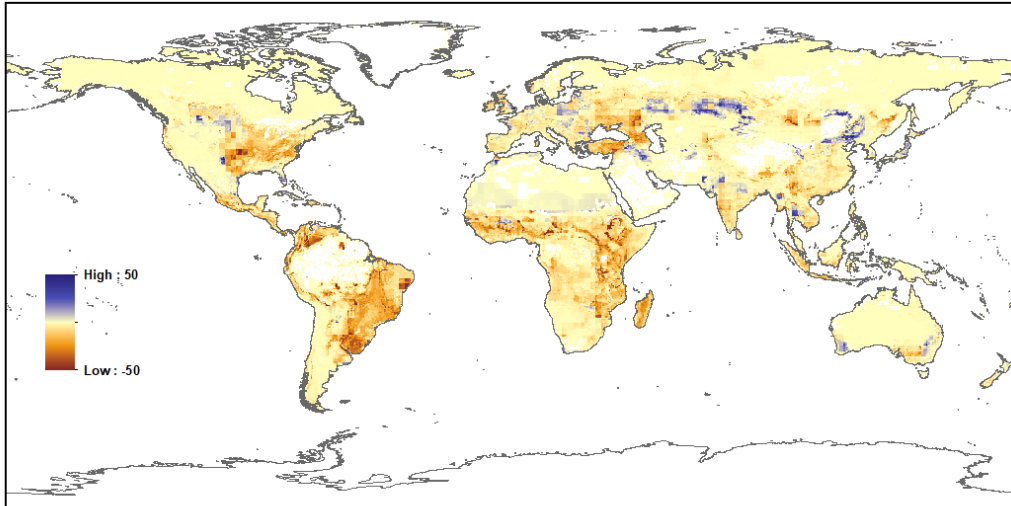**(c)** CLM5 SSP1-2.6 Change in Crop CFTs 2100 - 2015   (%)

7.7 CLM5 SSP 2-4.5
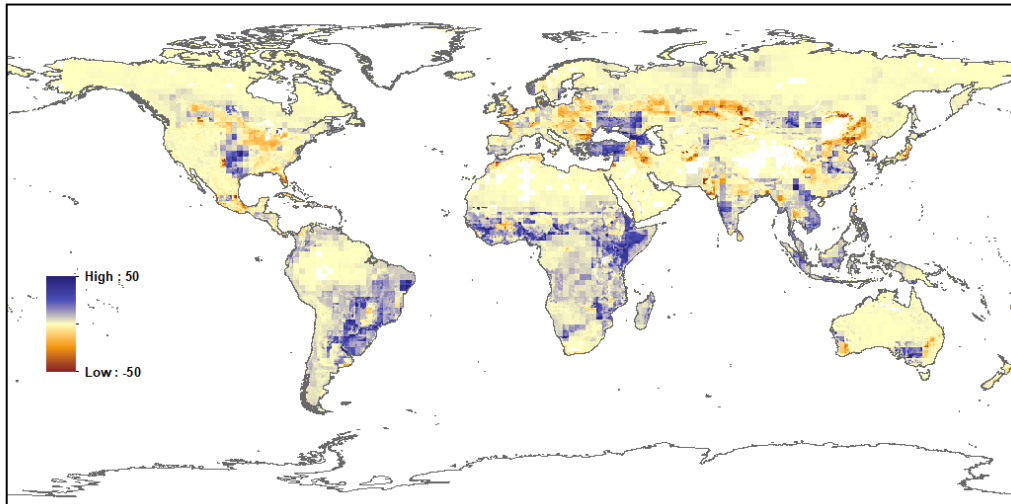
(a) CLM5 SSP2-4.5 Change in Tree PFTs 2100 - 2015   (%)

(b) CLM5 SSP2-4.5 Change in Grass PFTs 2100 - 2015   (%)

(c) CLM5 SSP2-4.5 Change in Crop CFTs 2100 - 2015   (%)

7.8 CLM5 SSP 3-7.0

(a) CLM5 SSP3-7.0 Change in Tree PFTs 2100 - 2015   (%)

(b) CLM5 SSP3-7.0 Change in Grass PFTs 2100 - 2015   (%)

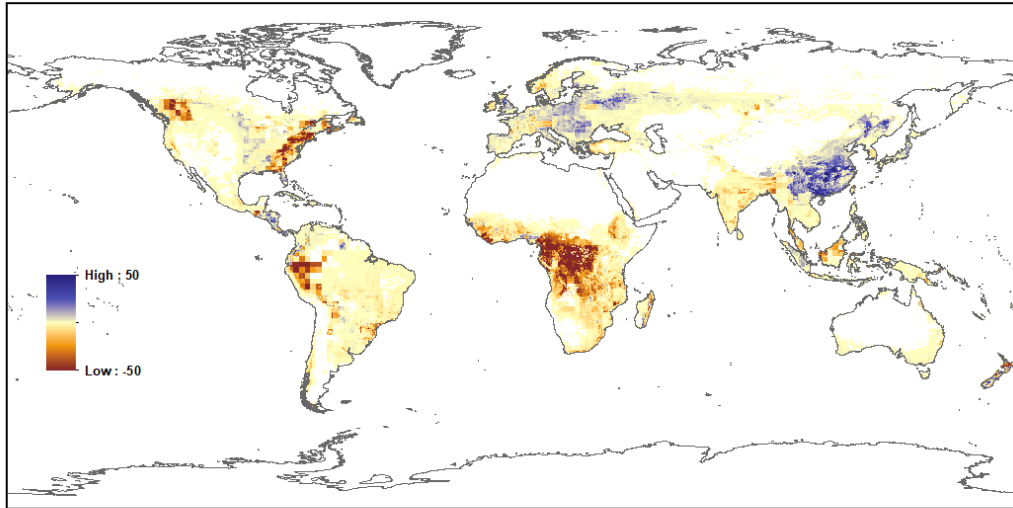**(c)** CLM5 SSP3-7.0 Change in Crop CFTs 2100 - 2015   (%)

7.9 CLM5 SSP 4-6.0

(a) CLM5 SSP4-6.0 Change in Tree PFTs 2100 - 2015   (%)

(b) CLM5 SSP4-6.0 Change in Grass PFTs 2100 - 2015   (%)

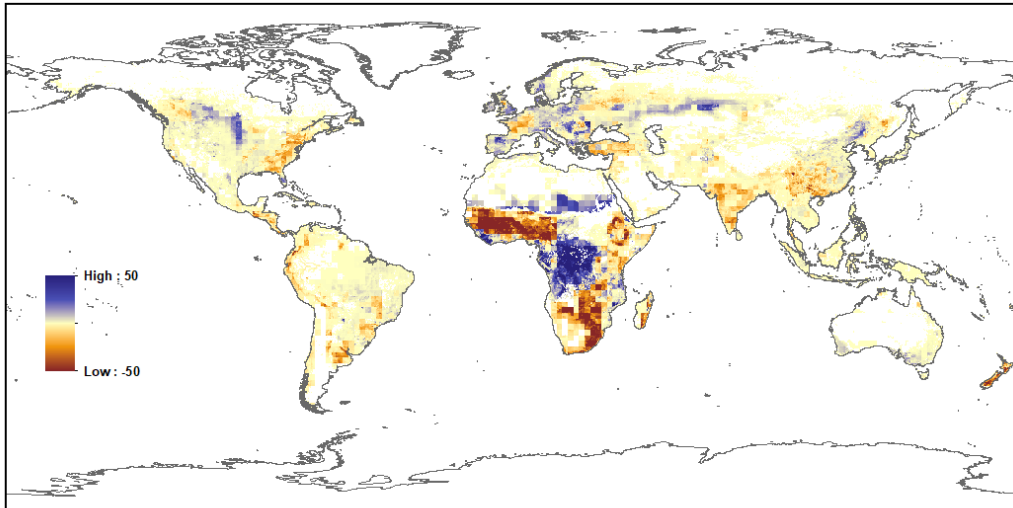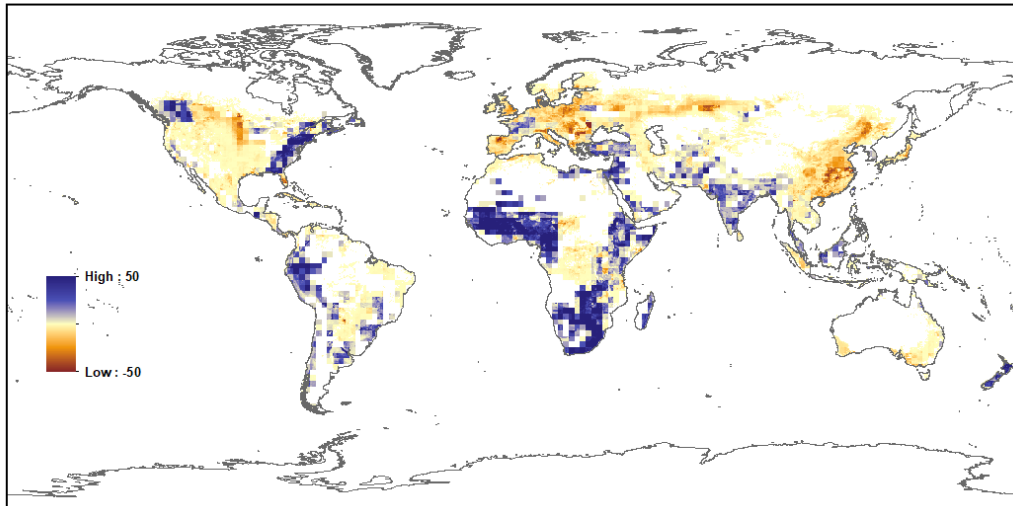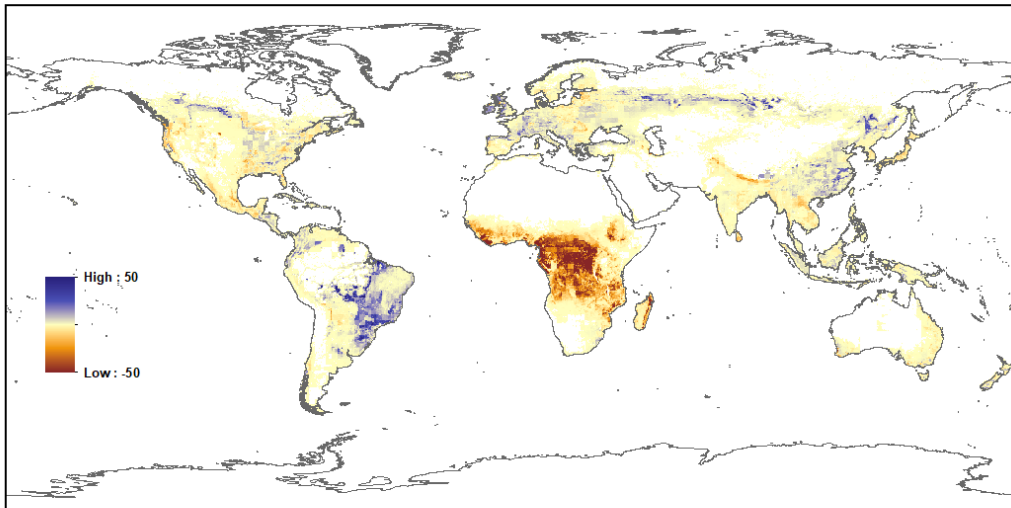(c) CLM5 SSP4-6.0 Change in Crop CFTs 2100 - 2015   (%)
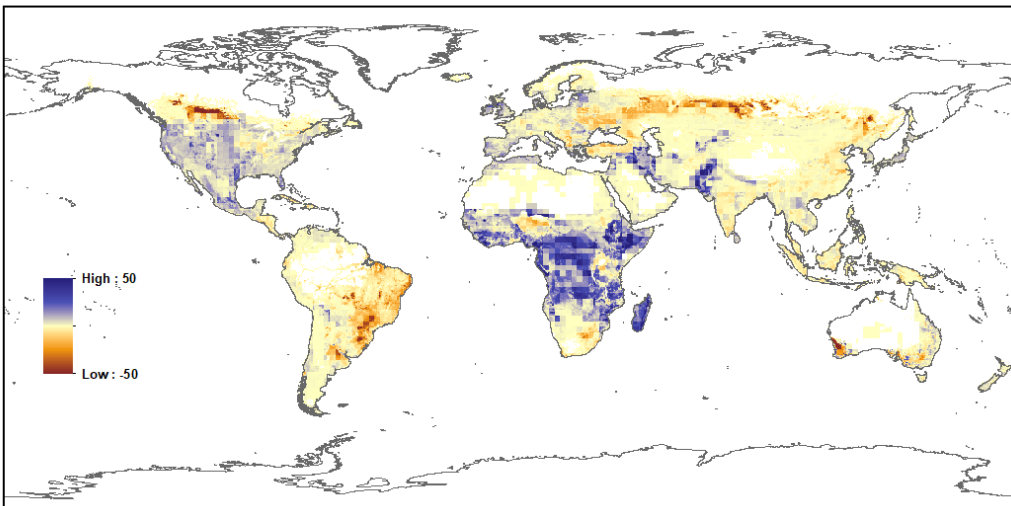
7.10 CLM5 SSP 5-8.5

(a) CLM5 SSP5-8.5 Change in Tree PFTs 2100 - 2015 (%)
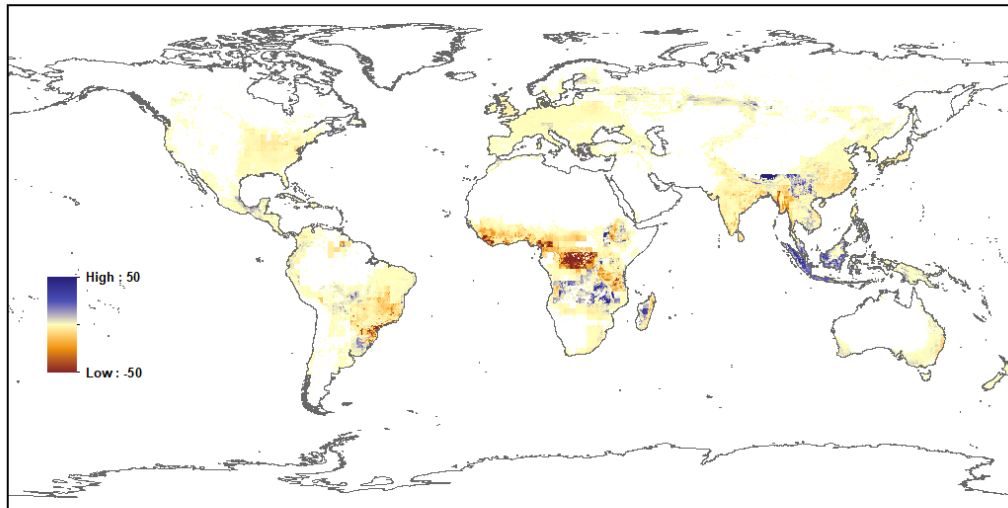
(b) CLM5 SSP5-8.5 Change in Grass PFTs 2100 - 2015 (%)

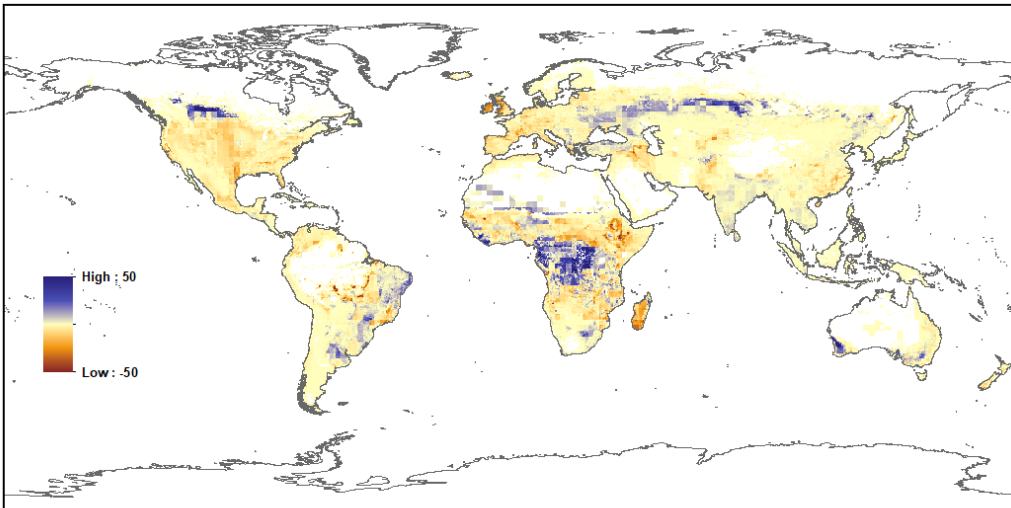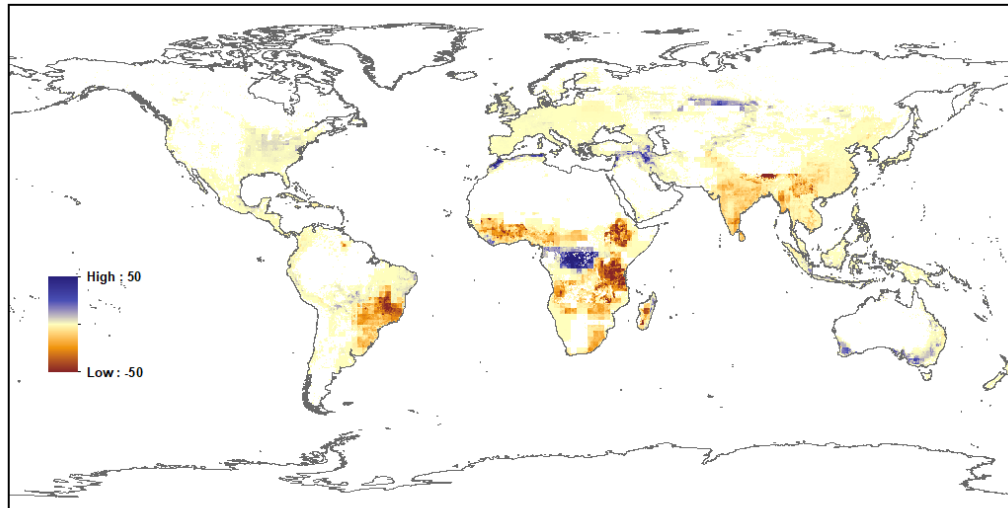**(c)** CLM5 SSP5-8.5 Change in Crop CFTs 2100 - 2015 (%)
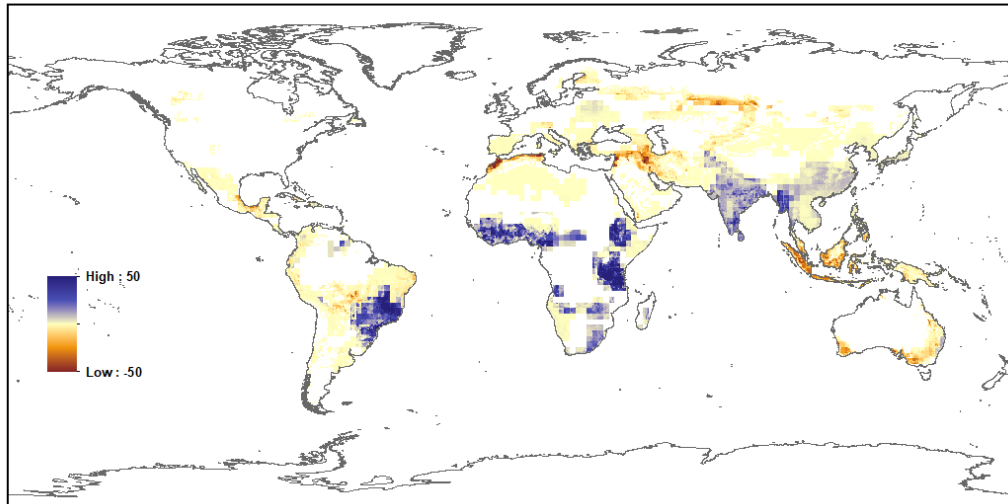
CHAPTER 8.

CLM5 LAND DATA TOOLS

8.1 CLM5 Land Data Tools Overview

The CLM5 model provides

Describe Raw Data File

Describe Surface Data File

Describe Land Use Time Series File

High level description of the process of generating the raw data

8.2 CLM5 – LUH2 Land Use Data Preparation

Combining the LUH2 current day state distribution for PFTs and CFTs to describe LUH2 Forests, Non Forests, Pastures, Rangelands, and each Crop Type.

(a) Global LUH2 2014 Landcover CLM5 PFT Percent
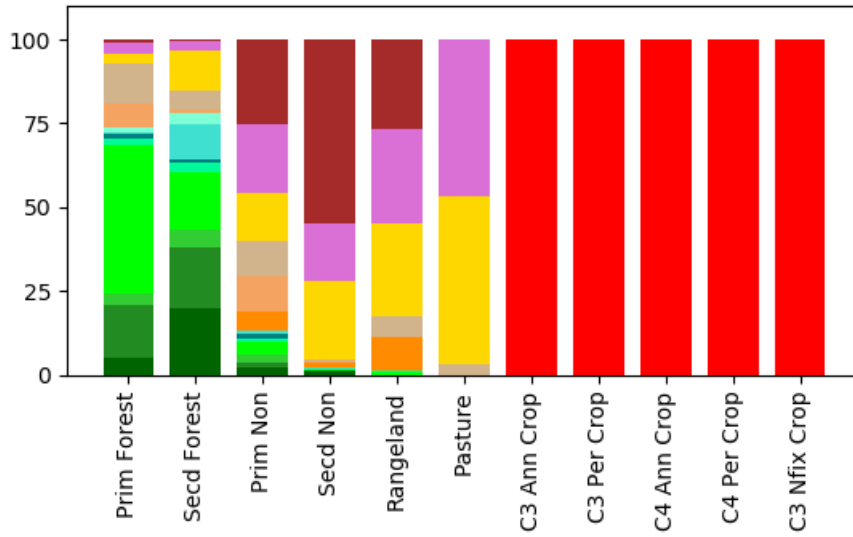
(b) Global LUH2 850 Landcover CLM5 PFT Percent

Legend:
NdlEvgTemp, NdlEvgBorl, NdlDecBorl, BrdEvgTrop, BrdEvgTemp, BrdDecTrop, BrdDecTemp, BrdDecBorl, ShrEvgTemp, ShrDecTemp, ShrDecBorl, GrsC3Arc, GrsC3, GrsC4, Crop
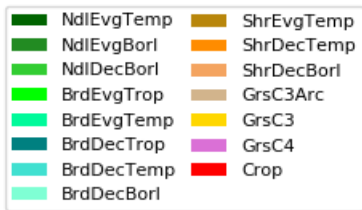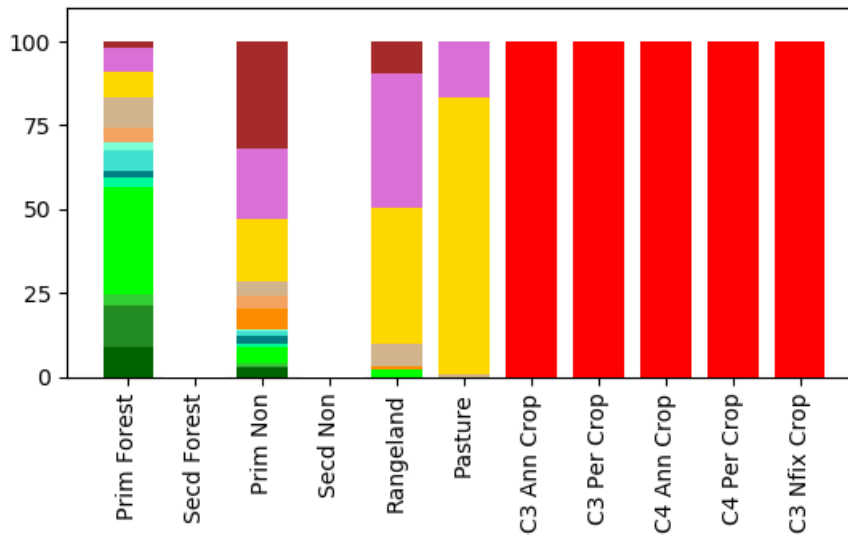
Redo This Graph with updated Secondary Forest and Non Forest!!

(a) Global CLM5 PFT 2014 LUH2 Landcover Percent
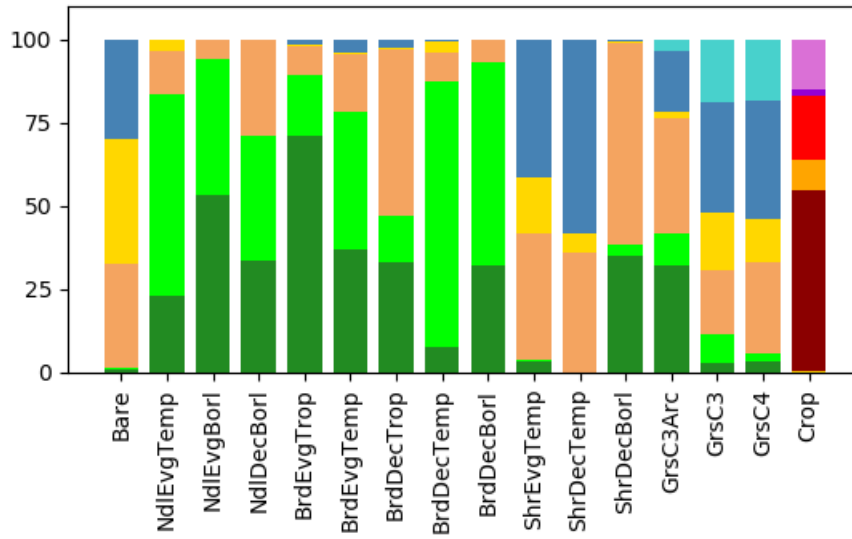
(b) Global CLM5 PFT 850 LUH2 Landcover Percent

Prim Forest
Secd Forest
Prim Non
Secd Non
Rangeland
Pasture
C3 Ann Crop
C3 Per Crop
C4 Ann Crop
C4 Per Crop
C3 NFix Crop
Urban
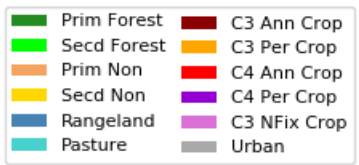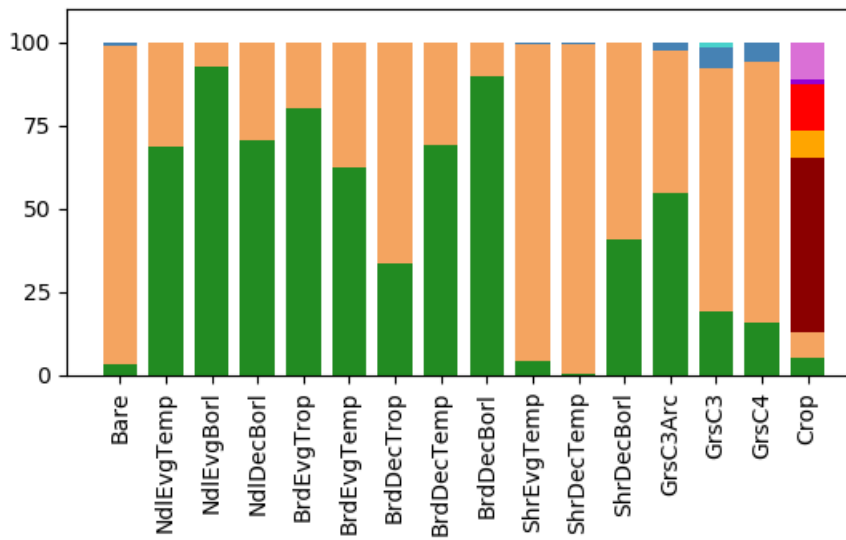
(a) Global CLM5 CFT 2014 LUH2 Landcover Percent
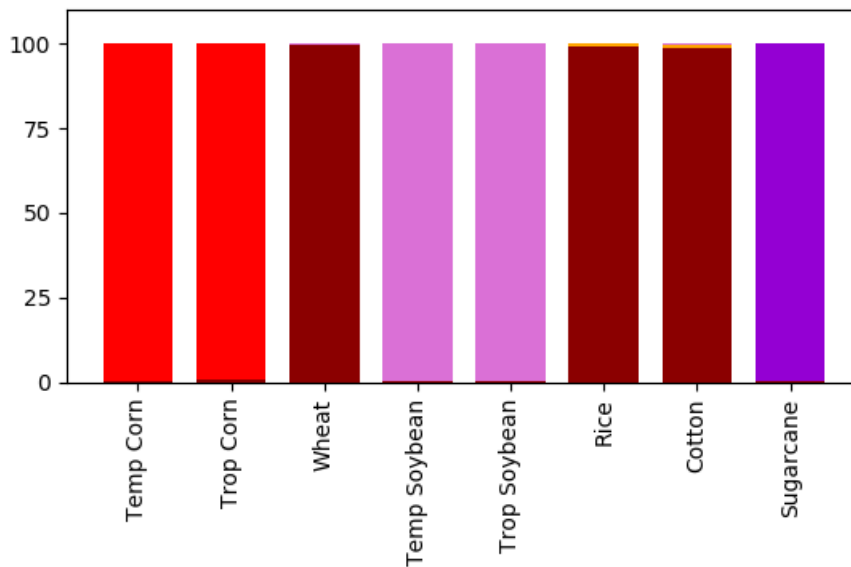
(b) Global CLM5 Crop 850 LUH2 Landcover Percent

Prim Forest
Secd Forest
Prim Non
Secd Non
Rangeland
Pasture
C3 Ann Crop
C3 Per Crop
C4 Ann Crop
C4 Per Crop
C3 NFix Crop
Urban
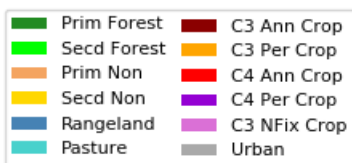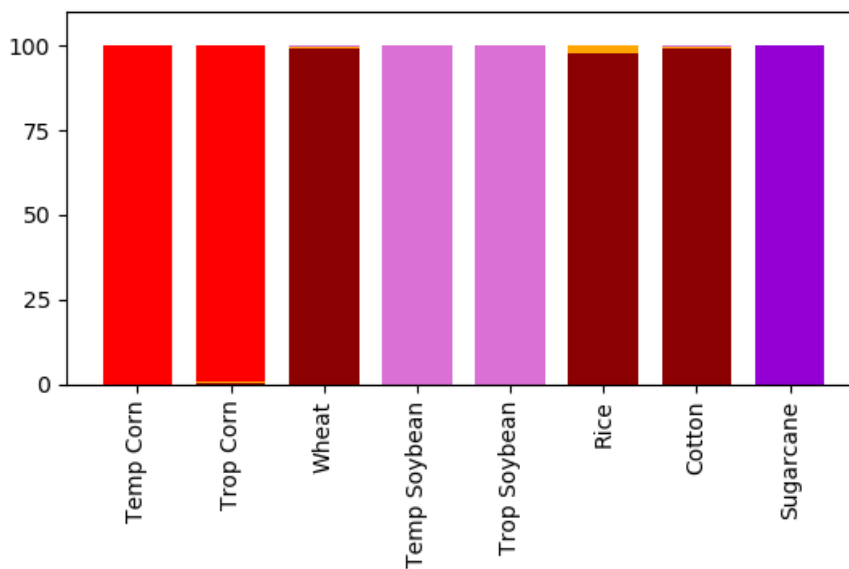
(a) Global CLM5 Other CFT 2014 LUH2 Landcover Percent
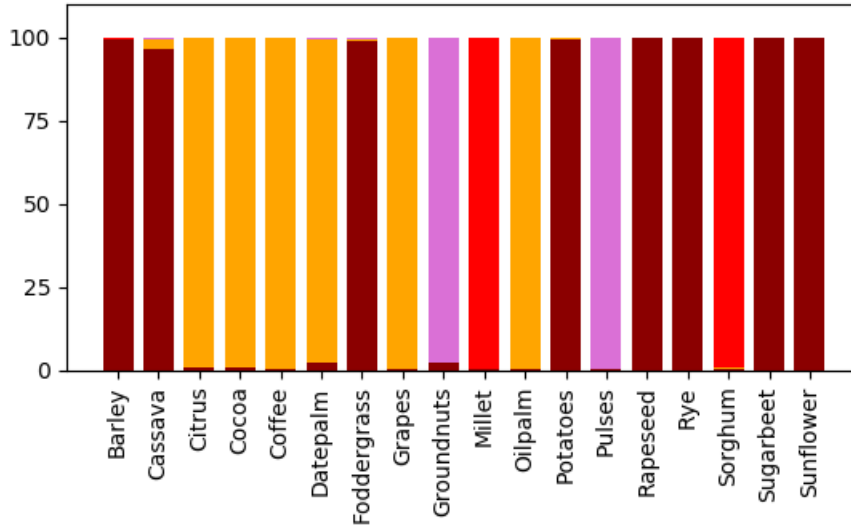
(b) Global CLM5 Other CFT 850 LUH2 Landcover Percent

Legend:
Prim Forest, Secd Forest, Prim Non, Secd Non, Rangeland, Pasture, C3 Ann Crop, C3 Per Crop, C4 Ann Crop, C4 Per Crop, C3 NFix Crop, Urban
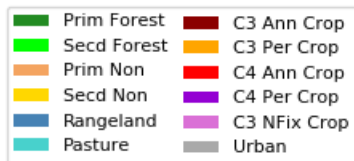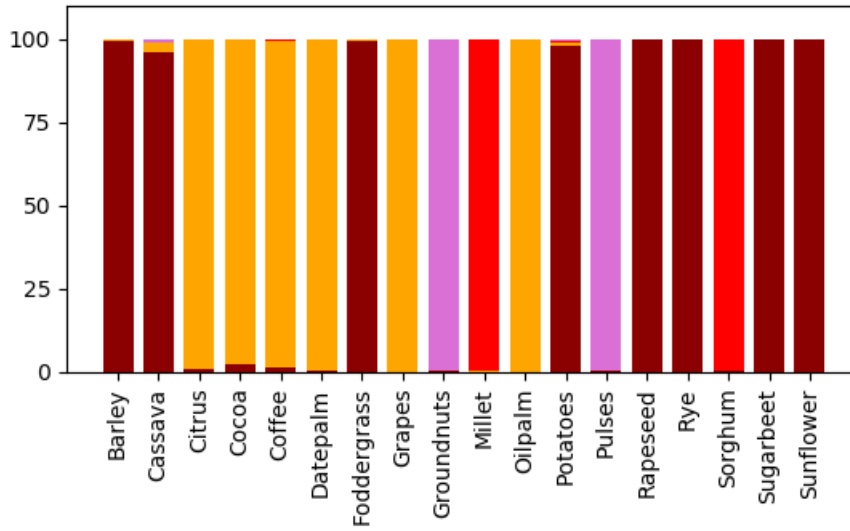
8.3 CLM5 Land Use Data Tool

Generating forest state data for a year from the current forest data and the change in annual LUH2 forest cover for that year relative to the reference current day.

Generating non forest and rangeland data for a year

Generating crop data for a year based on annual LUH2 crop distributions and the EarthSTAT / FAOSTAT distribution for the nearest year

Generating shifting cultivation data from the transitions relative to the changes in state for a year relative to the previous year

Generating wood harvest

Generating irrigation and fertilizer data

8.4 CLM5 Other Land Cover Data

Generating the LAI, SAI, Canopy Height and Soil Color

8.5 CLM5 MkSurfData Tool

Building mksurfdata_map and using mksurfdata.pl files

Time slice versus land use time series files.

Creating surface data namelist file

Creating land use time series text file

Running

CHAPTER 9.

REFERENCES

Ali, A.A., Xu, C., Rogers, A., Fisher, R.A., Wullschleger, S.D., Massoud, E., Vrugt, J.A ., Muss, J.D., McDowell, N., and Fisher, J. 2016. A global scale mechanistic model of photosynthetic capacity (LUNA V1. 0). Geosci. Mod. Dev. 9:587-606. DOI: 10.5194/gmd-9-587-2016.

Arendt, A., et al. 2012. Randolph Glacier Inventory: A Dataset of Global Glacier Outlines Version: 1.0, Global Land Ice Measurements from Space, Boulder Colorado, USA. Digital Media.

Badger, A.M., and Dirmeyer, P.A., 2015. Climate response to Amazon forest replacement by heterogeneous crop cover. Hydrol. Earth. Syst. Sci. 19:4547- 4557. DOI: 10.5194/hess-19-4547-2015

Batjes, N.H., 2006. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid. Report 2006/02 (available through: http://www.isric.org)

Bonan, G. B. 1995. Land–atmosphere CO_2 exchange simulated by a land surface process model coupled to an atmospheric general circulation model. J. Geophys. Res., 100D, 2817–2831.

Bonan, G.B. 1996. A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide. NCAR Technical Note NCAR/TN-417+STR, National Center for Atmospheric Research, Boulder, CO, 150 pp.

Bonan, G.B. 1998. The land surface climatology of the NCAR Land Surface Model coupled to the NCAR Community Climate Model. J. Climate 11:1307-1326.

Bonan, G.B., Oleson, K.W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., Dickinson, R.E., and Yang, Z.-L. 2002a. The land surface climatology of the Community Land Model coupled to the NCAR Community Climate Model. J. Climate 15: 3123-3149.

Bonan, G.B., Levis, S., Kergoat, L., and Oleson, K.W. 2002b. Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. Global Biogeochem. Cycles 16: 5.1-5.23.

Bonan, G.B., Lawrence P.J., Oleson K.W., Levis S., Jung M., Reichstein M., Lawrence, D.M., and Swenson, S.C. 2011. Improving canopy processes in the Community Land Model (CLM4) using global flux fields empirically inferred from FLUXNET data. J. Geophys. Res. 116, G02014. DOI:10.1029/2010JG001593.

Broxton, P. D., X. Zeng, D. Sulla-Menashe, and P. A. Troch, 2014. A Global Land Cover Climatology Using MODIS Data. Journal of Applied Meteorology and Climatology, 53,6, 1593-1605. doi:10.1175/JAMC-D-13-0270.1

Brunke, M. A., P. Broxton, J. Pelletier, D. Gochis, P. Hazenberg, D. M. Lawrence, L. R. Leung, G.-Y. Niu, P. A. Troch, and X. Zeng, 2016. Implementing and Evaluating Variable Soil Thickness in the Community Land Model, Version 4.5 (CLM4.5). J. Clim. 29:3441-3461.

Brzostek, E. R., J. B. Fisher, and R. P. Phillips, 2014. Modeling the carbon cost of plant nitrogen acquisition: Mycorrhizal trade-offs and multipath resistance uptake improve predictions of retranslocation. J. Geophys. Res. Biogeosci., 119, 1684–1697, doi:10.1002/2014JG002660.

Buzan, J.R., Oleson, K., and Huber, M. 2015: Implementation and comparison of a suite of heat stress metrics within the Community Land Model version 4.5, Geosci. Model Dev., 8, 151-170, doi:10.5194/gmd-8-151-2015.

CIESIN: Gridded population of the world version 3 (GPWv3), 2005. Population density grids, Technical report, Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, New York, USA.

Cleveland, C.C., Townsend, A.R., Schimel, D.S., Fisher, H., Howarth, R.W., Hedin, L.O., Perakis, S.S., Latty, E.F., Von Fischer, J.C., Elseroad, A., and Wasson, M.F. 1999. Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. Global Biogeochem. Cycles 13:623-645. DOI: 10.1029/1999GB900014.

Cogley, J.G. 1991. GGHYDRO – Global Hydrographic Data Release 2.0. Trent Climate Note 91-1, Dept. Geography, Trent University, Peterborough, Ontario.

Dahlin, K., Fisher, R., and Lawrence, P. 2015. Environmental drivers of drought deciduous phenology in the Community Land Model. Biogeosciences, 12,16:5061-5074. DOI: 10.5194/bg-12-5061-2015

Dai, Y., and Zeng, Q. 1997. A land surface model (IAP94) for climate studies. Part I: formulation and validation in off-line experiments. Adv. Atmos. Sci. 14:433-460.

Dai, Y., et al. 2001. Common Land Model: Technical documentation and user's guide [Available online at http://climate.eas.gatech.edu/dai/clmdoc.pdf].

Dai, Y., Zeng, X., Dickinson, R.E., Baker, I., Bonan, G.B., Bosilovich, M.G., Denning, A.S., Dirmeyer, P.A., Houser, P.R., Niu, G., Oleson, K.W., Schlosser, C.A., and Yang, Z.-L. 2003. The Common Land Model. Bull. Amer. Meteor. Soc. 84:1013-1023.

Dahlin, K. M., Fisher, R.A., and Lawrence, P.J. 2015. Environmental drivers of drought deciduous phenology in the Community Land Model. Biogeosciences 12(16): 5061-5074, doi:5010.5194/bg-5012-5061-2015.

DeFries, R. S., Townshend, J. R. G., and Hansen, M. C. 1999. Continuous fields of vegetation characteristics at the global scale at 1-km resolution. *J. Geophys. Res.,* **104D,** 16 911–16 923.

DeFries, R. S, Hansen, M. C., and Townshend, J. R. G. 2000a. Global continuous fields of vegetation characteristics: A linear mixture model applied to multi-year 8 km AVHRR data. Int. J. Remote Sens., **21,** 1389–1414.

DeFries, R. S, Townshend, J. R. G., Hansen, M. C., Janetos, A. C., and Loveland, T. R. 2000b. A new global 1-km dataset of percentage tree cover derived from remote sensing. Global Change Biol., **6,** 247–254.

Dickinson, R.E., Henderson-Sellers, A., and Kennedy, P.J. 1993. Biosphere-Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Climate Model. NCAR Technical Note NCAR/TN-387+STR. National Center for Atmospheric Research, Boulder, CO.

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., and Worley, B.A. 2000. LandScan: A global population database for estimating populations at risk. Photogramm. Eng. Rem. Sens. 66:849-857.

Dorman, J.L., and Sellers, P.J. 1989. A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the simple biosphere model (SiB). J. Appl. Meteor. 28:833-855.

Drewniak, B., Song, J., Prell, J., Kotamarthi, V.R., and Jacob, R. 2013. Modeling agriculture in the Community Land Model. Geosci. Model Dev. 6:495-515. DOI:10.5194/gmd-6-495-2013.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development. 9(5). DOI: 10.5194/gmd-9-1937-2016.

FAO, 2000. Global Forest Resources Assessment 2000 – Main Report. FAO Forestry Paper 140, Food and Agriculture Organization of the United Nations, Rome, Italy

FAO, 2020a. Land Use data, FAOSTAT Database, Food and Agriculture Organization of the United Nations, Rome, Italy, available at: http://www.fao.org/faostat/en/#data/RL (last access: 28 July 2016).

FAO, 2020b. Fertilizer data, FAOSTAT Database, Food and Agriculture Organization of the United Nations, Rome, Italy, available at: http://www.fao.org/faostat/en/#data/RL (last access: 28 July 2016).

FAO, 2020c. Forestry data, FAOSTAT Database, Food and Agriculture Organization of the United Nations, Rome, Italy, available at: http://www.fao.org/faostat/en/#data/RL (last access: 28 July 2016).

FAOSTAT, 2016. Food and Agriculture Organization (FAO) of the United Nations FAOSTAT Statistical Databases. [Available online at http:// http://www.fao.org/faostat].

FAO/UNESCO, 1974. Soil Map of the World. FAO, Paris, France.

Fisher, J. B., S. Sitch, Y. Malhi, R. A. Fisher, C. Huntingford, and S.-Y. Tan, 2010. Carbon cost of plant nitrogen acquisition: A mechanistic, globally applicable model of plant nitrogen uptake, retranslocation, and fixation. Global Biogeochem. Cycles, 24, GB1014, doi:10.1029/2009GB003621.

Fisher, R. A., Muszala, S., Verteinstein, M., Lawrence, P., Xu, C., McDowell, N. G., Knox, R. G., Koven, C., Holm, J., Rogers, B. M., Spessa, A., Lawrence, D., and Bonan, G. 2015. Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes, CLM4.5(ED). Geosci. Model Dev., 8: 3593-3619, DOI:10.5194/gmd-8-3593-2015.

Fricko, O., Havlik, P., Rogelj, J., Klimont, Z., Gusti, M., Johnson, N., Kolp, P., Strubegger, M., Valin, H., Amann, M., Ermolieva, T., Forsell, N., Herrero, M., Heyes, C., Kindermann, G., Krey, V., McCollum, D. L., Obersteiner, M., Pachauri, S., Rao, S., Schmid, E., Schoepp, W., and Riahi, K. 2017. The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century. Global Environmental Change. DOI: 10.1016/j.gloenvcha.2016.06.004.

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C. 2002. Global land cover mapping from MODIS: Algorithms and early results, Remote Sens. Environ., 83, 287– 302.

Fujimori, S., Hasegawa, T., Masui, T., Takahashi, K., Herran, D. S., Dai, H., Hijioka, Y., and Kainuma, M. 2017. SSP3: AIM implementation of Shared Socioeconomic Pathways. Global Environmental Change. 2017. DOI: 10.1016/j.gloenvcha.2016.06.009.

Ghimire, B., Riley, W.J., Koven, C.D., Mu, M., and Randerson, J.T. 2016: Representing leaf and root physiological traits in CLM improves global carbon and nitrogen cycling predictions. J. Adv. Mod. Earth Sys. 8: 598-613. DOI: 10.1002/2015MS000538.

Graham, S.T., Famiglietti, J.S., Maidment, D.R. 1999. Five-minute, 1/2º, and 1º data sets of continental watersheds and river networks for use in regional and global hydrologic and climate system modeling studies. Water Resour. Res. 35:583-587.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X., 2012. The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471–1492. DOI: 10.5194/gmd-5-1471-2012.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Dimiceli, C., Carroll, M., and Sohlberg R. 2003. Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm, Earth Interact., 7(10), 1 – 15.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., and Townshend, J.R.G. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change, Science, 342, 850-853. DOI: 10.1126/science.1244693.

Harris, I., Jones, P.D., Osborn, T.J., and Lister, D.H. 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. Int. Journal of Climatology, 34,3. DOI: 10.1002/joc.3711.

Hastings, D.A., Dunbar, P.K., Elphingstone, G.M., Bootz, M., Murakami, H., Maruyama, H., Masaharu, H., Holland, P., Payne, J., Bryant, N.A., Logan, T.L., Muller, J.-P., Schreier, G., and MacDonald, J.S., eds., 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80305-3328, U.S.A.

Heinimann, A., Mertz, 0., Frolking, S., Egelund Christensen, A., Hurni, K., Sedano, F., Chini, L.P., Sahajpal, R., Hansen, M., and Hurtt, G. 2017. A global view of shifting cultivation: Recent, current, and future extent. PLOS One. 12. DOI: 10.1371/journal.pone.0184479.

Houghton, R.A., and Hackler, J.L. 2000. Changes in terrestrial carbon storage in the United States. 1. The roles of agriculture and forestry. Global Ecology and Biology. 9, 125-144. DOI: 10.1046/J.1365-2699.2000.00166.X.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J.G., Kuhry, P., and Swanson, D.K. 2013. The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. Earth Syst. Sci. Data, 5, 3–13. DOI: 10.5194/essd-5-3-2013.

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

Hurtt, G.C., Chini, L.P., Frolking, S., Betts, R.A., Feddema, J., Fischer, G., Fisk, J.P., Hibbard, K., Houghton, R.A., Janetos, A., Jones, C.D., Kindermann, G., Kinoshita, T., Klein Goldeweijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D.P., and Wang, Y.P. 2011. Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change 109:117-161. DOI:10.1007/s10584-011-0153-2.

Hurtt, G.C., Chini, L.P., Sahajpal, R., Frolking, S., Bodirsky, B.L., Calvin, K., Doelman, C., Fisk, J.P., Fujimori, S., Klein Goldeweijk, K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J.O., Kennedy, J., Krisztin, T., Lawrence, D., Lawrence, P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F.N., van Vuuren, D.P., and Zhang, X. 2020. Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6. Geosci. Model Dev., 13, 5425–5464. DOI: 10.5194/gmd-13-5425-2020.

IFA 2015: Statistics, International Fertilizer Association, IFA Database, available at https://www.ifastat.org, last access: 21 January 2015.

Jackson, R.B., Canadell, J., Ehleringer, J.R, Mooney, H.A., Sala, O.E., and Schulze, E.D. 1996. A global analysis of root distributions for terrestrial biomes. Oecologia 108:389–411. DOI:10.1007/BF00333714.

Jackson, T.L., Feddema, J.J., Oleson, K.W., Bonan, G.B., and Bauer, J.T. 2010. Parameterization of urban characteristics for global climate modeling. Annals of the Association of American Geographers 4: 848-865. DOI:10.1080/00045608.2010.497328.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Peterschmitt, J.-Y., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., Crucifix, M., Dolan, A., Fernandez-Donado, L., Fischer, H., Hopcroft, P. O., Ivanovic, R. F., Lambert, F., Lunt, D. J., Mahowald, N. M., Peltier, W. R., Phipps, S. J., Roche, D. M., Schmidt, G. A., Tarasov, L., Valdes, P. J., Zhang, Q., and Zhou, T. 2018. The PMIP4 contribution to CMIP6 – Part 1: Overview and over-arching analysis plan, Geosci. Model Dev., 11, 1033–1057, DOI:10.5194/gmd-11-1033-2018, 2018

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D., 1996. The NCEP/NCAR 40-Year Reannalysis Project. Bull. Am. Met. Soc. 77,3. DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kaplan, J.O., Krumhardt, K.M., Gaillard, M.-J., Sugita, S., Trondman, A.-K., Fyfe, R., Marquer, L., Mazier, F., and Nielsen, A.B. 2017. Constraining the Deforestation History of Europe: Evaluation of Historical Land Use Scenarios with Pollen-Based Land Cover Reconstructions. Land. 6, 91. DOI: 10.3390/land6040091.

Kennedy, D., Swenson, S., Oleson, K., Lawrence, D.M., Fisher, R., Carlos Lola da Costa, A., and Gentine, P. 2019. Implementing Plant Hydraulics in the Community Land Model, Version 5. J. Adv. Model. Earth Sys. 11,2:485-513. DOI:10.1029/2018MS001500

Klein Goldewijk, K., Beusen, A., van Drecht, G., and de Vos, M. 2011. The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years. Global Ecology and Biogeography. 20:73-86. DOI:10.1111/j.1466-8238.2010.00587.x

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E. 2017. Anthropogenic land use estimates for the Holocene – HYDE 3.2. Earth System Science Data. 9, 927-953. DOI: 10.5194/essd-9-927-2017

Kourzeneva, E., Asensio, H., Martin, E. and Faroux, S., 2012. Global gridded dataset of lake coverage and lake depth for use in numerical weather prediction and climate modelling. Tellus A 64.

Koven, C.D. et al. 2013. The effect of vertically-resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. Biogeosciences 10, 7109–7131 DOI: 10.5194/bg-10-7109-2013.

Koven, C.D., G. Hugelius, D.M. Lawrence, and W.R. Wieder, 2017: Higher climatological temperature sensitivity of soil carbon in cold than warm climates. Nature Clim. Change, 7, DOI:10.1038/nclimate3421.

Kriegler, E., Bauer, N., Popp, A., Humpenöder, F., Leimbach, M., Strefler, J., Baumstark, L., Bodirsky, B. L., Hilaire, J., Klein, D., Mouratiadou, I., Weindl, I., Bertram, C., Dietrich, J.-P., Luderer, G., Pehl, M., Pietzcker, R., Piontek, F., Lotze-Campen, H., Biewald, A., Bonsch, M., Giannousakis, A., Kreidenweis, U., Müller, C., Rolinski, S., Schultes, A., Schwanitz, J., Stevanovic, M., Calvin, K., Emmerling, J., Fujimori, S., and Edenhofer, O. 2017. Fossil-fueled development (SSP5): An energy and resource intensive scenario for the 21st century Glob. Environ. Change 42 297–315. DOI:10.1016/j.gloenvcha.2016.05.015

Landrum, L., Otto-Bliesner, B.L., Wahl, E.R., Conley, A., Lawrence, P.J., Rosenbloom, N., and Teng, H. 2013. Last Millennium Climate and Its Variability in CCSM4. J. Climate 26,4: 1085–1111. DOI:10.1175/JCLI-D-11-00326.1

Lawrence, D.M., and Slater, A.G. 2008. Incorporating organic soil into a global climate model. Clim. Dyn. 30. DOI:10.1007/s00382-007-0278-1.

Lawrence, D.M., Oleson, K.W., Flanner, M.G., Thornton, P.E., Swenson, S.C., Lawrence, P.J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G.B., and Slater, A.G. 2011. Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. J. Adv. Model. Earth Sys. 3. DOI:10.1029/2011MS000045.

Lawrence, D.M., Hurtt, G.C., Arneth, A., Brovkin, V., Calvin, K.V., Jones, A.D., Jones, C.D., Lawrence, P.J., de Noblet-Ducoudre, N., Pongratz, J., Seneviratne, S.I., Shevliakova, E. 2016. The Land Use Model Intercomparison Project (LUMIP): Rationale and experimental design. Geoscientific Model Development. 9(9). DOI: 10.5194/gmd-9-2973-2016

Lawrence, D.M, et al. 2019. The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty. J. Adv. Model. Earth Sys. 3. DOI:10.1029/2011MS000045.

Lawrence, D.M., et al. 2018. Technical description of version 5.0 of the Community Land Model (CLM). NCAR Technical Note, National Center for Atmospheric Research, Boulder, CO, 329 pp.

Lawrence, P.J., and Chase, T.N. 2007. Representing a MODIS consistent land surface in the Community Land Model (CLM 3.0). J. Geophys. Res. 112:G01023.

Lawrence, P.J., and Chase, T.N. 2010. Investigating the climate impacts of global land cover change in the Community Climate System Model. Int. J. Climatol. 30:2066-2087. DOI:10.1002/joc.2061.

Lawrence, P.J., Feddema, J.J., Bonan, G.B., Meehl, G.A., O'Neill, B.C., Oleson, K.W., Levis, S., Lawrence, D.M., Kluzek, E., Lindsay, K., and Thornton, P.E. 2012. Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. J. Climate 25:3071-3095. DOI:10.1175/JCLI-D-11-00256.1

Lawrence, P.J., Lawrence, D.M., and Hurtt, G.C. 2018. Attributing the Carbon Cycle Impacts of CMIP5 Historical and Future Land Use and Land Cover Change in the Community Earth System Model (CESM1). J. Geophys. Res. Biogeo., 123:5. DOI:10.1029/2017JG004348

Lehner, B. and Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands, J. Hydrol., 296, 1–22. DOI:10.1016/j.jhydrol.2004.03.028.

Lehner, B., Verdin, K. and Jarvis, A., 2008. New global hydrograhy derived from spaceborne elevation data. Eos Trans., AGU, 89, 93 – 94. DOI:10.1029/2008EO100001.

Levis, S., Thornton, P., Bonan, G., and Kucharik, C. 2009. Modeling land use and land management with the Community Land Model. iLeaps newsletter, No. 7.

Levis, S., Bonan, G., Kluzek, E., Thornton, P., Jones, A., Sacks, W., and Kucharik, C. 2012. Interactive crop management in the Community Earth System Model (CESM1): Seasonal influences on land-atmosphere fluxes. J. Climate 25: 4839-4859. DOI:10.1175/JCLI-D-11-00446.1.

Levis S, Badger A, Drewniak B, Nevison C, Ren, X. 2016. CLMcrop yields and water requirements: avoided impacts by choosing RCP 4.5 over 8.5, Climatic Change, doi:10.1007/s10584-016-1654-9

Li, F., Zeng, X.-D., and Levis, S. 2012a. A process-based fire parameterization of intermediate complexity in a Dynamic Global Vegetation Model. Biogeosciences 9:2761-2780.

Li, F., Zeng, X. D., and Levis, S. 2012b. Corrigendum to "A process-based fire parameterization of intermediate complexity in a Dynamic Global Vegetation Model" published in Biogeosciences, 9, 2761–2780, 2012". Biogeosciences 9: 4771-4772.

Li, F., Levis, S., and Ward, D. S. 2013a. Quantifying the role of fire in the Earth system – Part 1: Improved global fire modeling in the Community Earth System Model (CESM1). Biogeosciences 10:2293-2314.

Li, F., and Lawrence, D. 2017. Role of fire in the global land water budget during the 20th century through changing ecosystems. J. Clim. 30,6: 1894-1908. DOI:10.1175/JCLI-D-16-0460.1

Li, H.-Y., Huang, M., Tesfa, T., Ke, Y., Sun, Y., Liu, Y., and Leung, L. R. 2013b. A subbasin-based framework to represent land surface processes in an Earth System Model, Geosci. Model Dev. Discuss. 6:2699-2730. DOI:10.5194/gmdd-6-2699-2013.

Li, H.-Y., Huang, M., Tesfa, T., Ke, Y., Sun, Y., Liu, Y., and Leung, L. R. 2014. A subbasin-based framework to represent land surface processes in an Earth System Model. Geosci. Model Dev. 7, 947–963. DOI: 10.5194/gmd-7-947-2014.

Lipscomb, W., and Sacks, W. 2012. The CESM land ice model documentation and user's guide. 46 pp. [Available online at http://www.cesm.ucar.edu/models/cesm1.1/cism/].

Lombardozzi, D.L., Lu, Y., Lawrence, P.J., Lawrence, D.M., Swenson, S., Oleson, K.W., Wieder, W.R., and Ainsworth, E.A. 2019. Simulating Agriculture in the Community Land Model Version 5. J. Geophys. Res. Biogeo., 125, DOI: 10.1029/2019JG005529

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

Medlyn, B.E., Duursma, R.A., Eamus, D., Ellsworth, D.S., Prentice, I.C., Barton, C.V.M., Crous, K.Y., De Angelis, P., Freeman, M., and Wingate, L. 2011. Reconciling the optimal and empirical approaches to modelling stomatal conductance. Global Change Biology, 17: 2134–2144. DOI:10.1111/j.1365-2486.2010.02375.x

Monfreda, C., Ramankutty, N., and Foley, J.A. 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochem. Cycles, 22, GB1022, doi:10.1029/2007GB002947.

Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., and Running, S.W. 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ., 83, 214– 2. DOI: 10.1016/S0034-4257(02)00074-3

Negron-Juarez, R. Koven, C.D., Riley, W.J., Knox, R.G., Chambers, J.Q. 2015. Observed allocations of productivity and biomass, and turnover times in tropical forests are not accurately represented in CMIP5 Earth system models. Environmental Research Letters 10:064017. DOI:10.1088/1748-9326/10/6/064017.

Nemani, R., and Running, S. W. 1996. Implementation of a hierarchical global vegetation classification in ecosystem function models. J. Veg. Sci., 7, 337– 346. DOI: doi.org/10.2307/3236277

Niu, G.-Y., Yang, Z.-L., Dickinson, R.E., and Gulden, L.E. 2005. A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. J. Geophys. Res. 110:D21106. DOI: 10.1029/2005JD006111.

Niu, G.-Y., and Yang, Z.-L. 2006. Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. J. Hydrometeor. 7:937-952.

Niu, G.-Y., Yang, Z.-L., Dickinson, R.E., Gulden, L.E., and Su, H. 2007. Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. J. Geophys. Res. 112:D07103. DOI:10.1029/2006JD007522.

Oke, T.R., 1987. Boundary Layer Climates (2nd edition). London, Routledge, 435 pp.

Olson, J.S., Watts, J.A., and Allison, L.J., 1983. Carbon in live vegetation of major world Ecosystems, Rep. ORNL-5862, Oak Ridge Natl. Lab., Oak Ridge, Tenn.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Heda, P., and Kassem, K. R., 2001. Terrestrial ecoregions of the world a new map of life on earth, Bioscience, 51, 933–938.

Oleson, K.W., and Bonan, G.B., 2000. The effects of remotely-sensed plant functional type and leaf area index on simulations of boreal forest surface fluxes by the NCAR land surface model. J. Hydrometeor. 1:431-446.

Oleson, K.W., Niu, G.-Y., Yang, Z.-L., Lawrence, D.M., Thornton, P.E., Lawrence, P.J., Stöckli, R., Dickinson, R.E., Bonan, G.B., Levis, S., Dai, A., and Qian, T. 2008. Improvements to the Community Land Model and their impact on the hydrological cycle. J. Geophys. Res. 113:G01021. DOI:10.1029/2007JG000563

Oleson, K.W., et al. 2010a. Technical description of version 4.0 of the Community Land model (CLM). NCAR Technical Note NCAR/TN-478+STR, National Center for Atmospheric Research, Boulder, CO, 257 pp.

Oleson, K.W., Bonan, G.B., Feddema, J., Vertenstein, M., and Kluzek, E. 2010b. Technical description of an urban parameterization for the Community Land Model (CLMU). NCAR Technical Note NCAR/TN-480+STR, National Center for Atmospheric Research, Boulder, CO, 169 pp.

Oleson, K.W., et al. 2013. Technical description of version 4.5 of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-503+STR, National Center for Atmospheric Research, Boulder, CO, 420 pp.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M. 2016. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geosci. Model Dev., 9, 3461–3482, DOI:10.5194/gmd-9-3461-2016

Pelletier, J. D., Broxton, P.D., Hazenberg, P., Zeng, X., Troch, P.A., Niu, G.Y., Williams, Z., Brunke, M.A., and Gochis, D. 2016. A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. J. Adv. Mod. Earth Sys. 8,1:41-65. DOI:10.1002/2015MS000526.

Pfister, G.G., Emmons, L.K., Hess, P.G., Lamarque, J.-F., Orlando, J.J., Walters, S., Guenther, A., Palmer, I., and Lawrence, P.J. 2008. Contribution of isoprene to chemical budgets: A model tracer study with the NCAR CTM MOZART-4. J. Geophys. Res. Atmos., 113:D5, DOI: 10.1029/2007JD008948

Popp, A., Calvin, K., Fujimori, S., Havlik, P., Humpenöder, F., Stehfest, E., Bodirsky, B. L., Dietrich, J. P., Doelmann, J. C., Gusti, M., Hasegawa, T., Kyle, P., Obersteiner, M., Tabeau, A., Takahashi, K., Valin, H., Waldhoff, S., Weindl, I., Wise, M., Kriegler, E., Lotze-Campen, H., Fricko, O., Riahi, K., and van Vuuren, D. P. 2017. Land-use futures in the shared socio-economic pathways, Glob. Environ. Change, 42, 331–345, DOI:10.1016/j.gloenvcha.2016.10.002.

Portmann, F.T., Siebert, S., and Döll, P. 2010. MIRCA2000 - Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cycles. 24, GB1011. DOI:10.1029/2008GB003435.

Ramankutty, N., and Foley, J.A. 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992, Global Biogeochem. Cycles, 13(4), 997– 1027.

Rastner, P., Bolch, T., Mölg, N., Machguth, H., and Paul, F., 2012. The first complete glacier inventory for the whole of Greenland, The Cryosphere, 6, 1483–1495, DOI: 10.5194/tc-6-1483-2012.

Ray, D K., Ramankutty, N., Mueller, N.D., West, P.C., and Foley, J.A. 2012. Recent patterns of crop yield growth and stagnation. Nature Communications, 3, 1293. DOI: 10.1038/ncomms2296.

Riahi, K., Grübler, A. and Nakicenovic, N. 2007. Scenarios of long-term socio-economic and environmental development under climate stabilization. Technol. Forecasting Soc. Change 74, 887–935. DOI:10.1016/j.techfore.2006.05.026.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change, 42, 153-168, DOI: 10.1016/j.gloenvcha.2016.05.009.

Riley, W. J., Z. M. Subin, D. M. Lawrence, S. C. Swenson, M. S. Torn, L. Meng, N. Mahowald, and P. Hess, 2011. Barriers to predicting global terrestrial methane fluxes: Analyses using a methane biogeochemistry model integrated in CESM. Biogeosciences, 8, 1925–1953. DOI:10.5194/bg-8-1925-2011.

Rogers, A., Medlyn, B.E., Dukes, J.S., Bonan, G., Caemmerer, S., Dietze, M.C., Kattge, J., Leakey, A.D., Mercado, L.M., and Niinemets, U. 2017. A roadmap for improving the representation of photosynthesis in Earth system models. New Phytologist, 213,1:22-42. DOI:10.1111/nph.14283

Sacks, W. J., Cook, B. I., Buenning, N., Levis, S., and Helkowski, J. H. 2009. Effects of global irrigation on the near-surface climate. Climate Dyn., 33, 159–175. DOI:10.1007/s00382-008-0445-z.

Schaaf, C. B., Gao, F., Strahler, A.H., Lucht, W., Li, X., Tsang, T., Strugnell, N.C., Zhang, X., Jin, Y., and Muller, J.-P. 2002. First operational BRDF, albedo nadir reflectance products from MODIS. Remote Sens. Environ., 83, 135– 148.

Shi, M., Fisher, J.B., Brzostek, E.R., and Phillips, R.P., 2016: Carbon cost of plant nitrogen acquisition: global carbon cycle impact from an improved plant nitrogen cycle in the Community Land Model. Glob. Change Biol., 22:1299-1314. DOI: 10.1111/gcb.13131.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.M., Frenken, K., Feick, S., 2005. Development and validation of the global map of irrigation areas. Hydrol Earth Syst Sc 9:535–547

Simard, M., Pinto, N., Fisher, J.B., and Baccini, A., 2011. Mapping forest canopy height globally with spaceborne lidar. J. Geophys. Res. 116:G04021. DOI:10.1029/2011JG001708

Sitch, S., B. Smith, I.C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J.O. Kaplan, S. Levis, W. Lucht, M.T. Sykes, K. Thonicke, and S. Venevsky, 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biol., 9, 161– 185.

Smil, V., 2001. Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production. MIT Press.

Stehfest, E., van Vuuren, D., Kram, T., Bouwman, L., Alkemade, R., Bakkenes, M., Biemans, H., Bouwman, A., den Elzen, M., Janse, J., Lucas, P., van Minnen, J., Müller, C., and Prins, A. 2014. Integrated Assessment of Global Environmental Change with IMAGE 3.0. Model description and policy applications, The Hague.

Still, C.J., Berry, J.A., Collatz, G.J., and DeFries, R.S. 2003. Global distribution of C3 and C4 vegetation: Carbon cycle implications, Global Biogeochem. Cycles, 17(1), 1006, doi:10.1029/2001GB001807.

Subin, Z.M., Riley, W.J. and Mironov, D. 2012a. Improved lake model for climate simulations, J. Adv. Model. Earth Syst., 4, M02001. DOI:10.1029/2011MS000072.

Subin, Z.M., Murphy, L.N., Li, F., Bonfils, C. and Riley,W.J., 2012b. Boreal lakes moderate seasonal and diurnal temperature variation and perturb atmospheric circulation: analyses in the Community Earth System Model 1 (CESM1). Tellus A, North America, 64.

Sun, Y., Gu, L., and Dickinson, R. E. 2012. A numerical issue in calculating the coupled carbon and water fluxes in a climate model, J. Geophys. Res., 117, D22103. DOI:10.1029/2012JD018059.

Swenson, S.C., Lawrence, D.M., and Lee, H. 2012. Improved Simulation of the Terrestrial Hydrological Cycle in Permafrost Regions by the Community Land Model. JAMES, 4, M08002. DOI:10.1029/2012MS000165.

Swenson, S.C. and Lawrence, D.M. 2012. A New Fractional Snow Covered Area Parameterization for the Community Land Model and its Effect on the Surface Energy Balance. JGR, 117, D21107. DOI:10.1029/2012JD018178.

Swenson, S.C., and D. M. Lawrence. 2014. Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data. JGR, 119, 10, 299–10,312, DOI:10.1002/2014JD022314.

Swenson, S.C., and D. M. Lawrence. 2015. A GRACE-based assessment of interannual groundwater dynamics in the Community Land Model. WRR, 51, DOI:10.1002/2015WR017582.

Tarnocai, C., Kettles, I. M., and Lacelle, B., 2011. Peatlands of Canada, Geological Survey of Canada, Open File 6561, CD-ROM. DOI:10.495/288786.

Thornton, P.E., and Zimmermann, N.E. 2007. An improved canopy integration scheme for a land surface model with prognostic canopy structure. J. Climate 20:3902-3923.

Thornton, P.E., Lamarque, J.-F., Rosenbloom, N.A., and Mahowald, N.M. 2007. Influence of carbon-nitrogen cycle coupling on land model response to CO2 fertilization and climate variability. Global Biogeochem. Cycles 21:GB4018.

UNSTAT, 2005. National Accounts Main Aggregates Database, United Nations Statistics Division.

van der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Mu, M., Kasibhatla, S.P., Morton, D.C., DeFries, R.S., Jin, Y., van Leeuwen, T.T. 2010. Global fire emissions and the

contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009) Atmos. Chem. Phys. 10:11707-11735.

van Kampenhout, L., J.T.M. Lenaerts, W.H. Lipscomb, W.J. Sacks, D.M. Lawrence, A.G. Slater, and M.R. van den Broeke, 2017. Improving the Representation of Polar Snow and Firn in the Community Earth System Model. Journal of Advances in Modeling Earth Systems 9, no. 7: 2583–2600. DOI: 10.1002/2017MS000988.

van Vuuren, D.P., Lucas, P.S., and Hilderink, H.B.M., 2006. Downscaling drivers of global environmental change: enabling use of global SRES scenarios at the national and grid levels, Report 550025001, Netherlands Environmental Assessment Agency, 45 pp.

Verdin, K.L., and Greenlee, S.K. 1996. Development of continental scale digital elevation models and extraction of hydrographic features, paper presented at the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, 21–26 January, Natl. Cent. for Geogr. Inf. and Anal., Santa Barbara, Calif.

White, M.A., Thornton, P.E., Running, S.W., and Nemani, R.R. 2000. Parameterization and sensitivity analysis of the Biome-BGC terrestrial ecosystem model: net primary production controls. Earth Interactions 4:1-85.

Willmott, C. J., and Matsuura, K. 2000. Terrestrial air temperature and precipitation: Monthly and annual climatologies. [Available online at http://climate.geog.udel.edu/;climate.]

Whittaker, R. H. 1970. Communities and ecosystems. Communities and ecosystems.

Wilson, M. F. 1984. The construction and use of land surface information in a general circulation climate model. Unpublished Ph.D. thesis, University of Liverpool, United Kingdom, 346 pp.

Wise, M., Calvin, K., Kyle, P., Luckow, P., and Edmonds, J. 2014. Economic and Physical Modeling of Land Use in GCAM 3.0 and an Application to Agricultural Productivity, Land, and Terrestrial Carbon. Climate Change Economics 5(2). DOI:10.1142/S2010007814500031.

World Bank, 2004. World development indicators 2004, Oxford University Press, New York, 416 pp.

Wu, H., Kimball, J.S., Mantua, N., and Stanford, J. 2011. Automated upscaling of river networks for macroscale hydrological modeling. Water Resour. Res., 47, W03517, DOI:10.1029/2009WR008871.

Wu, H., Kimball, J.S., Li, H., Huang, M., Leung, L.R., and Adler, R.F. 2012. A New Global River Network Database for Macroscale Hydrologic modeling, Water Resour. Res., 48, W09701, DOI:10.1029/2012WR012313.

Xu, C., R. Fisher, S. D. Wullschleger, C. J. Wilson, M. Cai, and N. G. McDowell, 2012: Toward a mechanistic modeling of nitrogen limitation on vegetation dynamics. PLoS ONE 7(5): e37914. DOI:10.1371/journal.pone.0037914

Zeng, X., M. Shaikh, Y. Dai, R. E. Dickinson, and R. Myneni, 2002. Coupling of the Common Land Model to the NCAR Community Climate Model. *J. Climate,* **15,** 1832–1854.

Zon, R., and Sparhawk, W.N. 1923. Forest Resources of the World. Volume I. McGraw-Hill. NY. 493 pp.