https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Towards an ensemble-based evaluation of land surface models in light of uncertain forcings and observations

Vivek. K. Arora[1], Christian Seiler[2], Libo Wang[2], and Sian Kou-Giesbrecht[1]

[1]Canadian Centre for Climate Modelling and Analysis, Climate Research Division, Environment Canada, Victoria, BC, Canada
[2]Climate Processes Section, Climate Research Division, Environment and Climate Change Canada, Toronto, ON, Canada

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Abstract

Quantification of uncertainty in fluxes of energy, water, and CO_2 simulated by land surface models (LSMs) remains a challenge. LSMs are typically driven with, and tuned for, a specified meteorological forcing data set and a specified set of geophysical fields. Here, using two data sets each for meteorological forcing and historical land cover reconstruction, as well as two model structures (with and without coupling of carbon and nitrogen cycles), the uncertainty in simulated results over the historical period is quantified for the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC) model. The resulting eight (2 x 2 x 2) equally probable model simulations are evaluated using an in-house model evaluation framework that uses multiple observations-based data sets for a range of quantities. Among the primary global energy, water, and carbon related fluxes and state variables, simulated area burned, fire CO_2 emissions, soil carbon mass, vegetation biomass, runoff, heterotrophic respiration, gross primary productivity, and sensible heat flux show the largest spread across the eight simulations relative to their mean. Simulated net atmosphere-land CO_2 flux, which is considered a critical determinant of the performance of LSMs, is found to be largely independent of the simulated pre-industrial vegetation and soil carbon mass. This indicates that models can provide reliable estimates of the strength of the land carbon sink despite some biases in carbon stocks. Results show that evaluating an ensemble of model results against multiple observations allows to disentangle model deficiencies from uncertainties in model inputs, observation-based data, and model configuration.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

1. Introduction

The current generation land surface models (LSMs) explicitly simulate the fluxes of energy, water, momentum, and trace gases (including CO_2, CH_4, and N_2O) between the atmosphere and the land surface. These models have become an essential tool in understanding what role the land surface plays in the global climate system under current and projected future changes in environmental conditions, including atmospheric CO_2 concentration. Land models are also an essential component of climate and Earth system models (ESMs), together with their ocean and atmosphere components. Within the framework of ESMs, land models are coupled interactively to their atmospheric component through the fluxes of energy, momentum, and matter.

The complexity of land models has increased over time as more physical and biogeochemical processes have been included in their framework. This increased complexity combined with the uncertainty in our understanding of the physical and biogeochemical processes implies that different models respond differently even when driven with the same external forcings. One estimate of uncertainty in our understanding of land surface physical and biogeochemical processes is obtained by evaluating the inter-model spread in a given quantity when models are forced in an exact same manner. Other than the uncertainty among models due to differences in their model structures and parameterizations of various processes, uncertainty also exists due to at least three other reasons. These include uncertainty 1) in

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

parameter values[1] of processes, 2) in driving meteorological data, and 3) in the specification of the geophysical fields. LSMs are typically driven with meteorological data consisting of seven primary variables (incoming long and shortwave radiation, temperature, precipitation, specific humidity, wind speed, and pressure). In addition, the geophysical fields of land cover, soil texture, and soil permeable depth are also required. Driving data for LSMs also consist of atmospheric CO_2 concentration and other model specific external forcings such as nitrogen deposition and fertilizer application rates for models that include a representation of the terrestrial nitrogen cycle, and lightning, population density, and gross domestic product (GDP) for models that simulate wildfires.

Every year more than 15 land surface modelling groups participate in the TRENDY (trends in net land atmosphere carbon exchanges) project where they perform a set of simulations that are driven with specified external forcings. The simulations are performed from year 1700 to the present day. These simulations contribute to the annual Global Carbon Project's (GCP) analysis of the land carbon sink together with its analysis for anthropogenic CO_2 emissions and the ocean carbon sink (Friedlingstein et al., 2019). The external forcings used to drive LSMs in the TRENDY intercomparison include, 1) six hourly meteorological data from 1901 to present day (the most recent 2020 TRENDY intercomparison used the CRU-JRA forcing obtained by blending the climate research unit (CRU) monthly data and the Japanese reanalysis (JRA)); 2) atmospheric CO_2 concentration; and 3) information about changes in crop area and other land use changes (LUC) from the land use harmonization (LUH) product (Hurtt et al., 2020a). The information about

[1] Changes in parameter values do not constitute different parameterizations. For example, two models may use the same parameterization, say $y=mx+b$, but different values of its parameters m and b. However, $y=mx + b$ and $y = mx$ are considered to be two different parameterizations.

changes in crop area and other LUC is used by land surface modelling groups to reconstruct historical land cover from the year 1700 to the present day consistent with the number of the plant functional types (PFTs) a given model represents. The protocol also provides nitrogen deposition and fertilization application rates for models including nitrogen cycling.

Models participating in the TRENDY simulations are thus driven with common meteorological and land use change forcings as part of its protocol. The resulting spread across models participating in the TRENDY project thus provides a measure of inter-model uncertainty, as mentioned earlier. Traditionally this uncertainty associated with model structure has gained the most attention and the scientific community has responded to this by performing model intercomparison projects (MIPs) where models are driven according to a common protocol. The coupled model intercomparison project (CMIP) in the climate community together with its various sub-projects (Eyring et al., 2016) is another prominent example. MIPs now routinely form the basis of evaluating models against observations and against multi-model means of various quantities. Multi-model means are also considered a best estimate for a given quantity (Tebaldi and Knutti, 2007).

The modelling community has been long aware of the uncertainty associated with parameter values, since a large fraction of physical and biogeochemical model processes are parameterized, and such uncertainty analysis dates back to the early hydrological models (e.g. Hornberger and Spear, 1981; Beven and Binley, 1992). More recent examples of parameter uncertainty in context of a given LSM include Poulter et al. (2010), Booth et al. (2012), and Li et al. (2018a). The land surface modelling community, however, has only recently begun to address and quantify uncertainty related with driving meteorological data. Wu et al. (2017), for example,

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

illustrate the uncertainty in gross primary productivity (GPP) simulated by the Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS) model when driven by six different meteorological data sets. Bonan et al. (2019) analyze the uncertainty in simulated carbon cycle related variables using three versions of the community land model (CLM) when driven with two meteorological data sets over the historical period. Slevin et al. (2017) assess the uncertainty in simulated GPP by the JULES land surface model when driven by three different climate data sets. Studies that evaluates the effect of different land cover representations on model performance are even fewer. Tian et al. (2004) and Lawrence and Chase (2007) study the effect of new land surface boundary conditions, including leaf area index and fractional vegetation cover, based on the MODIS satellite data as implemented in the CLM in Community Atmosphere Model (CAM2) and Community Climate System Model (CCSM 3.0), respectively.

Here, we drive the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC) with two sets of historical meteorological forcings but also two sets of historical land cover reconstructions in order to quantify the uncertainty associated with both these forcings. Other than these, we also use two versions of the CLASSIC model: one that represents the interactions between the carbon (C) and nitrogen (N) cycles and the other in which these interactions are turned off. Seiler et al. (2021a) have evaluated how well the CLASSIC model performs when forced with three different meteorological data sets using the model version without the N cycle. CLASSIC has contributed to the simulations for the TRENDY intercomparison, and the GCP, since 2016 (formerly under the CLASS-CTEM name). Using the two meteorological forcing data sets, two land cover reconstructions, and two versions of the model we are able to perform eight simulations over the historical period since 1700. All of these simulations are

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

equally likely representations of the state of the land surface over the historical period. Yet, they all have their own distinct biases since simulate land surface states and fluxes are different. We use these simulations to illustrate the uncertainty associated with meteorological forcing and the two different reconstructions of land cover that are used to drive the model. We also use an in-house open-source benchmarking system to evaluate these different simulations against observations-based data sets: AMBER (Automated Benchmarking R Package) (Seiler et al., 2021b) uses gridded and in-situ observation-based estimates of 19 energy, water, and C cycle related variables to evaluate land models.

Section 2 of this paper describes the framework of the CLASSIC land model and the forcing data that are required to drive the model. Section 3 describes the two meteorological data sets, the two reconstructions of the land cover that are used to drive the model, and the simulations performed for this study. Section 4 analyses the results from the simulations to illustrate their different states, and reports results from the AMBER benchmarking exercise. Finally, discussion and conclusions are presented in Section 5. The use of more than one meteorological forcing data sets and land covers yields a conundrum since tuning of model parameters for a given forcing data set is not a useful exercise anymore. We also report a new finding that despite different land C states (characterized in terms of vegetation and soil carbon mass) in the eight simulations considered here, the net atmosphere-land CO_2 flux over the historical period in these simulations is consistent with estimates from the GCP. This and the discussion about the broader question of model tuning are also presented in Section 5.

2. The CLASSIC land modelling framework

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

2.1 The physical and carbon biogeochemical processes

The CLASSIC land model is the successor to, and based on, the coupled Canadian Land Surface Scheme (CLASS; (Verseghy, 1991; Verseghy et al., 1993)) and the Canadian Terrestrial Ecosystem Model (CTEM; (Arora and Boer, 2005; Melton and Arora, 2016b). CLASSIC also serves as the land component in the family of Canadian Earth System Models (Arora et al., 2009, 2011; Swart et al., 2019). Melton et al. (2019) provide an overview of the CLASSIC land model and launched it as a community model. The basis of the modelling of physical and biogeochemical processes in CLASSIC, comes from CLASS and CTEM, respectively, both of which have a long history of development. CLASSIC simulates land-atmosphere fluxes of water, energy, and momentum based on its physics, and fluxes of CO_2, CH_4, N_2O, NO_x, and NH_3 based on its biogeochemical process. The representation of the terrestrial N cycle is a new addition to CLASSIC (Asaadi and Arora, 2021; Kou-Giesbrecht and Arora, 2022) and allows to simulate interactions between the C and N cycles explicitly.

The CLASSIC model simulations can be performed over a spatial domain, that may be global or regional, using gridded data or at a point scale, e.g. using meteorological and geophysical data from a FluxNet site. The primary physical and carbon biogeochemical processes of CLASSIC are briefly summarized in the next two sections.

2.1.1 Physical processes

The calculations for physical processes in CLASSIC are performed over vegetated, snow, and bare fractions in each model grid cell typically at a time step of 30 minutes. In the version used here, the fractional coverage of the four plant functional types (PFTs) (needleleaf trees,

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

broadleaf trees, crops, and grasses) characterizes vegetation for each grid cell. The fractional coverage of these four PFTs are specified over the historical period in this study. The structure of vegetation is characterized by leaf area index (LAI), vegetation height, canopy mass, and rooting distribution through the soil layers all of which are dynamically simulated by the biogeochemical module of CLASSIC. Twenty ground layers, represent the soil profile, starting with 10 layers of 0.1 m thickness. The thickness of layers gradually increases to 30 m for a total ground depth of over 61 m. The depth of permeable soil layers and thus the depth to bedrock varies geographically and is specified based on the SoilGrids250m data set (Hengl et al., 2017). Liquid and frozen soil moisture contents, and soil temperature, are determined prognostically for permeable soil layers. The temperature, albedo, mass, and density of a single layer snow pack (when the climate permits snow to exist) are also prognostically modelled. The result of physics calculations yield fluxes of energy (primarily net radiation, ground heat flux, and latent and sensible heat fluxes) and water (primarily evapotranspiration and runoff) at the land-atmosphere boundary.

2.1.2 Biogeochemical processes

The biogeochemical processes in CLASSIC, based on CTEM, are described in detail in the appendix of Melton and Arora (2016). The biogeochemical processes simulate the land-atmosphere exchange of CO_2 and as a result simulate vegetation as a dynamic component depending on the environmental conditions.

The biogeochemical module of CLASSIC prognostically calculates the amount of carbon in the model's three live (leaves, stem, and root) and two dead (litter and soil) carbon pools for each PFT. The live vegetation pools are separated into their structural and non-structural components.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

The C amount in these pools is represented per unit land area (kg C/m^2). The amount of carbon in the live and dead carbon pools and all terrestrial ecosystem processes in the biogeochemical module in this study are modelled for nine PFTs that map directly onto the four base PFTs used in the physics module of CLASSIC. Needleleaf trees are divided into their deciduous and evergreen phenotypes, broadleaf trees are divided into cold deciduous, drought deciduous, and evergreen phenotypes, and crops and grasses are divided based on their photosynthetic pathways into C_3 and C_4 versions. The sub-division of PFTs is essential for modelling biogeochemical processes. For instance, simulating the onset and offset of leaves is different between evergreen and deciduous phenotypes of needleleaf and broadleaf trees. However, once leaf area index (LAI) is known a physical process (such as the interception of rain and snow by canopy leaves) does not need the information about the underlying deciduous or evergreen nature of leaf phenology.

The litter and soil carbon pools are tracked for each soil layer but the movement of C between the soil layers is not yet modelled. Other than photosynthesis and leaf respiration which are modelled at a time step of 30 minutes all other biogeochemical processes are modelled at a daily time step. These include: 1) allocation of C from leaves to stem and roots, 2) autotrophic respiration from the live and heterotrophic respirations the dead carbon pools, 3) leaf phenology, 4) turnover of live vegetation components that generates litter, 5) mortality, 6) land use change (LUC), and 7) fire (Arora and Melton, 2018). Competition between PFTs for space is not modelled in this study and fractional coverage of the nine PFTs are specified based on reconstruction of the historical land cover as explained in the next section.

When the N cycle is turned on, land-atmosphere fluxes of N_2O, NO, and NH_3, and N leaching are also modelled in response to biological N fixation, N fertilizer inputs, and N deposition from the atmosphere. In particular, when the N cycle interacts with the C cycle the maximum photosynthetic capacities of model PFTs ($V_{c,max}$) are determined prognostically as a function of their leaf N content (Asaadi and Arora, 2021; Kou-Giesbrecht and Arora, 2022). When N cycle is turned off, prescribed PFT-specific $V_{c,max}$ rates are used (Melton and Arora, 2016a) and an empirical downregulation parameterization is used to emulate the effect of nutrient constraints as atmospheric CO_2 increases (Arora et al., 2009). N in all model components (leaves, stem, roots, litter, and soil organic matter) is prognostically tracked and therefore C:N ratio of all components are prognostically modelled except for soil organic matter for which a C:N ratio is 13 is specified. In addition, N in the soil mineral pools of nitrate (NO_3–) and ammonium (NH_4+) is also prognostically modelled.

3. Driving data for CLASSIC and model simulations

3.1 Land cover

Land cover is one of the most important geophysical fields that is required by LSMs and at its most basic level provides information about fractional vegetation cover in each of model grid cell for a given regional or global domain. Vegetation in LSMs is typically represented in terms of PFTs. Models may choose to represent a basic set of few PFTs (trees, grasses, shrubs, and crops) or a more elaborate set that distinguishes PFTs on the basis of their stature (trees, grasses, or shrubs), leaf form (needleleaf or broadleaf), leaf phenology (evergreen or deciduous), photosynthetic pathway (C_3 or C_4), and geographical location (tropical, temperate, or boreal).

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

The version of CLASSIC in this study uses a somewhat smaller set of nine PFTs for biogeochemical processes as mentioned in the previous section. The fractional coverage of PFTs in a model may be dynamically simulated based on competition between PFTs or prescribed based on observation-based land cover information. While CLASSIC does have a parameterization of competition between its PFTs (Arora and Boer, 2006) for the historical simulations considered here, and simulations which contribute to the TRENDY ensemble, prescribed fractional coverage of PFTs are used.

For the process of generating a historical reconstruction of land cover consisting of time-varying fractional coverage of a model's PFTs two observation-based data sets are used. The first data set is remotely-sensed land cover product that represents the geographical distribution of land cover at a point in time. The second data set is that of a spatially and temporally varying cropland (and pasture) area, which in this case represents the data set provided by the land use harmonization (LUH) product as part of the TRENDY protocol. The LUH product is fairly comprehensive (Hurtt et al., 2020b). For example, not all models use the pasture area and other information provided in the LUH product.

The process of generating land cover for a given model's PFTs is at least a three-step process. First, the fractional coverage of model PFTs are obtained from a remotely sensed land cover product that represents the snap shot of land cover for a given point in time. This requires typically mapping 20 – 40 land cover classes that exist in a remotely-sensed land cover product to a given model's PFTs. This step introduces the largest uncertainty in the entire process. The original land cover in the CLASSIC model is based on the GLC 2000 land cover product (https://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php). Table 2 of Wang et al. (2006)

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

summarizes the mapping/reclassification of the 22 GLC 2000 land cover categories to the nine PFTs used in CLASSIC. Each land cover class was split into either one or more of the nine CLASSIC PFTs based on the class description and knowledge of global biomes. For example, the discrete "broadleaf deciduous open tree cover" category of the GLC 2000 product is assumed to consist of 60% deciduous trees, 20% grasses, and 20% bare ground. This first step yields a snap shot of land cover expressed in terms of the fractional coverage of CLASSIC's nine PFTs. The second step of generating fractional coverage of PFTs for a given snap shot in time requires replacing the fractional area of crop categories with values from the LUH data set for the same year. For example, when using the GLC 2000 land cover product, the area of C_3 and C_4 crops from the LUH data set for the year 2000 are used and the fractional coverage of other seven non-crop CLASSIC PFTs are adjusted such that the total vegetation fraction in each grid cell stays the same. Finally, in the last step the temporally varying crop area from the LUH product is used to go backward in time to 1700 from year 2000 with typically decreasing crop area while the area of other non-crop PFTs is adjusted in proportion to their existing fractional coverage such that the total fractional vegetation cover stays the same . Similarly, area of C_3 and C_4 crops from the LUH product is used from year 2000 onwards to present day. All these steps yield a reconstruction of historical land cover, expressed in terms of fractional coverage of CLASSIC's nine PFTs (as interpreted from the GLC 2000 land cover product), from 1700 to 2019, in which crop area changes spatially and temporally according to the LUH product.

GLC 2000 is an older land cover product and more recent land cover products are now available. Here, in addition to the GLC 2000 based land cover for the CLASSIC model we also use the European Space Agency (ESA) Climate Change Initiative (CCI) land cover product. The ESA CCI

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

land cover product is available at 300 m spatial resolution for the period 1992-2018 and contains 37 land cover categories (ESA, 2017). Although a default mapping/reclassification table for converting the ESA CCI classes into PFTs is provided in its user guide (ESA, 2017), it overestimates tree cover along the taiga-tundra transition zone and underestimates it elsewhere in Canada (Wang et al., 2018). Wang et al. (2022, in preparation, Mapping of ESA CCI land cover data to plant functional types for use in the CLASSIC land model) have developed a new reclassification table for converting the 37 ESA CCI land cover categories to CLASSIC's nine PFTs which is used in this study. A high resolution land cover map over Canada and a tree cover fraction data at 30 m resolution are used to compute the sub-pixel fractional composition of each class in the ESA CCI dataset, which is then used to inform the cross-walking reclassification procedure (Wang et al., 2022, in preparation).

Figure 1 illustrates the uncertainty in land cover by comparing zonally summed areas of total vegetation, tree, and grass cover in CLASSIC when model land cover is based on the GLC 2000 (blue line) and ESA CCI (dark red line) land cover products. These two estimates are also compared to selected other models that participated in the 2020 TRENDY intercomparison (grey lines) for which land cover information was available, and to Li et al. (2018b) (dotted blackline) who analyzed the ESA CCI data based on the default reclassification table from the ESA CCI user guide. Figure 1 shows while there is relatively good agreement across TRENDY models in terms of total vegetation cover there's a much larger uncertainty in its split between tree and grass PFTs. This is because the current process of mapping/reclassifying 20-40 land cover classes of a land cover product to a model's PFTs is mainly based on class description and expert judgement that introduces some subjectiveness in the process. Compared to the GLC 2000 based land cover

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

in the CLASSIC model, the newer ESA CCI based land cover yields a somewhat higher total vegetation cover, a higher grass cover, and a somewhat lower tree cover area. Unlike the older GLC 2000 based land cover used in CLASSIC, the newer ESA CCI based grass and tree cover area are within the range of the TRENDY models reported here. Finally, Figure 1 also allows to compare the results from analysis of Li et al. (2018b) for the ESA CCI land cover (doted black line) to ESA CCI reclassification for CLASSIC (dark red line) by Wang et al. (2022, in preparation). Li et al. (2018b) used the default mapping/reclassification table for converting the ESA CCI classes into PFTs. This comparison illustrates that the remapping of the ESA CCI land cover classes to CLASSIC's PFTs yields total vegetation, tree, and grass coverage that is broadly comparable to Li et al. (2018b) although some differences remain for the grasses.

3.2 Meteorological data

As a land surface component of an ESM, CLASSIC requires meteorological forcing at a sub-daily temporal resolution. In the offline simulations reported here, the model is run with half hourly values of meteorological data (incoming long and shortwave radiation, temperature, precipitation, specific humidity, wind speed, and pressure). The first meteorological data set used to drive CLASSIC is from the TRENDY protocol for the year 2020, CRU-JRA v2.1.5, which provides 6 hourly values of the seven variables from the Japanese reanalysis (JRA) with monthly values adjusted to the climate research unit's data (CRU, https://crudata.uea.ac.uk/cru/data/hrg/). This yields a blended product from year January 1901 to December 2019 with the 6-hourly temporal resolution of a reanalysis but without the biases that may be present in reanalysis data (Harris, 2020). The second meteorological data set used here to drive CLASSIC is from the Global Soil Wetness Project 3 (GSWP3). The GSWP3 forcing data are based on a dynamical downscaling of

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

the 20th century reanalysis (Compo et al., 2011) using a Global Spectral Model (GSM) run at about 50 km resolution. GSM is nudged towards the vertical structures of 20th century (20CR) zonal and meridional air temperature and winds so that the synoptic features are retained at its higher spatial resolution. Additional bias corrections are also performed as explained in van den Hurk et al. (2016). The GSWP3 forcing is available for the 1901-2016 period. The 6-hourly values from both the CRU-JRA and GSWP3 forcings are further disaggregated to half hourly values for use by CLASSIC.

Figure 2 compares the two meteorological forcings data sets, over the 1996-2016 period, to illustrate that although these two data sets are very similar there are differences between the two. Global precipitation over land (excluding Greenland and Antarctica) in the GSWP3 data set (857 mm/year) is somewhat higher than in the CRU-JRA data set (820 mm/year). The global near-surface air temperature over land (excluding Greenland and Antarctica) is also slightly higher in the GSWP3 data set (14.22 °C) compared to the CRU-JRA data set (14.08 °C). The largest difference in temperature occur between the two data sets over northern tropics (panel h) where the GSWP3 data set is about 0.93 °C warmer than the CRU-JRA data set. The geographical distribution of mean annual temperature is very similar between the two data sets but there are some differences in the geographical distribution of precipitation (not shown). Despite very similar total precipitation amounts and their seasonality over large global regions in the two data sets, differences exist in the frequency distribution of precipitation. Figure A1 illustrates this over three broad regions, the Amazon, the Sahel, and the Midwest United States, which shows the frequency distribution of daily precipitation amounts (mm/day) over the 2001-2010 period from the two data sets. Figure A1 shows that the frequency of precipitation events greater than about

5-10 mm/day is higher in the GSWP3 data set compared to the CRU-JRA data set for the Amazonian, the Sahel, and the Midwest United States regions.

3.3 Other forcings

Other than land cover and meteorological forcings CLASSIC requires globally averaged atmospheric CO_2 concentration, and geographically varying time-invariant soil texture and soil permeable depth, population density, time-invariant monthly lightning, and geographically and time varying N fertilizer application rates and atmospheric N deposition rates. The atmospheric CO_2 concentration values are provided by the TRENDY protocol. The soil texture information consists of percentage of sand, clay, and organic matter and derived from Shangguan et al. (2014). N fertilizer is specified according to the TRENDY protocol and based on Lu and Tian (2017). N deposition is also specified according to the TRENDY protocol and based on model forcings provided for sixth phase of CMIP (CMIP6) through input4MIPs (Hegglin et al., 2016). N deposition for the historical (1850-2014) period is used as is provided while that for the period 2015-2019 is specified on the basis of N deposition from the SSP5-85 scenario. For the period 1700-1849, N deposition values from year 1850 are used.

3.4 Model simulations

Using the two reconstructions of the historical land cover (based on the GLC 2000 and ESA CCI land cover products), the two sets of meteorological data (CRU-JRA and GSWP3), and the two versions of the CLASSIC model (with and without interactions between the C and N cycles) we perform eight sets of pre-industrial and historical simulations as summarized in Table 1. Pre-industrial simulations that correspond to year 1700 are required prior to doing the historical

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

simulations (from which we analyse the model results) so that model pools can be spun up to near equilibrium for each combination of land cover, meteorological forcing, and model version. The pre-industrial simulations use 1901-1925 meteorological data repeatedly since this period shows little trends in meteorological variables. Global thresholds of net atmosphere-land C flux of 0.05 Pg/yr and net atmosphere-land N flux of 0.5 Tg N/yr, in simulations with the N cycle turned on, are used to ensure the model pools have reached equilibrium. Each historical simulation is then initialized from its corresponding pre-industrial simulation after it has reached equilibrium. Simulations driven with the CRU-JRA meteorological data are performed for the period 1701-2019, and for the period 1701-2016 for simulations driven with the GSWP3 meteorological data. Similar to the pre-industrial simulations, meteorological data from 1901-1925 is used repeatedly for the period 1701-1900. The global model simulations are performed at a spatial resolution of about 2.81° and the size of the spatial longitude-latitude grid is 128 × 64. All model forcings are regridded to this common spatial resolution. The model is run over about 1900 land grid cells at this resolution excluding glacial cells in Greenland and Antarctica.

3.5 Automated benchmarking

The results from the eight CLASSIC simulations reported here are evaluated using an in-house model benchmarking system called the Automated Model Benchmarking R package (AMBER) (Seiler et al., 2021b). AMBER is based on a skill score system originally developed by (Collier et al., 2018) which is used to quantify model performance. Five scores are used that assess a model's bias (S_{bias}), root-mean-square error (S_{rmse}), seasonality (S_{phase}), interannual variability (S_{iav}), and spatial distribution (S_{dist}) against globally gridded and in-situ data set(s) of

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

observation-based estimates for a given quantity. A score is computed by first calculating a dimensionless statistical metric, that is then scaled onto a unit interval, and finally calculating its spatial mean. Scores range from 0 to 1 and are dimensionless. Higher values indicate better performance. Finally, an overall score $S_{overall}$ is calculated as follows by giving twice as much weight to S_{rmse} given its importance

$$S_{overall} = \frac{S_{bias}+2S_{rmse}+S_{phase}+S_{iav}+S_{dist}}{1+2+1+1+1}. \quad (1)$$

The scores are calculated by comparing gridded and in-situ observation-based estimates, referred to as reference data sets in Seiler et al. (2021b), of 19 energy (surface albedo, net shortwave and longwave radiation, total net radiation, latent heat flux, sensible heat flux, ground heat flux), water (soil moisture, snow, and runoff), and C cycle (GPP, net ecosystem exchange, net biome productivity, aboveground biomass, soil C, LAI, area burnt, and fire CO_2 emissions) related variables to model simulated quantities. Table 2 summarizes the source of these observation-based data sets. The resulting model scores express to what extent simulated and observation-based data agree. A low score does not necessarily indicate poor model performance. Uncertainties in the meteorological forcing data and geophysical fields used to drive the model, and/or in the observation-based data itself are possible reasons for lack of agreement. One way to assess uncertainties in observation-based data sets is to quantify the skill score by comparing two independently-derived observation-based data sets (Seiler et al., 2022). The resulting scores are referred to as benchmark scores and quantify the level of agreement among the observation-based data sets themselves provided, of course, there are at least two

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

sets of observation-based data for a given quantity. The comparison of model scores against benchmark scores then shows how well a model-simulated quantity compares to the references data sets relative to the agreement between the observation-based data sets themselves.

4. Results

Figures 3 through 9 show the physical and biogeochemical states of the land surface and primary physical fluxes of water and energy, and primary biogeochemical fluxes of CO_2 simulated by CLASSIC at the land-atmosphere boundary for the eight simulations considered here. The objective is to illustrate how the simulated physical and biogeochemical states and fluxes vary across the eight simulations. Supplementary Figures A2 through A16, which are complementary to Figures 3 through 9, show the time series and/or zonally-averaged values of annual values of a variable of interest when averaged across four ensemble members each according to whether N cycle is turned on or not (panel a), whether GLC 2000 or ESA CCI based land cover is used (panel b) and whether model simulations are driven by CRU-JRA or GSWP3 meteorological data (panel c). While Figures 3 to 9 illustrate the range across the eight simulations, the supplementary figures evaluate the effect of model structure, meteorological forcing, and land cover on a given quantity. We also quantify the spread across the eight simulations using the coefficient of variation (cv= standard deviation/mean) calculated using annual global values for a given quantity averaged over last 20 years of each simulation.

4.1 Physical land surface state and fluxes

Figure 3 shows the globally-averaged simulated soil moisture and temperature in the top 1 m soil layer. While simulated soil temperature in the top 1 m is fairly similar across the eight

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

simulations, the simulated soil moisture is distinctly separated into two groups. This separation into these two groups is caused by the driving meteorological data as shown in Figure S2. The coefficient of variation for soil moisture and temperature values averaged over the last 20 years of each simulation are 0.02 and 0.004, respectively, indicating that overall the variation in these quantities is relatively small compared to their means. The use of the GSWP3 meteorological dataset yields slightly higher (~4%) globally-averaged soil moisture compared to the CRU-JRA meteorological data set (236.5 mm vs. 227.1 mm, Figure S2).

Figure 4 shows the simulated fluxes of global evapotranspiration and runoff. Similar to soil moisture, evapotranspiration and runoff also fall broadly into two groups and the reason for this again is the driving meteorological data. Figure A3 and A4 show that while interactive N cycle also affects evapotranspiration and runoff fluxes, the biggest factor is the difference in driving meteorological data. Neither evapotranspiration nor runoff are significantly affected by the choice of land cover. The reason an interactive N cycle affects evapotranspiration is that the N cycle in CLASSIC affects the rate of photosynthesis through prognostic determination of leaf N content. Photosynthesis in turns affects canopy conductance, which affects transpiration through the canopy leaves. Average evapotranspiration over the last 20 years of the simulations driven with GSWP3 meteorological data (1997-2016) is about 9% lower than in simulations driven with CRU-JRA meteorological data (1999-2018) (65.89 vs. 72.1 ×1000 km^3/year, Figure S4, panel c). Interactive N cycle reduces evapotranspiration by about 2% due to lower photosynthesis rates as shown later (Figure S4, panel a). Average runoff is about 27% higher in simulations driven with GSWP3 compared to simulations driven with CRU-JRA meteorological data (52.6 vs 41.3 ×1000 km^3/year, Figure S3, panel c). This is due to high slightly precipitation in the GSWP3

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

meteorological data set (Figure 2) but more so the simulated lower evapotranspiration when using the GSWP3 data (Figure S4, panel c). The coefficient of variation for evapotranspiration and runoff values averaged over the last 20 years of each simulation are 0.05 and 0.13, respectively.

Figure 5 shows the primary energy fluxes from the eight simulations. These include net downward shortwave and longwave radiation, and latent and sensible heat fluxes. Incoming shortwave and longwave radiation are part of the driving meteorological data. Similar to water fluxes, the differences in energy fluxes in CLASSIC are also primary driven by differences in meteorological data (Figure S5). Net shortwave radiation (Figure 5a) is equal to incoming shortwave radiation minus the fraction that is reflected back. Net longwave radiation (Figure 5b) is equal to incoming longwave radiation minus the longwave radiation emitted by the land based on its surface temperature following the Stefan-Boltzmann law. The difference in net shortwave radiation is also affected among other things by simulated vegetation biomass and leaf area index. The latter affects surface albedo that determines what fraction of incoming shortwave radiation is reflected back. This is the reason why an interactive N cycle affects shortwave radiation since the N cycle affects photosynthesis, and in turn simulated vegetation biomass and leaf area index (Figure S5). Latent heat flux is affected primarily by meteorological data but also if N cycle is interactive or not since it is essentially evapotranspiration but in energy units. Finally, differences in sensible heat fluxes are strongly affected by differences in driving meteorological data. Globally-averaged sensible heat flux in the simulations driven with GSWP3 data is ~14% higher compared to CRU-JRA driven simulations (40 vs. 35 W/m^2). The coefficient of variation for sensible heat flux values averaged over the last 20 years of each simulation is 0.07. Net shortwave (cv=0.006) and longwave (cv=0.03) radiative fluxes vary little across the eight simulations..

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Overall runoff (cv=0.13), sensible heat flux (cv=0.07), and evapotranspiration (latent heat flux) (cv=0.05) are most affected by the driving meteorological data but soil moisture and temperature not as much.

4.2 Biogeochemical land surface state and fluxes

4.2.1 Primary CO_2 fluxes and C pools

Figure 6 shows the simulated C state of the land surface expressed in terms of vegetation and soil C pools. Panels a and b show the annual time series of global vegetation and soil C mass from the eight simulations, and panels c and d show their zonally-averaged distributions averaged over the last 20 years of each simulation. The biggest difference in time series of global vegetation (cv=0.16) and soil (cv=0.21) C mass compared to soil moisture and temperature, which characterized the physical land surface state, is the large spread across the eight simulations as indicated by their high cv values. The zonally-averaged values further provide insight into the reasons for this spread and show that the largest differences between simulated vegetation and soil C occur at northern high latitudes (north of about 40°N). Panels c and d of Figure 6 also show observation-based zonally-averaged values of vegetation and soil carbon mass based on the Reusch and Gibbs (2008) and the Harmonized World Soils Database (v1.2) (Fischer et al., 2008), respectively, to provide a reference. A more thorough comparison with observations is provided in Section 4.3.

Differences in vegetation biomass are caused primarily when the N cycle is interactive or not (Figures A6 and S8). Both land cover and the driving meteorological data play a smaller role in the simulated spread in vegetation biomass (Figure S6). The ESA CCI based land cover has larger

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

vegetated area but most of this increase comes from an increase in the area of grasses that do not store a lot of C in their vegetation biomass. The spread in simulated soil carbon is caused due to N cycle but also by the choice of land cover (Figures A7 and S9). Since CLASSIC assumes that litter from grasses is more recalcitrant than that from trees the choice of ESA CCI based land cover leads to higher soil C mass. The choice of meteorological data doesn't affect the magnitude of simulated globally-summed soil C mass significantly but does affect its change over the historical period. In Figure A7 (panel c) the decrease in soil C mass over the 1700-2016 historical period is higher when using the GSWP3 (28 Pg C) compared to when using the CRU-JRA (12 Pg C) meteorological data.

The reason why an interactive N cycle in CLASSIC affects vegetation biomass and soil carbon is seen in Figure 7 which shows the spread of primary C fluxes including gross primary productivity (GPP) (cv=0.07), and autotrophic (cv=0.04) and heterotrophic (cv=0.10) respiratory fluxes, across the eight simulations. Since GPP is lower in the runs with the N cycle, both vegetation biomass (Figure S6a) and soil C mass (Figure S7a) are also lower. The lower GPP in the runs with N cycle is due primarily to lower GPP at high latitudes (Figure 7d), as mentioned earlier, which yields low vegetation biomass at high latitudes (Figure S8a). Low GPP at high latitudes translates to even larger relative differences in soil C given the longer turnover time scales of soil C at high latitudes (Figure 6d, Figure S8a).

Overall, while the primary biogeochemical fluxes (cv values vary from 0.04 to 0.10) vary as much as the water and energy fluxes the resulting spread in vegetation biomass (cv=0.16) and soil carbon mass (cv=0.21) across the eight simulations is much larger and driven primarily if N cycle is interactive or not and the difference in land cover.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

4.2.2 Area burned and fire CO_2 emissions

Figure 8 shows the time series of global area burned and global fire CO_2 emissions, and their zonally-averaged values. We chose area burned (cv=0.24) and fire CO_2 emissions (cv=0.21) in addition to the primary biogeochemical fluxes since fire shows large variability both in space and in time, and both these variables yield the largest spread across the eight simulations, among all the fluxes and simulated quantities considered here. Figures 8c and 8d also show observation-based estimates for area burned and fire CO_2 emissions based on GFED 4s (Giglio et al., 2013) to provide an observation-based context. Figure A13 and A14 help understand which factors contribute to this large variability. The variability in area burned is caused primarily by the choice of land cover and meteorological data and the variability is higher in the southern hemisphere (Figure A13). An interactive N cycle does not affect the zonal distribution area burned and fire CO_2 emissions (Figure A14) as much. The reason both area burned and fire CO_2 emissions are affected by the choice of land cover is because the ESA CCI land cover has higher grass area and as a result it yields higher area burned and fire CO_2 emissions since a larger area is burned for grasses than for trees in the model. The choice of driving meteorological data is a factor in area burned and our simulations show that the use of GSWP3 meteorological forcing yields to a higher area burned than the CRU-JRA data. In particular wind speed, which determines rate of spread of fire in CLASSIC, is much higher in the GWSP3 than in the CRU-JRA meteorological data. Globally-averaged land wind speed (excluding Greenland and Antarctica) in GSWP3 data is 6.1 m/s compared to 3.4 m/s in the CRU-JRA data for the period 2000-2016.

Table 3 shows the energy, water, and carbon related quantities considered so far and lists them from the most variable at the top to the least variable at the bottom according to their

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

coefficient of variation. Area burned is found to be the most variable quantity and net shortwave radiation the least variable.

4.2.3 Net biome productivity

Figure 9 shows the spread in the time series of annual global net biome productivity (NBP) values and their zonally-averaged values across the eight simulations averaged over the last 20 years of each simulation. The global NBP or the net atmosphere-land CO_2 flux is considered a critical determinant of the performance of land models, and is treated as such by TRENDY, because this flux ultimately affects changes in atmospheric CO_2 burden. TRENDY requires that land models simulate a terrestrial C sink for the decades of 1990s to present to be considered for inclusion in the TRENDY ensemble. Figure 9a also shows the estimates of NBP from the participating TRENDY models in grey boxes with mean and shaded range for the decades from 1960s to 2010s from the Global Carbon Project (Friedlingstein et al., 2022). Positive values in Figure 9 indicate a C sink over land and negative values a C source to the atmosphere.

From Figure 9a, all eight simulations reported here would qualify for inclusion in the TRENDY ensemble since they all simulate a terrestrial C sink during 1990s to the present day. In addition, the time series of global NBP from all eight simulations lie within the uncertainty range of reported estimates from the Global Carbon Project. Figure 9a suggests on the basis of global NBP, at least, it is not possible to exclude any of the eight simulations. In Figure 9b, zonally-averaged NBP averaged over the last 20 years from each of the eight simulations mostly lie within the range of NBP simulated by models that participated in TRENDY 2020. CLASSIC simulates a C sink at northern high-latitudes consistent with TRENDY models but it simulates a C sink on the

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

stronger side of TRENDY models in the southern tropics (0° - 20°S). This is likely because CLASSIC is known to simulate low C emissions associated with land use change most of which are generated in tropical regions (Asaadi and Arora, 2021).

Figures A15 and A16 provide additional insights into the effect of different forcings on the simulated NBP. In Figure A15, over the last 20 years of the simulations, an interactive N cycle leads to somewhat weaker C sink (panel a, 0.98 vs. 1.11 Pg C/yr), the choice of the ESA CCI based land cover leads to a somewhat stronger C sink (panel b, 1.14 vs 0.94 Pg C/yr), and the choice of the GSWP3 meteorological data leads to a much weaker C sink (panel c, 0.74 vs 1.33 Pg C/yr) than the CRU-JRA meteorological data. In Figure A16, panel a, the largest difference between the model versions with and without the N cycle occurs in the tropics (~ 5°N - 20°S) where an interactive N cycle leads to a weaker C sink. There are difference in zonally-averaged NBP with and without the N cycle south of 45°S but the land area below this latitude is small so the averages are calculated over only a few grid cells. The choice of land cover (Figure A16, panel b) doesn't substantially change the distribution of the zonally-averaged values of NBP although as noted above the choice of ESA CCI based land cover leads to a somewhat stronger sink. Finally, the choice of the GSWP3 meteorological forcing leads to a weaker C sink at most latitudes (Figure A16, panel c).

4.3 Automated benchmarking

Figure 10 plots the overall score, $S_{overall}$, against benchmark scores for several of the energy, water, and C cycle related variables. AMBER does not yet evaluate N cycle related variables for which observations are more scarce than for C cycle related variables. The whiskers

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

show the range in the overall score both for the benchmark and model scores. The range in model scores comes from the eight simulations, and the range in benchmark scores comes from the different observation-based data sets. Figure 10 shows that typically as the benchmark scores increase so do the overall model scores for a given quantity. This indicates that uncertainty in observation-based estimates themselves leads to a poor agreement between observations and model-simulated quantities. For energy and water fluxes score (panels a and b) the model overall scores lie around the 1:1 line indicating that model scores are generally as good as the benchmark scores, except for surface albedo (ALBS), runoff (MRRO), ground heat flux (HFG), and comparison against one observation-based estimate of snow water equivalent which lie below the 1:1 line. For C cycle related variables most scores lie somewhat below the 1:1 line indicating that simulated quantities do not agree as well with observations as observations agree among themselves. The lower benchmark scores for soil C (panel c) is due to the fact that the SoilGrids250m (SG250m) data and the Harmonized World Soil Database (HWSD) do not agree well amongst themselves because the SG250m soil C data includes peatlands and permafrost carbon at high latitudes while the HWSD data does not (see Figure 11b). Since the version of CLASSIC used here does not represent peatlands and permafrost C it compares better with the HWSD data than with the SG250m data.

Figure 11 shows the zonal distribution of vegetation biomass, LAI, area burnt, GPP, and fire CO_2 emissions (which constitute standard output from AMBER) and illustrates how AMBER compares the spread across the simulations indicated by 50%, 80%, and 100% shading against observation-based estimates. The red, orange, and yellow colours indicate the model mean and the spread across the eight model simulations and the thick lines in other colours show the mean

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

values of observation-based estimates. The time period over which observations and model-quantities are averaged are chosen to be the same. Figure A17 and A18 compare zonally averaged values of other simulated quantities with observation-based estimates used in the AMBER framework. Together Figures 11, A17, and A18 illustrate that the model is overall able to capture the latitudinal distribution of most land surface quantities.

Since overall scores are available for all eight simulations for model quantities that are compared to observations it is possible to evaluate how an interactive N cycle, and the choice of meteorological data and land cover data affects model performance. Figure 12 summarizes the difference in overall scores for model quantities and combinations for which the differences are statistically significant at the 5% level based on Tukey's test (Tukey, 1977). The score indicated in parentheses for each quantity is the average score across the eight simulations and provides context. For example, when evaluating the effect of change in land cover for NEE the use of the GLC 2000 based land cover, compared to the use of the ESA CCI based land cover, degrades the average score for net ecosystem exchange by about 0.02 given that the average score for net ecosystem exchange in 0.53. The use of the GLC 2000 based land cover on the other hand slightly improves scores for ecosystem respiration and liquid soil moisture. The use of GSWP3 data improves model scores for net shortwave, longwave, and total radiation, for sensible and ground heat flux but degrades the overall score for area burned, soil moisture, and more so for snow water equivalent. Finally, an interactive N cycle slightly improves model performance for area burned and fire CO_2 emissions (due to improved aboveground biomass in the tropics) but degrades it for ecosystem respiration, GPP, and net ecosystem exchange. Overall, the largest effect on model performance is due to the driving meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

The full suite of results from AMBER for the eight simulations presented in this study can be found at https://cseiler.shinyapps.io/ShinyCLASSIC/.

5. Conclusions

The results presented in this paper help draw three primary conclusions. First, even if the observations and models were perfect (including their structure and their parameterizations) the uncertainty associated with driving meteorological data and geophysical fields make it difficult to evaluate land models. The uncertainty in global scale driving data implies that a model can never be truly evaluated to its fullest extent. Model results can only be as good as the data that are used to force them and therefore even a perfect model cannot yield perfect results.

Second, model tuning when driving the model with a single set of forcings and evaluating it against a single set of observations is likely not a fruitful exercise. Models should not be tuned to a single set of driving data and rather their performance must be evaluated against a range of available observations in light of the uncertainty associated with driving data and the uncertainty associated with observations. A model's ability to reproduce a given single set of observations when driven with a single set of driving data is not a true measure of its success. Here again, a perfect model driven by perfect forcing data cannot be truly evaluated to its fullest extent since observations themselves have uncertainties.

Third, the response of a land model expressed in terms of net atmosphere-land CO_2 flux to perturbation in meteorological, CO_2, and land use change forcing over the historical period appears to be largely independent of its pre-industrial state as simulated here. The pre-industrial soil and vegetation C mass for the eight simulations considered here vary between 1035 ± 195

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Pg C and 405 ± 58 Pg C (mean ± standard deviation), respectively. Both pre-industrial and present-day vegetation and soil carbon pools (Figure 7) explain only about 2% to 7% of the variability in simulated net atmosphere-land CO_2 flux (Figure 10) over the last 20-year of each of the eight simulations. The net atmosphere-CO_2 flux from all eight simulations for the period 1960s to 2000s is found to lie within the uncertainty range provided by the GCP (Friedlingstein et al., 2022). Given the current uncertainty in net atmosphere-land CO_2 flux it is therefore not possible to exclude any of the eight simulations at least on this basis. The finding that a transient response of a model is independent of its preindustrial state is also consistent with land components of CMIP6 models. Arora et al. (2020) analyzed results from CMIP6 simulations in which atmospheric CO_2 increases at a rate of 1% per year from year 1850 until CO_2 quadruples from ~285 to ~1140 ppm. They found that the carbon-concentration and carbon-climate feedback parameters for the land component of CMIP6 models do not depend on the absolute values of their vegetation and soil carbon pools but rather how a given model responds to changes in atmospheric CO_2 and the associated change in temperature. This conclusion is perhaps somewhat comforting in that while pre-industrial states of land models may be different from their true observed states they still have the ability to reproduce net atmosphere-land CO_2 flux over the historical period that is consistent with current observation-based estimates. Clearly, this reasoning does not apply if pre-industrial vegetation or soil C mass are zero. However, successful reproduction of atmosphere-land CO_2 fluxes over the historical period is no guarantee that future projections from land models are reliable.

The ensemble-based approach used here also allows for the evaluation of the effect of an interactive N cycle on model simulated quantities in a robust manner. By comparing simulations

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

with and without the N cycle averaged over all meteorological data and land cover combinations we are able to clearly identify the effect of N cycle. In particular, we found that the somewhat low productivity at high latitudes, when the N cycle is turned on, leads to relatively large differences in soil carbon at high latitudes regardless of the meteorological data or land cover being used to drive the model. Although, this is not the reason for differences in net atmosphere-land CO_2 flux between models with and without N cycling: as mentioned above present-day net atmosphere-land CO_2 flux is independent of both the pre-industrial and present-day vegetation and soil carbon pools. Given the knowledge about the effect of N cycling on model behaviour, the reasons can now be investigated to further improve the N cycle component of CLASSIC.

It is logical to assume that the results presented here are sensitive to the horizontal resolution of the model. Both forcing data that are used to drive the model, and observations against which model results are compared, are regridded to be consistent with the model's spatial resolution. For example, at the scale of a few meters, meteorological variables measured at a given site will indeed be less uncertain than their spatially-averaged values say for a 2.81° grid cell. Similarly, observations at a scale of a few meters for soil carbon and/or vegetation biomass will also likely be more certain than their values at large spatial scales. This is one reason why AMBER uses both gridded and in-situ observation-based estimates to calculate its scores. Fluxes of latent and sensible heat, on the other hand, may not be any more certain at a given site than over large spatial scales. This is because of the problems associated with energy budget closure (Mauder et al., 2020) which, at the point scale, prevent the sum of annual latent and sensible heat flux to be equal to net radiation (average of ground heat fluxes is close to zero at an annual time scale).

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Land models have become increasing complex over the years and so has the requirement for forcing data to drive these models. The evaluation of land models has also become complex as the models now generate a multitude of variables which must be evaluated against their observation-based estimates. Estimates of observation-based data to evaluate models, and the availability of forcing data, have also increased. Given the uncertainties associated with model inputs, model structure, and observation-based data, it is unrealistic to expect land models to perfectly reproduce observations for large-scale global simulations. Rather a more robust model evaluation must take into account the uncertainties both in the forcing and observations-based data. A comprehensive and robust model evaluation can be performed by comparing multiple model realizations against multiple observation-based data sets.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Appendix

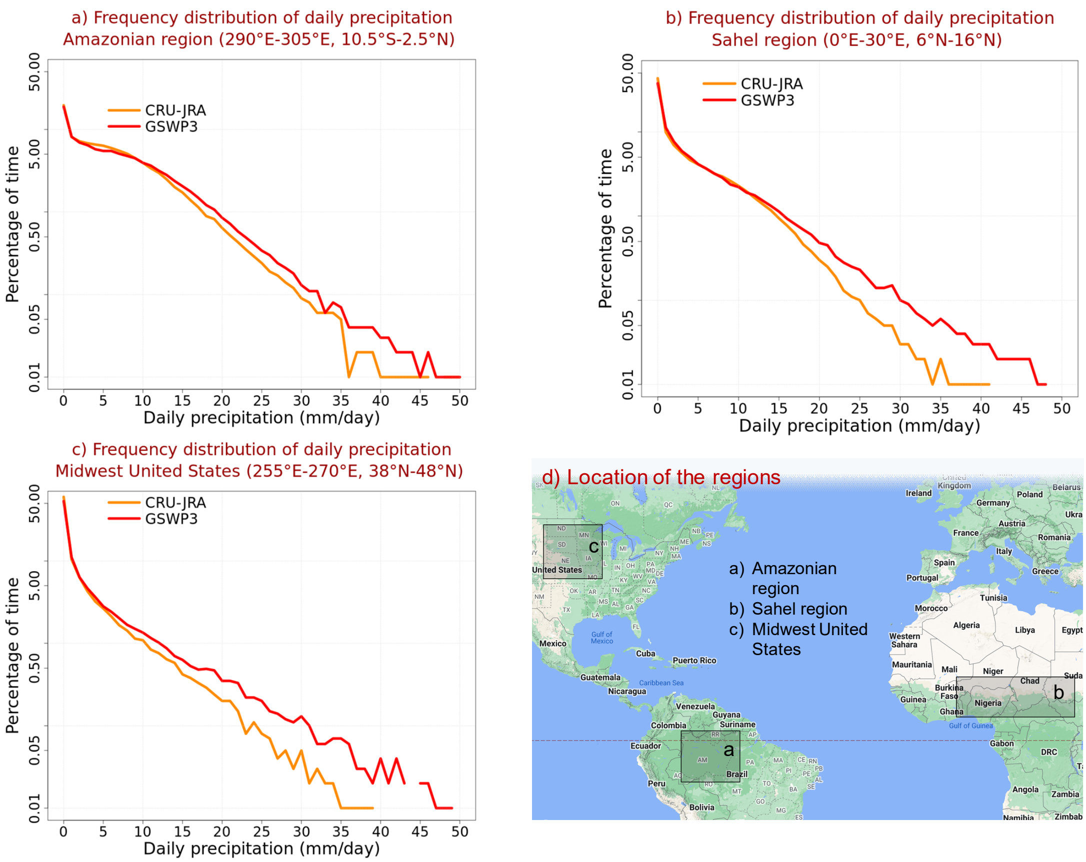


Figure A1: Comparison of frequency distribution of daily precipitation between the CRU-JRA and GSWP3 meteorological data sets for three broad regions and for the period 2001-2010: a) the Amazonian region, b) the Sahel region, and c) the Midwest United States. The frequency is represented as percentage of time daily precipitation is between *x* and *x*+1 mm/day, where x is the value on the x-axis. Panel (d) shows the location of these broad regions. The underlying map in panel (d) is from Google Maps.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

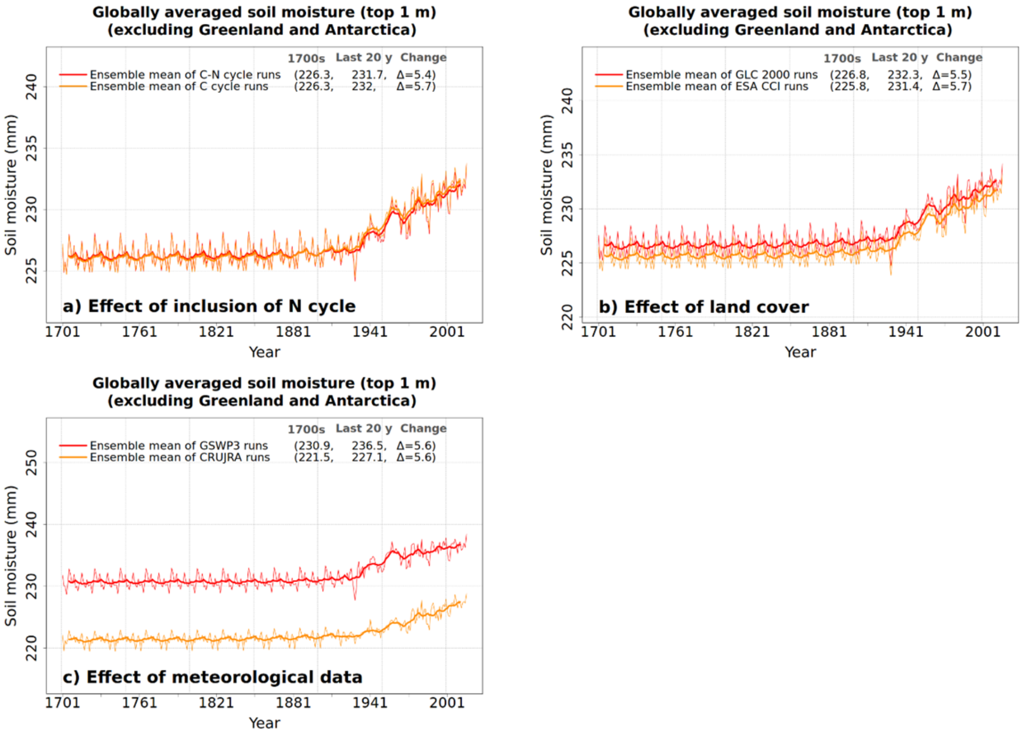


Figure A2: Comparison of time series of annual globally-averaged soil moisture in the top 1m averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

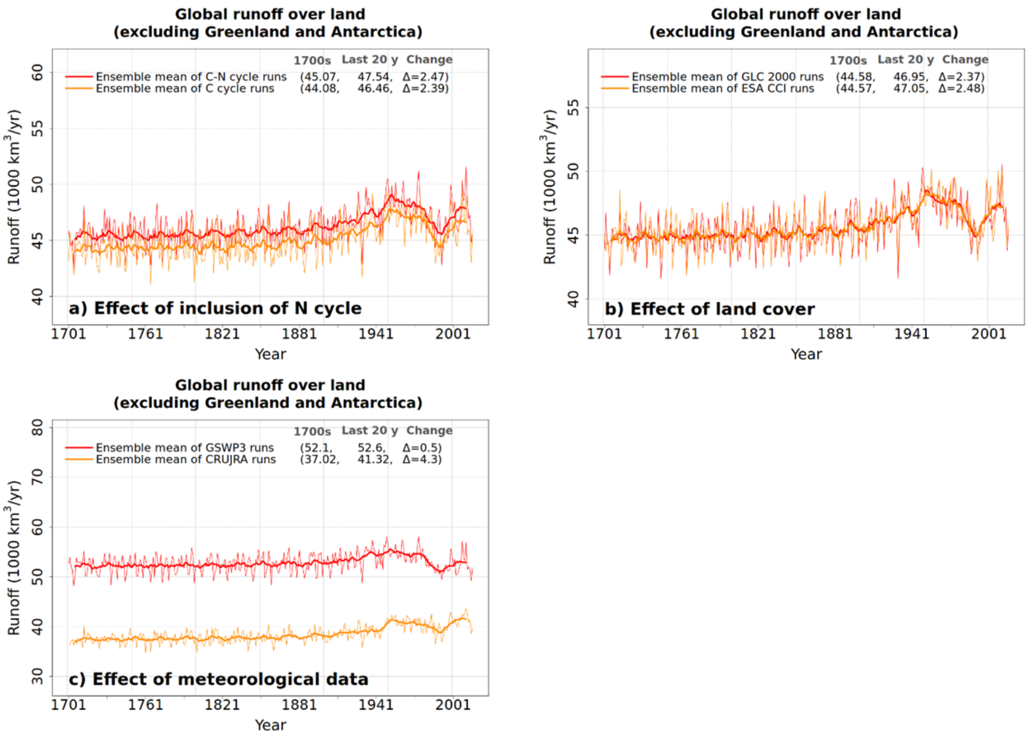


Figure A3: Comparison of time series of annual global runoff values averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

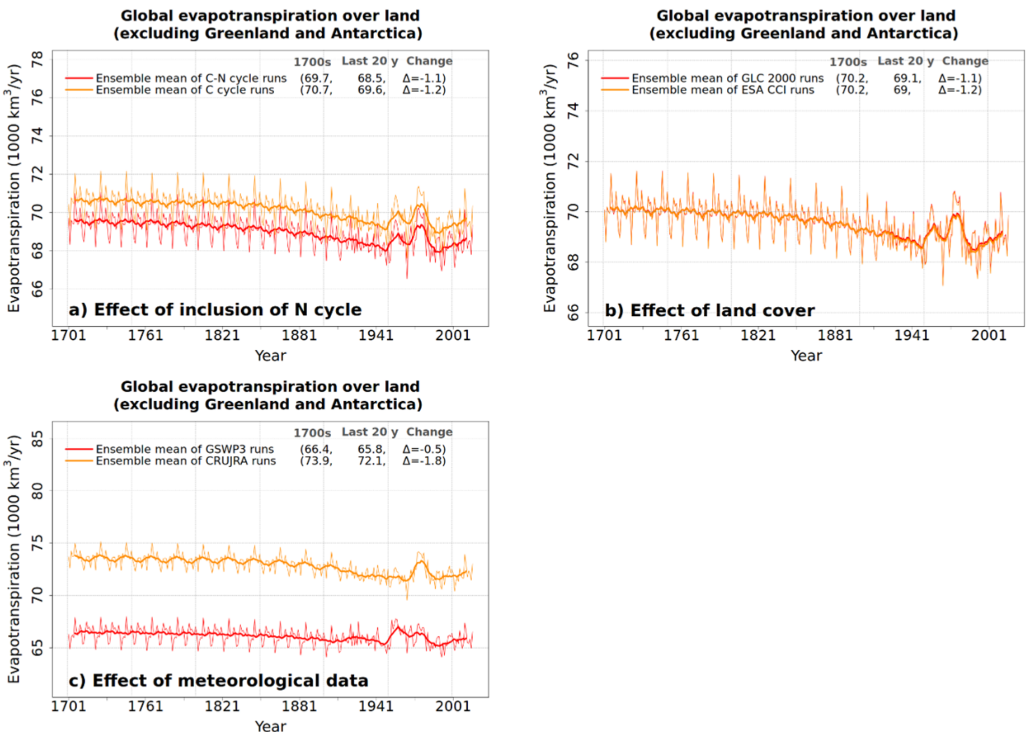


Figure A4: Comparison of time series of annual global evapotranspiration (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

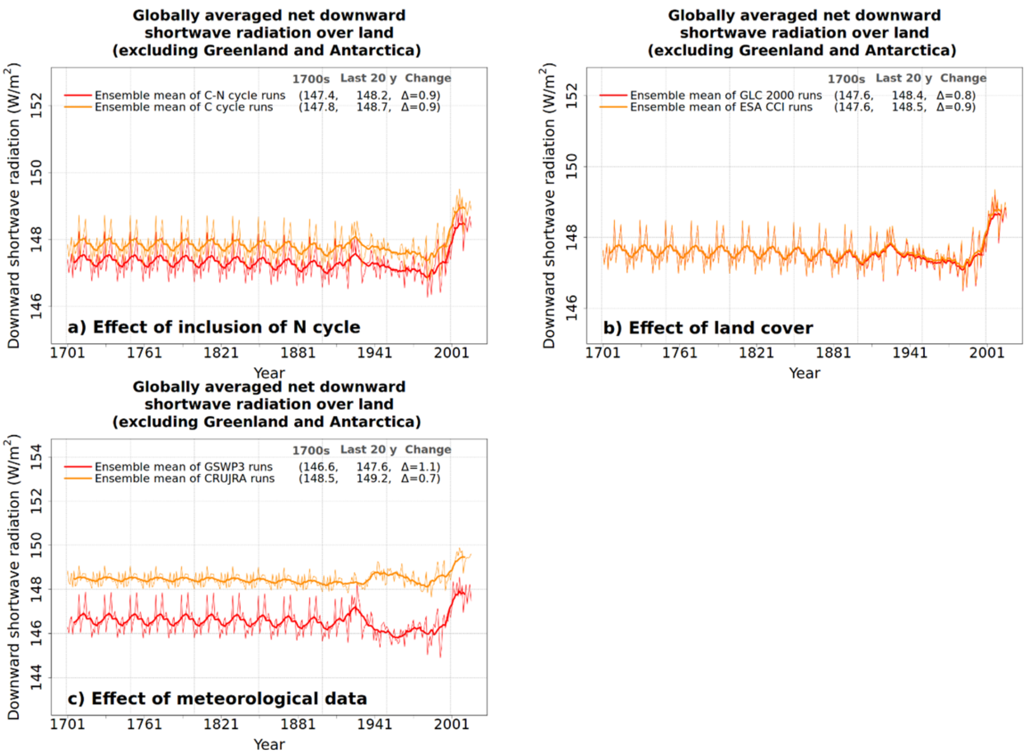


Figure A5: Comparison of time series of annual globally-averaged net downward shortwave radiation (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

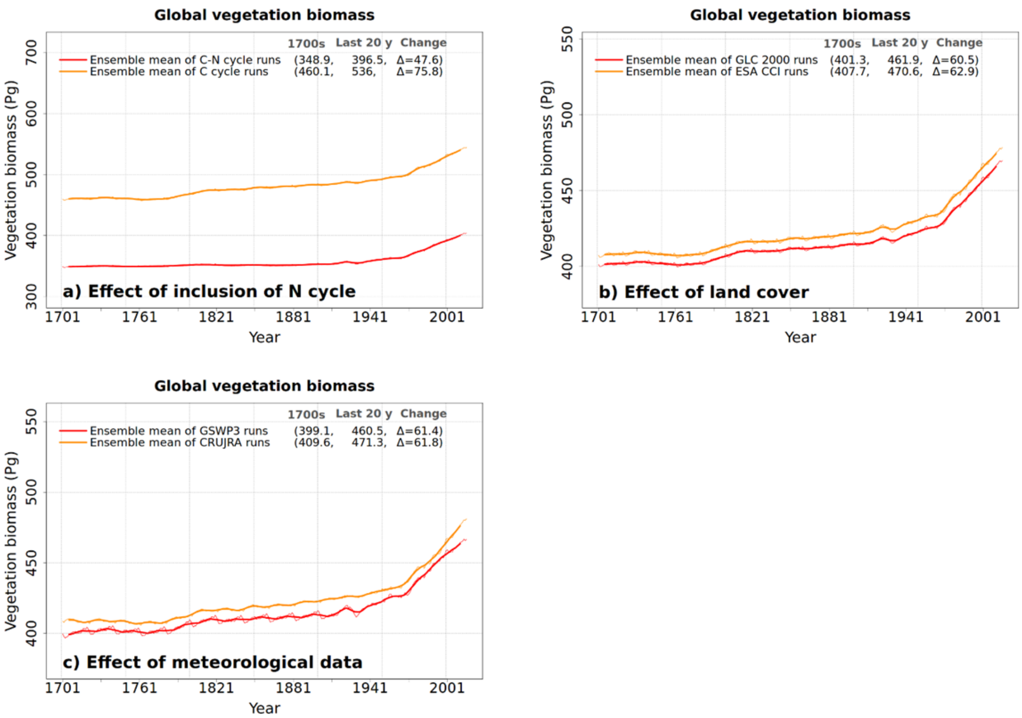


Figure A6: Comparison of time series of annual global vegetation biomass (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

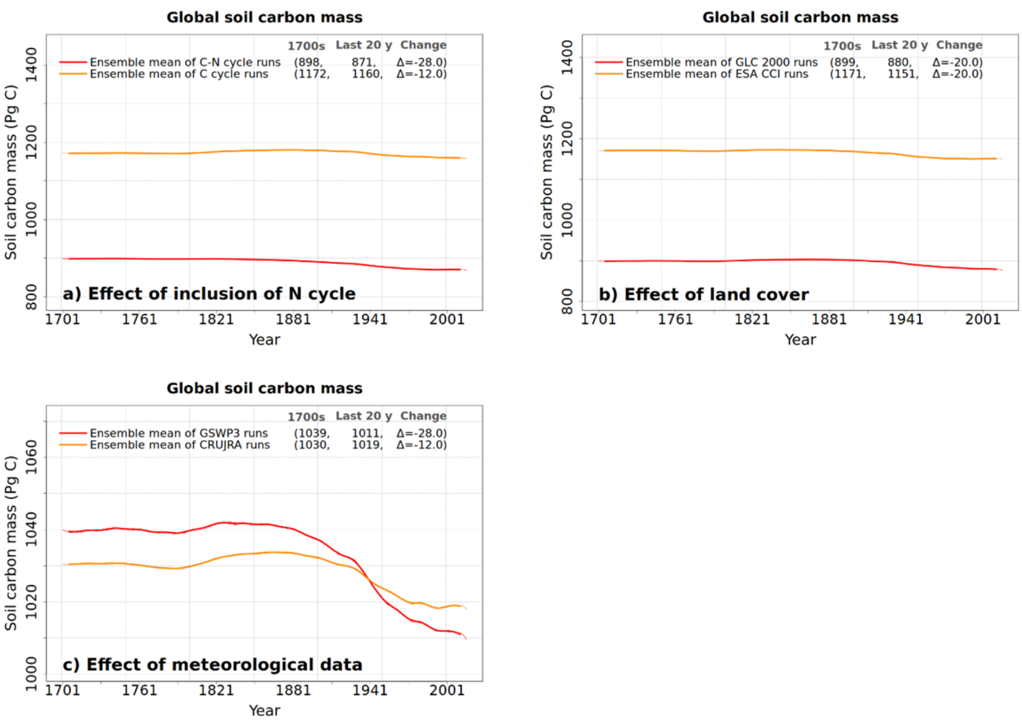


Figure A7: Comparison of time series of annual global soil carbon mass (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

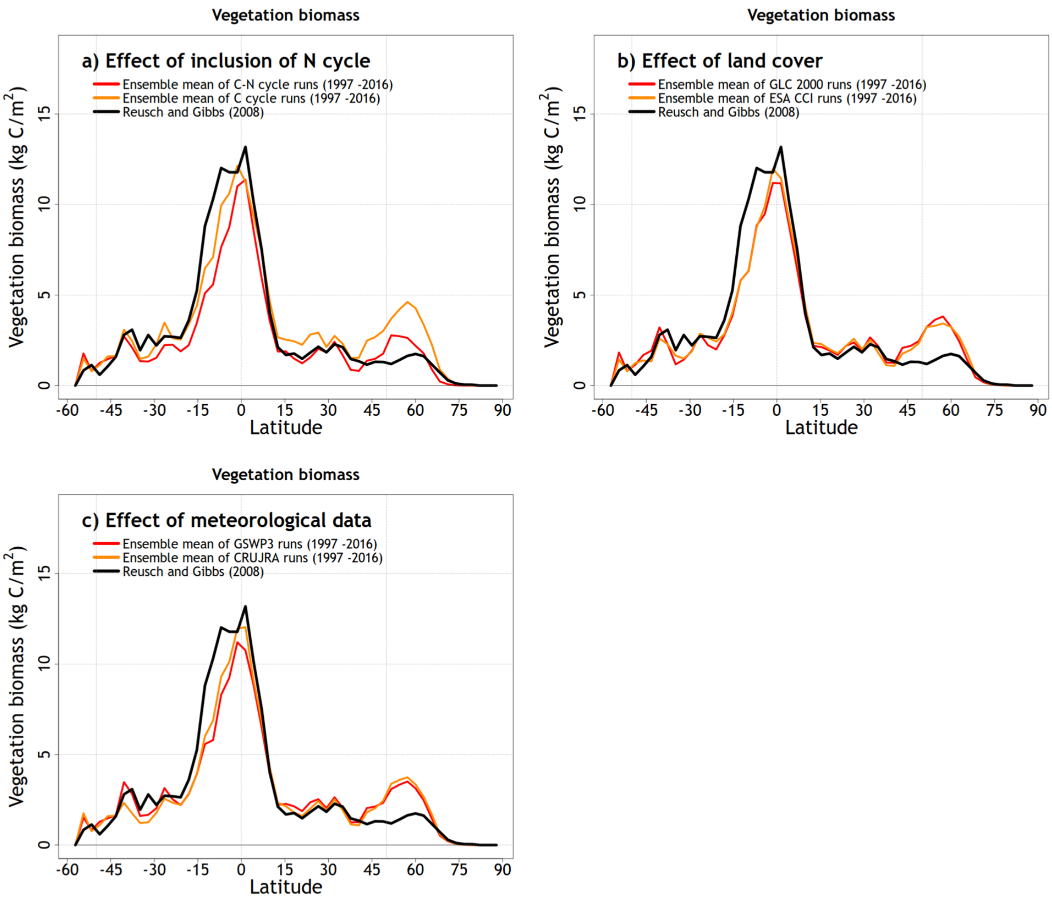


Figure A8: Comparison of zonally-averaged vegetation biomass (over all land area excluding Greenland and Antarctica) time-averaged over last 20-years of each simulation and then averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

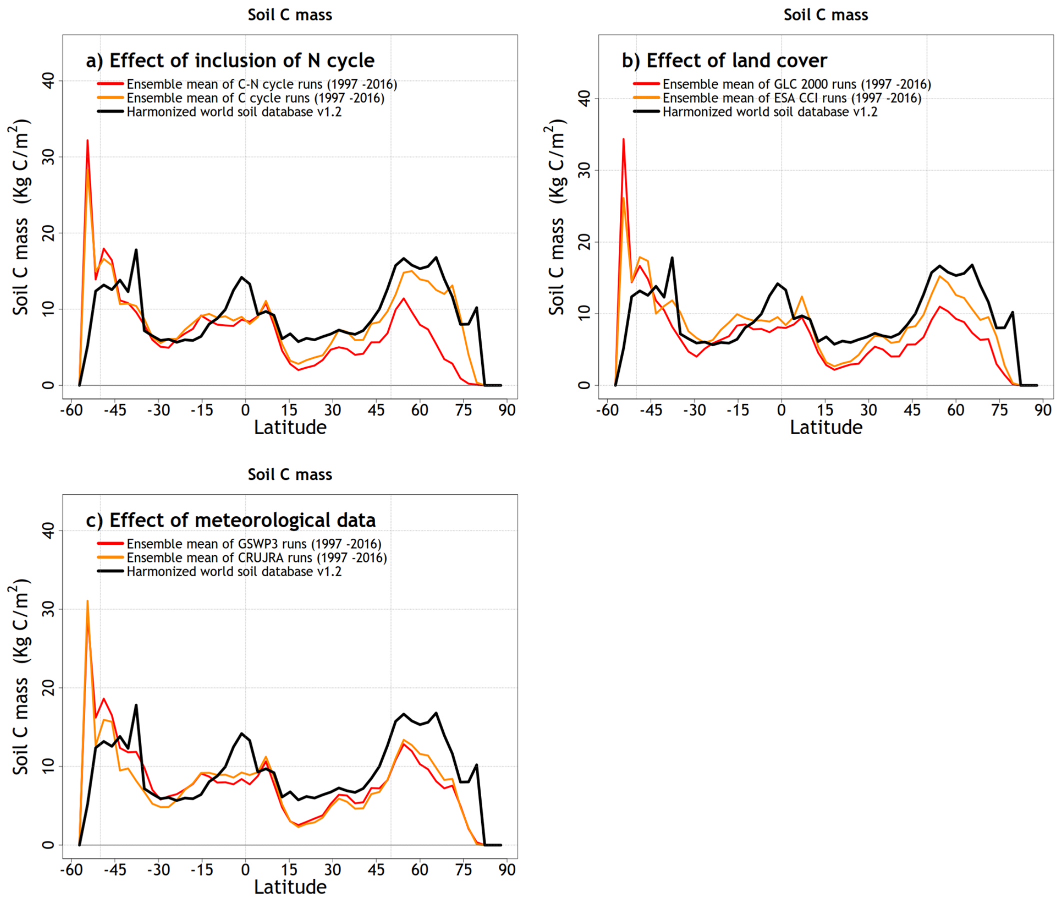


Figure A9: Comparison of zonally-averaged soil carbon mass (over all land area excluding Greenland and Antarctica) time-averaged over last 20-years of each simulation and then averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

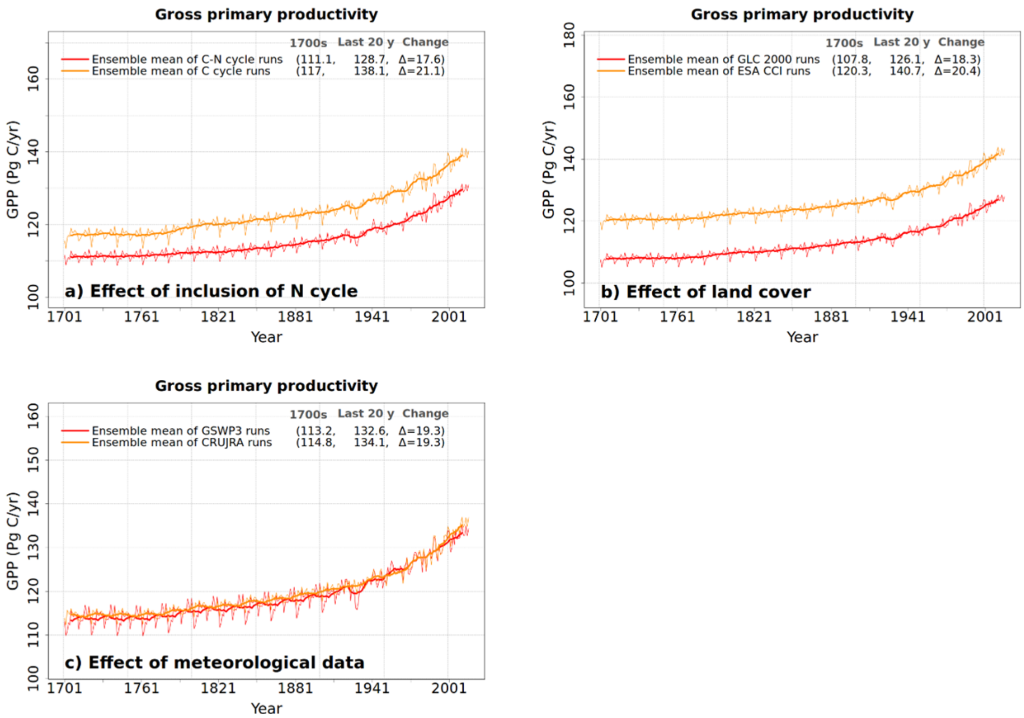


Figure A10: Comparison of time series of annual gross primary productivity (GPP) (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

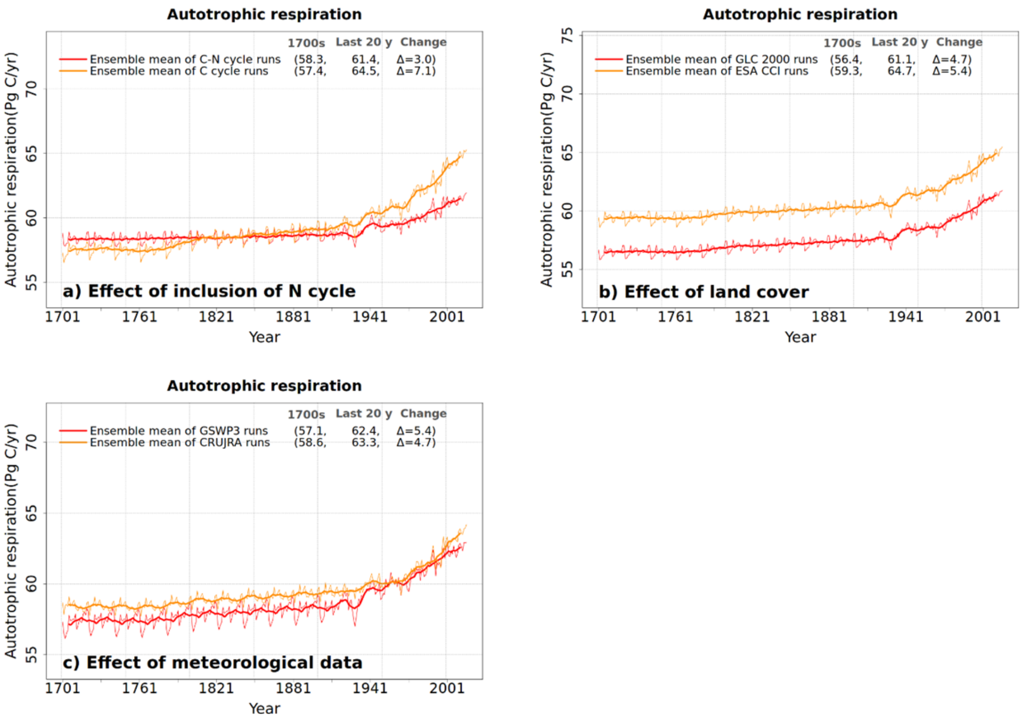


Figure A11: Comparison of time series of annual global autotrophic respiration (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

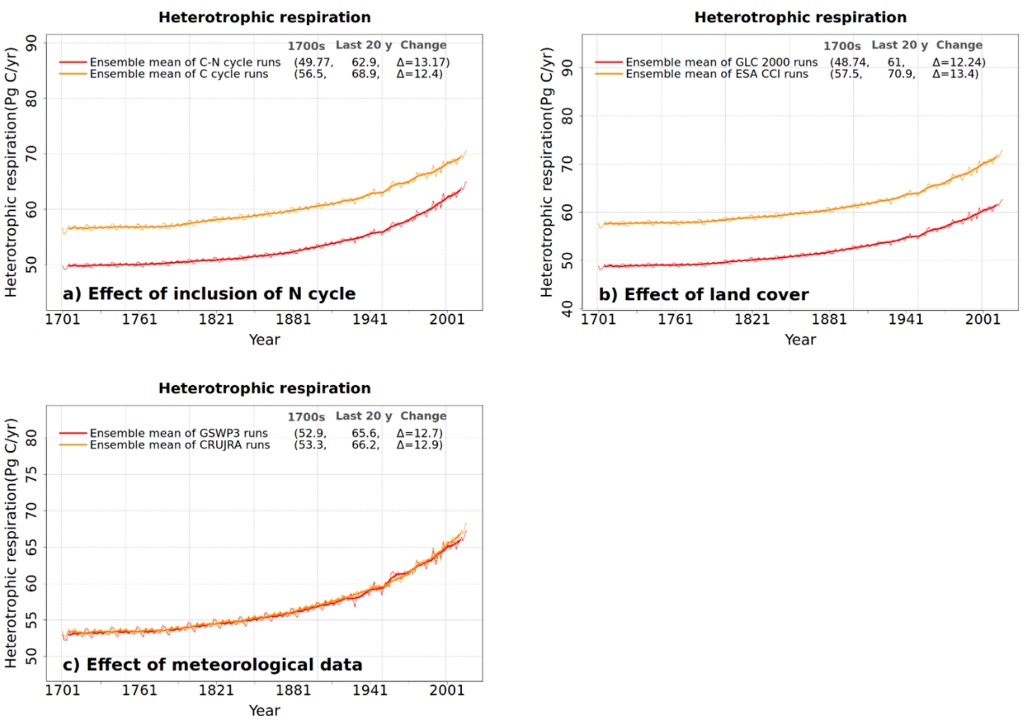


Figure A12: Comparison of time series of annual global heterotrophic respiration (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

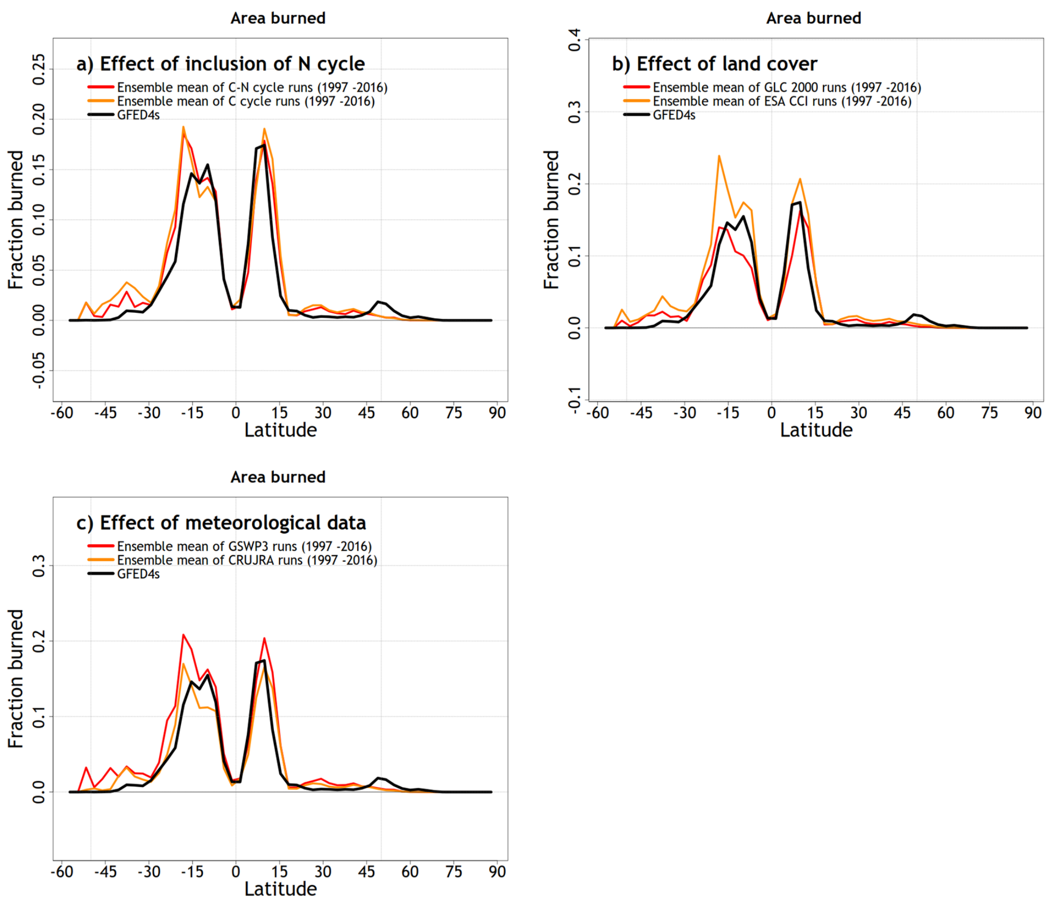


Figure A13: Comparison of zonally-averaged area burned (over all land area excluding Greenland and Antarctica) time-averaged over last 20-years of each simulation and then averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

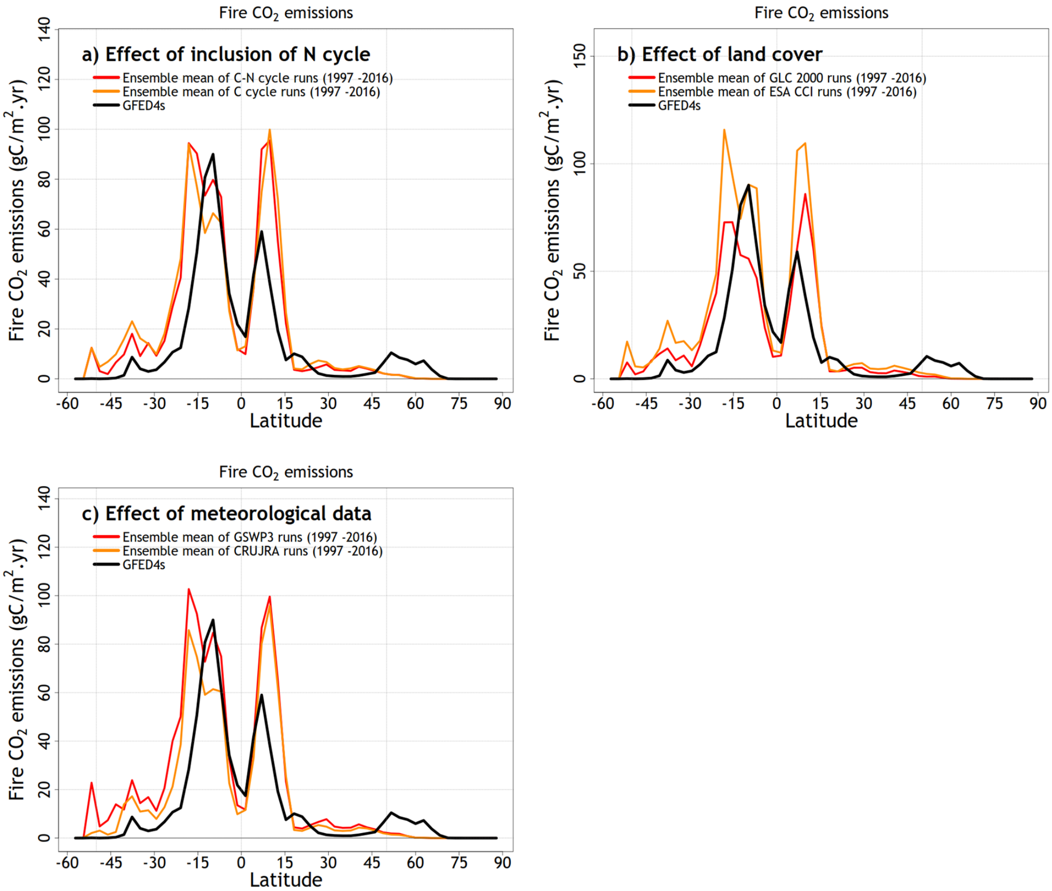


Figure A14: Comparison of zonally-averaged fire CO_2 emissions (over all land area excluding Greenland and Antarctica) time-averaged over last 20-years of each simulation and then averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

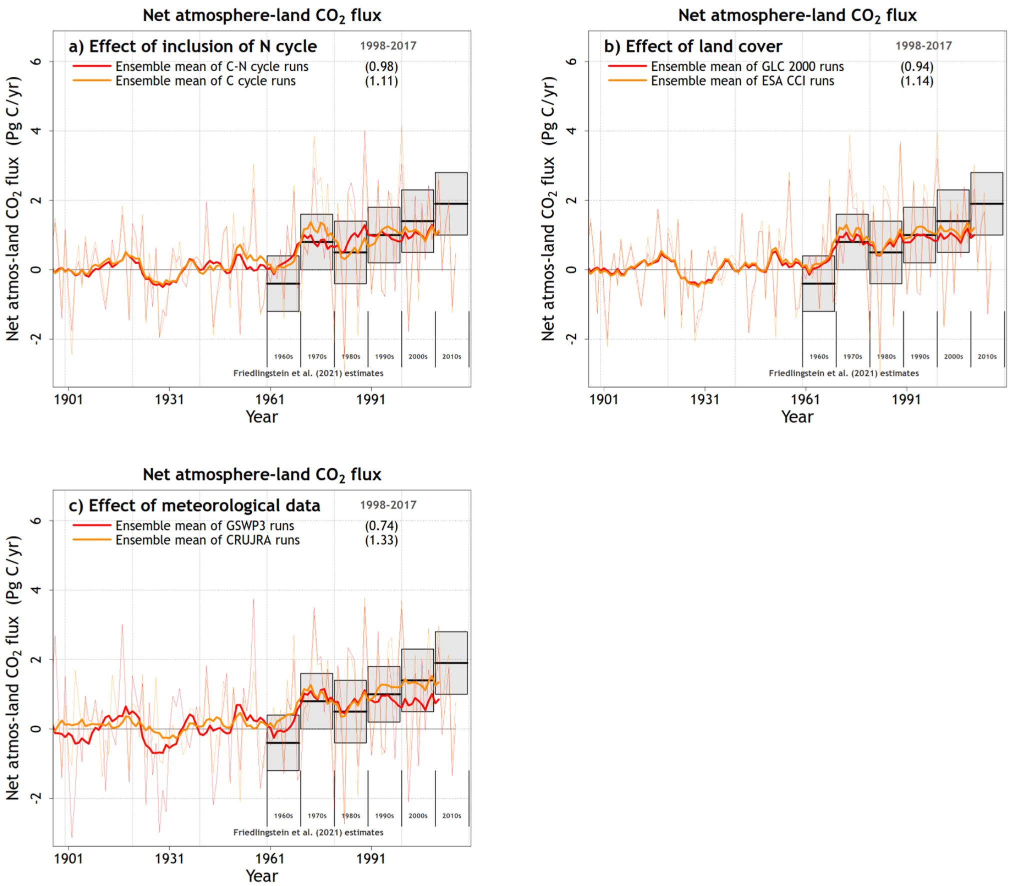


Figure A15: Comparison of time series of global net atmosphere-land CO_2 flux (over all land area excluding Greenland and Antarctica) averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

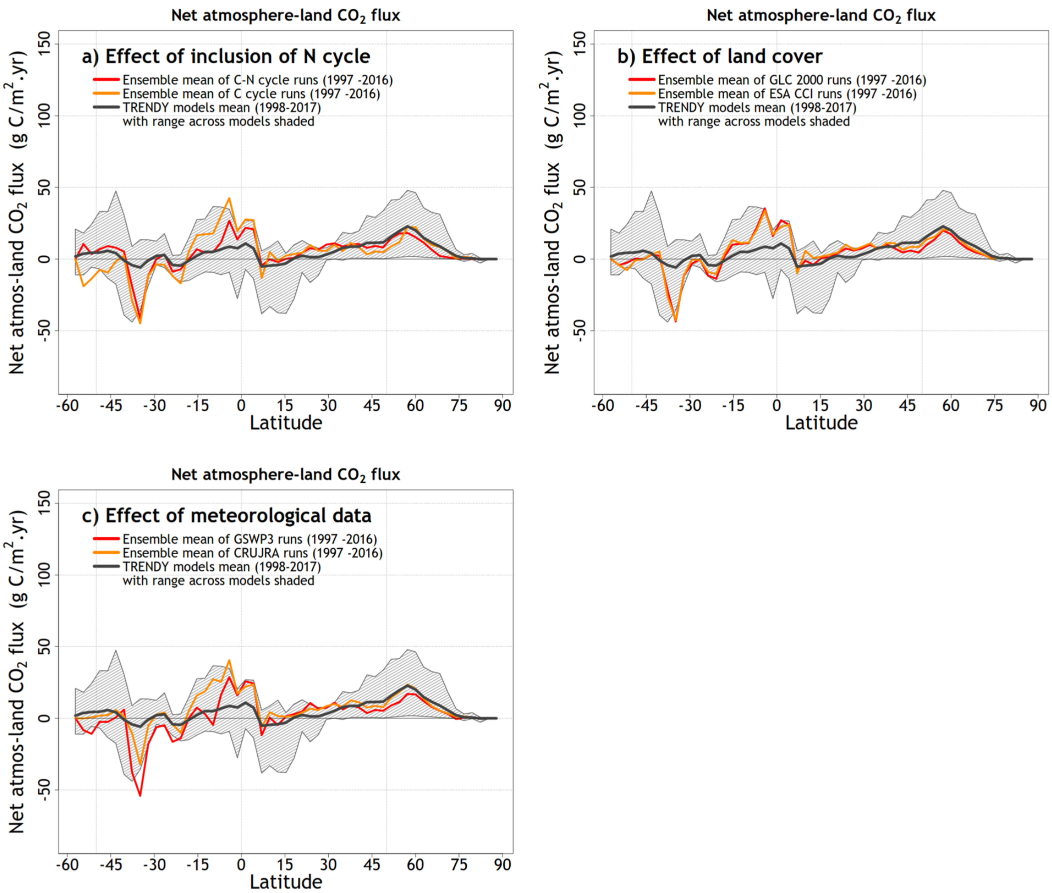


Figure A16: Comparison of zonally-averaged net atmosphere-land CO_2 flux (over all land area excluding Greenland and Antarctica) time-averaged over last 20-years of each simulation and then averaged over four ensemble members each that are driven with and without N cycle (panel a), driven with GLC 2000 and ESA CCI based land cover (panel b), and driven with GSWP3 and CRU-JRA meteorological data.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

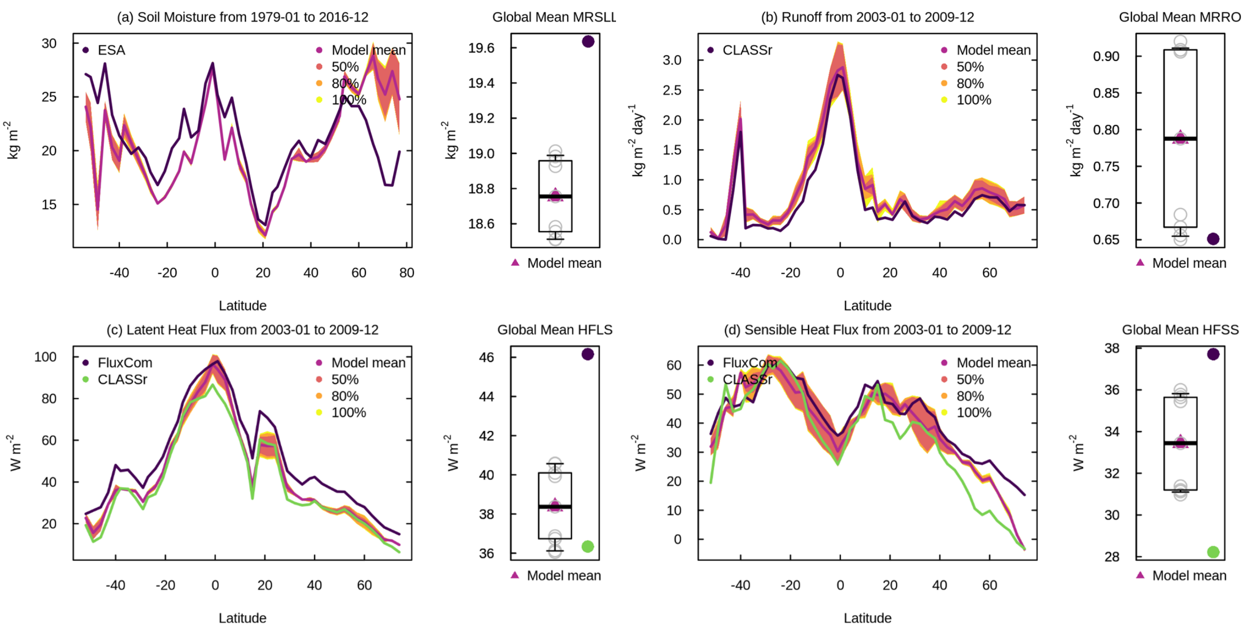


Figure A17: Zonally-averaged values of soil moisture (a), runoff (b), latent heat flux (c), and sensible heat flux (d) from the eight simulations summarized in Table 1 shown as their mean (dark purple line) and the spread across the simulations indicated by 50%, 80%, and 100% shading. The observation-based estimates used in AMBER to calculate scores are shown in black, blue, and green colours depending on how many observation-based datasets are available.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

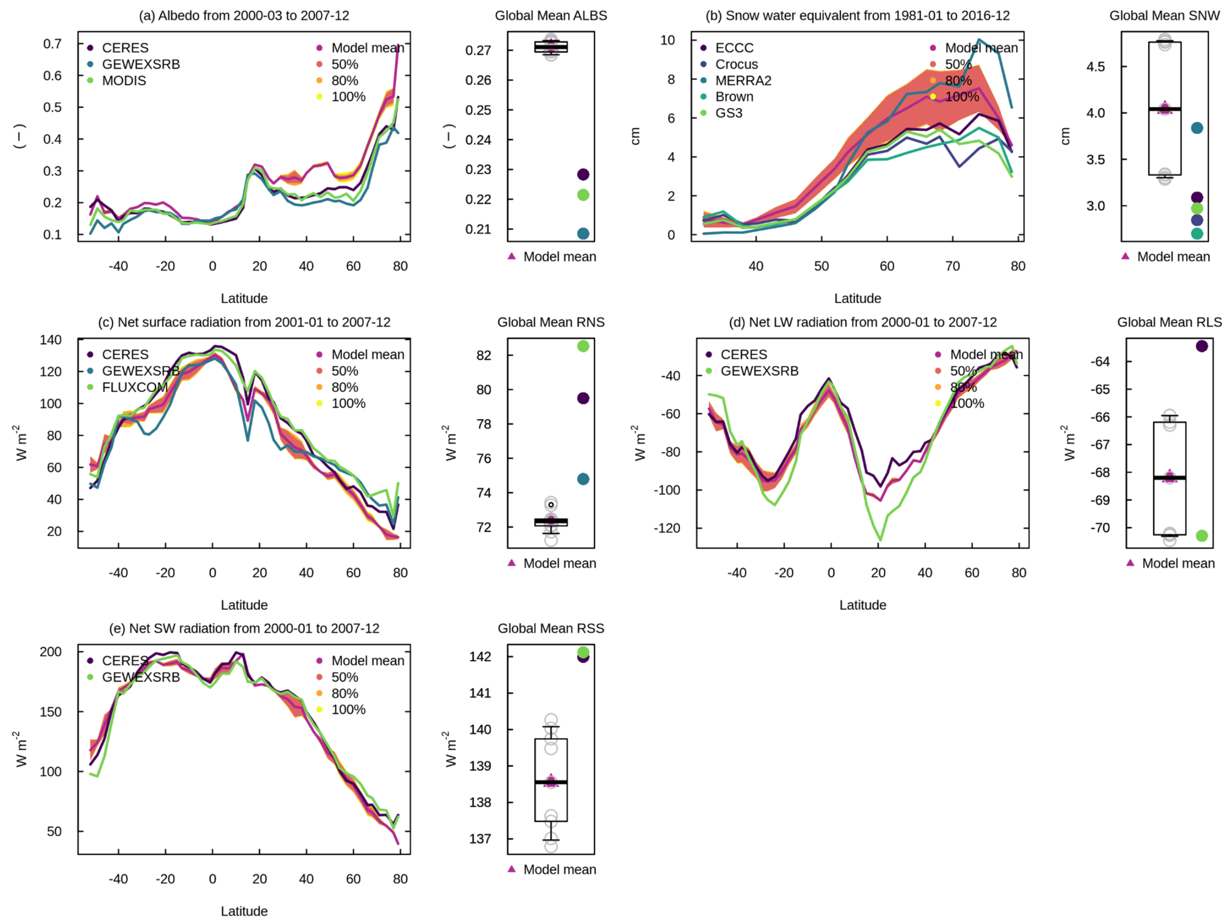


Figure A18: Zonally-averaged values of surface albedo (a), snow water equivalent (b), net surface radiation (c), net longwave radiation (d), and net shortwave radiation (e) from the eight simulations summarized in Table 1 shown as their mean (dark purple line) and the spread across the simulations indicated by 50%, 80%, and 100% shading. The observation-based estimates used in AMBER to calculate scores are shown in black, blue, and green colours depending on how many observation-based datasets are available.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

References

Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz-Sabater, J., Barré, J., Curcoll, R., Engelen, R., Langerock, B., Law, R. M., Loh, Z., Morguí, J. A., Parrington, M., Peuch, V.-H., Ramonet, M., Roehl, C., Vermeulen, A. T., Warneke, T., and Wunch, D.: Modelling CO_2 weather – why horizontal resolution matters, Atmos Chem Phys, 19, 7347–7376, 2019.

Arora, V. K. and Boer, G. J.: A parameterization of leaf phenology for the terrestrial ecosystem component of climate models, Glob. Change Biol., 11, 39–59, https://doi.org/10.1111/j.1365-2486.2004.00890.x, 2005.

Arora, V. K. and Boer, G. J.: Simulating Competition and Coexistence between Plant Functional Types in a Dynamic Vegetation Model, Earth Interact., 10, 1–30, 2006.

Arora, V. K. and Melton, J. R.: Reduction in global area burned and wildfire emissions since 1930s enhances carbon uptake by land, Nat. Commun., 9, 1326, https://doi.org/10.1038/s41467-018-03838-0, 2018.

Arora, V. K., Boer, G. J., Christian, J. R., Curry, C. L., Denman, K. L., Zahariev, K., Flato, G. M., Scinocca, J. F., Merryfield, W. J., and Lee, W. G.: The Effect of Terrestrial Photosynthesis Down Regulation on the Twentieth-Century Carbon Budget Simulated with the CCCma Earth System Model, J. Clim., 22, 6066–6088, https://doi.org/10.1175/2009JCLI3037.1, 2009.

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., and Merryfield, W. J.: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, Geophys. Res. Lett., 38, https://doi.org/10.1029/2010GL046270, 2011.

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., Krasting, J. P., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models, Biogeosciences, 17, 4173–4222, https://doi.org/10.5194/bg-17-4173-2020, 2020.

Asaadi, A. and Arora, V. K.: Implementation of nitrogen cycle in the CLASSIC land model, Biogeosciences, 18, 669–706, https://doi.org/10.5194/bg-18-669-2021, 2021.

Avitabile, V., Herold, M., Heuvelink, G. B. M., and others: An integrated pan tropical biomass map using multiple reference datasets, Glob Chang Biol, 2016.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, https://doi.org/10.1002/hyp.3360060305, 1992.

Bonan, G. B., Lombardozzi, D. L., Wieder, W. R., Oleson, K. W., Lawrence, D. M., Hoffman, F. M., and Collier, N.: Model Structure and Climate Data Uncertainty in Historical Simulations of the Terrestrial Carbon Cycle (1850–2014), Glob. Biogeochem. Cycles, 33, 1310–1326, https://doi.org/10.1029/2019GB006175, 2019.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Booth, B. B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., Huntingford, C., Betts, R. A., Harris, G. R., and Lloyd, J.: High sensitivity of future global warming to land carbon cycle processes, Environ. Res. Lett., 7, 024002, https://doi.org/10.1088/1748-9326/7/2/024002, 2012.

Chuvieco, E., Lizundia-Loiola, J., Pettinari, M. L., Ramo, R., Padilla, M., Tansey, K., Mouillot, F., Laurent, P., Storm, T., Heil, A., and Others: Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies, Earth Syst. Sci. Data, 10, 2015–2031, 2018.

Claverie, M., Matthews, J. L., Vermote, E. F., and Justice, C. O.: A 30+ Year AVHRR LAI and FAPAR Climate Data Record: Algorithm Description and Validation, Remote Sens., 8, 263, 2016.

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., and Randerson, J. T.: The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation, J. Adv. Model. Earth Syst., 10, 2731–2754, https://doi.org/10.1029/2018MS001354, 2018.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Q. J. R. Meteorol. Soc., 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Dai, A. and Trenberth, K. E.: Estimates of Freshwater Discharge from Continents: Latitudinal and Seasonal Variations, J Hydrometeorol, 3, 660–687, 2002.

ESA: Land Cover CCI Product User Guide Version 2 Technical Report, European Space Agency. Available at http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf, 2017.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H. T., Verelst, L., and Wiberg, D.: Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008), IIASA and FAO, Laxenburg, Austria and Rome, Italy, 2008.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth Syst. Sci. Data, 11, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019, 2019.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang, L. M., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Luijkx, I. T., Jain, A., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G. R., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J.: Global Carbon Budget 2021, Earth Syst. Sci. Data, 14, 1917–2005, https://doi.org/10.5194/essd-14-1917-2022, 2022.

Garrigues, S., Lacaze, R., Baret, F., Morisette, J. T., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., and Others: Validation and intercomparison of global Leaf Area Index products derived from remote sensing data, J. Geophys. Res. Biogeosciences, 113, 2008.

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

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), J. Geophys. Res. Biogeosciences, 118, 317–328, https://doi.org/10.1002/jgrg.20042, 2013.

Harris, I. C.: CRU JRA v2.1: A forcings dataset of gridded land surface blend of Climatic Research Unit (CRU) and Japanese reanalysis (JRA) data; Jan. 1901 - Dec. 2019, Centre for Environmental Data Analysis, University of East Anglia Climatic Research Unit, https://catalogue.ceda.ac.uk/uuid/10d2c73e5a7d46f4ada08b0a26302ef7, 2020.

Hegglin, M., Kinnison, D., and Lamarque, J.-F.: Wet and dry NHx and NOy deposition data, input4MIPs.CMIP6.CMIP.NCAR. Version 2016-11-15, Earth System Grid Federation, https://doi.org/10.22033/ESGF/input4MIPs.10448, 2016.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, 1–40, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hobeichi, S., Abramowitz, G., and Evans, J.: Conserving Land-Atmosphere Synthesis Suite (CLASS), J Clim, 2019.

Hornberger, G. M. and Spear, R. C.: Approach to the preliminary analysis of environmental systems, J Env. Manage U. S., 12:1, 1981.

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M.,

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Materia, S., Law, R. M., and Sheffield, J.: LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome, Geosci. Model Dev., 9, 2809–2832, https://doi.org/10.5194/gmd-9-2809-2016, 2016.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, 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.: Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6, Geosci Model Dev, 13, 5425–5464, 2020a.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, 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.: Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6, Geosci. Model Dev., 13, 5425–5464, https://doi.org/10.5194/gmd-13-5425-2020, 2020b.

Jacobson, A. R., Schuldt, K. N., Miller, J. B., and Oda, T.: CarbonTracker Documentation CT2019 release, 2020.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Sci Data, 6, 74, 2019.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., and others: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, 2020.

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., Yu, L., and Weller, R. A.: Surface Irradiances Consistent with CERES-Derived Top-of-Atmosphere Shortwave and Longwave Irradiances, J Clim, 26, 2719–2740, 2013.

Kou-Giesbrecht, S. and Arora, V. K.: Representing the Dynamic Response of Vegetation to Nitrogen Limitation via Biological Nitrogen Fixation in the CLASSIC Land Model, Glob. Biogeochem. Cycles, 36, e2022GB007341, https://doi.org/10.1029/2022GB007341, 2022.

Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), J. Geophys. Res. Biogeosciences, 112, https://doi.org/10.1029/2006JG000168, 2007.

Li, J., Duan, Q., Wang, Y.-P., Gong, W., Gan, Y., and Wang, C.: Parameter optimization for carbon and water fluxes in two global land surface models based on surrogate modelling, Int. J. Climatol., 38, e1016–e1031, https://doi.org/10.1002/joc.5428, 2018a.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R. A., and Peng, S.: Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

land cover maps (1992–2015), Earth Syst. Sci. Data, 10, 219–234, https://doi.org/10.5194/essd-10-219-2018, 2018b.

Li, X. and Xiao, J.: Mapping Photosynthesis Solely from Solar-Induced Chlorophyll Fluorescence: A Global, Fine-Resolution Dataset of Gross Primary Production Derived from OCO-2, Remote Sens., 11, 2563, 2019.

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., Van Dijk, A., McCabe, M. F., Evans, J., and Others: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, 2011.

Lu, C. and Tian, H.: Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance, Earth Syst. Sci. Data, 9, 181–192, https://doi.org/10.5194/essd-9-181-2017, 2017.

Mauder, M., Foken, T., and Cuxart, J.: Surface-Energy-Balance Closure over Land: A Review, Bound.-Layer Meteorol., 177, 395–426, https://doi.org/10.1007/s10546-020-00529-6, 2020.

Melton, J. R. and Arora, V. K.: Competition between plant functional types in the Canadian Terrestrial Ecosystem Model (CTEM) v. 2.0, Geosci. Model Dev., 9, 323–361, 2016a.

Melton, J. R. and Arora, V. K.: Competition between plant functional types in the Canadian Terrestrial Ecosystem Model (CTEM) v. 2.0, Geosci Model Dev, 9, 323–361, https://doi.org/10.5194/gmd-9-323-2016, 2016b.

Melton, J. R., Arora, V. K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E., and Teckentrup, L.: CLASSIC v1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Part 1: Model framework and site-level performance, Geosci. Model Dev. Discuss., 2019, 1–40, https://doi.org/10.5194/gmd-2019-329, 2019.

Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., and Tedesco, M.: Evaluation of long-term Northern Hemisphere snow water equivalent products, The Cryosphere, 14, 1579–1594, 2020.

Mudryk, L.: Historical gridded snow water equivalent and snow cover fraction over Canada from remote sensing and land surface models, 2020.

Myneni, R. B., 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.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sens Env., 83, 214–231, 2002.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Ribeca, A., van Ingen, C., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti,

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., De Cinti, B., de Grandcourt, A., De Ligne, A., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., di Tommasi, P., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci Data, 7, 225, 2020.

Poulter, B., Hattermann, F., Hawkins, E., Zaehle, S., Sitch, S., Restrepo-Coupe, N., Heyder, U., and Cramer, W.: Robust dynamics of Amazon dieback to climate change with perturbed ecosystem model parameters, Glob. Change Biol., 16, 2476–2495, https://doi.org/10.1111/j.1365-2486.2009.02157.x, 2010.

Reusch, A. and Gibbs, H. K.: New IPCC Tier-1 Global Biomass Carbon Map For the Year 2000, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 2008.

Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: How does the terrestrial carbon exchange respond to inter-annual climatic variations? A quantification based on atmospheric CO_2 data, Biogeosciences, 15, 2481–2498, 2018.

Santoro, M., Beaudoin, A., Beer, C., Cartus, O., Fransson, J. E. S., Hall, R. J., Pathe, C., Schmullius, C., Schepaschenko, D., Shvidenko, A., Thurner, M., and Wegmüller, U.: Forest growing stock volume of the northern hemisphere: Spatially explicit estimates for 2010 derived from Envisat ASAR, Remote Sens Env., 168, 316–334, 2015.

Schepaschenko, D., Chave, J., Phillips, O. L., Lewis, S. L., Davies, S. J., Réjou-Méchain, M., Sist, P., Scipal, K., Perger, C., Herault, B., Labrière, N., Hofhansl, F., Affum-Baffoe, K., Aleinikov, A., Alonso, A., Amani, C., Araujo-Murakami, A., Armston, J., Arroyo, L., Ascarrunz, N., Azevedo, C., Baker, T., Ba\lazy, R., Bedeau, C., Berry, N., Bilous, A. M., Bilous, S. Y., Bissiengou, P., Blanc, L., Bobkova, K. S., Braslavskaya, T., Brienen, R., Burslem, D. F. R. P., Condit, R., Cuni-Sanchez, A., Danilina, D., Del Castillo Torres, D., Derroire, G., Descroix, L., Sotta, E. D., d'Oliveira, M. V. N., Dresel, C., Erwin, T., Evdokimenko, M. D., Falck, J., Feldpausch, T. R., Foli, E. G., Foster, R., Fritz, S., Garcia-Abril, A. D., Gornov, A., Gornova, M., Gothard-Bassébé, E., Gourlet-Fleury, S., Guedes, M., Hamer, K. C., Susanty, F. H., Higuchi, N., Coronado, E. N. H., Hubau, W., Hubbell, S., Ilstedt, U., Ivanov, V. V., Kanashiro, M., Karlsson, A., Karminov, V. N., Killeen, T., Koffi, J.-C. K., Konovalova, M., Kraxner, F., Krejza, J., Krisnawati, H., Krivobokov, L. V., Kuznetsov, M. A., Lakyda, I., Lakyda, P. I., Licona, J. C., Lucas, R. M., Lukina, N., Lussetti, D., Malhi, Y., Manzanera, J. A., Marimon, B., Junior, B. H. M., Martinez, R. V., Martynenko, O. V., Matsala, M., Matyashuk, R. K., Mazzei, L., Memiaghe, H., Mendoza, C., Mendoza, A. M., Moroziuk, O. V., Mukhortova, L., Musa, S., Nazimova, D. I., Okuda, T., Oliveira, L. C., Ontikov, P. V., et al.: The Forest Observation System, building a global reference dataset for remote sensing of forest biomass, Sci Data, 6, 198, 2019.

Seiler, C., Melton, J. R., Arora, V. K., and Wang, L.: CLASSIC v1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Part 2: Global benchmarking, Geosci. Model Dev., 14, 2371–2417, 2021a.

Seiler, C., Melton, J. R., Arora, V. K., and Wang, L.: CLASSIC v1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Part 2: Global benchmarking, Geosci. Model Dev., 14, 2371–2417, https://doi.org/10.5194/gmd-14-2371-2021, 2021b.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Seiler, C., Melton, J. R., Arora, V. K., Sitch, S., Friedlingstein, P., Anthoni, P., Goll, D., Jain, A. K., Joetzjer, E., Lienert, S., Lombardozzi, D., Luyssaert, S., Nabel, J. E. M. S., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., and Zaehle, S.: Are Terrestrial Biosphere Models Fit for Simulating the Global Land Carbon Sink?, J. Adv. Model. Earth Syst., 14, e2021MS002946, https://doi.org/10.1029/2021MS002946, 2022.

Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling, J. Adv. Model. Earth Syst., 6, 249–263, https://doi.org/10.1002/2013MS000293, 2014.

Slevin, D., Tett, S. F. B., Exbrayat, J.-F., Bloom, A. A., and Williams, M.: Global evaluation of gross primary productivity in the JULES land surface model v3.4.1, Geosci. Model Dev., 10, 2651–2670, https://doi.org/10.5194/gmd-10-2651-2017, 2017.

Stackhouse, P. W., Jr, Gupta, S. K., Cox, S. J., Zhang, T., Mikovitz, J. C., and Hinkelman, L. M.: The NASA/GEWEX surface radiation budget release 3.0: 24.5-year dataset, Gewex News, 21, 10–12, 2011.

Strahler, A. H., Muller, J., Lucht, W., Schaaf, C., and others: MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0, MODIS, 1999.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 365, 2053–2075, https://doi.org/10.1098/rsta.2007.2076, 2007.

Tian, Y., Dickinson, R. E., Zhou, L., and Shaikh, M.: Impact of new land boundary conditions from Moderate Resolution Imaging Spectroradiometer (MODIS) data on the climatology of land surface variables, J. Geophys. Res. Atmospheres, 109, https://doi.org/10.1029/2003JD004499, 2004.

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, 10, 1717–1736, 2013.

Tukey, J. W.: Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977.

Verger, A., Baret, F., and Weiss, M.: Near real-time vegetation monitoring at global scale, IEEE J. Sel. Top. In, 2014.

Verseghy, D. L.: Class—A Canadian land surface scheme for GCMS. I. Soil model, Int. J. Climatol., 11, 111–133, https://doi.org/10.1002/joc.3370110202, 1991.

Verseghy, D. L., McFarlane, N. A., and Lazare, M.: Class—A Canadian land surface scheme for GCMS, II. Vegetation model and coupled runs, Int. J. Climatol., 13, 347–370, https://doi.org/10.1002/joc.3370130402, 1993.

Wang, A., Price, D. T., and Arora, V.: Estimating changes in global vegetation cover (1850–2100) for use in climate models, Glob. Biogeochem. Cycles, 20, https://doi.org/10.1029/2005GB002514, 2006.

Wang, L., Bartlett, P., Chan, E., and Xiao, M.: Mapping of Plant Functional Type from Satellite-Derived Land Cover Datasets for Climate Models, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 3416–3419, https://doi.org/10.1109/IGARSS.2018.8518046, 2018.

Wieder, W.: Regridded Harmonized World Soil Database v1.2, , https://doi.org/10.3334/ORNLDAAC/1247, 2014.

Wu, Z., Ahlström, A., Smith, B., Ardö, J., Eklundh, L., Fensholt, R., and Lehsten, V.: Climate data induced uncertainty in model-based estimations of terrestrial primary productivity, Environ. Res. Lett., 12, 064013, https://doi.org/10.1088/1748-9326/aa6fd8, 2017.

Xue, B.-L., Guo, Q., Hu, T., Wang, G., Wang, Y., Tao, S., Su, Y., Liu, J., and Zhao, X.: Evaluation of modeled global vegetation carbon dynamics: Analysis based on global carbon flux and above-ground biomass data, Ecol Modell, 355, 84–96, 2017.

Zhang, Y. and Liang, S.: Fusion of Multiple Gridded Biomass Datasets for Generating a Global Forest Aboveground Biomass Map, Remote Sens., 12, 2559, 2020.

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., and Dong, J.: A global moderate resolution dataset of gross primary production of vegetation for 2000–2016, Sci. Data, 4, 170165, 2017.

Code/data availability

More information about the CLASSIC land surface model and its Fortran code are available at https://cccma.gitlab.io/classic_pages/.

Author contribution

VA and SKG performed the simulations, and VA wrote majority of the manuscript. CS performed the AMBER related analysis. LW put together the ESA CCI land cover. CS, LW, and SKG provided comments on the entire manuscript and also wrote their respective sections.

Competing interests

There are no competing interests.

Acknowledgment

We thank Joe Melton for providing comments on an earlier version of this paper. We also thank Benjamin Bond-Lamberty for taking this paper on as an Associate Editor.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Table 1: Summary of simulations performed with two reconstructions of the historical land cover, two sets of meteorological data, and two versions of the CLASSIC land model.

Simulation	Land cover reconstruction	Meteorological forcing	N cycle interactions with the C cycle
A	based on GLC 2000	CRU-JRA v2.1.5	On
B	based on GLC 2000	GSWP3	On
C	based on GLC 2000	CRU-JRA v2.1.5	Off
D	based on GLC 2000	GSWP3	Off
E	based on ESA CCI	CRU-JRA v2.1.5	On
F	based on ESA CCI	GSWP3	On
G	based on ESA CCI	CRU-JRA v2.1.5	Off
H	based on ESA CCI	GSWP3	Off

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Table 2: Observation-based data sets used for model evaluation in AMBER.

Globally gridded variable(s)	**Source**	**Approach used**	**Reference**
Leaf area index	AVHRR	Artificial neural network	Claverie et al. (2016)
Net biome productivity	CAMS	Atmospheric inversion	Agustí-Panareda et al. (2019)
Net biome productivity	Carboscope	Atmospheric inversion	Rödenbeck et al. (2018)
Surface albedo, net shortwave and longwave radiation, net radiation	CERES	Radiative transfer model	Kato et al. (2013)
Net radiation, latent and sensible heat flux, ground heat flux, runoff	CLASSr	Blended product	Hobeichi et al. (2019)
Leaf area index	Copernicus	Artificial neural network	Verger et al. (2014)
Net biome productivity	CT2019	Atmospheric inversion	Jacobson et al. (2020)
Snow amount	ECCC	Blended product	Mudryk (2020)
Liquid soil moisture	ESA	Land surface model	Liu et al. (2011)
Area burnt	ESA CCI	Burned area mapping	Chuvieco et al. (2018)
Latent and sensible heat flux, gross primary productivity	FLUXCOM	Machine learning	Jung et al. (2019, 2020)
Above ground biomass	GEOCARBON	Machine learning	Avitabile et al. (2016); Santoro et al. (2015)
Surface albedo, net shortwave and longwave radiation, net radiation	GEWEXSRB	Radiative transfer model	Stackhouse et al. (2011)
Area burnt	GFED 4s	Burned area mapping	Giglio et al. (2010)
Gross primary productivity	GOSIF	Statistical model	Li and Xiao (2019)
Soil carbon	HWSD	Soil inventory	Wieder (2014); Todd-Brown et al. (2013)
Surface albedo	MODIS	Bidirectional Reflectance Distribution Function	Strahler et al. (1999)
Gross primary productivity	MODIS	Light use efficiency model	Zhang et al. (2017)
Leaf area index	MODIS	Radiative transfer model	Myneni et al. (2002)
Soil carbon	SGS250m	Machine learning	Hengl et al. (2017)
Above ground biomass	Zhang	Data fusion	Zhang and Liang (2020)
In situ variable(s)	**Source**	**Approach used (number of sites)**	**Reference**
Leaf area index	CEOS	Transfer function (141)	Garrigues et al. (2008)
Latent, sensible, and ground heat flux, gross primary productivity, ecosystem respiration, net ecosystem exchange	FLUXNET 2015	Eddy covariance (204)	Pastorello et al. (2020)
Above ground biomass	FOS	Allometry (274)	Schepaschenko et al. (2019)
Runoff	GRDC	Gauge records (50)	Dai and Trenberth (2002)
Snow amount	Mortimer	Gravimetry (3271)	Mortimer et al. (2020)
Above ground biomass	Xue	Allometry (1974)	Xue et al. (2017)

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

Table 3: Simulated energy, water, and carbon cycle quantities considered in this study sorted according to their coefficient of variation. The quantities are listed from the most variable at the top to the least variable at the bottom. The coefficient of variation is based on annual values averaged over the last 20 years across the eight simulations.

Energy, water, or carbon cycle quantities	**Coefficient of variation**
Area burned (million km^2)	0.24
Fire CO_2 emissions (Pg C/year)	0.21
Soil carbon mass (Pg C)	0.21
Vegetation biomass (Pg C)	0.16
Runoff (1000 km^3/year)	0.13
Heterotrophic respiration (Pg C/year)	0.10
Gross primary productivity (Pg C/year)	0.07
Autotrophic respiration (Pg C/year)	0.07
Sensible heat flux (W/m^2)	0.07
Latent heat flux (W/m^2) / Evapotranspiration (1000 km^3/year)	0.05
Net longwave radiation (W/m^2)	0.03
Soil moisture in the top 1m soil layer (mm)	0.02
Soil temperature in the top 1m soil layer (° C)	0.004
Net shortwave radiation (W/m^2)	0.006

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

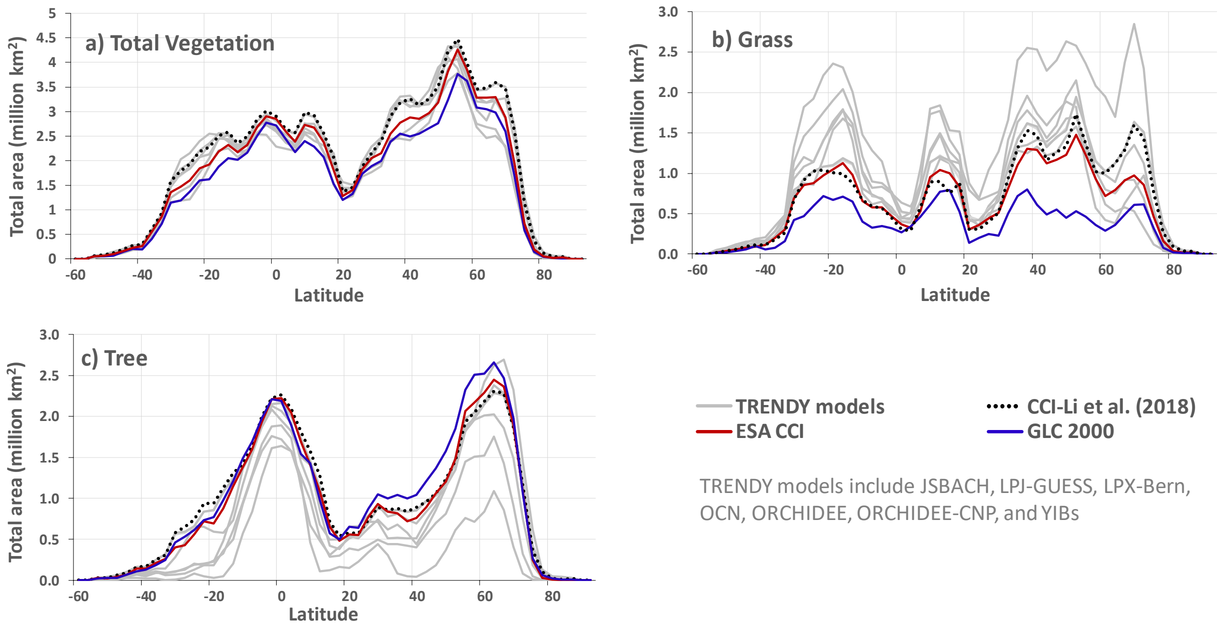
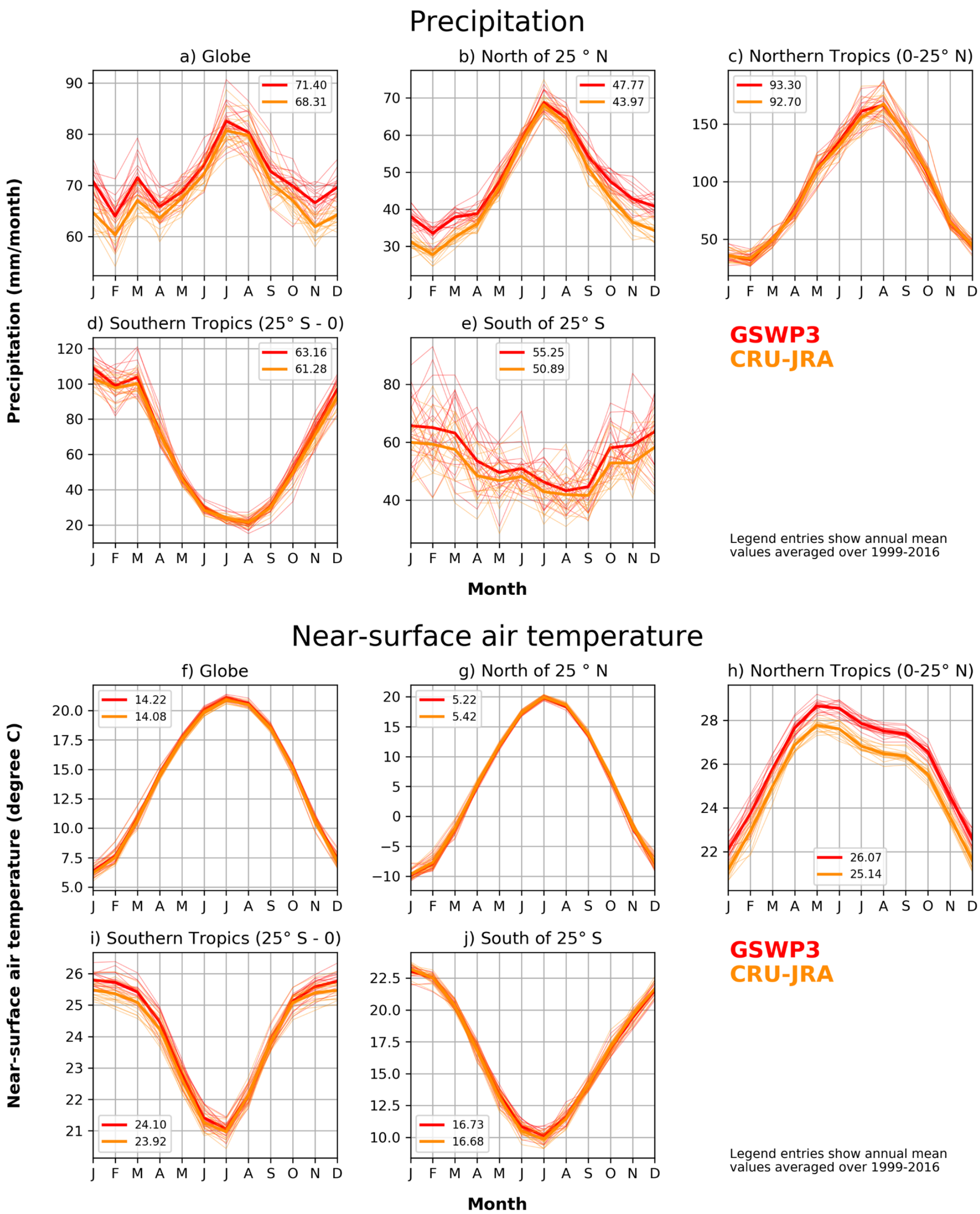


Figure 1: Comparison of zonally summed areas of total vegetation (a), grass (b), and tree (c) cover used in the CLASSIC model based on GLC 2000 (blue line) and ESA CCI (dark red line) land cover products to each other, to selected other models that participated in the 2020 TRENDY intercomparison (grey lines) for which land cover information was available, and to Li et al. (2018) (dotted black line) who analyzed the ESA CCI data. CLASSIC does not yet explicitly represents shrub PFTs. Tall shrubs are merged into tree PFTs in CLASSIC. For the Li et al. (2018) data plotted here the shrub PFTs are combined with the tree PFTs to be consistent with those in CLASSIC.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022



1201 Figure 2: Comparison of monthly precipitation (upper panel) and temperature (lower panel) for five
1202 global regions (global, north of 25 °N, northern and southern tropics, and south of 25 °S) from the CRU-
1203 JRA and GSWP3 meteorological forcing data sets that are used to drive the CLASSIC model. The global
1204 and regional averages exclude Greenland and Antarctica. The legend entries show the annual mean
1205 values averaged over the 1999-2016 period. The thin lines show individual years and the thick line their
1206 average.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

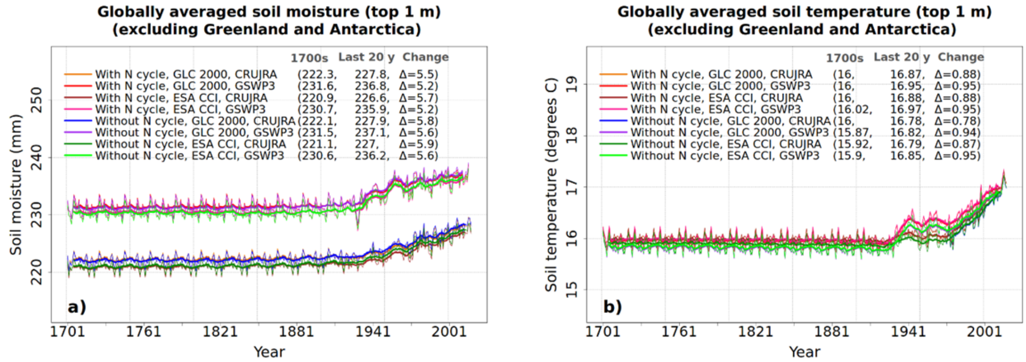


Figure 3: Comparison of time series of simulated globally-averaged annual soil moisture (a) and soil temperature (b) in the top 1m from the eight simulations summarized in Table 1.

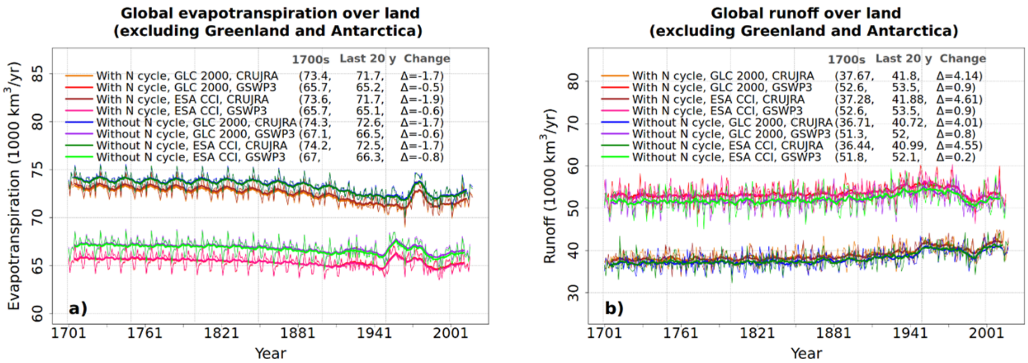


Figure 4: Comparison of time series of simulated global annual evapotranspiration (a) and runoff (b) from the eight simulations summarized in Table 1.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

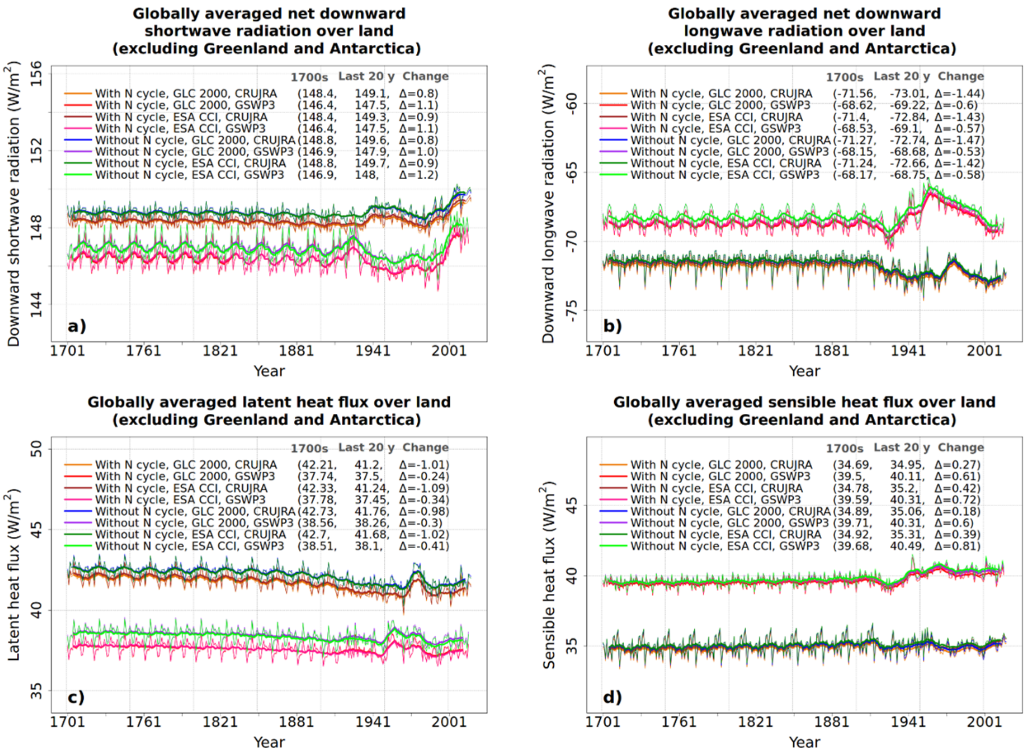


Figure 5: Comparison of time series of simulated globally-averaged annual energy fluxes from the eight simulations summarized in Table 1. Panel (a) shows net downward shortwave radiation, panel (b) shows net downward longwave radiation, panel (c) shows latent heat flux, and panel (d) shows sensible heat flux.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

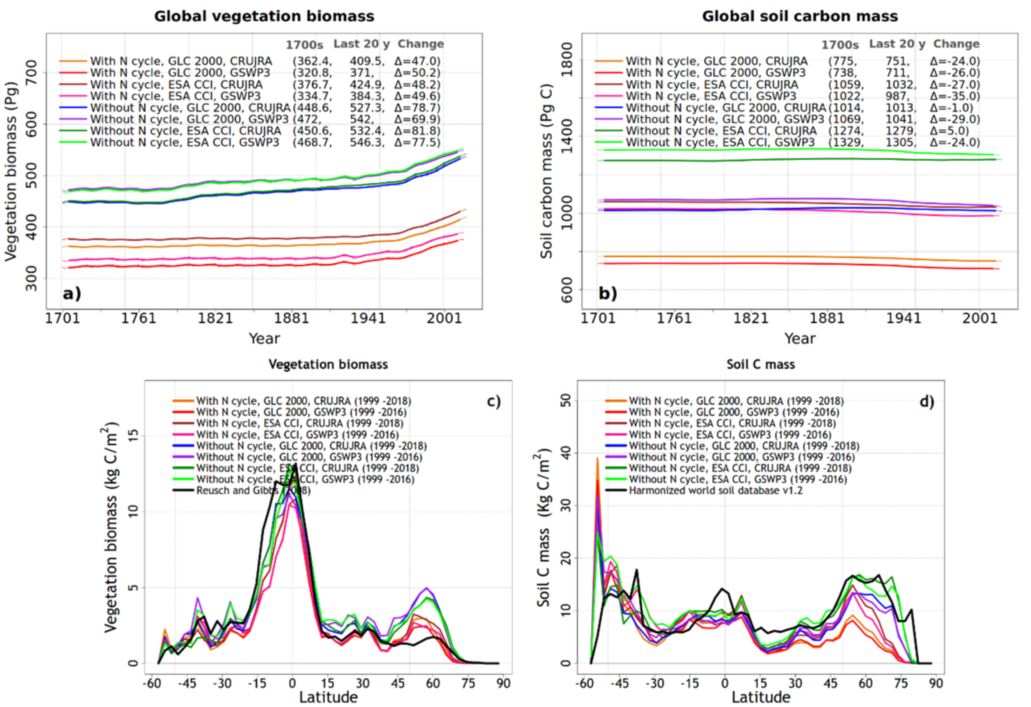


Figure 6: Comparison of time series of simulated global annual vegetation biomass (a) and soil carbon (b) from the eight simulations summarized in Table 1. Panels (c) and (d) show the zonally-averaged values of vegetation biomass and soil carbon mass from the eight simulations.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

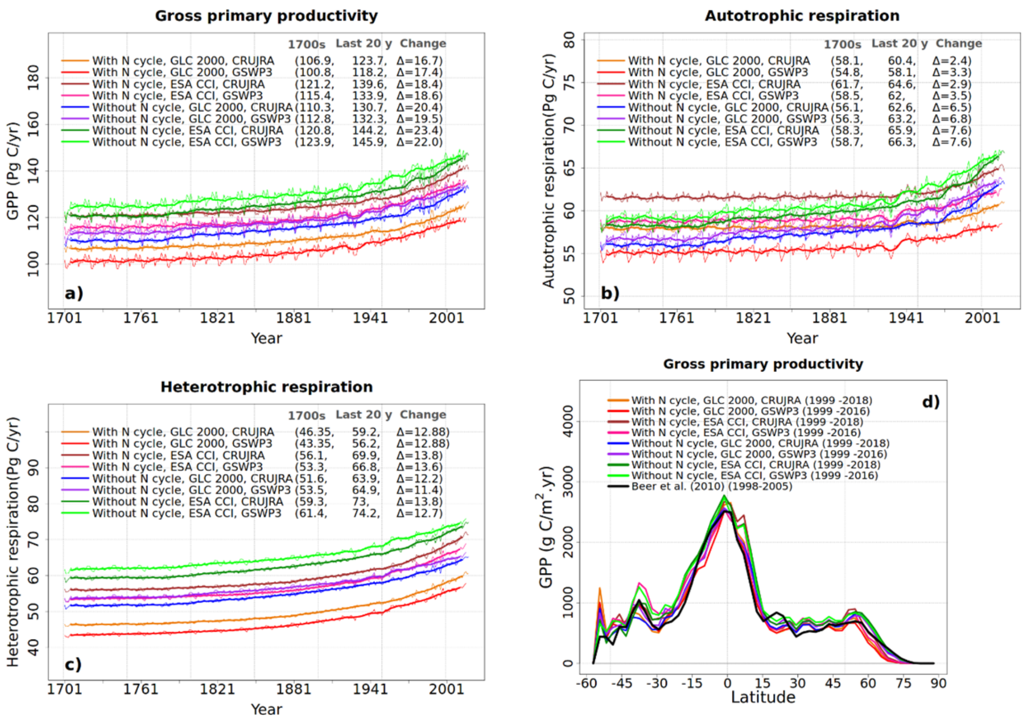


Figure 7: Comparison of time series of simulated global annual gross primary productivity (GPP) (a), autotrophic respiration (b), and heterotrophic respiration (c) from the eight simulations summarized in Table 1. Panel (d) shows the zonally-averaged values of GPP from the eight simulations averaged over last 20-years of each simulation.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

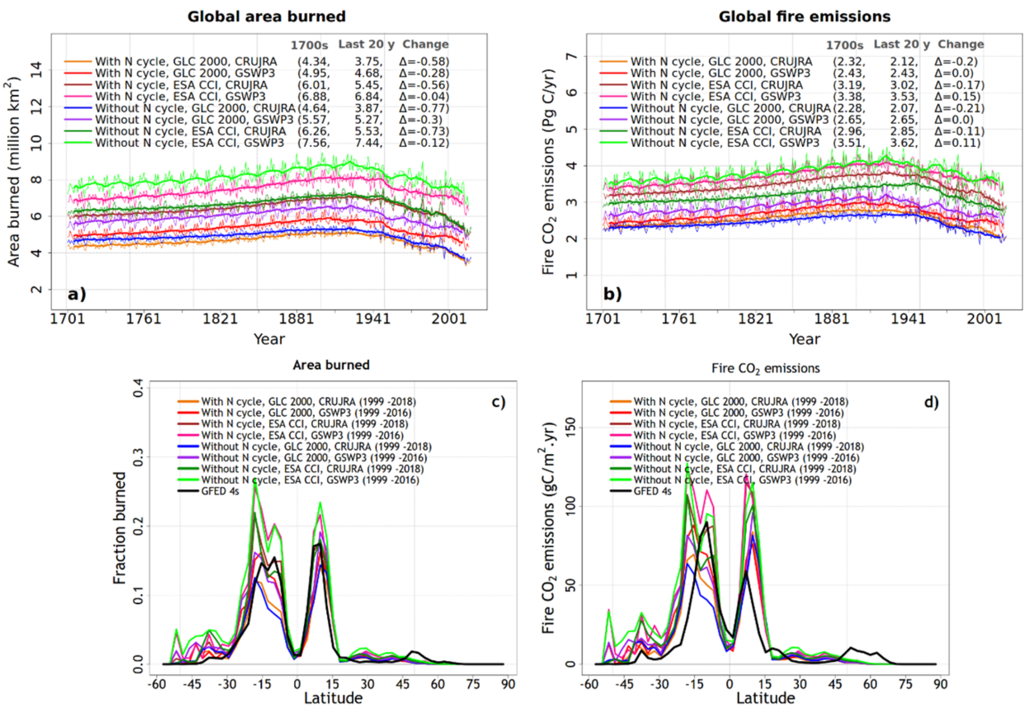


Figure 8: Comparison of time series of simulated global annual area burned (a) and fire CO_2 emissions (b) from the eight simulations summarized in Table 1. Panels (c) and (d) show the zonally-averaged area burned and fire CO_2 emissions from the right simulations averaged over last 20 years of each simulation.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

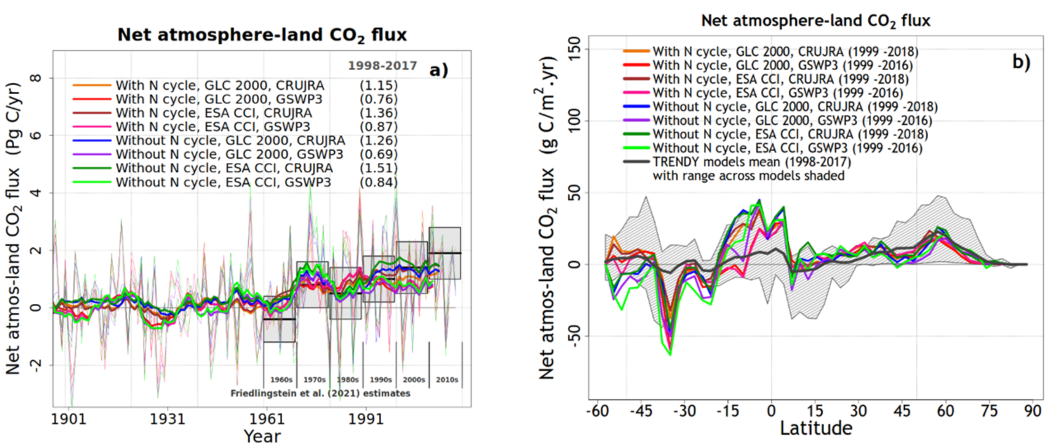


Figure 9: Comparison of time series of simulated global net atmosphere-land CO_2 flux (a) and its zonally-averaged values (b) from the eight simulations summarized in Table 1.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

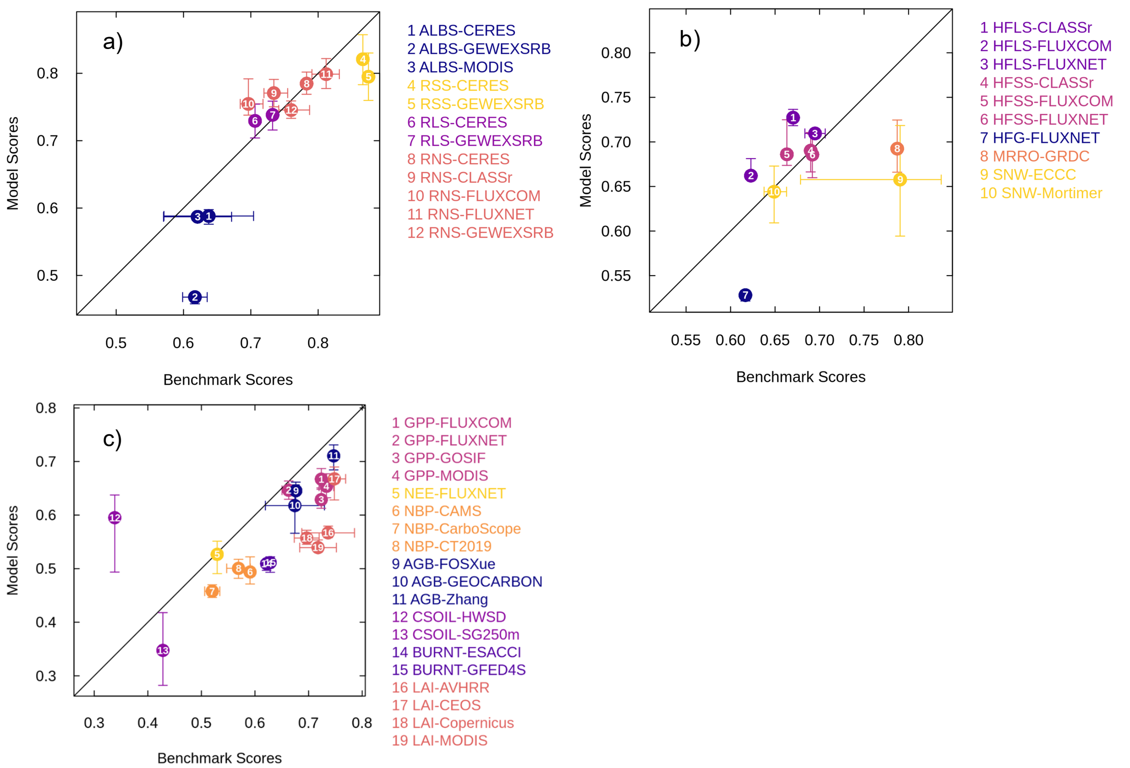


Figure 10: Comparison of benchmark scores with model overall scores for a range of energy, water, and carbon related quantities. The whiskers indicate the range for benchmark scores across different observation-based data sets and the range across the eight model simulations for the overall model scores. The quantities in in panel (a) are ALBS (surface albedo), RSS (net shortwave radiation), RLS (net longwave radiation), and RNS (net radiation). Quantities in panel (b) are HFLS (latent heat flux), HFSS (sensible heat flux), HFG (ground heat flux), MRRO (runoff), and SNW (snow water equivalent). Quantities in panel (c) are GPP (gross primary productivity), NEE (net ecosystem exchange), NBP (net biome productivity), AGB (above ground biomass), CSOIL (soil carbon mass), BURNT (area burned), and LAI (leaf area index).

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022

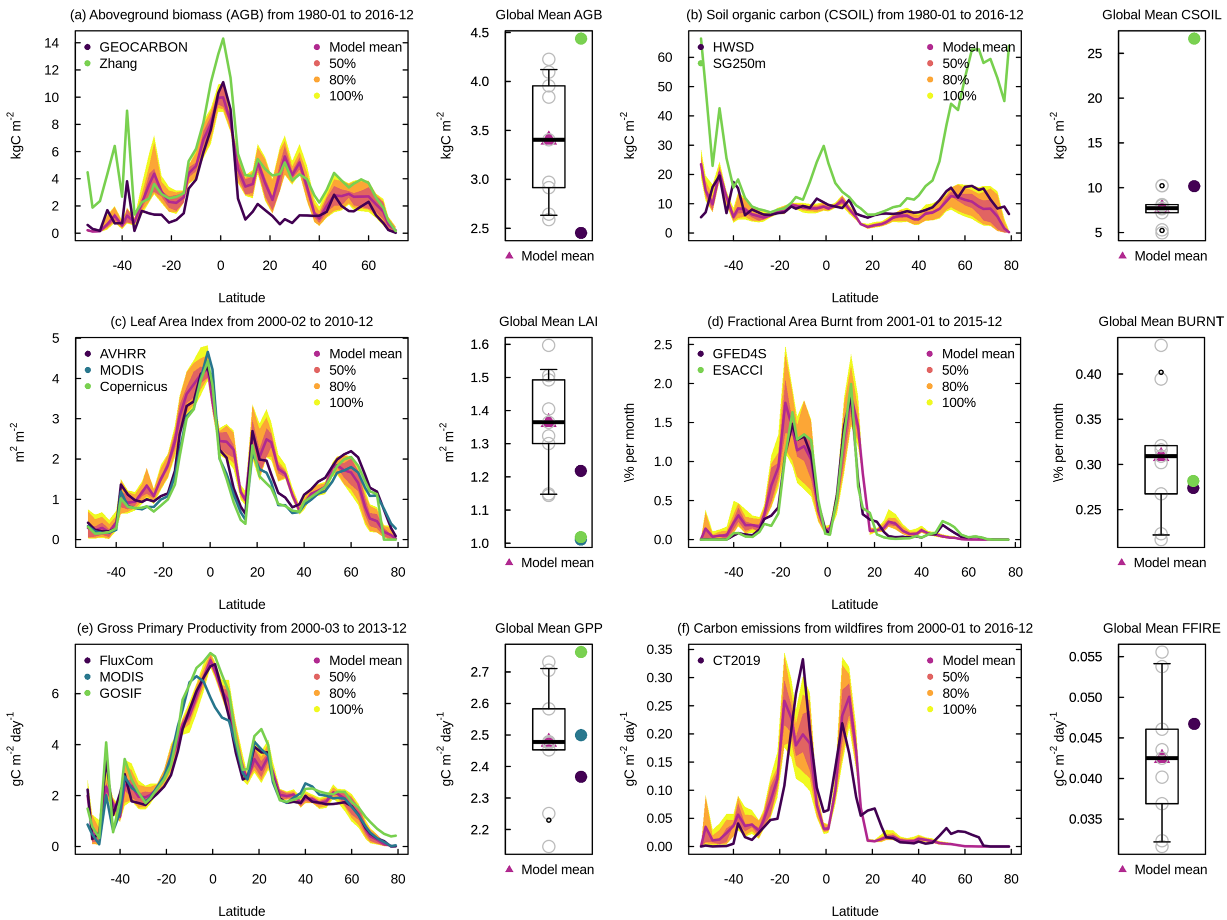
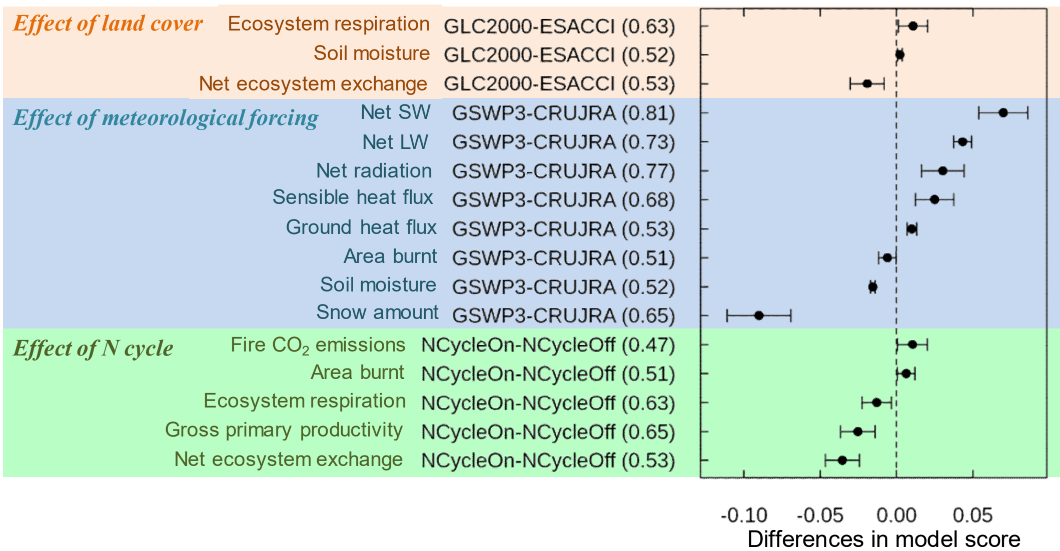


Figure 11: Zonally-averaged values of aboveground biomass (a), soil carbon mass (b), leaf area index (c), fractional area burnt (d), gross primary productivity (e), and fire CO_2 emissions (f) from the eight simulations summarized in Table 1 shown as their mean (dark purple line) and the spread across the simulations indicated by 50%, 80%, and 100% shading. The observation-based estimates used in AMBER to calculate scores are shown in black, blue, and green colours depending on how many observation-based datasets are available.

https://doi.org/10.5194/egusphere-2022-641
Preprint. Discussion started: 29 July 2022



1320

1321 Figure 12: Summary of difference in overall scores for model simulated quantities and combinations for
1322 which the differences are statistically significant. The scores in parentheses for each quantity are the
1323 average scores across the eight simulations and provide context.

1324

1325

1326