2021 Global Carbon Budget: Land modelling protocol (Trendy-v11)
Contact: Stephen Sitch (s.a.sitch@exeter.ac.uk) and Pierre Friedlingstein (p.friedlingstein@exeter.ac.uk)

https://sites.exeter.ac.uk/trendy

1. Deadline for submission of simulations:

GCB 2022 simulations: S0, S1, S2, S3 simulations latest: <mark>August 31st 2022</mark>

Goal: To provide the land components of the Global Carbon Project 2022 Budget, and an ensemble of land carbon cycle simulations to be used by the scientific community. Usual Trendy data-use policy will apply (https://sites.exeter.ac.uk/trendy/data-policy).

2. Model simulations

Full transient simulations are needed from DGVMs for the standard set of S0-S3 as all forcings have changed (not just one year extension):

Climate dataset: As last year, we include the impact of historical changes in aerosol on radiation fields used to drive DGVMs. We encourage models that can use fields of both radiation quantity and quality (diffuse radiation) to include them in simulations.

CO_2 file: The global CO_2 dataset is extended by one year.

Land use and Land cover changes: Like LUH2-GCB2021, LUH2-GCB2022 will be based on the new HYDE3.3 cropland/grazing land dataset which now in addition to FAO country-level statistics is constrained spatially based on multi-year satellite land cover maps from ESA CCI LC. However, the approach to extrapolate cropland and pasture areas beyond FAO year 2017 using the trend function from the previous 5 years does not capture the recent upturn in deforestation in Brazil. For this reason, this year we replace the full FAO timeseries for Brazil (1960-2017) with one based on mapbiomas state-level area totals (1985-2020. For the pre-1985 period, the cropland area -per-capita was calculated for 1985 for each state of Brazil (based on MapBiomas), then these numbers (fixed) were used to multiply with the historical population numbers for each state of Brazil. We extend forward from 2020, using the trend

FAO has retrospectively updated their data for DRC – this modification is in response to a query regarding a large LU change in 2011 identified in GCB2021/Trendy-v10 and also accounted for in the LUH2 update.

The complete LUH2-GCB2022 forcing timeseries has been updated to use the new HYDE3.3 data for all years, including the extra year 2022. LUH2-GCB2022 uses wood harvest data for based on the 2021 FAO national wood harvest production data along with new HYDE3.3 population data.

To enable alignment with national invenvories we need DGVMs to output NBP on a PFT basis where possible. This is in order to separate fluxes on managed versus unmanaged forested land.

As last year, we will also explore the option to include DGVMs in the GCB Eluc term that are fairly comprehensive w.r.t. land management practices and that are close to observed biomass in addition to the bookkeeping models.

Models can have static or dynamic natural vegetation but all will use prescribed cropland and grazing land (=managed pasture+rangeland) distribution. The models will be forced over the 1700-2021 period with changing CO_2, climate and land use according to the following simulations.

2.1. GCB 2021 simulations (see more detailed protocol below)

S0: Control. No forcing change (time-invariant "pre-industrial" CO_2, climate and land use mask). S0 is needed to diagnose any "cold start" issues or model drift

S1: CO_2 only (time-invariant "pre-industrial" climate and land use mask)

S2: CO_2 and climate only (time-invariant "pre-industrial" land use mask)

S3: CO_2, climate and land use (all forcing time-varying)

Models with N cycle should have time-varying N inputs for S1, S2 and S3 (see Annex 3).

3. Criteria for budget inclusion

As in the past, we will apply three criteria for minimum model realism by including only those models with :

(1) steady state after spin-up. Diagnosed from S0 run. Steady-State defined as an offset < 0.10 GtC/yr, drift < 0.05 GtC/yr per century (i.e. first is the average over 1700-2021, second is the slope x 100).

(2) net annual land flux (Sland-Eluc) is a carbon sink over the 1990s and 2000s as constrained by global atmospheric and oceanic observations (Keeling and Manning 2014). Diagnosed from S3 run.

(3) global net annual land use flux (Eluc) is a carbon source over the 1990s. Diagnosed from S3-S2 runs.

Note- as last year, DGVM results will be evaluated in the iLamb benchmarking system and summary statistics will be given for each model (in summary table/figures) and included in the supplementary material of the ESSD paper. This will enable us to document model improvement each year, and to identify possible issues / model deficiencies to aid model development. We do not envisage using the benchmarking results as criteria for budget inclusion at the moment, but potentially in future years after further consultation among participating groups.

4. Dataset provided and data access

4.1 CRU Climate forcing:

0.5 degree CRU monthly historical forcing over 1901-2021.

Monthly CRU data are provided by Ian Harris at UEA 1901-2021 and available from the following website:

https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.06/

4.2 CRU-JRA climate forcing

0.5 degree CRU-JRA55 6-hourly historical forcing over 1901- 2021

6 hourly CRU-JRA55 climatology provided by Ian Harris at UEA 1901-2021 and available from from Exeter ftp site.

4.2.1 *Revised Radiation fields*

As in GCB2021 we have revised radiation fields: A diffuse fraction dataset offers 6-hourly distributions of the diffuse fraction of surface shortwave fluxes over the period 1901-2021, as described in O'Sullivan et al., https://doi.org/10.1088/1748-9326/ac3b77.

These new radiation fields are in ./input/CRUJRA2022/Radiation-fields

fd_v11_year (diffuse fraction fields), tswrf_v11_year (total downward shortwave radiation at the surface)

4.3 Global atmospheric CO_2

1700-2021 annual time-series, derived from from ice core CO_2 data merged with NOAA annual resolution from 1958 onwards. Prepared by C Le Quéré / Matthew Jones for the Global Carbon Project. Most small differences with the 2017 data are from the revisions of the trend between MLO and SPO which is used to fill missing SPO data. This dataset is intended to be used as atmospheric forcing for modelling the evolution of carbon sinks. Data from March 1958 are monthly average from MLO and SPO provided by NOAA's Earth System Research Laboratory http://www.esrl.noaa.gov/gmd/ccgg/trends/. When no SPO data are available (including prior to 1975), SPO is constructed from the 1976-2017 average MLO-SPO trend and average monthly departure. Data for 2016-2021 are preliminary values. The data from 1980 through 2006 were reprocessed in 2011 to bring them into the WMO X2007 scale. Data prior to March 1958 are estimated with a cubic spline fit to ice core data from Joos and Spahni 2008, "Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years", PNAS (https://www.pnas.org/content/105/5/1425).

Annual mean fields are generated from these monthly data. DGVMs may also wish to run directly with monthly CO_2 fields.

CO_2 data are available from Exeter ftp site.

4.4 Land use change:

Land-use Harmonization (LUH) data for GCB 2022 is provided in 3 separate files, which can be downloaded from the following links (for the states, transitions, and management data layers respectively):

http://luh.umd.edu/LUH2/LUH2_GCB_2022/states.nc
http://luh.umd.edu/LUH2/LUH2_GCB_2022/transitions.nc
http://luh.umd.edu/LUH2/LUH2_GCB_2022/management.nc

These files are based on the new HYDE3.3, as well as the 2021 FAO wood harvest data, for all years 850-2022. A summary of the methods we used are described in annex two.

The data files are for the years 850-2022, which keeps the file format consistent with the LUH2 data produced for CMIP6, hence the start year of 850. The LUH2-GCB2022 data will be different from the LUH2 v2h data used for CMIP6 for all years, due to the use of the new HYDE3.3 crop/grazing land dataset.

4.5 Misc. Datasets

Each group will use its own data source for soil properties etc.

5. Experiment protocol

- Model spin up:

    - 1700 CO_2 concentration (276.59ppm).

    - recycling climate mean and variability from the early decades of the 20th century (i.e. 1901-1920).

    - constant 1700 LUC (crops and pasture distribution).

- 1701-1900 transient simulation:

    - varying CO_2 (S1, S2, S3). 1700 CO_2 (S0)

    - continue recycling spin up climate (all simulations)

    - varying LUC (S3). 1700 LUC, as in spin-up (S0, S1, S2)

- 1901-2021 transient simulation:

    - varying CO_2 (S1, S2, S3). 1700 CO_2 (S0).

    - varying climate (S2, S3). Continue recycling spin up climate (1901-1920: S0, S1)

    - varying LUC (S3). 1700 LUC, as in spin-up (S0, S1, S2)

Models having a nitrogen cycle should use time varying Nitrogen inputs (see annex 3)

6. Required outputs

- For all simulations (S0 to S3): Ascii file with five columns: year, annual global NBP, annual northern extra tropics NBP, annual tropical NBP, annual southern extra-tropics NBP (see excel file for definition and sign convention); one row per year, 1700-2021. Name convention: Model_zonalNBP.txt, e.g. JULES_zonalNBP.txt. Units are $PgCyr^{-1}$. One dataset per simulation S0-S3, four in total. First row use the following column headings and order: "Year, Global, North, Tropics, South". Row 2 values for the year 1700, Row 3 for year 1701 … North = north of 30°N; Tropics = 30°N to 30°S; South = south of 30°S.

- List of gridded output variables: See companion Excel file.

    - Level 1 variables: essential

    - Level 2 variables: desirable for additional analysis/studies

    - Additional N-cycle variables where applicable (see end of excel file)

- Time period: 1700-2021

- Time resolution: as specified in the file

- Spatial resolution: 0.5x0.5 (or at a coarser resolution if necessary; ideally at 0.5 or 1 degree)

- Format netcdf (see Excel file). ***Important*** See annex 5 for netcdf formats developed with input from iLamb team.

- Please define PFTs in the header of Vegtype level netcdf files , e.g. PFT 1 = broadleaf tree, PFT2 = ... Please supply Fractional Land Cover [0-1] of PFT for each simulation as requested (1=total land). If Dynamic Vegetation is not enabled in your DGVM (i.e. changing natural PFT fraction in response to climate) please indicate (e.g. include information in an associated README file). Note the ocean fraction of any given gridcell may not be zero (e.g. at coastal gridcells). Please provide your gridbox fluxes in units per m2 of land fraction, PFT fluxes should be per m2 of PFT, and the PFT land cover fraction should be provided. Please upload the land-sea mask that you are applying.

- **Note**- in previous years we have received identical outputs for different experiments (e.g., same S1 and S2 outputs), different units for different experiments – please double check before submission.

- **Note-** in previous years there has been an order of magnitude size difference in the same output from different DGVMs, e.g. PFT level LAI ranges from ~ 6 to 60 GB – this is likely due nc version (it makes a massive difference). Please if you are generating massive nc files perhaps consider changing nc version.

7. Output file name convention

One file per variable, entire time-series

Model_Simulation_variable.nc (e.g. JULES_S1_mrso.nc)

Please see Annex 5 for an example netcdf header for variable nomenclature

8. ftp Instructions for data access:

Request access through

https://globalcarbonbudget.org

Ian Harris (UEA) merged the "new generation" reanalysis from JRA-55 (Japanese *55*-year Reanalysis) with the CRU TS dataset.

1. All JRA-55 data are regridded to the CRU 0.5° grid using appropriate NCL routines based on the Spherepack package, and masked to give a land-only (excluding Antarctica) dataset.

2. For the four variables tmp, dswrf, shum and pre, JRA-55 is aligned to CRU TS (v4.03) tmp, cld, vap and pre (also wet) respectively over land, using the same transformations as previously. The other four variables (pres, ugrd, vgrd, dlwrf) pass through without further modification.

3. For years between 1958 and 2021, JRA-55 is used. Alignment to CRU TS occurs where appropriate.

4. For years between 1901 and 1957, random (but fixed) years from JRA-55 for 1958-1967 are used to fill. Alignment to CRU TS applies separately to each instance, as appropriate (ie, using the appropriate CRU TS year).

The resolution of JRA is 0.5 degree. This means that now resolution of reanalysis is compatible with resolution of the CRU dataset. This will not change the monthly fields that are still aligned to CRU TS but obviously it will change the spatial and high frequency temporal variability of the fields.

Land-use states for all years 850-2022.

Land-use states are based on updated HYDE 3.3 land use and population data and FAO wood harvest data. In addition the time series has been extended to include land-use states in the year 2022 (and land-use transitions during the year 2021). LUH2 algorithms and methodology remains the same, and other inputs to the LUH2 model also remain the same.

HYDE inputs: Data from HYDE3.3, prepared for GCB 2021, is based on a FAO release (bulkdownload February 19th 2020), which includes yearly data from 1961 up to and including the year 2017. After the year 2017 HYDE extrapolates the cropland, pasture, and urban data, based on the trend over the previous 5 years, to generate data until the year 2021 HYDE also uses satellite imagery from ESA-CCI from 1992 – 2018 for more detailed yearly allocation of cropland and grazing land. The 2018 map is also used for the 2019-2021 period. The original 300 meter resolution data from ESA was aggregated to a 5 arc minute resolution according to the classification scheme as described in Klein Goldewijk et al (2017). For Brazil we replace FAO state-level data for cropland and grazing land by those from in-country land cover dataset MapBiomas for 1985-2020. ESA-CCI is used to spatially disaggregate as described above. Similarly, an estimate for year 2021 is based on the MapBiomas trend 2015-2020. The pre-1985 period is scaled with the per capita numbers from 1985 from MapBiomas, so this transition is smooth. FAO has retrospectively updated their data for DRC – this modification is in response to a query regarding a large LU change in 2011 identified in GCB2021/Trendy-v10 and is also been accounted for in the GCB 2022 update. All other countries remain as for GCB 2021, apart from the extrapolation by the additional year, 2022.

Wood harvest inputs: The version of wood harvest data used for LUH2 v2h was based on a previous FAO release that included data up to and including the year 2014 – those inputs have been updated for this GCB dataset to use the 2021 FAO wood harvest dataset for all years from 1961 to 2020. After the year 2019 we extrapolated the wood harvest data until the year 2022. The HYDE3.3 population data is also used to extend the wood harvest timeseries back in time. Other wood harvest inputs (for years prior to 1961) remain the same in LUH2.

Conversion to pasture/rangeland

The LUH2 methodology uses the cropland, urban, managed pasture, and rangeland layers from HYDE. DGVM groups in the past have requested more information on whether natural vegetation is lost in conversion to pasture and rangeland.

Following LUH2 simple guidelines (on their website): "***all natural vegetation should be cleared for managed pasture, and only cleared for rangeland if it is forested***".

Using this rule/guideline gives maps of forest area, carbon density, and carbon emissions that are consistent with other published maps.

The "staticData_quarterdeg.nc" file on the LUH website contains a layer named fstnf which is 1 when the potential vegetation is forested, and 0 when it is not. This layer can be used to designate whether any rangeland increases should imply clearing of natural vegetation (yes, if fstnf is 1 and no if fstnf is 0).

Users can download this file from here:

http://gsweb1vh2.umd.edu/LUH2/LUH2_v2h/staticData_quarterdeg.nc
Annex 3 Nitrogen cycle

Models having a nitrogen cycle should use time varying Nitrogen inputs as follows:

S0 none (PI CO_2, PI climate, PI LUC, PI Ndep, PI Nfert)

S1 CO_2 + Ndep (PI Nfert)

S2 CO_2 + climate + Ndep (PI Nfert)

S3 CO_2 + climate + LUC + Nfert + Ndep

Note: PI = 1700 for LUC, PI = 1850 for Nfert, PI= 1850 Ndep.

Nitrogen fertiliser input datasets are available via the NMIP2 project

(https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-17-0212.1)

Note, N fertiliser data is available until 2018 from NMIP2. Values for 2019 and 2020 will be available by ca. 17th June. As NMIP2 assume these N input data remain unchanged in year 2021. N fertiliser is available only from 1860, please assume N Fertiliser at the 1860 value for years 1700-1860.

Manure is an organic fertiliser (animal waste put on fields). It's fairly important from the N cycle perspective, because it's one of the important pre-artificial fertiliser sources. However, as it's based on organic N, it causes a problem with the model mass balance (you need to take the C and N from land, respire some of the C, and then add the remaining C:N onto the cropland). Doing this wrongly will have an effect on the C cycle simulation. For TRENDY, we recommend to not include it (however if you use it, you must tell us where you take the C and N from...). Note: If models choose to include manure, against our recommendation, then we need a manure application rate for S0-S3 Nfert.

In terms of artificial fertiliser, it's fairly safe to assume that the per area rates haven't changed much between 1700 and 1850. For manure, this would not be so easy.

N deposition (search for "N deposition" from):

https://esgf-node.llnl.gov/search/input4mips/

Please use the historical N-deposition database (1850-2014) then transition onto the Future RCP8.5 N-deposition databases (2015-2100) for years 2015, 2016, 2017, 2018, 2019, 2020 and 2021.

N deposition is available only from 1850, please assume N deposition at the 1850 value for years 1700-1850.

NOTE: Peter Anthoni has kindly downloaded and regridded these N deposition files and uploaded them onto the Exeter server:

./input/ndep

Annex 4. Lightning ignition and population density

Given uncertainty around lightning datasets, scaling factors, and potential need for model recalibration, and the fact in TRENDY we want models to supply their best C-cycle representation, groups are free to choose the lightning dataset they use.

Gridded *population density* based on HYDE3.2 is available on the Exeter site.

There is also included the total land fraction per gridcell (from HYDE), as this might be important for some models.

For fire-enabled DGVMs please use *varying* population density in simulations S1-S3. Our simplified logic is there is LUC and its direct consequences (Nfert) that go together in Eluc, and all other environmental changes (Ndep, population, climate, CO_2) in SLAND.

Annex 5 Output netcdf formats

The aim is to be more consistent with CMIP, LUMIP, LS3MIP in our format/variable requests to aid analysis:

1. Please follow the protocol (or explicitly state why not).
2. All modelling teams provide a methodology (in a README file) of how to calculate global annual nbp from gridded monthly files (grid and pft level). This will avoid confusion of whether to use landmasks/landcover/gridareas/etc.
3. In the past "lai" has not been consistent between models. We have changed "lai" to gridcell mean lai and include new variable laipft for the pft level.
4. Order of dimensions should be consistent. Eg [lon,lat,PFT,time]. When using ncdump this reads as [time,PFT,lat,lon].
5. Please provide a list of variables that are not applicable for your model. E.g. cSoilpft might not exist. This gives us an idea of what variables we can request/expect.
6. Using cf-complient units. Remove "C" for carbon and "N" for nitrogen from the units and don't measure time in years or months, e.g. All CO_2 stocks and fluxes were previously requested in units kgC m-2, and kgC m-2 s-1, respectively, please remove the letter C to be cf-compliant in the netcdf files.
7. Gridbox fluxes should be per m2 of land
8. PFT fluxes should be per m2 of PFT
9. Pools, coverages, LAI etc should be per m2 of land
10. All models to provide a land fraction file if using regular lat-lon grids, or a land fraction and grid area if using non regular grids.
11. All models should use a consistent file naming (e.g. JULES_S1_mrso.nc). Eg. do not include annual/monthly/perpft tag.

12. Following this, PFT labels are different among DGVMs (pft, PFT, vegtype…). Please all use nomenclature, PFT.
13. Consistent latitude/longitude use (e.g. do not use lat/lon)
14. Consistent fill value of -99999 to be used (e.g. not -9999)
15. All data from -180 -> 180 and -90 -> 90.
16. All models output over the same time period, 1700-2018, e.g. until now some supply from 1700, others 1840, 1850, 1900, 1901.

To ensure accessibility by broad users, avoid to format netcdf files with netcdf library 4.4.0 or earlier, combined with libhdf5 1.10.0 or greater. There is a known issue with netcdf formatted by this set of libraries.