2020 Global Carbon Budget: Land modelling protocol (Trendy-v9)

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 2020 simulations: S0, S1, S2, S3 simulations latest: August 16th 2020

Goal: To provide the land components of the Global Carbon Project 2020 Budget, and an ensemble of land carbon cycle simulations to be used by the scientific community. The Trendy-v9 data will be made available to GCP's REgional Carbon Cycle Assessment Project phase 2 (RECCAP2), and represent Tier 1 products (regional C balance & component fluxes based on global products). Usual Trendy data-use policy will apply for RECCAP2 (https://sites.exeter.ac.uk/trendy/data-policy).

2. Model simulations

Given the current circumstances (Covid-19) we appreciate the extra burden to continue to contribute to Trendy. We therefore attempt to minimize the work for groups this year and are not requesting any extra simulations above the standard S0-S3. For example, the S4-S6 from Trendy-v8 will be used to adjust the land-use flux (Eluc) from book-keeping models for environmental effects. 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 for this or one of the coming budgets (at present we only use DGVMs to characterise uncertainty in Eluc to better characterise our understanding).

The LUH2 forcing has been updated to include the extra year. There are no retrospective changes to the HYDE cropland/grazing land data but the FAO wood harvest data has changed for the years 2015 onwards and so those are now being used in this year's LUH-GCB dataset. This means the LUH-GCB data will be identical to last year's dataset for all years up to 2015 and will differ slightly in terms of wood harvest and resulting secondary area/age/biomass for years after 2015. We'll postpone significant LU forcing changes until next year. Although CRUJRA has been retrospectively changed, changes are minimal for all variables except vapour pressure (VAP) which is converted to specific humidity (SPFH). Therefore groups who are planning to use their 2019 model version and do not use SPFH as a driver, may consider simply applying their model for the extra years (2015-2019) rather than redoing for the full timeseries, 1700-2019. However we expect all groups to upload full timeseries datasets to the ftp server.

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 1700-2019 period with changing CO_2, climate and land use according to the following simulations.

2.1. GCB 2020 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 1700-2019, 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 1901-2019.

Monthly CRU data are provided by Ian Harris at UEA 1901-2019 and available from the following website:

https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.04/

4.2 CRU-JRA climate forcing

0.5 degree CRU-JRA55 6-hourly historical forcing over 1901- 2019

6 hourly CRU-JRA55 climatology provided by Ian Harris at UEA 1901-2019 and available from from Exeter ftp site (see instructions below) in directory:

./input/CRUJRA2020

Groups needing daily forcing can start from either monthly CRU or 6-hourly CRU-JRA55. In any case they have to make sure that monthly totals are conserved. Groups aggregating from 6-hourly to daily might find the following repository of bash+cdo+nco scripts useful: https://github.com/MagicForrest/CRUJRA_4_GUESS. See annex 1 for description of JRA-55 and differences with NCEP previously used up to TRENDYv6).

4.3 Global atmospheric CO_2

1700-2019 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 2016-2019 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.

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 (see instructions below) in directory:

./input/CO2field/global_co2_ann_1700_2019.txt

4.4 Land use change:

Land-use Harmonization (LUH) data for GCB 2020 is provided in 3 separate files, which can be downloaded from the following links (for the states, transitions, and management data layers respectively):

http://gsweb1vh2.umd.edu/LUH2/LUH2_GCB_2020/states.nc
http://gsweb1vh2.umd.edu/LUH2/LUH2_GCB_2020/transitions.nc
http://gsweb1vh2.umd.edu/LUH2/LUH2_GCB_2020/management.nc

These files are based on updated data from HYDE for the years 1960-2020, as well as the latest FAO wood harvest data. A summary of the methods we used are described in annex two.

The data files are for the years 850-2020. We thought it would be helpful to keep the file format consistent with the LUH2 data produced for CMIP6 though, hence the start year of 850. The land-use states for the years 850-1950 will be identical to the LUH2 v2h dataset, and the states for the years 1951-2020 will be different from the LUH2 v2h dataset, primarily in Brazil.

Note, for Trendy-v9 the LUH2 dataset for GCB 2020 is supplemented for the one additional year of HYDE crop/grazing forcing (i.e. HYDE has **not** been retrospectively changed), but the FAO wood harvest data has changed for the years 2015 onwards and so those are now being used in this year's LUH-GCB dataset. This means the LUH-GCB data will be identical to last year's dataset for all years up to 2015 and will differ slightly in terms of wood harvest and resulting secondary area/age/biomass for years after 2015.

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)
- 1901-2018 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, 1700-2019. Name convention: Model_zonalNBP.txt, e.g. JULES_zonalNBP.txt. Units are $PgCyr^{-1}$. One dataset per simulation S0-S3, four in total. Global; 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: 1700-2019
- 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. This is particularly relevant for the regional analysis as part of RECCAP2.

- **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 Output

sftp trendy-v9@trendy.ex.ac.uk

password: gcb-2020

sftp> cd output

sftp> mkdir YOUR_MODEL

sftp> cd YOUR_MODEL

sftp> mkdir Sx (S0 to S6)

sftp> cd Sx

sftp> put Model_Simulation_variable.nc

9. PLEASE update model description tables and nominate author

Please report any new info (for new models) or changes from Tables 4 and A1 in the ESSD paper, and identify who the ESSD nominated author should be for each model, and to identify specific sources of funding if relevant (Table A5).

GCP2020_Table4_Model_specific_changes	https://drive.google.com/file/d/1bdXV-J2CK_TNp0NVFrTFx4PMVHLf5UoG/view?usp=sharing
GCP2020_TableA1_DGVM_processes_setup	https://drive.google.com/file/d/1ChTmgbgjgQ8y7toWDk1w44JJptkUhpTE/view?usp=sharing
GCP2020_TableA5_Funders	https://drive.google.com/file/d/1UguwYmAD1INXYPh_cWF94rptnUMd5w2B/view?usp=sharing

Annex 1 Description of CRU-JRA55

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 2019, 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.

Note although CRU retrospectively update their monthly climate data each year, as additional data become available, the differences are considered small except for VAP (vapour pressure), which is converted to specific humidity (SPFH), where a new algorithm has been employed, see:

https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.04/comparisons_with_other_releases/

Ideally all DGVM groups will make full simulations from 1700-2019, but we understand, given the exceptional circumstances, if DGVMs who do not use SPFH as input, opt to submit last years runs for years 1700-2014 supplemented with new output for years 2015-2019.

Annex 2 LULCC forcing

Land-use states for the years 850-1950, and land-use transitions for the years 850-1949, are the same as LUH2 v2h (released for CMIP6).

Land-use states for the years 1951-2020, and land-use transitions for the years 1950-2019, are based on the forcing supplied in Trendy-v8, with the LUH2-GCB2020 forcing updated to include the extra year. There are no retrospective changes to the HYDE cropland/grazing land data but the FAO wood harvest data has changed for the years 2015 onwards. This means the LUH2-GCB2020 data will be identical to last year's dataset for all years up to 2015 and will differ slightly in terms of wood harvest and resulting secondary area/age/biomass for years after 2015.

The majority of differences between the LUH2-GCB2020 data and the LUH2 v2h data will be seen in the country of Brazil, where a correction based on new FAO data was made.

HYDE inputs: Data from HYDE, prepared for GCB 2020, is based on an FAO release, which includes data up to and including the year 2015 (HYDE applied annual changes in FAO data to the year 2012 data from a previous release to get the new 2013-2015 data used for GCB 2020). After the year 2015 HYDE extrapolates the cropland, pasture, and urban data, based on a moving window of the previous 5 years, to generate data until the year 2020.

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 remained the same in this GCB dataset. The new wood harvest data, prepared for GCB 2020, is based on the most recent FAO release, which includes data up to and including the year 2018 (we applied annual changes in FAO WH data to the year 2014 data from the previous release to get the new 2015-2018 data used for GCB 2020). After the year 2018 we extrapolated the wood harvest data until the year 2019.

Conversion to pasture/rangeland
The LUH2 methodology uses the cropland, 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= 1860 Ndep.

Nitrogen fertiliser input datasets are available via the NMIP project

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

Note, N fertiliser data is available until 2014. As NMIP assume these N input data remain unchanged in years 2015, 2016, 2017, 2018 and 2019. 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 and 2019.

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 TRENDYv9 server:

./input/pop_dens

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,vegtype,time]. When using ncdump this reads as [time,vegtype,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-complient 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 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.