

Moirai Land Data System (Moirai LDS) User Guide

Version: 3.0

Alan Di Vittorio
January, 2019

NOTICE:

The repository [<https://github.com/JGCRI/moirai/>] uses Git Large File Storage (LFS). Please download and install from here: [<https://github.com/git-lfs/git-lfs/wiki/Installation>]

Once installed, run the following command before cloning this repository: ``git lfs install``

Copyright

Moirai Land Data System (Moirai) Copyright (c) 2019, The Regents of the University of California, through Lawrence Berkeley National Laboratory (subject to receipt of any required approvals from the U.S. Dept. of Energy). All rights reserved.

If you have questions about your rights to use or distribute this software, please contact Berkeley Lab's Intellectual Property Office at IPO@lbl.gov.

NOTICE. This Software was developed under funding from the U.S. Department of Energy and the U.S. Government consequently retains certain rights. As such, the U.S. Government has been granted for itself and others acting on its behalf a paid-up, nonexclusive, irrevocable, worldwide license in the Software to reproduce, distribute copies to the public, prepare derivative works, and perform publicly and display publicly, and to permit other to do so.

License

Moirai (moi'-rĪ) is free software: you can use it under the terms of the modified BSD-3 license (.../moirai/license.txt)

Citations

Please cite the associated paper in the Journal of Open Research Software (forthcoming). Also please cite the appropriate input data and associated papers (see Required downloads and installs below, Inputs below, and also .../Moirai/docs/third_party_contributions.pdf).

Overview

The Moirai Land Data System (Moirai LDS) is designed to produce recent historical land data inputs for the AgLU module of GCAM data system¹, but the Moirai LDS outputs could also be used by other models/applications. The Moirai are the Greek Fates, and this software is named Moirai to represent the fundamental influence of land data inputs on model outcomes. The primary function of the Moirai LDS is to combine spatially explicit input data (e.g., raster images) with tabular input data (e.g., crop price table) to generate tabular output data for a suite of variables. Some of these outputs replace the data provide by the Global Trade Analysis Project (GTAP), and other data replace and augment the original GCAM GIS processing. The Moirai LDS output data are aggregated by Geographic Land Unit (GLU)² within each country. The GLU coverage is an input to the Moirai LDS (as a thematic raster image and an associated CSV file that maps the thematic integers to names), and the GLU boundaries can be determined arbitrarily. Previous versions of GCAM (and Moirai LDS) used only bioclimatic Agro-Ecological Zones (AEZs) and corresponding data that were provided by GTAP, as the GLUs. As a result, some AEZ terminology still exists in the code, but this terminology now refers more generally to GLUs. The Moirai LDS now enables any set of boundaries to be used as GLUs (including AEZs), allowing for more flexible generation of land use region boundaries (defined as the intersection of GLUs with geopolitical regions). The current default set of GLUs is the same set of 235 global watersheds as used by the GCAM water module. The GCAM 5.1 geopolitical regions (32 or 14) are included and used as inputs to Moirai to generate a mapping file between the Moirai outputs, which are at the level of the intersection between the GLUs and the country boundaries, and the geopolitical regions. The diagnostics scripts use this geopolitical region mapping in some cases. Moirai can also recalibrate three of the outputs (crop production, harvested area, and land rent) to a specified year that is the center of a five-year averaging window. No recalibration retains the circa 2000, 7-year average of the source data. The currency-year for land rent can also be specified, and the default is 2001 to match the GTAP data.

Moirai LDS Framework

¹ Moirai LDS development started with the GCAM data system in a new branch created by Page Kyle on 18 Sep. 2015 (<https://128.8.246.24/svn/branches/lds-workspace>; r6376). The initial commit of the LDS (r6408) on 23 Sep. 2015 was the previous version used to generate data for 18 new AEZs, and development of the current Moirai LDS proceeded from this point. The actual code and most of the required inputs for version 2 are in `.../lds-workspace/input/gcam-data-system/aglu-processing-code/lds`. The outputs used by the GCAM data system are copied by the version 2 LDS into `.../aglu-data/LDS` and `.../aglu-data/mappings`, and for the current Moirai LDS these directories are specified in the Moirai LDS input file.

² All output files refer to GLUs, but not all the terminology (i.e., comments, variables) within the code has been converted from AEZ to GLU.

This section focuses on the meta-structure of the Moirai LDS framework with the aim of providing a background for using the system. Complementarily, the basic processing flow is depicted in Figures 1 and 2. The Moirai LDS framework consists primarily of C code and is contained within the `.../moirai` project directory (in `src` and `include` directories), along with all input data. Five publically available data sets are also included that can be moved or downloaded separately because their location is set in the Moirai LDS input file (e.g., `.../moirai/input_files/moirai_input_basins235.txt`). The user will need to download and install the C NetCDF library to read one of these data sets. The input data are in `.../moirai/indata` (this directory is set in the Moirai LDS input data file), including the two files that specify the GLUs (which are also set in the Moirai LDS input file). The Moirai LDS outputs (main and diagnostic) go into a directory within the `.../moirai` directory that is specified in the Moirai LDS input file (e.g., `.../moirai/outputs/basins235`). A runtime log file (e.g., `moirai_log_basins235.txt`), the name of which is also set in the Moirai LDS input file, is also written to the specified output directory. The ten main output files used by the GCAM data system are also copied by the Moirai LDS into directories specified in the Moirai LDS input file (one directory for data and one for mappings).

Five R scripts are also included in the framework. The four R scripts in `.../moirai/diagnostics` generate various diagnostic outputs, and need some of the Moirai LDS diagnostic output files in order to run. The `.../moirai/indata/WaterFootprint/convert_wfgrids2binary.r` script was used to convert the water footprint files from ARC binary grids to simple binary raster images for input to the Moirai LDS (because the linked GDAL library in the C code would not recognize the original files).

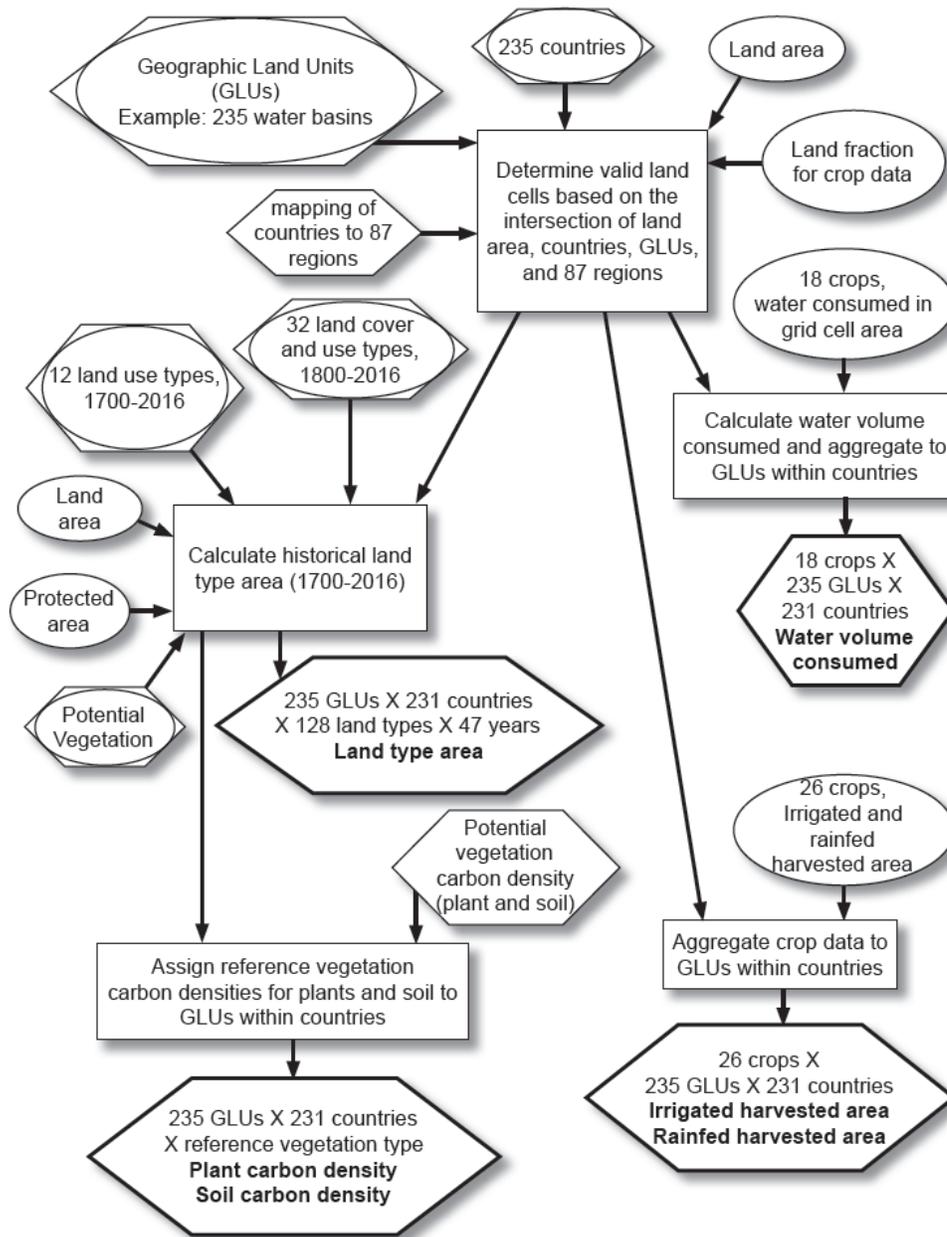


Figure 1. Spatial basis of the Moirai land data system and pathways to five outputs (the irrigated and rainfed data are output as separate files). The Geographic Land Units (GLUs) are defined by the user. Ovals are raster data (all at 5 arcmin except for the land cover/use data, which are at half-degree), hexagons are tabular data, boxes are processes, and the outputs are in underlined, bold lettering within bold hexagons.

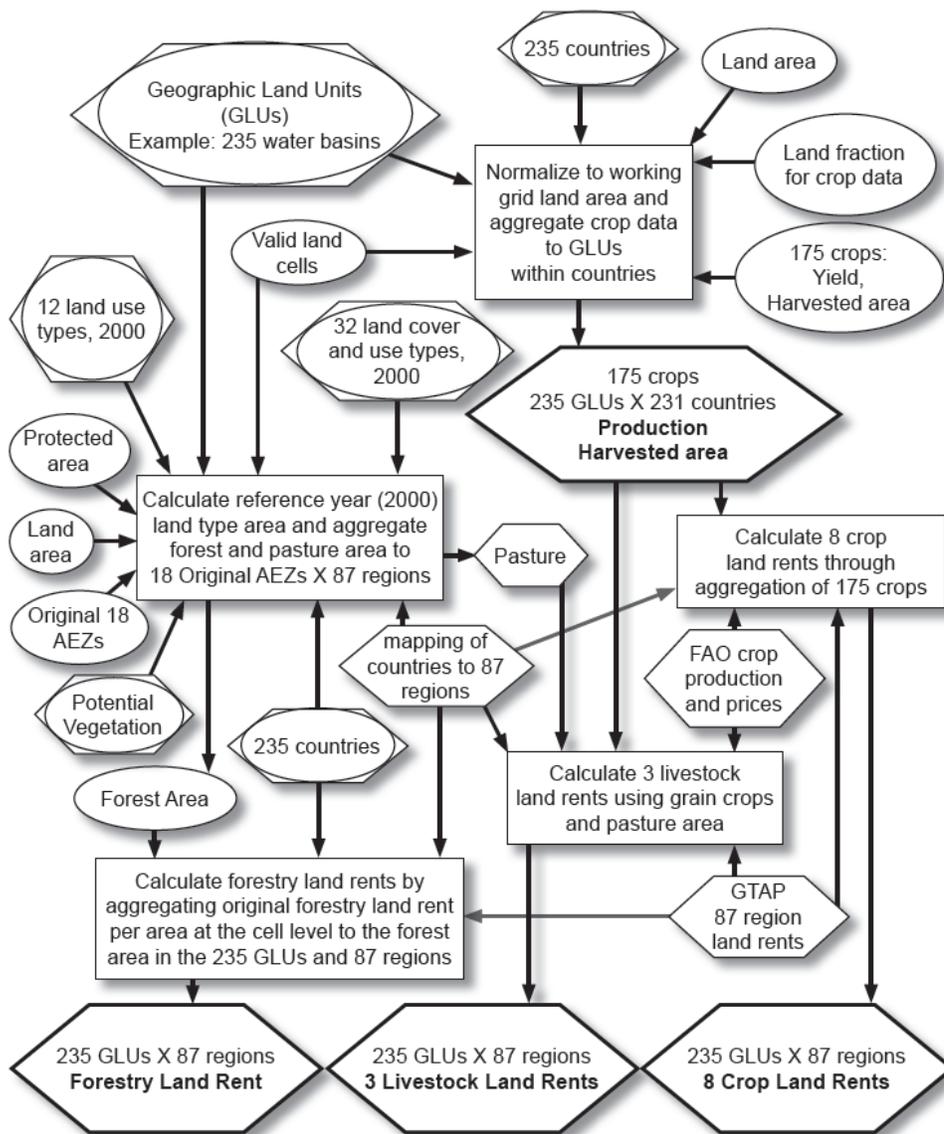


Figure 2. Processing three primary Moirai outputs. Recalibration of crop data year occurs before aggregation, and recalibration of land rent price year occurs before land rent calculation. Ovals are raster data (all at 5 arcmin except for the land cover/use data, which are at half-degree), hexagons are tabular data, boxes are processes, and the outputs are in underlined, bold lettering within bold hexagons.

The Moirai LDS is a command line tool that takes the name of the Moirai LDS input file as the only argument (two examples are provided in `.../moirai/input_files`, and can be run immediately after the code is built). The Moirai LDS can be run within Xcode on a Mac, or compiled anywhere (hopefully) using the `.../moirai/makefile` (which is for Linux and has been tested on the Mac terminal). The easier way to use the Moirai LDS is by opening the `.../moirai/moirai.xcodeproj` file in Xcode, then compiling and running it directly within Xcode, mainly because most of the conditions for running it are already established (the location of the NetCDF library may have to be changed in the build settings, and also in `moirai.h`; the project directory is the working directory, and the name of the input file needs to be set in the scheme). The current default setting for the location of the executable is `.../moirai/Build/Products/Debug`. The makefile, however, puts the executable in `.../moirai/bin`, and the objects in `.../moirai/obj`, but this executable needs to be called from `.../moirai` because the example input file path entries are based on this project directory as the working directory (these can be changed, as needed). The user must first change the location of the NetCDF library in both the makefile and in `moirai.h`. Regardless of how the code is run, the user must also correctly specify the input and output directories (and any other input data files that they may want to substitute) in the Moirai LDS input file before running the code.

The Moirai LDS production outputs are independent of the GCAM regions. However, there are diagnostic outputs that do aggregate data to the GCAM land use regions based on either 14 or 32 GCAM regions. These aggregated diagnostic outputs are used by some of the diagnostic R scripts. The default is 32 regions in the example input files, but the user can specify which GCAM region set to use in the Moirai LDS input file, by pointing to two files (one for the region codes and names and another for mapping iso country codes to region codes) originally copied directly from `.../gcam-data-system/_common/mappings`.

Required downloads and installs

Only the NetCDF library has to be downloaded and installed by the user, as the five data sets below are now included in the repository through the LFS system.

Associated licenses and ownership are included in `.../moirai/docs/third_party_contributions.docx`.

C NetCDF library (v4.1.1)

The user must have the C NetCDF library installed (available at <http://www.unidata.ucar.edu/software/netcdf>). The header and library search paths for compiling Moirai LDS must include the location of the installed C NetCDF library. The linking flags must also be set (`-L<header_location> -lnetcdf`) for successful compilation of the Moirai LDS.

An archived version of this library is now available here:

[https://stash.pnnl.gov/projects/JGCRI/repos/moirai/browse/required_libs].

SAGE 175 crop harvested area and yield data, circa 2000

These data are now available at [<http://www.earthstat.org/data-download/>], labeled as "Harvested Area and Yield for 175 Crops." Put all of the zipped NetCDF files (one for each crop) in a single directory, then set this directory in the Moirai LDS input file. The Moirai LDS will automatically unzip the files and leave the both the zipped and unzipped files in the directory. Alternatively, the user can download the ascii grid files and rewrite the read function accordingly so that the NetCDF library is not necessary. The metadata file is included for reference, and the corresponding journal article is cited on the download page. Please cite these data when using Moirai: Monfreda, C., Ramankutty, N. & Foley, J. A. 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, *Global Biogeochem. Cycles*, 22, GB1022. Harvested area units are the fraction of land area within each grid cell, and yield units are metric tonnes per ha.

MIRCA2000 crop irrigated and rainfed harvested area data, circa 2000

These data are available at [https://www.uni-frankfurt.de/45218031/data_download/]. The specific data are labeled "Annual harvested area grids for 26 irrigated and rainfed crop classes." Login as a guest, and put all of the 5 arcmin individual crop files (ANNUAL_AREA_HARVESTED_IRC_CROP#_HA.ASC.gz and ANNUAL_AREA_HARVESTED_RFC_CROP#_HA.ASC.gz) into a single directory, gunzip them (use "gunzip -k" if you want to retain the gzipped files), then set this directory in the Moirai LDS input file. The Moirai LDS will NOT automatically unzip these files (because the included files are already unzipped). A metadata file is included for reference, and the corresponding journal article is also available. Please also cite the MIRCA journal article when using Moirai: PORTMANN, F. T., SIEBERT, S. & DÖLL, P. 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, 24, GB1011, doi: 10.1029/2008GB003435. Units are hectares.

HYDE 3.2.000 baseline land use data

These data are available at [ftp://ftp.pbl.nl/hyde/hyde3.2/2017_beta_release/]. Only 1700-2016 baseline land use data are included here, and the Moirai LDS works only with "AD" era years (the "BC" era years are not supported). Note that there is a newer version (3.2.1) of these data available at [<ftp://ftp.pbl.nl/hyde/hyde3.2/>], which can also be used as input to the Moirai LDS, but we include 3.2.000 here because it is the same version used to generate the included ISAM land cover data (see below). Put all of the zipped files in a single directory, then set this directory in the Moirai LDS input file. The Moirai LDS will automatically unzip these files. The corresponding README file is included for reference. Please cite these data when using Moirai: Klein-Goldewijk, K., Beusen, A., Doelman, J. & Stehfest, E. 2017. Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth Syst. Sci. Data*, 9, 927-953. Units are square kilometers.

ISAM land cover data

These data have been generated specifically for the Moirai LDS and are based on the HYDE 3.2.000 baseline data. The full dataset is available at [<http://climate.atmos.uiuc.edu/atuljain/availabledata.html>], and previous versions of these data with associated documentation are available at [<https://www.atmos.illinois.edu/~meiyapp2/datasets.htm>]. Only the years corresponding to the HYDE 3.2 years (from 1800-2016) are included here. Put all of the gzipped files in a single directory, then set this directory in the Moirai LDS input file. The Moirai LDS will automatically gunzip the files and retain the gzipped files. A data document for the public version is included for reference. Please also cite these data and the forthcoming ISAM data paper when using Moirai. Units are fraction of grid cell for land cover, and square meters for grid cell area.

Water footprint data, circa 2000

These data are available at [<https://waterfootprint.org/en/resources/waterstat/product-water-footprint-statistics/>], labeled as “Product water footprint statistics: Water footprints of crops and derived crop products.” Select the Rastermap download link, unzip the file, and then run `.../moirai/indata/WaterFootprint/convert_wfgrids2binary.r` (with the proper paths, of course) to convert the files to simple binary raster files. This R script writes the new files into the same, newly unzipped directory, so the user can set this directory in the Moirai LDS input file (the current default is the name already given to this directory). The corresponding journal article is also available. Please cite these data when using Moirai: Mekonnen, M.M. & Hoekstra, A.Y. (2011) The green, blue and grey water footprint of crops and derived crop products, *Hydrology and Earth System Sciences*, 15(5): 1577-1600. Units are average annual mm over the entire grid cell area (1996-2005).

Outputs

The Moirai LDS takes a given set of GLU boundaries and generates **ten production output files** for use by the GCAM data system. It also generates many diagnostic files (if specified in the Moirai LDS input file) that will not be described here. Eight of the production output files contain data and two of them contain mapping values between countries or land types. The Moirai LDS production data files contain several “type” columns and a single “value” column, with no zero-value records. The first two columns of each file contain the country iso3 code and the GLU integer, in that order. The rightmost type column varies the fastest, and the value column is last. The values are rounded to the integer to represent an appropriate level of precision based on the input data. The data files contain six header lines, and the mapping files contain five header lines, the last of which contains the column labels.

The Moirai LDS generates the following three files and copies them to a user-specified destination directory (e.g., `.../moirai/outputs/basins235/aglu-data`; these files replace the GTAP data previously stored in `.../aglu-data/level0`):

- *MOIRAI_ag_HA_ha.csv* = harvested area, country X GLU X 175 crop (hectares)
- *MOIRAI_ag_Prod_t.csv* = production, country X GLU X 175 crop (metric tonnes)
- *MOIRAI_value_milUSD.csv* = land rent value, country X GLU X 12 use sector (million USD)

Note that these data are based on the 1997-2003 annual average unless otherwise specified in the Moirai LDS input file.

These names and the destination directory are set in the Moirai LDS input file.

The Moirai LDS also generates these files (some of which were previously stored in either the deprecated .../aglu-data/GIS directory or .../aglu-data/level0, may have been previously produced from the old "GIS" code), and copies them to user-specified destination directory (e.g., .../moirai/outputs/basins235/aglu-data):

- *MIRCA_irrHA_ha.csv* = irrigated harvested area, country X GLU X 26 crop classes (hectares)
- *MIRCA_rfdHA_ha.csv* = rainfed harvested area, country X GLU X 26 crop classes (hectares)
- *Land_type_area_ha.csv* = land type area, country X GLU X land type X year (hectares)
- *Ref_veg_carbon_Mg_per_ha.csv* = soil and veg C density for reference vegetation land types, country X GLU X land type X soil/veg C (Megagrams per hectare)
- *Water_footprint_m3.csv* = average annual water volume consumed (1996-2005), country X GLU X 18 crop X water type (m3), blue = surface and groundwater irrigation, green = rain, gray = needed to dilute pollutant runoff, total = the sum, but is slightly different than summing the individual type outputs due to rounding

These names and the destination directory are set in the Moirai LDS input file.

These two mapping files are generated and copied to a user-specified mappings destination directory (e.g., .../moirai/outputs/basins235/aglu-data):

- *MOIRAI_ctry_GLU.csv* = maps the VMAP0 raster integer country codes to iso3 code and FAO country name for each GLU within each country. This is based on the .../moirai/indata/iso_GCAM_regID_#reg.csv file specified in the Moirai LDS input file. The mapping works because the GCAM file lists the FAO country names, which can be matched to the Moirai LDS input FAO/VMAP0 country list.
- *LDS_land_types.csv* = mapping of land type code to description for area and carbon outputs

These names and the destination directory are set in the Moirai LDS input file.

Inputs

This section focuses on the inputs listed in the Moirai LDS input file. Of the many inputs to the Moirai LDS, those listed in the Moirai LDS input file are the most important as they include the primary source data in addition to the GLU definition that determines how to aggregate the source data. The extent and resolution of the default raster input data determine the working resolution of the Moirai LDS, which is defined in the Moirai LDS header file (.../moirai/include/moirai.h; global extent, 5 arcmin resolution). When substituting input data the user must ensure that the new data are read in and resampled to the working grid. An alternative grid can also be defined in the Moirai LDS header file. Each input has its own read function that can be rewritten to accommodate input dataset substitution. Input raster data sources are listed in Table 1 (.../moirai/docs/moirai_v3_table_1.pdf) and input text data sources in Table 2 (.../moirai/docs/moirai_v3_table_2.pdf).

Moirai LDS input file (e.g., .../moirai/input_files/moirai_input_basins235.txt)

The Moirai LDS input file specifies the input and output paths, the file names of the primary input and output files, and whether additional diagnostic files are output. The output year for production, harvested area, and land rent outputs, is specified, as well as the input year of the required crop data to determine whether or not recalibration is necessary. Similarly, the output USD value year for land rent is specified along with the input USD value year of the FAO price data in order to perform the correct price calibration. The input file code variables are filled based on the order of the uncommented lines in the input file, rather than by keyword (# is the comment character), and there are 50 input values read from the input file. Thus, the following input descriptions follow the order in the input file.

Flags

- diagnostics: 0 = no, 1 = output diagnostics files

Data years for recalibration

- out_year_prod_ha_lr: output year for crop production, harvest area, and land rent; 0 = no recalibration (retain circa 2000 sage source data); (valid recalibration years 1995 - 2014 based on the FAO inputs specified below)
- in_year_sage_crops: input year of the 175 crop harvest area and yield data
- out_year_usd: the output US dollar value year for land rent (valid years 1970 - 2017 based on the cpi input specified below)
- in_year_lr_usd: the US dollar value year for the input land rent data

File paths (must include final "/")

- inpath: path to directory containing all input files except for the SAGE 175 crops, HYDE, ISAM land cover, MIRCA2000, and water footprint data (.../moirai/indata)
- outpath: path to directory where all output files will be written (e.g., .../moirai/outputs/<user_specified>)
- sagepath: path to directory containing the SAGE 175 crop netCDF files (<user_specified>)

- hydepath: path to directory containing the unzipped hyde land use files (<user_specified>)
- lulcpath: path to directory containing the ISAM land cover files (<user_specified>)
- mircapath: path to directory containing the MIRCA2000 ascii grid files (<user_specified>)
- wfpath: path to directory containing the water footprint simple binary raster files (<user_specified>)
- ldsdestpath: path to directory where the eight Moirai LDS output data files will be copied to (e.g., .../aglu-data/moirai)
- mapdestpath: path to directory where the two Moirai LDS output mapping files will be copied to (e.g., .../aglu-data/mappings)

Raster input data

- HYDE grid cell area: area of whole grid cell in km², based on a spherical earth (WGS84 mean radius), with valid data only in cells with valid HYDE land area data
 - Default data are based on the HYDE 3.2.000 described above, with data cells added to include Greenland and several arctic islands that have been added to the HYDE land area data set based on the SAGE land fraction and potential vegetation data (hyde_cell_plus.bil)
 - Alternatively, the user can select the original HYDE 3.2.000 grid cell area data (cell_area_hyde.bil)
- SAGE land fraction: land area fraction of grid cell, with valid data only in cells with valid SAGE potential vegetation or crop data (sage_land_frac.bil)
 - These fractions assume a spherical earth (WGS84 mean radius), so the corresponding grid cell area is calculated for the entire grid
 - SAGE land area is then calculated by multiplying the land fraction data with these calculated grid cell data
 - The SAGE land area data are used in conjunction with other data sets that utilize the SAGE data, such as the SAGE crop data, the MIRCA2000 data, and the water footprint data
 - These data are now included in the SAGE cropland datasets
- HYDE land area: available land area within grid cell in km², with valid data only in cells that are not completely ocean. As such, these data constitute the effective land mask for HYDE data. The HYDE land area data are the bases for land type area calculations and are considered as the working grid land area data
 - Default data are based on the HYDE 3.2.000 data described above, with data cells added to include Greenland and several arctic islands that have been added based on the SAGE land fraction and potential vegetation data (hyde_area_plus.bil). The original data do not include major glacier area in the available land area, so zero area land cells that are classified as “Polar Desert/Rock/Ice” and have non-zero land area in the SAGE potential vegetation and land fraction data sets are

set to the SAGE land area. Additionally, all Greenland cells have been set so that the land area equals the entire grid cell area, in order to capture the ice area (only 141828 km² of additional area is added by this method over the method of just adding entire grid cell area for the zero area land cells, so capturing the partial ice cells is worth potential overestimation of coastal ice-free land)

- Alternatively, the user can select the original HYDE 3.2.000 land area data (land_area_hyde.bil)
- GLU thematic map: integer codes representing the coverage of the output GLUs, with valid data only in cells determined to be part of a GLU. The GLUs can be arbitrarily defined by the user, and are enumerated starting at 1
 - Default data are the 235 water basins used by the GCAM water module (Global235_CLM_5arcmin.bil). There is a corresponding csv file that maps the integer values to basin names. A shapefile of these data is available in .../moirai/ancillary/shp/shp/ (once shp.zip is unzipped)
 - Alternatively, the user can select the original 18 AEZs (AEZ_orig_lds.gri), which have also been tested. There is a corresponding csv file that maps the integer values to AEZ names
 - Also, the user can select the ECHAM 2100 projected 18 AEZs (AEZ_echam_2100_lds.gri), which have not been evaluated in this version, but were used in a previous version of Moirai LDS. There is a corresponding csv file that maps the integer values to AEZ names
- Original AEZ thematic map: integer codes representing the coverage of the original AEZs, with valid data only in cells determined to be part of an AEZ
 - These data are the 18 original AEZs previously used by GCAM (AEZ_orig_lds.gri). These data are needed to remap the forest land rent to the new GLUs. There is a corresponding csv file that maps the integer values to AEZ names
 - Please cite these data when using Moirai: Lee, H.-L., Hhertel, T. W., Rose, S., Avetisyan, M. An integrated global land use data base for CGE analysis of climate policy options. Chapter 4, pp. 72-88, in Hertel, T. W., S. Rose and R. Tol (eds.) (2009). Economic Analysis of Land Use in Global Climate Change Policy. Abingdon: Routledge.
- SAGE potential vegetation: integer codes representing the potential vegetation circa 2000 if no land use change had occurred, with valid data only in cells that have been assigned to one of 15 classes
 - Default data have been updated to include Greenland, which is absent from the SAGE data sets (potveg_plus.bil). In Greenland, if the original HYDE land area equaled zero then the SAGE potential vegetation was set to "Polar Desert/Rock/Ice" (15), and if the original HYDE land area was greater than zero then the SAGE potential vegetation was set to "Tundra" (13). There is a corresponding csv file that maps the integer values to the potential vegetation type names

- Alternatively, the user can select the original SAGE potential vegetation file (potveg_thematic.bil), which does not include Greenland and uses the same csv mapping file as the default data
 - Please cite these data when using Moirai: Ramankutty, N. & Foley, J. A. 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochem. Cycles*, 13, 997-1027.
- Country thematic map: integer codes representing the country coverage, with valid data only in cells that have been assigned to a country (fao_etry_rast.bil)
 - The base for these data is the VAMPO spatial data set, which has been used by the FAO and is thus labeled with FAO country names. The FAO integer country codes from FAOSTAT have been assigned to the countries where possible, with some additional values for VMAP0 countries not listed in the FAOSTAT database. Furthermore, East Timor has been added based on a map, and the edges of the country raster data were “grown” to ensure better coverage of coastal land cells. See .../moirai/indata/FAO_iso_VMAP0_etry_readme.txt for details on this data set and how it maps to the tabular country data. A shapefile of these data is available in .../moirai/ancillary/countries/ (once countries.zip is unzipped)
- Protected area thematic map: integer codes representing whether a cell is protected (1) or not (255) (ProArea_0408_ENVI.img). Every cell contains one of these two values
 - These data were imported from the old GCAM “GIS” processing system, and the original source is the World Database on Protected Areas, although the version and date are unknown. They are used to segregate the land type area and the potential vegetation carbon outputs
- Nitrogen application rate: deprecated. rate of nitrogen application in kg/ha (Nfert_0083d.img)
 - These data were imported from the old GCAM “GIS” processing system, and the original source is Potter et al., 2010, but how they were aggregated to this file is unknown. Furthermore, the data were not processed correctly and subsequently not used by GCAM. The read and process functions are included in the source code, but they are not part of the build target and they are not called by the main program. This input line still exists, along with the code, in case one wants to figure out how to process the data correctly or substitute another data set. The user would have to hardcode the output file name or add it to the input file and corresponding data structure.
- SAGE physical cropland area, circa 2000: Physical cropland area circa 2000, as fraction of cell area (Cropland2000_5min.nc)
 - These data are used to normalize the SAGE individual crop harvested area values to each grid cell

- Please cite these data when using Moirai: Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochem. Cycles*, 22, GB1003.

CSV input data (filename without path)

- GTAP_value_milUSD.csv: original GTAP LU2.1 land rent data. Please cite these data when using Moirai: Lee, H.-L., Hertel, T. W., Rose, S., Avetisyan, M. An integrated global land use data base for CGE analysis of climate policy options. Chapter 4, pp. 72-88, in Hertel, T. W., S. Rose and R. Tol (eds.) (2009). *Economic Analysis of Land Use in Global Climate Change Policy*. Abingdon: Routledge.
- GTAP_GCAM_ctry87.csv: GTAP 87 country list, in GCAM order (alphabetical by iso)
- FAO_ctry_GCAM_ctry87.csv: FAO country to 87 country mapping list. This file determines which economic region each country is assigned to, and whether it is included in the outputs. See FAO_ctry_GCAM_ctry87_readme.txt for details
- FAO_iso_VMAP0_ctry.csv: FAO country to VMAP0 country and iso mapping list. The FAO_iso_VMAP0_ctry_readme.txt file contains info about how this country mapping list was developed. The integer codes correspond to the country thematic map as described above (fao_ctry_rast.bil)
- Global235_CLM_5arcmin.csv: the list of output GLU names mapped to the raster integer codes of Global235_CLM_5arcmin.bil
- Alternatively, the user can select the original 18 AEZs (AEZ_orig_lds.csv)
- Also, the user can select the ECHAM 2100 projected 18 AEZs (AEZ_echam_2100_lds.csv)
- iso_GCAM_regID_32reg.csv: iso countries to gcam region mapping list. This determines aggregation in the diagnostic output files, and must be consistent with the GCAM_region_names_nreg.csv file
- Alternatively, the user can select the iso_GCAM_regID_14reg.csv file
- GCAM_region_names_32reg.csv: GCAM region list (with integer codes). This determines aggregation in the diagnostic output files, and must be consistent with the iso_GCAM_regID_nreg.csv file
- Alternatively, the user can select the GCAM_region_names_14reg.csv file
- GTAP_use.csv: GTAP product use categories, with integer codes, and abbreviations
- SAGE_PVLT.csv: maps SAGE land cover type names for potential vegetation to integer codes in the SAGE potential vegetation raster file
- hyde32_lu.csv: assigns integer codes to HYDE 3.2.000 land use types
- isam_2_sage_hyde_mapping.csv: maps ISAM land cover types to SAGE land cover and HYDE land use
- SAGE_gtap_fao_crop2use.csv: maps the 175 sage crops to GTAP use sectors and FAO crops. The fourth column (gtap_crop_name) is the crop label used by

Moirai LDS and GCAM. See SAGE_gtap_fao_crop2use_readme.txt for additional details

- `FAO_production_1993_2016.csv`: FAO production data used in land rent calculations. This is also a source for recalibration to a different year. Format and years must match `FAO_ag_yield`, `HA`, and `prodprice` files. FAO Downloaded July 2018 from www.fao.org/faostat/
- `FAO_yield_1993_2016.csv`: Source of FAO yield data for diagnostics. Format and years must match `FAO_ag_prod`, `HA`, and `prodprice` files. Downloaded July 2018 from www.fao.org/faostat/
- `FAO_harvarea_1993_2016.csv`: Source of FAO harvested area data for recalibration to a different year. Format and years must match `FAO_ag_yield`, `prod`, and `prodprice` files. Downloaded July 2018 from www.fao.org/faostat/
- `FAO_producerprice_1993_2016.csv`: FAO price data used in land rent calculations. Format and years must match `FAO_ag_yield`, `HA`, and `prod` files. Currency year is equal to data year. Downloaded July 2018 from www.fao.org/faostat/
- `cpi_all_1970_2017_bls_june2018_annual.csv`: consumer price index table to convert from one US dollar year to another. Downloaded from <https://www.bls.gov/cpi/> June 2018
- `veg_carbon.csv`: vegetation (including root) carbon density (kg/m²) for each SAGE potential vegetation type, based on literature
- `soil_carbon.csv`: soil carbon density (kg/m²) for each SAGE potential vegetation type, based on literature

Output file names (filename without path)

- `moirai_log_basins235.txt`: Moirai LDS runtime log file (this should be case-specific)
- `MOIRAI_ag_HA_ha.csv`: the crop harvested area output file
- `MOIRAI_ag_Prod_t.csv`: the crop production output file
- `MOIRAI_value_milUSD.csv`: the land rent output file
- `MIRCA_irrHA_ha.csv`: the crop irrigated harvested area output file
- `MIRCA_rfdHA_ha.csv`: the crop rainfed harvested area output file
- `Land_type_area_ha.csv`: the historical land type area output file
- `Ref_veg_carbon_Mg_per_ha.csv`: the reference vegetation type carbon density output file
- `Water_footprint_m3.csv`: the crop water volume consumption output file
- `MOIRAI_etry_GLU.csv`: the country X GLU mapping output file
- `MOIRAI_land_types.csv`: the land type mapping output file

Diagnostics

There are four R scripts in `.../moirai/outputs` that generate a variety of figures. Make sure that `.../moirai/outputs` is the R working directory before running the scripts. Each script writes to a user-specified directory within the outputs directory. Set this diagnostic output directory within each script. Each script has a detailed

description at the beginning, and comments identifying the relevant directories, files, and flags that the user can change to customize the outputs. To diagnose country-level harvested area and production outputs, use "plot_moirai_crop_ctry.r." To diagnose land use region level harvested area and production outputs, use "plot_moirai_crop_rglu.r." Note that the difference statistics and the KS tests in "plot_moirai_crop_rglu.r" are invalid when the geographic land units are not identical among the different data sets. To diagnose the land type area output, use "plot_moirai_land_type_area.r." To diagnose the land rent output (at different aggregation levels), use "plot_moirai_landrent_glu.r". Note that the difference statistics and KS tests in "plot_moirai_landrent_aez.r" are invalid when the geographic land units are not identical among the different data sets.