GCAM boundary spatial products from moirai v3.1

Date- 14th April 2021

Summary- These data products present vector files for different representations of land area from the moirai land data system. Vector files are generated at 3 main spatial levels, namely country, region, basin. In addition to this, files are generated for different intersections for the 3 main categories, intersections for country and basin boundaries (country_basin), region and basin boundaries (region_basin) and region and country boundaries (region_country). Since the land data system does not generate land area information for all cells within the above mentioned boundaries (for water bodies for example), the vectors are presented for 3 main classes for each spatial category, land cells, cells with no land and combined. With all of the above mentioned combinations, the data products contain 18 different vector files.

Methodology- In generating these vector files, we used the land outputs from moirai as inputs along with separate inputs for the boundaries for the main spatial levels (country, basin and region). Combining the spatial boundaries with land inputs we generated 3 raster outputs (land, no land and combined) for each of the main spatial levels along with all the intersections. A unique key is assigned for each unique spatial boundary. We then converted these rasters to vectors through a process of polygonization where polygons were dissolved using the key and finally added all metadata (basin names, region names, country names) to each of the vector files. We check and correct geometry errors in the polygons themselves. We also added various validation tests to the code to account for completeness and accuracy.

CONTENTS:

gcam_boundaries_moirai_3p1_0p5arcmin_wgs84 folder contains the following,

spatial_input_files contain the following,

- *valid_land_area.bil:* raster file containing actual land area by grid cell globally.
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins (Generated by the moirai LDS)
- *country_out.bil:* raster file containing valid land cells matched to country codes.
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins (Generated by the moirai LDS)
- *country_out_noland.bil:* raster file containing non-land cells matched to country codes.
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins (Generated by the moirai LDS)
- *glu_raster.bil:* raster file containing valid land cells matched to GLU codes.
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins (Generated by the moirai LDS)

- *glu_raster_noland.bil:* raster file containing non-land cells matched to GLU codes.
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins (Generated by the moirai LDS)
- *region_gcam_out.bil:* raster file containing valid land cells matched to GCAM region codes.
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins (Generated by the moirai LDS)
- *region_gcam_out_noland.bil:* raster file containing non-land cells matched to GCAM region codes.
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins (Generated by the moirai LDS)
- *Global235_CLM_5arcmin.bil*: raster file containing basin boundaries for all cells output by moirai
  crs: EPSG:4326 WGS84 - World Geodetic System 1984
  resolution: 0.5 arc mins
- *GCAM_region_names.csv*- Mapping file with details on GCAM region names (Used to fill in metadata)
- *iso_GCAM_regID.csv*- Mapping file containing details on individual country names by iso code. (Used to fill in metadata)
- *basin_to_country_mapping.csv* – Mapping file containing details on basin names by country and region. (Used to fill in metadata)

main_outputs contain the following,

Contains files for each spatial boundary (country, region, country_basin etc), for each land category (land cells, no land and combined)

<spatial>_boundaries_moirai_<land_category>_3p1_0p5arcmin.shp

column names in outputs:

- *key***:** Unique identifier for feature
- *reg_id*: Unique identifier for region (region number)
- *ctry_id:* Unique identifier for country (country number)
- *glu_id:* Unique identifier for basin (basin number)
- *reg_nm*: Region name
- *ctry_nm:* Country name
- *glu_nm:* Basin name

List of files along with details

file name	spatial boundary	land category	number of unique features	metadata (column names)
basin_boundaries_moirai_combined_3p1_0p5arcmin.shp	GLUs	combined	232	key,glu_id, glu_nm
glu_boundaries_moirai_landcells_3p1_0p5arcmin.shp	GLUs	land cells	232	key,glu_id, glu_nm
glu_boundaries_moirai_noland_3p1_0p5arcmin.shp	GLUs	no land	232	key,glu_id, glu_nm
country_boundaries_moirai_combined_3p1_0p5arcmin.shp	Country	combined	201	key,ctry_id, ctry_nm
country_boundaries_moirai_landcells_3p1_0p5arcmin.shp	Country	land cells	201	key,ctry_id, ctry_nm
country_boundaries_moirai_noland_3p1_0p5arcmin.shp	Country	no land	201	key,ctry_id, ctry_nm
region_boundaries_moirai_combined_3p1_0p5arcmin.shp	GCAM regions	combined	32	key, reg_id, reg_nm
region_boundaries_moirai_landcells_3p1_0p5arcmin.shp	GCAM regions	land cells	32	key, reg_id, reg_nm
region_boundaries_moirai_no_land_3p1_0p5arcmin.shp	GCAM regions	no land	32	key, reg_id, reg_nm
ctry_glu_boundaries_moirai_combined_3p1_0p5arcmin.shp	country glu intersections	combined	686	key,ctry_id,glu_id,ctry_nm,glu_nm
ctry_glu_boundaries_moirai_landcells_3p1_0p5arcmin.shp	country glu intersections	land cells	686	key,ctry_id,glu_id,ctry_nm,glu_nm
ctry_glu_boundaries_moirai_noland_3p1_0p5arcmin.shp	country glu intersections	no land	686	key,ctry_id,glu_id,ctry_nm,glu_nm
reg_glu_boundaries_moirai_combined_3p1_0p5arcmin.shp	region glu intersections	combined	394	key,reg_id,glu_id,ctry_nm,reg_nm, glu_nm
reg_glu_boundaries_moirai_landcells_3p1_0p5arcmin.shp	region glu intersections	land cells	394	key,reg_id,glu_id,ctry_nm,reg_nm, glu_nm
reg_glu_boundaries_moirai_noland_3p1_0p5arcmin.shp	region glu intersections	no land	394	key,reg_id,glu_id,ctry_nm,reg_nm, glu_nm
reg_ctry_boundaries_moirai_combined_3p1_0p5arcmin.shp	country region intersections	combined	201	key,reg_id,ctry_id,ctry_nm,reg_nm, ctry_nm
reg_ctry_boundaries_moirai_landcells_3p1_0p5arcmin.shp	country region intersections	land cells	201	key,reg_id,ctry_id,ctry_nm,reg_nm, ctry_nm
reg_ctry_boundaries_moirai_noland_3p1_0p5arcmin.shp	country region intersections	no land	201	key,reg_id,ctry_id,ctry_nm,reg_nm, ctry_nm

Additional outputs generated by code (if required by user):

The code also generates the following outputs in case required by the user:

- raster files for each of the 18 files described above
- mapping files for each of the 6 spatial boundaries
- raw shape files i.e. shapefiles without any metadata (in case user would like to add their own metadata)
- diagnostics related to area calculated from the shape files.

Details for software used/required to reproduce outputs

1. Link to R code used to generate outputs – The code file *generate_moirai_spatial.R* was used to generate these shape files. The same is available here - https://github.com/JGCRI/moirai/blob/master/diagnostics/generate_moirai_spatial.R
- R version: 4.0.0
- Packages required (with version number): rgdal (1.4-8), raster (3.1.5), sp (1.4-1), sf (0.9-2), dplyr (0.8.5), data.table(1.12.8), nngeo(0.3.7) and smoothr (0.1.2)
2. Version of moirai used to generate land outputs – moirai v3.1 (https://github.com/JGCRI/moirai)
3. OSGeo4w for windows/mac (required for polygonization)- Latest version can be downloaded here - https://trac.osgeo.org/osgeo4w/

Validation

In order to validate our results, we compared the area calculated for each polygon in the shape files (calculated using *st_area*) with the actual area from moirai. Below is a scatterplot that presents this comparison at the glu level. The small differences we observe are related to pixels at coastal terrain for islands where moirai generates a smaller value of area per pixel (this happens because this terrain contains both land and water). For these locations the geometric area calculated from the shapefiles does not exactly match the actual area on account of the terrain.

Scatterplot showing area calculated from polygons and actual area for all basins