

Data and code to accompany

“Climate-driven limits to future carbon storage in California’s wildland ecosystems”

In review at *AGU Advances*, 2021

Corresponding author: Shane Coffield scoffiel@uci.edu

This archive contains input data, model output, Python scripts, and Google Earth Engine (GEE) scripts.

Python and GEE scripts can also be accessed directly via

- https://github.com/scoffiel/carbon_projections
- https://code.earthengine.google.com/?accept_repo=users/scoffiel/carbon_projections
- git clone https://earthengine.googlesource.com/users/scoffiel/carbon_projections

Data overview:

- **input_data**: contains all processed data needed to run models in Python. All were derived from public sources. Processed raster layers are 1/8-degree resolution in EPSG:4326 projection.
 - **Climate** – Bias-Corrected Spatial Downscaled (BCSD) CMIP5 data for 2006-2099 which has been compiled into 6 different netcdf files in Python scripts #1-2. For RCP4.5 and RCP8.5 scenarios, we generated a “climate_present” file for 2006-2099 average and “climate_future” file for 2090-2099 average with dimensions for 32 models (in order of most drying to wetting), 8 variables (4 seasons of T & P), and lat/lon. An additional “climate_present_10yrs” file maintains all 10 years of data needed for calculating interannual variability in the climate analogs approach (script #9). These climate data are the driver variables for all models.
 - Units: °C for mean daily temperature and mm/day for precipitation
 - Web source: https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html
 - Citation: Brekke et al., 2013; Maurer et al., 2007
 - **carbon_eighth.tif** – aboveground live wildland carbon for California for 2014. Rescaled from raw 30m data obtained from the California Air Resources Board (30m data available upon request from CARB). This is generated by GEE script #1 and is the target dataset for training RF regression models of carbon density in Python script #5.
 - Units: ton C/ha
 - Web source: <https://ww2.arb.ca.gov/nwl-inventory>
 - Citation: CARB 2018, Gonzalez et al., 2015

- **landcover_eighth.tif** – dominant land cover class generated from 30m National Land Cover Database (NLCD) for 2016 in GEE script #3. This is the target dataset for training RF classification models of vegetation type in Python script #8
 - Units: 1 for shrub/grass, 2 for forest
 - Web source: <https://www.mrlc.gov/data/nlcd-2016-land-cover-conus>
 - Citation: Homer et al., 2020
- **valid_fraction.tif** – fraction of each 1/8-degree pixel which is comprised of herbaceous, shrub, or forest landcover, also derived from the NLCD dataset. Generated in GEE script #2.
- **landcover_mask_eighth.tif** – Mask layer with “1” for all areas of the Western US that are at least 50% wildland cover, also derived from the NLCD dataset. Generated in GEE script #5
- **elev_eighth.tif** – Derived from 30m USGS elevation data in GEE script #3
 - Units: m
 - Web source: <https://lpdaac.usgs.gov/products/srtmgl1v003/>
 - Citation: NASA JPL 2013
- **offsets.zip** – shapefiles of 32 forest carbon offset project polygons in California, collected from <https://webmaps.arb.ca.gov/ARBOCIssuanceMap/> . Used in Python script #7 to assess vulnerability of these areas.
- **ecoregion_carbon_densities.tiff** – forest carbon density averaged by EPA Level III ecoregions using CARB AGL carbon layer; generated in GEE script #4 and used in Python script #8 to estimate carbon change associated with vegetation type conversions. Units: ton C/ha
- **cci_eighth.tif** – aboveground live carbon density for the western US and Mexico for 2017, derived from the ESA Climate Change Initiative global biomass dataset. Generated in GEE script #6 and used for climate analogs approach in Python script #9.
 - Units: ton biomass/ha (converted to carbon in Python)
 - Web source: <https://catalogue.ceda.ac.uk/uuid/bedc59f37c9545c981a839eb552e4084>
 - Citation: Santoro & Cartus, 2019
- **lemma_39spp_eighth.tif** – aboveground live carbon density for 39 species in California, compiled from 30m data from Oregon State for 2012 via GEE script #7 and used as target variables in species niche models (Python script #11). Each band is one species, ordered by most total biomass to least.
 - Units: ton biomass/ha (converted to carbon in Python)
 - Web source: <https://lemma.forestry.oregonstate.edu/data>
 - Citation: Kennedy et al., 2018
- **model_output:** contains subfolders corresponding to the four approaches discussed in the manuscript. For all approaches, we provide projections of carbon change (ton C/ha) for 6 scenarios: RCP4.5 & RCP8.5 x dry/mean/wet.
 - Random forest regression of carbon density

- Random forest classification of dominant vegetation type
- Climate analogs
- Random forest regression of 20 individual species' carbon density

Google Earth Engine code overview

1. Carbon_data: rescales 30m CARB carbon data layer (available upon request from CARB) to 1/8-degree to match the BCSD climate dataset, including masking out water/ag/urban landcover
2. Valid_land_fraction: calculates the fraction of sub-gridcell area that is allowed to support aboveground carbon (excludes water/ag/urban/barren cover)
3. Elevation: rescales 30m USGS elevation data to 1/8-degree to match the BCSD climate dataset
4. Landcover: rescales 30m NLCD land cover data to 1/8 degree (forest, shrub/grass, null)
5. Landcover_mask: creates a 1/8-degree layer masking out any areas of the western US that are not 50% wildland cover (for climate analogs analysis)
6. Cci_biomass: rescales 100m CCI biomass data to 1/8-degree for US and Mexico
7. Lemma_spp: reformats LEMMA species-level data into one raster layer with one band for each species' density at 1/8 degree

Python code overview

1. Process_climate.py: processes raw BCSD monthly climate data into combined netcdf files
2. Process_climate_10yrs.py: duplicate of script 1 to process raw BCSD climate data, but modified slightly to maintain all 10 years of data in the present. This is needed for calculating the interannual variability in the climate analogs approach.
3. Plot_clim_change.py: generates maps of mean annual T & P change for RCP4.5 & RCP8.5 (Fig 1)
4. Plot_clim_spread.py: generates maps of spread of precipitation across 32 models for RCP8.5 (FigS1) and dry vs. wet models averages for RCP8.5 (FigS2)
5. RF_carbon_model.py: approach #1. Models present-day distribution of CARB carbon layer based on climate data. Project future carbon and change
6. RF_model_spread.py: rebuilds RF regression models from script 5, for each of the 32 climate models. Compares 3 different runs: T & P both change, T only (P constant), and P only (T constant).
7. Offsets_change.py: compares RF regression model results from script 5 for all forests vs. carbon offset projects
8. RF_veg_class_model.py: approach #2. Models present-day distribution of NLCD forest-vs-shrub layer based on climate data. Projects future land cover, change, and associated carbon change
9. Climate_analogs.py. approach #3. Matches future climate pixels with their present analogue using Mahalanobis distance. Projects future carbon density by assigning that of

the present analog. 3 runs: full domain (25-49 lat and -125 - -100 lon), 500 km radius, 100 km radius

10. Analog_whittaker_plots.py: approach #3 supplementary figure - Whittaker scatter plots of mean annual P vs. T, showing how CA's gridcells shift
11. Species_models.py: approach #4. Fits RF regression models to each of the top 20 tree spp in California. Project future carbon and change. Applies restrictions on distance between spp present and future locations (migration scenarios).