Data and code to accompany
"Using remote sensing to quantify the additional climate benefits of California forest carbon offset projects"
In review at *Global Change Biology*, 2022

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_offsets
- https://code.earthengine.google.com/?accept_repo=users/scoffiel/california_offsets_final

Data overview:

- **input_data**: contains raw data needed to run Python and GEE scripts. Raster and vector data are also available directly as public assets on GEE.
    - **CalLands** – shapefile showing land ownership across California. Used to identify lands owned by Green Diamond Resource Company and Sierra Pacific Industries in GEE script #4 and Python script #7
        - Web source: https://callands.ucanr.edu/about.html
        - Citation: Macaulay & Butsic, 2017
    - **disturbance** – geotiffs of annual Landsat-derived disturbance data for California at 30m for 1985-2021. Values of "2" indicate harvest occurred for each pixel. Used to calculate the annual fraction of area harvested for projects, surroundings, and regions.
        - Citation: Wang et al., 2022
    - **lemma_species** – data extracted from LEMMA plot database with species composition for 2012. One file is a geotiff with IDs for each gridcell used with the second file, a csv lookup table providing columns for biomass breakdowns by tree species. Both files are used in GEE script #0 to build a raster with 39 bands representing the biomass of the top 39 species in California at 30m.
        - Units: kg biomass/ha
        - Web source: https://lemma.forestry.oregonstate.edu/data/plot-database
        - Citation: Bell et al., 2018
    - **offsets**
        - shapefile for 37 IFM compliance projects in California, collected from https://webmaps.arb.ca.gov/ARBOCIssuanceMap/

        - "all_projects" csv table with data we extracted from projects' documentation and reports, including yearly carbon stocks, owner information, baselines, and buffer pool allocations.
        - ARB offset credit issuance table as of January 2022, downloaded from https://ww2.arb.ca.gov/our-work/programs/compliance-offset-program/arb-offset-credit-issuance
    - **prism_climate** – geotiffs of 800m temperature and precipitation normals for 1990-2020. Used to identify spatial control groups in GEE script #2 and Python script #9
        - Units: deg C and mm/y
        - Web source: https://prism.oregonstate.edu/normals/
        - Citation: Daly et al., 2008
    - **public_lands** – shapefile indicating public lands in California. Used to filter out public lands in GEE script #1 and only compare projects against other private lands
        - Web source: California State Geoportal https://gis.data.ca.gov/datasets/f73858e200634ca888b19ca8c78e3aed_0/explore
    - **shapefiles** – miscellaneous geographic outlines used for mapping of Figures 2 & 3 in Python scripts
        - US State outlines from US Census https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html
        - California Air Resources Board supersections https://ww2.arb.ca.gov/our-work/programs/compliance-offset-program/compliance-offset-protocols/us-forest-projects/2015
    - **Wilson_site_class** – geodatabase of "site class", a metric of forest productivity at 100m determined by USFS FIA for California forests. Used in GEE script #2 and Python script #9 as a covariate for matching projects to spatial control areas, along with PRISM climate normals.
        - More information: https://www.srs.fs.usda.gov/pubs/rn/rn_srs025.pdf
        - Citation: Obtained from B. Wilson at USFS, cited in Tubbesing et al., 2020
    - eMapR and LEMMA 30m biomass layers for 1986-2017 were available by special request from their original sources (below). We have also provided them as GEE assets
        - eMapR: http://emapr.ceoas.oregonstate.edu/getData.html
            - https://code.earthengine.google.com/?asset=projects/ca-ecs/eMapR/eMapR_biomass_CA_ARD_all
            - Units: ton biomass/ha

        - Citation: Kennedy et al., 2018
    - LEMMA https://lemma.forestry.oregonstate.edu/projects/ca-biomass
        - https://code.earthengine.google.com/?asset=projects/ca-ecs/LEMMA
        - Units: kg biomass/ha
        - Citation: Bell et al., 2018
- **processed_data**: contains outputs from GEE scripts, with subfolders for 6 variables extracted for different regions of interest (in "shapefile" subfolder). The regions of interest include projects, their surroundings, the coast and interior regions, and properties owned by Sierra Pacific Industries (SPI) and Green Diamond Resource Company (GD).
    - carbon_emapr
    - carbon_lemma
    - harvest
    - prism_climate
    - site_class
    - species

Google Earth Engine code overview

Pre-processing and extracting of geospatial data prior to statistical analysis in Python

- 0_species_data: Convert LEMMA plot database to geotiff asset with bands for biomass by 39 tree species at 30m for California
- 1_process_extract_data: Extract variables (carbon, harvest, species) for various regions of interest (projects, surroundings, regions)
- 2_process_extract_data_800m: Rescale all data (carbon, harvest, species, climate, site class) to 800m resolution matching the PRISM climate scale in order to match projects to their most similar pixels as control groups in Python script #9
- 3_verification_example: Extract 30m raster data of annual carbon and harvest for an example project, CAR1066, used in Python script #8 to generate Supporting Information Fig S2.
- 4_SPI_GD_properties: Extract carbon, harvest, and species-level data for Sierra Pacific Industries (SPI) and Green Diamond Resource Company (GD) offset holdings versus other landholdings. Processed data are used in Python script #7 for these case studies shown in Fig 7.

Python code overview

1. study_area: Plot study area with offset projects and supersections in California (Fig 2)
2. data_comparisons: Compare carbon stocks and accumulation rates from reports, eMapR, and LEMMA. Generate Fig 3 with timeseries for 9 projects, Fig S1 validation plots, and Table S1 with full report/eMapR/LEMMA data by project

3. payoff_times: Compare 30 projects' reported accumulation rates, baselines, and credits issued. Generate Fig S3 illustrating trade-off between high accumulation rate and high initial stocking above baseline.
4. carbon_harvest_timeseries: Generate Fig 4 with spatial comparisons framework, showing timeseries of carbon and harvest for projects, surroundings, and regions.
5. beforeafter_barcharts: Generate Fig 5 bar charts contrasting before-and-after carbon accumulation and harvest for projects vs. surroundings.
6. spps_comparison: Generate Fig 6 with tanoak/redwood species comparison and pie chart for Northern Coast projects.
7. SPI_GD_casestudy: Generate Fig 7 with harvest and species for SPI and GD offsets vs. other landholdings.
8. verification_example: Generate Fig S2 showing harvest, eMapR, and LEMMA maps + timeseries for example project CAR1066
9. mahalanobis_matching: Provide supplemental analysis matching projects to "similar forests" as defined by minimized Mahalanobis distance among climate and productivity. Generates Fig S4-S6 timeseries and bar charts comparing projects against original "surroundings" control group and these new "matched controls".