

Data and code to accompany

“Using remote sensing to quantify the additional climate benefits of California forest carbon offset projects.”
Global Change Biology, 2022

Corresponding author contact: 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. **In several cases due to incompatible licensing we can only provide the link to the original data source**
 - **CalLands** – source of geospatial land ownership data in California, including by Green Diamond Resource Company and Sierra Pacific Industries. CalLands provides R data files which can be exported as a shapefile from R and then filtered based on landowner name. Used 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
 - Also available at <https://doi.org/10.7910/DVN/CVTNLY>
 - **lemma_species** – source of plot database with species composition for 2012. We used the geotiff FID file which provides IDs for each gridcell into the lookup table in the Access database. Final gridded species composition data can be shared upon request.
 - 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** – source 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 can also share them as GEE assets upon request
 - 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 “shapefiles” 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
 - files
 - coast_emapr.csv: yearly eMapR aboveground biomass data for the California coast region
 - coast_projects_emapr.csv: yearly eMapR aboveground biomass data for combined area of 33 coastal projects
 - coast_surroundings_emapr.csv: yearly eMapR aboveground biomass data for combined area of 33 coastal projects’ 2km surrounding rings
 - emapr_800m.tiff: yearly eMapR aboveground biomass data with 32 bands for years 1986-2017 at 800m resolution, output from GEE

- emapr_car1066.tiff: yearly eMapR aboveground biomass data with 32 bands for years 1986-2017 at 30m resolution for one sample project, CAR1066, output from GEE
- interior_emapr.csv: yearly eMapR aboveground biomass data for the interior region
- interior_projects_emapr.csv: yearly eMapR aboveground biomass data for the combined 4 interior projects' area
- interior_surroundings_emapr.csv: yearly eMapR aboveground biomass data for the combined 4 interior projects' surrounding 2km ring areas
- norcal_emapr.csv: yearly eMapR aboveground biomass data for the entire domain of northern California
- projects_emapr.csv: yearly eMapR aboveground biomass (columns) for the 37 IFM projects (rows). Also contains non-essential columns for project status ("ARBstatus"), area in m², and source of shapefile
- surroundings_emapr.csv: Same as previous, for the 37 IFM projects' 2km surrounding rings (columns)
- variables:
 - For CSV files: "b3"- "b34" representing aboveground live biomass for years 1986-2017, inclusive
 - For tiff files: 32 bands representing aboveground live biomass for years 1986-2017, inclusive
- units
 - ton biomass/ha
- carbon_lemma
 - files: see details above, but with LEMMA source of biomass data instead of eMapR
 - variables:
 - For CSV files: "1986_b1"- "2017_b1" representing aboveground live biomass for years 1986-2017, inclusive
 - For tiff files: 32 bands representing aboveground live biomass for years 1986-2017, inclusive
 - units
 - kg biomass/ha
- harvest
 - files: see details above for carbon_emapr, but with "harvest" as the variable of interest instead of eMapR biomass. Harvest is measured as described in the manuscript Methods and sourced from Wang et al, 2022. Files also exist for 4 additional regions of interest:
 - gd_harvest.csv: All combined non-offset project areas owned by Green Diamond Resource Company (GD) in California
 - gd_projects_harvest.csv: All combined IFM project areas owned by GD in California
 - spi_harvest.csv: All combined non-offset project areas owned by Sierra Pacific Industries (SPI) in California
 - spi_projects_harvest.csv: All combined IFM project areas owned by SPI in California

- variables:
 - For CSV files: “b3”-“b34” representing annual harvest for years 1986-2017, inclusive
 - For tiff files: 32 bands representing annual harvest for years 1986-2017, inclusive
- units
 - fraction of area harvested per year (multiply by 100 for percentage)
- shapefiles (also see manuscript “Methods” for full details)
 - all_projects: 37 features for the 37 IFM projects in California
 - coast: the northern California coast region
 - gd: non-project areas of the Green Diamond Resource Company lands
 - gd_projects: IFM project areas of the Green Diamond Resource Company lands
 - interior: the interior California region
 - norcal: the entire study domain of Northern California
 - spi: non-project areas of the Sierra Pacific Industries lands
 - spi_projects: IFM project (or proposed project) areas of the Sierra Pacific Industries lands
- site_class
 - files
 - projects_site_class.csv: “mean” column indicates each project’s (37 rows) mean site class, averaged over contained 100-m pixels
 - site_class_800m_nonproject.tiff: 800m raster of site class for non-project areas of Northern California
 - surround_site_class.csv: “mean” column indicates each project’s 2km surrounding ring area mean site class
 - variables: “site class” as defined by the USFS, a metric for productivity https://www.srs.fs.usda.gov/pubs/rn/rn_srs025.pdf
 - units: site class codes ranging 1-7, with 1 indicating highest productivity
- species
 - files: for the same regions described in previous subfolders, now with columns for biomass density by different tree species for year 2012.
 - variables: LEMMA aboveground biomass density for different tree species as described in column labels. The order of the bands in the tiff file matches the alphabetized order of the species in the CSV files’ columns.
 - units: kg biomass / ha

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 6.

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 7 bar charts contrasting before-and-after carbon accumulation and harvest for projects vs. surroundings.
6. spp_comparison: Generate Fig 5 with tanoak/redwood species comparison and pie chart for Northern Coast projects.
7. SPI_GD_casestudy: Generate Fig 6 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".