Climate change and ecosystem carbon in California

Description:
This is a fairly small dataset of 2258 points representing 2258 latitude/longitude gridcells in California (a 1/8 degree or 12.5km spatial resolution). The data are tabular (CSV) format, with each row representing one gridcell on the map. There are columns for latitude; longitude; 8 climate variables: 4 seasons of mean daily temperature (deg C) and 4 seasons of mean daily precipitation (mm/day); aboveground live carbon density (ton carbon per hectare); dominant vegetation type (1 for shrub or grass, 2 for forest). There are also 6 files with 6 different future climate scenarios: extreme RCP8.5 warming vs moderate RCP4.5 warming and either dry, mean, or wet moisture scenarios. For more details and links to the original data, see below.

Understanding how ecosystems are connected to climate and might respond to climate change is a major topic of research in Earth System Science, especially because plants can help mitigate climate change. During photosynthesis, plants take carbon out of the atmosphere, and some of that carbon ends up stored in their aboveground biomass (tree trunks are about 50% carbon) or in the soil. This makes them a "natural climate solution", so long as the ecosystems are stable and that carbon is withheld from the atmosphere for a long time.

This dataset could be used (1) to help understand the climatic controls on ecosystems - that is, how seasonal patterns of temperature and precipitation help determine how much vegetation can exist at a location and how much carbon can be stored - and (2) to make future projections of how much carbon we would expect to be stored at different locations under different climate scenarios. As Earth System Scientists we want to know whether it's realistic to expect increased carbon uptake by the land over this century to help mitigate climate change. We also want to know whether changes in temperature or precipitation are going to be more critical for determining future carbon storage, and to quantify uncertainty of any projections. With this dataset we could gain general scientific insight about the relationships between temperature, precipitation, and carbon, and identify patterns of vulnerability. In other words, be wary of overfitting models, and strive for interpretability.

Your task:

- Fit models to explain 2 different dependent variables (carbon density and vegetation type at a given location) as a function of 8 climate variables (4 seasons of temperature and precipitation)
- Evaluate models, explore error structure
- Apply models to make projections of future carbon and vegetation under 6 different future climate scenarios

Sources of original data:

Climate: Bias-Corrected Spatially Downscaled CMIP5 climate models

- https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html

- Monthly data 2006-2099 at 1/8-degree spatial resolution (approx. 12.5km)
- I have aggregated to “present” 2006-2015 averaged and aggregated by seasons and “future” 2090-2099
- Present-day data are in the “present_data.csv” file and future data are in the 6 different files for 6 different climate scenarios
- Units: degree C for temperature, mm/day for precipitation

Carbon: California Air Resources Board natural & working lands inventory

- Aboveground live carbon density
- See technical supporting documents here: https://ww2.arb.ca.gov/nwl-inventory
- See related academic paper here: https://www.sciencedirect.com/science/article/pii/S0378112715001796
- 30m data for 2014, covering all wildlands of California (grass, shrub, forested lands and excluding agricultural or urban areas)
- Units: tons per hectare (ton C/ha)
- I have spatially aggregated to 1/8 degree resolution to match the climate data and put it all together in CSV files

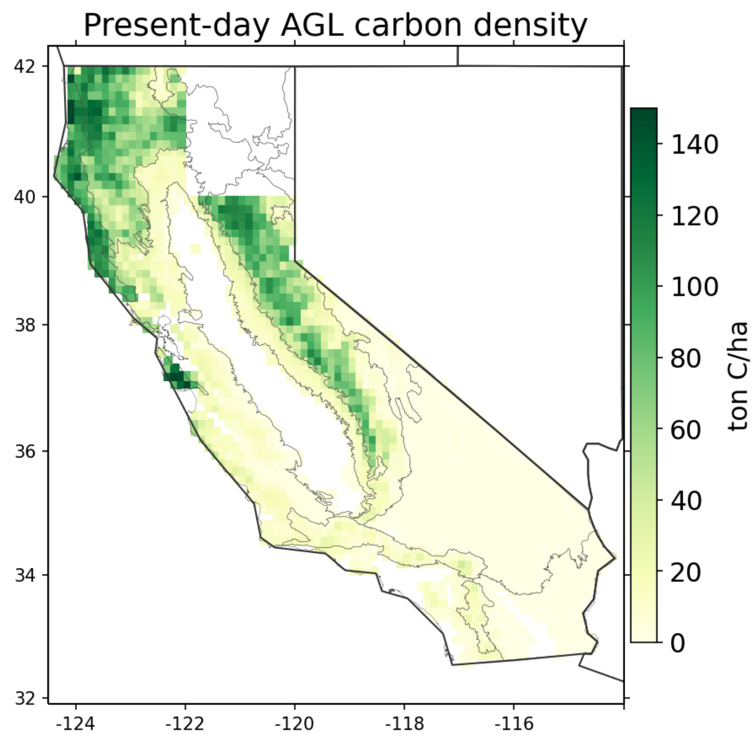
Vegetation type: National Land Cover Database

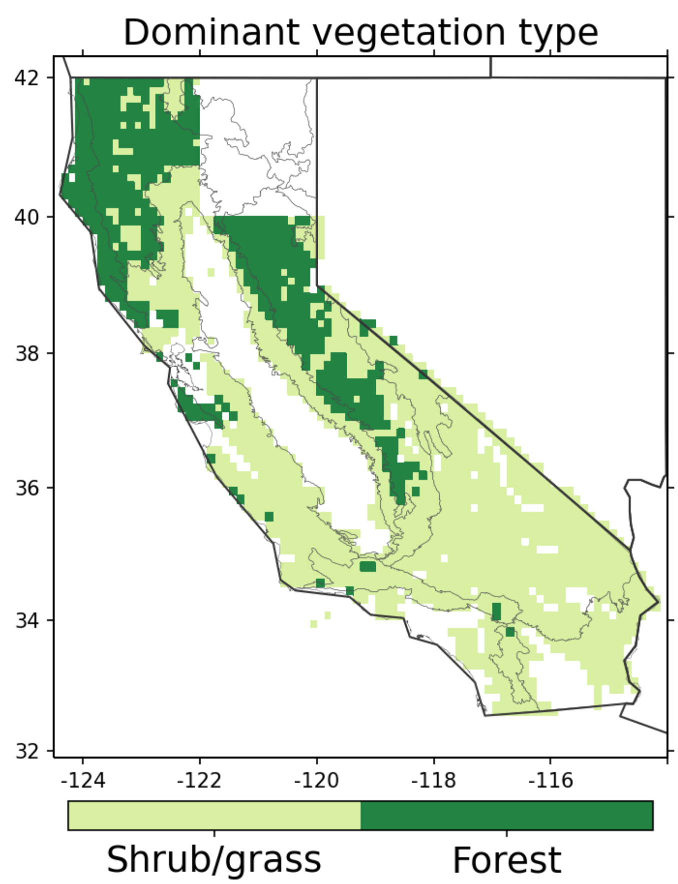
- https://www.mrlc.gov/data/nlcd-2016-land-cover-conus
- 30m data for 2016
- I have spatially aggregated to 1/8 degree like the other variables and added “veg” as a column
- Units: 1 for shrub or grass, 2 for forest
- Some rows are missing data in this column. That’s mostly where there is very low carbon and no vegetation. Don’t worry about those pixels if you try to model vegetation.

Note: I have replaced about 11% of rows with “NaN” in the carbon and vegetation column, for a chunk of the northeastern part of the state (see figure below). These are intended to be a hold-out data set that I will reveal to you later to test your models on.

Please reach out on Slack with questions! I can definitely talk more about the science or data processing if you are interested. Most of my programming experience is in Python, using packages like numpy, pandas, and scikit-learn. Good luck!

Shane





Lines represent EPA ecoregions