READ.ME file for code associated with: Forest carbon storage in the Western United States: distribution, drivers, and trends

This code is comprised of multiple R scripts that, together, perform all analyses reported in the associated publication.

*MainScript\_240227\_clean* is the master analysis script from which all other scripts are ordered and called as they relate to the analysis. Refer to this script first.

Auxillary script descriptions, ordered by organizational folder

**Section1\_SourceData**

This folder contains scripts to identify and export the project study area and read in relevant subsequent layers for further analysis.

*identify\_studyarea* – Process publicly available geospatial layers to identify the study area of the project

*shapefile\_readin* – Call processed layers for use in MainScript\_240227\_clean and all additional scripts

**Section2\_ProcessData**

This folder contains scripts to read in and process to FIA forest inventory data to plot-level carbon estimates, export data, and estimate recent trends in carbon by ecoregion using linear models.

*ReadFIA\_CalculateCarbon\_LiveDead* – Read in raw data tables from FIA forest inventory data, then estimate forest carbon in live and dead pools and export for later use.

*CTrends.\_ShapeSelection\_lm\_livedead* – Determine trends in carbon within live and dead pools for each ecoregion using linear models. Generate a table of trends to be visualized and exported in the main script.

**Section3\_Drivers**

This folder contains scripts to read in and process all driver data from publicly available datasets. This folder also contains a script to create maps of drivers used in the analysis as well as to select drivers to be included in subsequent random forest models.

Fire

*CBI\_prepare\_fires* – Prepare composite burn index (CBI) files exported from Google Earth Engine using the script and workflow developed by Parks et al. (2019).

*Drivers\_RS\_CBI\_fireseverity\_240214* and *Drivers\_WUMI\_wildfire\_updated\_230307* – Process datasets to restrict data to the project study area and to summarize values within 4km grid cells used in this study. Scripts process files from the CBI analysis (Parks et al. 2019) and the Western US MTBS-Interagency wildfire dataset (Juang et al., 2022), respectively.

Harvest

*Drivers\_NAFD\_ForestDisturbance* – Process forest harvest area and year up to 2000 within 4km grid cells from the North American Forest Dynamics Forest Loss Attribution product (Schleeweis et al., 2020).

*Harvest\_FIA* – Determine harvest year after 2000 within selected FIA plots using FIA data tables.

Insects and Disease

*BeetleDisease\_Processing* – Process area affected by insects and disease within 4km grid cells using the National Insect and Disease Aerial Detection Survey Summary Maps (Johnson & Wittwer, 2006; USDA Forest Service, 2024).

General driver processing methods

*ProcessDrivers*, *ProcessAllDrivers\_230202* – supplementary code not used in actual analysis to quickly summarize and compare methods taken to process drivers from original resolution into 4-km grid cells. No longer used in the main script.

Other scripts

*drivermaps* – Create maps of drivers used in the analysis which are displayed in the supplementary information in the original publication.

*select\_variables\_240227* – For each ecoregion and carbon pool, statistically select drivers by choosing drivers that are not correlated to one another. When pairwise correlation is found between drivers, drivers with the highest correlation to carbon are kept.

**Section4\_Models**

This folder contains scripts to run random forest models. Using results from random forest models, additional scripts calculate and plot the partial dependence relationships for a selected number of drivers.

*RandomForest\_function\_240227*, *RandomForest\_function\_240227\_dead* – Run random forest models for each ecoregion for live and dead carbon pools, respectively.

*CMYK\_legend* – create a color legend representing relative driver important to be used in the main script.

*PD\_CalcFunctions\_240227*, *PD\_CalcFunctions\_240227\_dead* – calculate partial dependence relationships for drivers within the top 4 most importance carbon drivers for each ecoregion. Scripts process live and dead carbon pools, respectively.

*PD\_Plots\_240227, PD\_Plots\_240227\_dead* – Create figures displaying partial dependence relationships. Figures are shown in the main text.

**reviewercomments**

This folder contains a script, *disturbancecomparison\_fia\_externaldatasets*, that performs a supplementary analysis recommended by reviewers. This analysis compares FIA estimates of disturbance and harvest to those taken from external datasets which were then processed to 4km grid cell values.

**ChecksBalances**

This folder contains a script, *Map\_FIALocation\_year*, that maps the reported coordinates of surveyed FIA plots annually.