Reforestation Opportunities Analysis Steps

**Short summary:**

*Methods:* See [here](https://www.sciencedirect.com/science/article/pii/S2590332220306035) for the original publication where these methods are described in detail, with a description on steps taken for the updated (June 2022, V2) analysis below. Please refer to Susan Cook-Patton ([susan.cook-patton@tnc.org](mailto:susan.cook-patton@tnc.org)) or Joe Fargione ([jfargione@tnc.org](mailto:jfargione@tnc.org)) if you have questions on methods.

*Scripts:* Scripts are available on <https://github.com/thenatureconservancy/ReforestationHub>

*Running the analysis:* Everything needed to run the analysis, including data inputs and scripts, is in a zipped folder. The only thing that needs to be done to rerun the analysis is to change the working directory that is specified in each of the R scripts (this is set using the function setwd()). The R scripts will need to be run in increasing numeric order, with the exception of the scripts that have a name beginning with “0.0” (these scripts only contain functions. To run the analysis, this means you need to first run the scripts with a name starting “0.1”, then the scripts with a name starting “1.x” in increasing order (1.0, then 1.1, then 1.2), then the script with a name that starts “2.0”, and finally the script that starts “3.0”.

*Modifying the reforestation opportunity:* The main scripts (the ones with name that start “1.x”) has a few places where assumptions used in the analysis can be modified (e.g. to include different types of conservation corridors, different flood plain durations, different assumptions on sequestration duration, etc). These all appear in the first 200 lines of the file, and are heavily commented. More information on how to change assumptions in the analysis are below.

*Updating the reforestation opportunity with new data sources:* Links to all data sources used in the analysis are below. The folder containing the R scripts that run the analysis also contain these links, and the R scripts should run on newer versions of the datasets when they become available.

**(Quick) methods:**

*Identifying reforestation opportunities:*

1. Potential forest, as per BPS with cross walk from Susan Cook-Patton. Also potential Tamalpan forest in the Rio Grande valley, as identified by TNC partners.
2. Not current forest, as per NAFD
3. Not excluded lands (barren, ice, urban built up areas), as per NLCD
4. Not within a 2x2 cell buffer of Class 1 or 2 roads, as per TIGER
5. Not in a wilderness area, as per US PAD
6. Not in cropland of good soil qualities (1-3), as per gSSURGO

*Estimating sequestration rates:*

Sequestration rates by forest type are from the USFS and available [here](https://usfs-public.app.box.com/v/NRS-GTR-202-Supplemental/file/765151685817). The default is to estimate annual benefits under a 10-year sequestration rate, using the above-ground biomass estimates from the USFS and a soil organic carbon estimate of 0.23 tonnes C per hectare per year. This default can be modified to estimate potential benefits under (a) different time horizons (10, 30, 50, and 100 year estimates are available), and/or (b) using different potential carbon pools (e.g. only in living biomass, living or dead biomass, etc).

*Opportunity classes:*

There are 9 main ones, which are described in depth in the original analysis. In short, these are:

1. Shrubland
2. Protected lands
3. Post burn landscapes
4. Croplands with poor quality soil
5. By land manager
6. Urban open space
7. One in five pluvial or fluvial floods
8. Riparian buffer
9. Conservation corridors

These are further divided by land manager (as from US PAD), and into current land cover (as from NLCD).

**Running the analysis:**

Re-running the analysis will require R and GIS (or an equivalent).

All of the scripts are in R. To rerun these scripts you need will also need to change the working directory that is set in each of the R scripts, so that it corresponds with the location of the reforestation hub redo folder. A quick description of the scripts is below:

* *0.0\_Reforestation\_Functions*: This contains the functions that are used in the other scripts. This should not need to be modified.
* *0.1 PAD Manipulation*: This creates a numeric ID for all of the US PAD shape objects, which is later called in the 1.0 files. This can also be done in GIS or an equivalent.
* *0.1\_Flood\_Tiles*: This creates tiles that contain locations identifies cells as a fluvial or pluvial flood risk (value of 9999) or not a flood risk (other value).
* *1.0\_Reforestation\_Deductions*: This is the main script that runs the analysis.
* *1.1\_Reforestation\_Deductions*: This double-checks whether all parts of the CONUS have properly run in the 1.0 script, and then run the analysis for these parts of CONUS as needed.
* *1.2\_Reforestation\_Deductions*: This is a second double-check whether all parts of the CONUS have properly run in the 1.0 and 1.1 scripts, and then run the analysis for these parts of CONUS as needed.
* *2.0\_Mosaicing\_Rasters*: This mosaics individual raster tiles to create a CONUS map.
* *3.0\_Manage\_CONUS\_CSVs*: This manages the CSV file outputs to create the county-level and CONUS-level reforestation results.

The only part of the analysis that requires GIS (or an equivalent) is to rasterize the vector shape files (e.g. the US PAD database, the roads database, and so on). This can also be done in R, but R is terribly inefficient at rasterizing vector data sets, both in terms of time and processing power required. Rasterizing the shape files needs to be done before running the R scripts. When rasterizing the data, please follow the below:

1. For roads and waterways, use any burn-in value that is >1
2. For the US PAD, use a burn-in value that is equal to the object id.
3. For BPS, use a burn-in value equal to the BPS “VALUE” (this ranges between 0 and ~3500)
4. For soil characteristics, exclude “good quality soils”, and then any burn-in value that is not negative
5. For counties, use a burn-in value equivalent to the county id number

The analysis will work if you use a different burn in value, but you’ll need to create a different cross walk between the values you rasterize and what objects they correspond with in the shapefiles.

To run the analysis after rasterizing the shapefiles:

1. Run the R scripts with a name starting “0.1”. One of these creates raster tiles of of the one in five year flood locations (or one in ten year or one in twenty year flood locations if these are instead specified), whilst the other creates numeric object ID values for the US PAD database (this can also be done in GIS if easier).
2. Run the R script named “1.0 Reforestation Analysis…”, then “1.1”, and then “1.2”. The Script starting “1.0” is the main script, with the “1.1” and “1.2” scripts being redundant to make sure the analysis has fully run (there are a few additional sense checks in these)
3. Run the R script named “2.0 …”
4. Run the R script named “3.0…”.

The analysis runs on individual tiles of the continental US (CONUS). I have done this to reduce the memory required to run the analysis. I’ve defaulted to 400 tiles, but this can be changed by changing L369. If you want to use a different number of tiles in the analysis, then you will need: (1) change L369 in the “1.0” script, and (2) change the number of tiles specified in L28 of the “0.1” script that creates the flood tile to correspond with the number of tiles you’ve specified. The resultant tiles from the “0.1” and “1.0” scripts are indexed by tile number, which means that you will run into an error if the number of tiles in the “1.0” and the “0.1” scripts are different.

Note that (1) the function that splits CONUS into multiple tiles is based on the square of the number specified in L369 of the “1.0” script (e.g. using a value of 20 results in 400 tiles, using a value of 10 would result in 100 tiles), and (2) splitting CONUS into 400 tiles requires ~15GB of RAM.

You can run the analysis in parallel by either making multiple copies of the script, and having each copy of the script loop over a different subset of the tiles. This subset of tiles is specified on L421 of the “1.0” script. Once this is done, the scripts can be run in parallel by using multiple R terminals, or a batch script that has multiple R scripts run simultaneously.

Note that if you change the number of tiles in the CONUS map, you will also need to update the number of tiles specified in the 1.1 and 1.2 scripts.

**Modifying the reforestation opportunities:**

The reforestation hub redo uses the same default assumptions and data inputs as the [initial analysis](https://usfs-public.app.box.com/v/NRS-GTR-202-Supplemental/file/765151685817), with the exception of (a) using newer datasets where available, (b) using a different dataset on conservation corridors, and (c) including specified regions of the Rio Grande river valley as potential forest opportunities.

There is capacity to change the underlying assumptions of the analysis. All of these changes require modifying the 1.x scripts. Alternatively, sensitivity around the results can be assessed by using different data sets as inputs into the reforestation hub analysis.

The below list includes information on how the reforestation hub analysis can be changed, with the line numbers used matching those in the “1.0” script:

***Sequestration rates:***

*Default:* To report annual sequestration, assuming a time horizon of 10 years, using all non-soil carbon pools.

*Options to Modify:*

(1) To change the time horizon--change L63 to 30, 50, or 100)

(2) To report total benefits over the total time horizon, rather than annual estimates--change L66 to “No”)

(3) To change the carbon pools incorporated in the analysis—change the values in the data frame created in L34-42. Carbon pools with a value of TRUE are included in the analysis, carbon pools with a value of FALSE are not.

***Wilderness areas:***

*Default:* To exclude wilderness areas from the analysis. These are designated as “WSA”, “WA”, or “SA” in the US PAD database.

*Options to Modify:* You will need to do three things: (1) L59 species which US PAD designation types are excluded from being counted as reforestation opportunities. Change this list to include or exclude different designation types. (2) Use the US PAD shapefile to create a binary raster, where a value of 1 specifies wilderness areas (or other designations to exclude), and 0s (or NAs) specifies other designation types. (3) Specify the file location of the new raster created above by changing L338.

***Conservation corridors:***

*Default:* To include linkage/concentrated flow areas as an opportunity class.

*Options to Modify:* Change the values in the data frame created at L46 to TRUE to include conservation corridors as a reforestation opportunity, or to FALSE to exclude them as a reforestation opportunity.

***Floodplains:***

*Default:* To include one-in-five year fluvial floodplains as a reforestation opportunity.

*Options to Modify:* Rerun the “0.1” script for the floodplain tiles, but using the one-in-ten year or one-in-twenty year floodplain data. Note that this will overwrite flood tiles that already exist (I’ve already run the 0.1 script for the one-in-five year flood events), so I recommend saving the already managed flood tiles into a new folder for later use.

***Including additional reforestation areas:***

*Default:* The reforestation hub redo includes certain parts of the Rio Grande valley (Tamalpan forest) as reforestation opportunities. These regions are not identified as potential forests by BPS, and are instead regions specified by TNC partners.

*Options to Modify:* The two easiest approaches to include other regions as potential reforestation opportunities are listed below:

Approach 1:

1. Get shapefiles of potential reforestation opportunities
2. Convert this into a raster, either in GIS or in R, where a value of anything but 0 or NA indicates a potential reforestation opportunity.
3. Copy the code used to incorporate the Tamalpan forest opportunities, but modifying it to use the other rasters created above. The code used to incorporate the Tamalpan forest opportunities is in L429-435 (this imports the raster if it isn’t already in R), L453-472 (this checks whether the CONUS tile overlaps with the Tamalpan reforestation opportunity, and if it does, it modifies the Tamalpan data to have the same extent/projection as the CONUS tile, and then identifies the Tamalpan forest as a reforestation opportunity), and L727-730 (this specifies the sequestration rates for the Tamalpan forest)

Approach 2:

1. Get shapefiles of potential reforestation opportunities
2. Merge the multiple smaller shapefiles into a single larger one
3. Convert this into a raster, either in GIS or in R, where a value of anything but 0 or NA indicates a potential reforestation opportunity.
4. Repurpose the code specified used to specify the Tamalpan forest opportunities.

Approach 1 will likely be easiest if incorporating one or two additional reforestation opportunities. Approach 2 will likely be easiest if incorporating many other reforestation opportunities. Note that for both of these approaches, the sequestration rate of these reforestation opportunities will need to be specified (as in L727-730 for the Tamalpan forest).

**Outputs:**

There two primary outputs from the reforestation hub redo:

***CONUS rasters:***

There is a different CONUS raster for (a) each reforestation opportunity, and (b) each land cover class. These rasters are provided both before and after the semi-spatial deductions are applied.

The data in these rasters is indicated as the annual reforestation opportunity, measured in tonnes C per hectare.

Some of these rasters are continuous, and contain the estimated sequestration per cell, with the value representing tonnes C (this is not CO2) per hectare per year.

Other rasters will contain a single numeric value. These rasters indicate the location of certain opportunity classes (mainly opportunity by land manager or by current land cover). The value in the raster isn’t important, as it is only used to differentiate cells that are not in the opportunity class vs cells that are.

***CSV files***

There are four .csv files. Two include county-by-county results, one of which contains results before the semi-spatial deductions and the other which contains results after the deductions are applied. Likewise, there are two CONUS level .csvs, one of which contains results before the deducations are applied, and the other which contains results after the deductions are applied.

**Data Inputs:**

Protected Areas (land managers, and wilderness areas): US Protected Area Database <https://www.usgs.gov/programs/gap-analysis-project/science/pad-us-data-download>

Biophysical potential: US LandFire; <https://landfire.gov/viewer/viewer.html> (biophysical data is available through the menu at the right, LF -> vegetation -> biophysical settings; this analysis uses 2016 data because 2019 data was not available)

Flood: FATHOM; obtained from TNC

Current Land Cover: US National Land Cover Data <https://www.mrlc.gov/data?f%5B0%5D=category%3Aland%20cover> (2019 data used; note this is not used for current forest cover)

Current Forest Cover: North American Forest Dynamics; <https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1290>

Roads: From US Census; <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>

Cropland soil classification: gSSURGO

Carbon sequestration rates: From US Forest Service; <https://www.fs.fed.us/nrs/pubs/gtr/gtr_nrs202.pdf>; specifically the tables in Appendix A

Waterways: US National Hydrography Dataset; <https://www.usgs.gov/national-hydrography/national-hydrography-dataset>

Habitat Corridors: From TNC; Resilience Lands; Connectivity and Climate Flow Categorical; <https://maps.tnc.org/resilientland/>

County and State Maps: From US Census; TIGER Data Products; <https://www.census.gov/programs-surveys/geography/guidance/tiger-data-products-guide.html>