----------Read me file ----------

Processing of Geospatial Data for the Habitat Risk Software

Cornell Wildlife Health Lab,
Animal Health Diagnostic Center,
College of Veterinary Medicine,
Cornell University
Ithaca, New York, USA

Contacts

Software programmers:
Corey I. Mitchell, coreyirene@gmail.com
W. David Walter, wdw12@psu.edu

Research collaborators:
Nicholas Hollingshead, nah88@cornell.edu

Research supervisor:
Krysten Schuler, ks833@cornell.edu

Overview

The ***Processing of Geospatial Data for the Habitat Risk Software*** allows the user to process publicly available geospatial data with the use of open-source software into the format necessary for inclusion in the Habitat Risk Software (Hanley et al. 2021; doi.org/10.7298/rcz8-nw50), which is the software that leverages Bayesian Hierarchical models (Clayton and Kaldor 1987; Banderjee et al. 2004; Gelman et al. 2004; Evans et al. 2016) with disease testing data to estimate the spatial risk of Chronic Wasting Disease in white-tailed deer (*Odocoileus virginianus*).

Origins of the USA-Sourced Geospatial Data

Publicly available geospatial data for the United States include the National Elevation Dataset 1 arc-second (30 m) Digital Elevation Model (DEM; U.S. Geological Survey 2020), the TIGER/Line Dataset (U.S. Census Bureau 2019), the National Land Cover Database (NLCD) 2016 Land Cover dataset for the conterminous United States (CONUS) (Dewitz 2019), the National Hydrography Dataset Plus Version 2 (USEPA and USGS 2012), and the Gridded National Soil Survey Geographic (gNATSGO) Database (Wieczorek 2014; Soil Survey Staff 2021).

Processing Notes for USA-Sourced Data

Processing Notes for the 0_USA_Setup & Extend State Boundaries Script

- All processing is done in R (R Core Team 2020a), see script for annotative details.
- Required packages for use of the software are installed.
- State boundaries are downloaded (500k resolution, 2019 dataset) directly in R using the tigris package (Walker 2020), which downloads TIGER/Line shapefiles directly from the United States Census Bureau (U.S. Census Bureau 2019).
- Outlying states & territories are removed and only states in the conterminous United States are retained.
- Each state area is extended by 50 km buffer and intersected with an outline of the US to remove areas that extended beyond the national shoreline (over oceans/Great Lakes where no spatial data are present, i.e., NLCD, Soils, etc.).
- Extended state areas are saved as a geopackage: *"States48_buff_intersect_50km.gpkg"*.

Processing Notes for the 1_USA_Select SPCS Projections Script

- All processing is done in R (R Core Team 2020a), see script for annotative details.
- State Plane Coordinate System (SPCS) in North American Datum 1983 (NAD83) is used for all states.
- When a single SPCS zone is available for a state it is selected.
- Central is selected when a state is split into multiple SPCS zones (>2).
- North is selected when a state is split into north/south SPCS zones.
- West is selected when a state is split into west/east SPCS zones.
- The output *SPCS_EPSG_SOP4CWD.csv* file contains a list of states and selected projections.
- The corresponding NAD83 EPSG code is used for further spatial processing in R.

Processing Notes for the 2_USA_Process NLCD Data Script

- All processing is done in R (R Core Team 2020a), see script for annotative details.
- The National Land Cover Database (NLCD) 2016 Land Cover dataset for CONUS (Dewitz 2019) must be downloaded externally from the Multi-Resolution Land Characteristics Consortium and saved to the working directory before processing (See **Step 0** below).
- Land cover layers are processed using a function that takes the extended state area, EPSG code, and NLCD 2016 Land Cover dataset as an input to crop, mask, re-project to the SPCS EPSG, and reclassify landcover raster layers.
- Output includes four raster files and two geopackage files all masked to the extended state's extent and projected to the selected State Plane Coordinate System (SPCS): (1) NLCD layer extent limited to state area, (2) NLCD reclassified to forest (land cover classes 41, 42, 43, 51, 52, 90, and 95), (3) NLCD reclassified to developed (land cover classes 21-24, 31), (4) NLCD layer reclassified to open (land cover classes 71-74, 81, 82), (5) Geopackage of the extended state's extent, (6) Geopackage of the state cartographic boundary

Processing Notes for the 3_USA_Process Dist to NHD Strahler4plus Streams Script

- All processing is done in QGIS 3.16 (QGIS Development Team 2020), R (R Core Team 2020a), and GRASS GIS 7.8 (GRASS Development Team 2020) within R using the rgrass7 (Bivand 2022) and fasterRaster (Smith 2022) packages, see script for annotative details.
- The National Hydrography Dataset Plus Version 2 (USEPA and USGS 2012) must be downloaded externally from the United States Environmental Protection Agency and saved to the working directory before processing (See **Step 0** below).
- Streams with Strahler order of at least 4 are extracted from the NHDFlowline_Network feature layer from the NHDPlusV21_National_Seamless_Flattened_Lower48 file geodatabase (USEPA and USGS 2012) in QGIS 3.16.
- Layers are processed using a function that takes the extended state outline, EPSG code, and extracted NHDFlowline_Network_Strahler4plus shapefile as an input to create a distance to streams raster layer.
- The processed NLCD layer for each state is used to define raster cell size, extent, and alignment.
- Output includes a raster and a geopackage file masked to the extended state's extent and projected to the selected SPCS: (1) Streams NHD Strahler order >= 4 distance raster, (2) Geopackage of NHD Strahler order >= 4 streams cropped to the extended state's extent.

Processing Notes for the 4_1_USA_Get DEM Layer Script

- All processing is done in R (R Core Team 2020a), see script for annotative details.
- Layers are processed using a function that extracts 1 arc-second (30 m) DEM tiles directly from the National Elevation Dataset (U.S Geological Survey 2020) using the FedData (Bocinsky 2019) package. Tiles are extracted based on an input of the extended state outline and EPSG code.
- Layers are processed in one of two ways due to memory restrictions in R: (1) Larger states (>40 tiles) can be split into two sections (divided in half to create N and S polygons), data for both polygons are extracted from NED, mosaicked, re-projected, then masked, while (2) Smaller states (<40 tiles) can be extracted as a complete set, re-projected, and then masked.
- The NLCD layer for each state is used to define raster cell size, extent, and alignment.

Processing Notes for the 4_2_USA_Mosaic Missing DEM Tiles Script

- All processing is done in R (R Core Team 2020a), see script for annotative details.
- Multiple states included in this packet (DE, IA, IL, MD, MO, PA, NJ, NY, VT) are missing one or more tiles after processing and these are added by modifying a function from the FedData (Bocinsky 2019) package to download the tiles, reproject, and mosaic to the statewide DEM created in the step above.

Processing Notes for the 5_USA_Create Slope & Roughness Layers Script

- All processing is done in R (R Core Team 2020a), see script for annotative details.
- Layers are processed using a function that takes the Digital Elevation Model (DEM) elevation raster created in the *4_1_USA_Get DEM Layer* script, the extended state outline, and EPSG code to process slope and roughness rasters based on the terrain function from the terra package (Hijmans 2022).

Processing Notes for the 6_USA_Process gNATSGO Average Clay Layer Script

- All processing is done in R (R Core Team 2020a), PostgreSQL 13.3 (PostgreSQL Global Development Group 2020), and QGIS 3.16 (QGIS Development Team 2020), see script for annotative details.
- Percent clay in the upper 50 cm of soil is calculated for each SSURGO map unit (MU) in PostgreSQL based on a method provided by Wieczorek (2014), and resulting values are stored in a csv file named *"*mapunit_clay.csv*"*.
- Average percent clay layers are processed using a function that takes the extended state outline, EPSG code, *"*gNATSGO-mukey.tif*"* raster, and *"*mapunit_clay.csv*"* file with mu_clay field as an input to create a raster layer of average clay using the ratify and deratify functions from the raster package (Hijmans 2021).
- The NLCD layer for each state is used to define raster cell size, extent, and alignment.
- Output includes a raster based on the AVG_CLAY field masked to the extended state's extent and projected to the selected State Plane Coordinate System (SPCS).

Procuring the Publicly Available Geospatial Data

Note: File/folder names and links to external repositories described herein represent the current version as of publication of the software. The processing instructions represent steps for processing the specific releases referenced here, releases of newer versions may require updated steps and/or additional processing.

Step 0: Procure USA-Sourced Geospatial Data.

1) The National Hydrography Dataset Plus Version 2 (USEPA and USGS 2012) is publicly available from the United States Environmental Protection Agency.
    a. Download the "NHDPlusV21_National_Seamless_Geodatabase_Lower48_07.7z" zip file and the "NHDPlusV2 User Guide" pdf at https://www.epa.gov/waterdata/nhdplus-national-data.
    b. Follow the instructions provided in the "NHDPlusV2 User Guide" to extract and install the data and save to the chosen location on your machine.
2) The National Land Cover Database (NLCD) 2016 Land Cover dataset for CONUS (Dewitz 2019) is publicly available from the Multi-Resolution Land Characteristics Consortium.
    a. Download the dataset from https://www.mrlc.gov/data/nlcd-2016-land-cover-conus by clicking on "Download".

3) The Gridded National Soil Survey Geographic Database (gNATSGO; Soil Survey Staff 2021) is publicly available from the United States Department of Agriculture.
    a. Download the "gNATSGO_CONUS.zip" zipped file in the "October 2021 gNATSGO CONUS" folder from the USDA Box site located at https://nrcs.app.box.com/v/soils.
    b. Download the "gNATSGO_mukey_grid.zip" zipped file in the "October 2021 MUKEY Grids (TIF)" folder from the USDA Box site linked above.

Note: State boundaries are downloaded (500k resolution, 2019 dataset) directly in R using the tigris package (Walker 2020), which downloads TIGER/Line shapefiles directly from the United States Census Bureau (U.S. Census Bureau 2019). The National Elevation Dataset 1 arc-second (30 m) Digital Elevation Models (DEM; U.S. Geological Survey 2020) are downloaded directly in R using the FedData package (Bocinsky 2019). No external downloads for these datasets are necessary.

Geospatial Processing

Step 1: Open R and verify that you are running the appropriate version of the R Software (R Core Team 2020a), which is `4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 [64-bit].`

Preparing and running the 0_USA_Setup & Extend State Boundaries Script

Step 2: Download the ***Processing of Geospatial Data for the Habitat Risk Software*** packet, unzip the downloaded folder, and save it in the chosen working directory on your machine. The new working directory should contain the following files:

1. "0_USA_Setup & Extend State Boundaries.R" script
2. "1_USA_Select SPCS Projections.R" script
3. "2_USA_Process NLCD Data.R" script
4. "3_USA_Process Dist to NHD Strahler4plus Streams.R" script
5. "4_1_USA_Get DEM Layer.R" script
6. "4_2_USA_Mosaic Missing DEM Tiles.R" script
7. "5_USA_Create Slope & Roughness Layers.R" script
8. "6_USA_Process gNATSGO Average Clay Layer.R" script
9. "renv.lock" file
10. ".Rprofile" file
11. A folder titled "renv" containing the script "activate.R"

Step 3: Open R and set the working directory to the chosen location on your machine.

Step 4: Open the "0_USA_Setup & Extend State Boundaries.R" script in R by clicking "File", "Open", navigating to the chosen working directory on your machine, clicking on the script, and then selecting "Open".

Step 5: Install the proper package versions onto your machine.

The commands necessary to install each package appear "commented out" (#) at the top of the script. **You must only install the versions of each package on your machine once,** after which - if you do not update them - you can load the versions of the packages using the `library` commands.

a. First install the `devtools` package and associated package dependencies by running the following code one time:
```
install.packages("devtools")
```

b. Then install `Rtools` by running the following lines of code one time:
```
devtools::install_version("installr", version="0.23.2")
installr::install.Rtools()
```
   i. When prompted: "`A newer R version is available, do you want to update R first?`"
   Select: "`No`"
   ii. When prompted: "`Do you want to allow this app to make changes to your device?`"
   Select: "`Yes`"
   iii. Click "`next`", "`next`", "`install`", and "`finish`" when prompted.

c. To install `fasterRaster`, download the zip file from https://github.com/adamlilith/fasterRaster/blob/master/zipTarFiles/fasterRaster_0.7.0.zip. Install manually by clicking "`Packages`", "`install package(s) from local files…`", then navigating to the downloaded file, and clicking "`open`".

d. Next install the `renv` package by running the following code one time:
```
devtools::install_version("renv", version="0.15.5")
```

e. Once `renv` is installed ensure the working directory includes the "`renv.lock`", "`.Rprofile`", and `renv/`"`activate.R`" files that were included in the packet. Then activate and restore the `renv` library:

NOTE: If the working directory contains the files listed above, this step will install the package versions and dependencies necessary to run the software and may take a while to complete. See https://rstudio.github.io/renv/ for more information on the renv package.

   i. Run the code: `renv::activate()`
   ii. When prompted: "`…Do you want to proceed? [y/N]:`"
   Type: `Y`
   iii. Run the code: `renv::restore()`
   iv. When prompted: "`The following package(s) were installed from an unknown source:....`"
   Type: `Y`
   v. When prompted: "`The following package(s) will be updated:....`"
   Type: `Y`

If you receive an error message at this step, make sure `fasterRaster` is installed and attempt to restore the `renv` library again.

f. If activating the `renv` library fails, uncomment and run the code on lines 67-84 one time to manually setup required packages. If prompted to update package versions enter 3 (None).

g. Re-comment out (#) any lines of code that were uncommented to install packages before continuing to the next step.

Step 6: Hit "Run all".

Preparing and running the 1_USA_Select SPCS Projections Script

Step 7: Open R and set the working directory to the chosen location on your machine.

Step 8: Open the "1_USA_Select SPCS Projections.R" script in R by clicking "File", "Open", navigating to the chosen working directory on your machine, clicking on the script, and then selecting "Open".

Step 9: Hit "Run all".

Preparing and running the 2_USA_Process NLCD Data Script

Step 10: Unzip and save the NLCD 2016 Land Cover CONUS dataset folder into your working directory.

Step 11: Open R and set the working directory to the chosen location on your machine.

Step 12: Open "2_USA_Process NLCD Data.R" script in R by clicking "File", "Open", navigating to the chosen working directory on your machine, clicking on the script, and then selecting "Open".

Step 13: Hit "Run all".

Step 14: Alter the state abbreviation in the `Process_NLCD_layers(StatePostalCode…` command with the state to process.

Step 15: Select that line of code and hit "Run".

Step 16: Repeat **Steps 14-15** until all states have been run.

*NOTE: This script is dependent on having the NLCD 2016 Land Cover CONUS dataset already downloaded, as well as the outputs from the "0_USA_Setup and Extend State Boundaries.R" and "1_USA_Select SPCS Projections.R" scripts, and should thus be run only after completion of the process in **Steps 0-9**.*

Preparing and running the 3_USA_Process Dist to NHD Strahler4plus Streams Script

Step 17: Create the "NHDFlowline_Network_Strahler4plus" shapefile from the National Hydrography Dataset Plus Version 2.

a. Open QGIS 3.16 and click "Project", then select "New" to create a new project.

b. In the QGIS File Browser pane navigate to the installed "NHDPlusV21_National_Seamless_Flattened_Lower48.gdb" file geodatabase on your machine. Click the arrow next to the geodatabase to view its contents.

c. Find the "NHDFlowline_Network" feature layer in the geodatabase and double click to add it to the project.

d. Use the Extract by Expression tool to extract the stream features with Strahler stream order greater than or equal to four. In the Processing Toolbox pane search

for “Extract by Expression” and double click the search result. Use the following parameters to perform the extract:

Input layer: NHDFlowline_Network
Expression: "StreamOrde" >=4
Matching features: …/NHDFlowline_Network_Strahler4plus.shp

Replace the ellipsis in the Matching features parameter with the absolute path of your chosen working directory
OR click the “…” button, select “Save to File…”, navigate to your working directory, enter the file name NHDFlowline_Network_Strahler4plus.shp in “File Name”, select SHP files (*.shp) for “Save as Type”, and hit “Save”.

Click “Run” and the extracted streams layer will save to your working directory.

e. Close QGIS.

Step 18: Open R and set the working directory to the chosen location on your machine.
NOTE: Your working directory must include the sub folder for the state being processed that was created in “2_USA_Process NLCD Data.R” script with the naming format "StatePostalCode_EPSG" (e.g. "NY_32116").

Step 19: Open the “3_USA_Process Dist to NHD Strahler4plus Streams.R” script in R by clicking “File”, “Open”, navigating to the chosen working directory on your machine, clicking on the script, and then selecting “Open”.

Step 20: Update the code on line 31 to the proper path of Grass GIS 7.8 on your machine.

Step 21: Update the code on line 55 to the proper path on your machine of the “NHDFlowline_Network_Strahler4plus.shp” layer created in **Step 17**.

Step 22: Hit “Run all”.

Step 23: Alter the state abbreviation in the `Process_dist_to_streams("…` command with the state to process.

Step 24: Select that line of code and hit “Run”.

Step 25: Repeat **Steps 23-24** until all states are run.
*NOTE: This script dependent on having the “NHDPlusV21_National_Seamless_Flattened_Lower48” file geodatabase already downloaded and installed, as well as the outputs from the “0_USA_Setup & Extend State Boundaries.R”, “1_USA_Select SPCS Projections.R”, and “2_USA_Process NLCD Data.R” scripts and should thus be run only after completion of the process in **Steps 0-16.***

Preparing and running the 4_1_USA_Get DEM Layer Script

Step 26: Open R and set the working directory to the chosen location on your machine.

NOTE: Your working directory must include the sub folder for the state being processed that was created in "2_USA_Process NLCD Data.R" script with the naming format "StatePostalCode_EPSG" (e.g. "NY_32116").

Step 27: Open the "4_1_USA_Get DEM Layer.R" script in R by clicking "File", "Open", navigating to the chosen working directory on your machine, clicking on the script, and then selecting "Open".

Step 28: Hit "Run all".

Step 29: Alter the state abbreviation in the `Get_DEM_layer_LRGstate("…` command with the next large state to process.

Step 30: Select that line of code and hit "Run".

Step 31: Repeat **Steps 29-30** until all large states have been run.

Step 32: Alter the state abbreviation in the `Get_DEM_layer_SMstate("…` command with the next small state to process.

Step 33: Select that line of code and hit "Run".

Step 34: Repeat **Steps 32-33** until all small states have been run.

*NOTE: This script is dependent on the outputs from the "0_USA_Setup & Extend State Boundaries.R", "1_USA_Select SPCS Projections.R", and "2_USA_Process NLCD Data.R" scripts and should thus be run only after completion of the process in **Steps 0-16**.*

Preparing and running the 4_2_USA_Mosaic Missing DEM Tiles Script

Step 35: Open R and set the working directory to the chosen location on your machine.

NOTE: This script is specific to states that didn't fully download DEM tiles in "4_1_USA_Get DEM Layer.R" and should only be run if processing DEM layers for those specific states (DE, IA, IL, MO, MD, PA, NJ, NY, VT). In addition, your working directory must include the sub folder for the state being processed that was created in "2_USA_Process NLCD Data.R" script with the naming format "StatePostalCode_EPSG" (e.g. "NY_32116").

Step 36: Open the "4_2_USA_Mosaic Missing DEM Tiles.R" script in R by clicking "File", "Open", navigating to the chosen working directory on your machine, clicking on the script, and then selecting "Open".

Step 37: Select the lines of code for the state you are processing and hit "Run".

*NOTE: This script is dependent on the outputs in the "4_1_USA_Get DEM Layer.R" script and should thus be run only after completion of the process in **Steps 26-34**.*

Preparing and running the 5_USA_Create Slope & Roughness Layers Script

Step 38: Open R and set the working directory to the chosen location on your machine.

NOTE: Your working directory must include the sub folder for the state being processed that was created in "2_USA_Process NLCD Data.R" script with the naming format "StatePostalCode_EPSG" (e.g. "NY_32116").

Step 39: Open the "5_USA_Create Slope & Roughness Layers.R" script in R by clicking "File", "Open", navigating to the chosen working directory on your machine, clicking on the script, and then selecting "Open".

Step 40: Hit "Run all".

Step 41: Alter the state abbreviation in the `Create_SlopeRough_rasts("`… command with the state to process.

Step 42: Select that line of code and hit "Run".

Step 43: Repeat **Steps 41-42** until all states are run.

*NOTE: This script is dependent on the outputs from the "0_USA_Setup & Extend State Boundaries.R", "1_USA_Select SPCS Projections.R", "2_USA_Process NLCD Data.R", "4_1_USA_Get DEM Layer.R", and "4_2_USA_Mosaic Missing DEM Tiles.R" scripts and should thus be run only after completion of the process in **Steps 0-16 and Steps 26-37**.*

Preparing and running the 6_USA_Process gNATSGO Average Clay Layer Script

Step 44: Unzip and save the "gNATSGO_mukey_grid" and "gNATSGO_CONUS" folder contents into your working directory.

Step 45: Calculate the percent clay in the upper 50 cm for all SSURGO mapping units in the gNATSGO dataset.

a. Run pgAdmin for PostgreSQL.

b. Create a new Server with the following parameters (leaving unspecified parameters as defaults):

Name: GIS
Hostname/Address: localhost

c. Create a database in the GIS Server using the following parameters (leaving unspecified parameters as defaults):

Database: gNATSGO

d. Open QGIS 3.16 and click "Project", then select "New" to create a new project.

e. In the QGIS File Browser pane create a new PostGIS Connection by right-clicking on PostGIS and selecting "New Connection". Use the following parameters for the connection:

Name: gNATSGO_database
Host: localhost
Port 5432
Database: gNATSGO
Authentication: Basic
enter the PostgreSQL GIS server username and password and test connection

f. After the connection is established, in the QGIS File Browser pane, navigate to the location of the extracted "gNATSGO_CONUS.gdb" on your machine and click the arrow next to the file geodatabase to view its contents.

g. Click-and-drag the "chorizon" table from the "gNATSGO_CONUS.gdb" to the public schema within the gNATSGO_database database connection to export the table into the PostgreSQL gNATSGO database.

Note: Adding the file geodatabase contents to the project via “Layer”, “Add Layer”, “Add Vector Layer…” will alter the names of the loaded layers and the query will not run as expected, be sure to use the Browser pane to add all files.

h. Repeat step g. for the “component” table.

i. In pgAdmin, in the browser pane, select the gNATSGO database that now contains the “chorizon” and “component” tables from the “gNATSGO_CONUS.gdb”.

j. Open the Query Tool and run the following command to generate PostgreSQL by copying the code below, pasting directly into the Query Editor, and clicking the play button:

```
/*
**************************
PostgreSQL - mapunit clay query
**************************
 */

WITH QComponent_TotalDepth as (
        select cokey,
            sum(
                case
                    when 50 - hzdept_r > 0 then 50 - hzdept_r
                    else 0
                END
                -
                CASE
                    WHEN 50- hzdepb_r > 0 then 50- hzdepb_r
                    ELSE 0
                END
            ) as total_depth
        from Chorizon
        group by cokey
),
QComponent_Clay as (
        select
        qComponent_TotalDepth.cokey,
        sum(
            CASE
                WHEN total_depth = 0 THEN 0
                ELSE
                    (
                        CASE
                            WHEN 50-hzdept_r > 0 THEN 50 - hzdept_r
                            ELSE 0
```

```
                                    END
                                    -
                                    CASE
                                        WHEN 50- hzdepb_r > 0 THEN 50 - hzdepb_r
                                        ELSE 0
                                    END
                                ) * claytotal_r / total_depth
                        END
                        ) as component_claytotal_r
                from Chorizon join qComponent_TotalDepth on Chorizon.cokey =
                qComponent_TotalDepth.cokey
                group by qComponent_TotalDepth.cokey
),
qMU_total as (
select mukey, sum(comppct_r) as total_comppct_r
from component
group by mukey
)
select component.mukey,
sum(component_claytotal_r*comppct_r/total_comppct_r) as MU_clay
into mapunit_clay
from qComponent_Clay
join (qMU_Total join Component on qMU_Total.mukey = Component.mukey) on
        qComponent_Clay.cokey = component.cokey
group by Component.mukey

/*
**************************
End of query
**************************
 */
```

k. In pgAdmin, export the “mapunit_clay” table from the gNATSGO server to your working directory by right-clicking on the “mapunit_clay” table and selecting “Import/Export…”. (Note: if the “mapunit_clay” table does not appear in the “Tables” list after successfully processing the query, right-click on “Tables” and select “Refresh…”.) Use the following parameters (leaving other parameter values as defaults) for the export:

Filename: mapunit_clay.csv
Format: csv
Encoding: UTF8
Header: Yes
Delimiter: ,
Quote: “

Step 46: Open R and set the working directory to the chosen location on your machine.
NOTE: Your working directory must include the sub folder for the state being processed that was created in "2_USA_Process NLCD Data.R" script with the naming format "StatePostalCode_EPSG" (e.g. "NY_32116").

Step 47: Open the "6_USA_Process gNATSGO Average Clay Layer.R" script in R by clicking "File", "Open", navigating to the chosen working directory on your machine, clicking on the script, and then selecting "Open".

Step 48: Update the code on line 51 to the proper path on your machine for the exported "gNATSGO-mukey.tif" layer extracted in **Step 44**.
NOTE: If the file is in your working directory, do not update the path.

Step 49: Update the code on line 52 to the proper path on your machine for the "mapuit_clay.csv" file created in **Step 45**.
NOTE: If the file is in your working directory, do not update the path.

Step 50: Hit "Run all".

Step 51: Alter the state abbreviation in the `Process_Clay_Layer("…` command with the state to process.

Step 52: Select that line and hit "Run".

Step 53: Repeat **Steps 51-52** until all states are run.
*NOTE: This script is dependent on having the "gNATSGO_CONUS" and"gNATSGO_mukey_grid" files already downloaded, as well as the outputs from the "0_USA_Setup & Extend State Boundaries.R", "1_USA_Select SPCS Projections.R", and "2_USA_Process NLCD Data.R" scripts and should thus be run only after completion of the process in **Steps 0-16.***

Technical Details for Software Processing USA-Sourced Data

The *0_USA_Setup & Extend State Boundaries.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires several packages: (1) "devtools" (Wickham et al. 2020), (2) "installr" version 0.23.2 (Galili 2021), (3) "renv" version 0.15.5 (Ushey 2022), (4) "sp" version 1.4-4 (Pebesma and Bivand 2005; Bivand et al. 2013), (5) "sf" version 0.9-8 (Pebesma 2018), (6) "rgdal" version 1.5-18 (Bivand et al. 2020), (7) "dplyr" version 1.0.6 (Wickham et al. 2021), (8) "tigris" version 1.0 (Walker 2020), (9) "USAboundaries" (Mullen and Bratt 2018), (10) "raster" version 3.4.9 (Hijmans 2021), (11) "terra" version 1.5-34 (Hijmans 2022), (12) "XML" version 3.99-0.9 (Temple Lang 2022), (13) "igraph" version 1.3.0 (Csárdi and Nepusz 2006), (14) "rgrass7" version 0.2.7 (Bivand 2022), (15) "foreach" version 1.5.1 (Microsoft and Weston 2020), (16) "FedData" version 2.5.7 (Bocinsky 2019), (17) "gdalUtils" version 2.0.3.2 (Greenberg and Mattiuzzi 2020), (18) "foreign" version 0.8.8 (R Core Team 2020b), (19) "plyr" version 1.8.6 (Wickham 2011), (20) "link2GI" version 0.4-5 (Reudenbach 2020), (21) "rgeos" version 0.5-5 (Bivand and Rundel 2020), and (22) "fasterRaster" version 0.7.0 (Smith 2022) available from https://github.com/adamlilith/fasterRaster/blob/master/zipTarFiles/fasterRaster_0.7.0.zip.

The *1_USA_Select SPCS Projections.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires five packages: (1) “renv” version 0.15.5 (Ushey 2022), (2) “rgdal” version 1.5-18 (Bivand et al. 2020), (3) “USAboundaries” (Mullen and Bratt 2018), (4) “sf” version 0.9.8 (Pebesma 2018), and (5) “tigris” version 1.0 (Walker 2020).

The *2_USA_Process NLCD Data.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires several packages: (1) “renv” version 0.15.5 (Ushey 2022), (2) “rgdal” version 1.5-18 (Bivand et al. 2020), (3) “terra” version 1.5-34 (Hijmans 2022), (4) “raster” version 3.4.9 (Hijmans 2021), (5) “sp” version 1.4-4 (Pebesma and Bivand 2005; Bivand et al. 2013), (6) “sf” version 0.9.8 (Pebesma 2018), and (7) “tigris” version 1.0 (Walker 2020).

The *3_USA_Process Dist to NHD Strahler4plus Streams.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires several packages: (1) “renv” version 0.15.5 (Ushey 2022), (2) “rgdal” version 1.5.18 (Bivand et al. 2020), (3) “terra” version 1.5-34 (Hijmans 2022), (4) “raster” version 3.4.9 (Hijmans 2021), (5) “sf” version 0.9.8 (Pebesma 2018), (6) “rgrass7” version 0.2.7 (Bivand 2022), which is available after the installation of Grass GIS 7.8 from https://grass.osgeo.org/download/, and (7) “fasterRaster” version 0.7.0 (Smith 2022) available from https://github.com/adamlilith/fasterRaster/blob/master/zipTarFiles/fasterRaster_0.7.0.zip.

The *4_1_USA_Get DEM Layer.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires several packages: (1) “renv” version 0.15.5 (Ushey 2022), (2) “rgdal” version 1.5.18 (Bivand et al. 2020), (3) “terra” version 1.5-34 (Hijmans 2022), (4) “raster” version 3.4.9 (Hijmans 2021), (5) “sp” version 1.4-4 (Pebesma and Bivand 2005; Bivand et al. 2013), and (6) “FedData” version 2.5.7 (Bocinsky 2019).

The *4_2_USA_Mosaic Missing DEM Tiles.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires several packages: (1) “renv” version 0.15.5 (Ushey 2022), (2) “rgdal” version 1.5.18 (Bivand et al. 2020), (3) “terra” version 1.5-34 (Hijmans 2022), (4) “raster” version 3.4.9 (Hijmans 2021), (5) “sp” version 1.4-4 (Pebesma and Bivand 2005; Bivand et al. 2013), (6) “foreach” version 1.5.1 (Microsoft and Weston 2020), and (7) “FedData” version 2.5.7 (Bocinsky 2019).

The *5_USA_Create Slope & Roughness Layers.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires several packages: (1) “renv” version 0.15.5 (Ushey 2022), (2) “rgdal” version 1.5.18 (Bivand et al. 2020), (3) “terra” version 1.5-34 (Hijmans 2022), and (4) “raster” version 3.4.9 (Hijmans 2021).

The *6_USA_Process gNATSGO Average Clay Layer.R* software was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit), and requires several packages: (1) "renv" version 0.15.5 (Ushey 2022), (2) "rgdal" version 1.5.18 (Bivand et al. 2020), (3) "terra" version 1.5-34 (Hijmans 2022), (4) "raster" version 3.4.9 (Hijmans 2021), and (5) "plyr" version 1.8.6 (Wickham 2011).

Financial support

Financial support comes from: (1) Michigan Disease Initiative - *Optimizing CWD Surveillance: Regional Synthesis of Demographic, Spatial, and Transmission-Risk Factors* (2019); (2) Tennessee Wildlife Resources Agency - *Modeling Risk of Infection for Individually Harvested Deer & Estimating Prevalence When Sampling is Limited* (2020); (3) Michigan Disease Initiative - *SOP4CWD Dashboard: A Web Application for Disease Visualization and Data-Driven Decisions* (2020); (2) Multistate Conservation Grant Program - *Surveillance Optimization Project for Chronic Wasting Disease: Streamlining a Web Application for Disease Visualization and Data-Driven Decisions* (2021).

Acknowledgements

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

License

Citations

Banerjee, S., Carlin, B. P., & Gelfand, A. E. 2004. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, New York.

Bivand, R., Pebesma, E., & Gomez-Rubio, V. 2013. Applied spatial data analysis with R, Second edition. Springer, NY. *asdar-book.org/*

Bivand, R., Keitt, T., & Rowlingson B. 2020. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.5-18. https://CRAN.R-project.org/package=rgdal.

Bivand, R., & Rundel, C. 2020. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). package version 0.5-5. https://CRAN.R-project.org/package=rgeos

Bivand, R. 2022. rgrass7: Interface Between GRASS 7 Geographical Information System and R. R package version 0.2-7. https://CRAN.R-project.org/package=rgrass7

Bocinsky, R. Kyle. 2019. FedData: Functions to Automate Downloading Geospatial Data Available from Several Federated Data Sources. R package version 2.5.7. https://CRAN.R-project.org/package=FedData

Clayton, D., & Kaldor, J. 1987. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43:671-681.

Csárdi, G., & Nepusz, T. 2006. The igraph software package for complex network research, InterJournal, Complex Systems 1695. https://igraph.org

Dewitz, J. 2019. National Land Cover Database (NLCD) 2016 Products: U.S. Geological Survey data release, https://doi.org/10.5066/P96HHBIE.

Evans, T. S., Kirchgessner, M. S., Eyler, B., Ryan, C. W., & Walter, W. D. 2016. Habitat Influences Distribution of Chronic Wasting Disease in White-tailed Deer. Journal of Wildlife Management, 80(2):284-291.

Galili, T. 2021. installr: Using R to Install Stuff on Windows OS (Such As: R, 'Rtools', 'RStudio', 'Git', and More!). R package version 0.23.2. https://CRAN.R-project.org/package=installr

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2004. Bayesian data analysis. New York: Chapman and Hall/CRC. 1-696 p.

GRASS Development Team, 2020. Geographic Resources Analysis Support System (GRASS) Software, Version 7.8. Open Source Geospatial Foundation. https://grass.osgeo.org.

Greenberg, J.A., & Mattiuzzi, M. 2020. gdalUtils: Wrappers for the Geospatial Data Abstraction Library (GDAL) Utilities. R package version 2.0.3.2. https://CRAN.R-project.org/package=gdalUtils.

Hanley, B., Mitchell, C., Walter, W. D., Kelly, J., Abbott, R., Hollingshead, N., & Schuler, K. 2021. Habitat Risk Software [Software]. Cornell University Library eCommons Repository, https://doi.org/10.7298/rcz8-nw50.

Hijmans, R.J. 2021. raster: Geographic Data Analysis and Modeling. R package version 3.4-9. https://CRAN.R-project.org/package=raster.

Hijmans, R.J. 2022. terra: Spatial Data Analysis. R package version 1.5-34. https://CRAN.R-project.org/package=terra

Microsoft & Weston, S. 2020. foreach: Provides Foreach Looping Construct. R package version 1.5.1. https://CRAN.R-project.org/package=foreach

Mullen, L.A., & Bratt, J. 2018. "USAboundaries: Historical and Contemporary Boundaries of the United States of America," Journal of Open Source Software 3, no. 23 (2018): 314. https://doi.org/10.21105/joss.00314.

Pebesma, E., & Bivand, R. 2005. Classes and methods for spatial data in R. R News 5 (2), *cran.r-project.org/doc/Rnews/*

Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10 (1), 439-446, https://doi.org/10.32614/RJ-2018-009

PostgreSQL Global Development Group. 2020. PostgreSQL Database Management System. Version 13.3. https://www.postgresql.org/.

QGIS Development Team. 2020. QGIS Geographic Information System. Version 3.16. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

R Core Team. 2020a. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

R Core Team. 2020b. foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ... R package version 0.8-81.

Reudenbach, C. 2020. link2GI: Linking Geographic Information Systems, Remote Sensing and Other Command Line Tools. R package version 0.4-5. https://CRAN.R-project.org/package=link2GI

Soil Survey Staff. 2021. Gridded National Soil Survey Geographic (gNATSGO) Database for the Conterminous United States. United States Department of Agriculture, Natural Resources Conservation Service. Available online at https://nrcs.app.box.com/v/soils. March 20, 2022 (FY2021 official release).

Smith, A.B. 2022. fasterRaster: Faster raster processing in R using GRASS GIS. R package version 0.7.0. http://www.earthSkySea.org.

Temple Lang, D. 2022. XML: Tools for Parsing and Generating XML Within R and S-Plus. R package version 3.99-0.9. https://CRAN.R-project.org/package=XML

U.S. Census Bureau. 2019. 2019 TIGER/Line Shapefiles (machine-readable data files).

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS). 2012. National Hydrography Dataset Plus – NHDPlus 2.10. https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus. Accessed 2020-11-1.

U.S. Geological Survey. 2020. 1 Arc-second Digital Elevation Models (DEMs) - USGS National Map 3DEP Downloadable Data Collection (published 20200101). Available online at https://www.usgs.gov/core-science-systems/ngp/3dep/data-tools.

Ushey, K. 2022. renv: Project Environments. R package version 0.15.5. https://CRAN.R-project.org/package=renv

Walker, K. 2020. tigris: Load Census TIGER/Line Shapefiles. R package version 1.0. https://CRAN.R-project.org/package=tigris.

Wickham, H. 2011. The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Wickham, H., Hester J., & Chang W. 2020. devtools: Tools to Make Developing R Packages Easier. R package version 2.3.2. https://CRAN.R-project.org/package=devtools.

Wickham, H., François, R., Henry, L., & Müller, K. 2021. dplyr: A Grammar of Data Manipulation. R package version 1.0.6. https://CRAN.R-project.org/package=dplyr

Wieczorek, M. 2014. Area- and depth- weighted averages of selected SSURGO variables for the conterminous United States and District of Columbia. http://pubs.er.usgs.gov/publication/ds866. doi:10.3133/ds866.

Suggested Citation for Re-use:

Mitchell, C.I., Walter, W. D., Hollingshead, N., & Schuler, K. 2021. Processing of Geospatial Data for the Habitat Risk Software [Software]. Cornell University Library eCommons Repository, https://doi.org/10.7298/2tt1-yy48.

Corrections Since Original Submission

7-1-2022. Clarified wording in readme file per peer review comments, updated broken code, updated packages, and updated gNATSGO dataset version and instructions.