Compiling Covariates

There are 36 "studies" for which we need covariates. In general, the goal is to make a .tif raster file (potentially time varying) for each covariate for each study.

1. EVI extraction: Top priority

Goals/End Products

Extract EVI for all studies into 250m by 250m rasters that span the spatial and temporal extent of the study. We want EVI on a monthly time-scale for each study.

Currently we are using NDVI for a proxy for plant productivity. However, EVI is better proxy, but is harder to get a hold of in a cleaned and formatted form. All the EVI data is freely available on MODIS, but to access it you need to query MODIS, stitch relevant MODIS tiles together to get the area that you want, and reproject the MODIS data. Ideally, we want to do this for each study site.

Some details

Example: For the study `txcamp` we want 8 EVI raster files on the 250 m by 250 m scale over then extent longmin = -98.62192, longmax=-98.49822, latmin=29.62192, latmax=29.79721, for the year 2016 and the months January-August. Each would be named `txcamp_evi_<month>_2016.tif`.

Steps

The generic steps will be something along the lines of

For each study

- Extract the duration of the study
- Extract the spatial extent of the study
- Use some MODIS tool to pull the required MODIS tiles on 16-day intervals over the duration of the study for the spatial extent.
- Use the MODIS tools/GDAL to stitch and reproject the MODIS tiles to WGS84.
- Convert to raster format
- If the resulting are much bigger than the extent, crop the raster to the spatial study extent.
- Save the resulting raster file.

Some EVI resources

MODIS extraction packages

R packages that seem useful

- `MODIStsp`
- `MODISTools`

They both require GDAL or the MODIS tool. They were both being a little finicky for me to get downloaded.

Python packages

- `pyMODIS`

This will take some fiddling around to figure out the best way to extract this from MODIS.

Websites

1. MODIS website

- We want the MOD13Q1 data at 350 m resolution
2. Information on how MODIS grids the world
   - Some of the MODIS extraction packages let you query the extents directly, but for others you need to know the MODIS tiles.

2. Distance to road metric: Second priority

Goal/End Products

For each study 1 arc second (~ 30m by 30m) resolution raster where each raster cells contains distance to nearest road in the extent of the study.

1. Come up with a recipe for "distance to nearest road" and ensure consistent road data across studies. Ideally, this would be the same data source for all studies. 1a. I think the recipe should be something like
   - Download road shp file (from where, I am not sure. Maybe here
   - Crop road shp file to the extent of each study
     - Given the size of these files, using features of GDAL could really speed this up.
   - Compute distance to nearest road point for each cell in raster.
     - I have already built a distance to nearest polygon function. It needs a few tweaks (see below).
2. Generate the distance to nearest road rasters for all studies.

Update "distance to nearest" metric: Second priority

Currently, this metric is being computed as distance to a centroid of a polygon. The function is in `code/covariate_scripts/covariate_fxns.R` and is named `dist_nearest_neighbor`. This really should be being computed as distance to the nearest perimeter point. Because there are many more points than polygons, this could be quite slow for some of the data. Would be useful if we could think about ways to speed up this calculation.

3. Covariates that need extracting: Third priority

The following covariates need to be extracted and formatted for each study.

1. NDVI/EVI (Plant productivity)
2. Distance to water
3. Landcover (Forest canopy cover)
4. Croptypes (Agriculture)
5. Drought (Drought index)
6. Elevation
7. Distance to roads

There are currently 5 studies for which these have been extracted (excluding distance to roads): txcamp, fl_raoul, srel_contact, tx_tyler_w2, and tejon. The file `covariate_table.xlsx` contains a description of how to extract each of these covariates under the column "Extraction Methodology".

Extraction overview

1. All of the covariates live in the directory `rsf_swine/data/covariate_data/`.
2. There is a separate folder for each covariate.
3. The structure of this folder is important in that when the raw covariate files are downloaded they need to be put in a specific place in a specific way.

4. After the covariate data is downloaded and placed in the correct directory in the correct format (see the excel file), the appropriate `extract_*.R` script should be run with the current working directory as `rsf_swine/code/covariate_scripts`.

Running an `extract_*.R` script for a study

1. Open the `extract_*.R` script
2. Look for a line similar to `if(studynm %in% c('srel_contact'))`.
3. Replace the `srel_contact`, in this example, with the study name you are considering (e.g. `tejon`, `txcamp`, etc.).

4. Covariates will be formatted for all studies that are named and this can take a long time for some studies. So I have been just running one, or a few, at a time.

Studies for which no covariates have been extracted as of March 21, 2018

1. cali0, . . . , cali4
2. canada
3. florida
4. ga_bill
5. ga_steve
6. judas_pig
   - Ignore until we figure out the capturing times
7. la_hartley
8. la_steve
9. michigan
10. mo_kurt0, . . . , mo_kurt8
11. sc_jim2
12. scjim1
13. srel_vacuum
14. srs_kilgo
15. tx_christy_14r
16. tx_chirsty_15r
17. tx_susan
18. tx_tyler_k1
19. tx_tyler_w1
    - Need NDVI for this study still. Site currently down.

Priorities for extraction

1. cali0, . . . , cali4
2. florida
3. michigan
4. mo_kurt0, . . . , mo_kurt9
5. tx_tyler_w1
6. tx_tyler_k1
7. tx_susan
8. ga_bill
9. ga_steve
10. la_hartley
11. scjim1
12. sc_jim2
13. canada
    - This is one could be tricky as a lot of the crop data isn't consistent into Canada. We might have to exclude this one. . .

14. srs_kilgo
15. tx_christy_14r
16. tx_chirsty_15r
17. Others