



How to Run the Model

Step 1

Add layers for making the model to folders in a base directory that is an R project (if you don't make the root an r project, you will need to specify absolute paths for all files).

The folder structure should be as follows:

1. `inputs/`

1. `raster/`

State of Colorado DEM

2. `shape/`

1. `distance/`

All files that you want to calculate the distance to. This can be either polygons or lines. NHD flowline will be created and placed here.

2. `cover_shapes/`

All shape files that cover the entire analysis area. For example soils is a coverage.

3. `extent/`

Polygon of the field office boundary or the extent of the model.

4. `flowlines/`

NHD Flowline shapefile. The shapefile in this folder will be subset into artificial, perennial and intermittent. These files should be output into the `distance/` folder.

2. `outputs/`

1. `Final_dataset/`

Contains the processed dataset that has not be subset to only observations that have been surveyed.

2. `final_layers/`

All input rasters that will go into the model. These have either been created by processing the layers in the inputs folder. Might be good for me to move this to a root folder in future iterations. It really isn't a input or an output.

3. `model_runs/`

These are save random forests models. They can be loaded and run.

1. `prediction/`

saved raster that have been predicted.

3. `1 - subset_flowlines.Rmd`

This file takes USGS NHD flowline data and subsets it into "Artificial", "Intermittent", "Perennial" FTypes.

4. `2.1 - distance_function.r`

Input Creation - Creates a blank raster based on a base raster, in this case the 100m DEM, and calculates the distance of each cell to line and polygon features. WARNING, calculating distance to the line features, if there are a lot of them like the case of the flowline layers, takes a really long time. The process can take up to 10 to 24 hours. for three layers depending on compute power. Outputs to a file that should be the same as the other input creation steps.

5. `2.2 - rasterize_polygon_inputs.Rmd`

Input Creation - Converts 100% coverage shapefiles (shapefiles that cover the entire area) and converts them to raster with the same dimensions as a base raster, in this case the 100m dem. Outputs to a file that should be the same as the other input creation steps.

6. `2.3 - topographic_calcs.Rmd`

Input Creation - Takes a digital elevation model and calculates slope, aspect, terrain roughness index, terrain roughness index, and flow direction. Outputs to a

7. file that should be the same as the other input creation steps.
3 - `make_training_dataset.Rmd`

Creates the training dataset by merging all rasters created in steps 2.x. It also loads and rasterizes the archy site shapefile, subsets all of the historic and archaeological district RES_TYPES out of the site data prior to rasterization, and loads and rasterizes the survey data.

The resulting dataset is saved to an output location to be loaded in the next step.

8. 4 - `run_random_forest.Rmd`

Loads the dataset created in the previous step, subsets it to only those observations that have been surveyed. Then creates a variable to be predicted out of the archy site raster layer added to the training dataset in step 3. The resulting dataset can be run through two random forest algorithms, one is from the `randomForest` package and is a base implementation of Random Forests. The other is the `Caret` package which is package for classification and regression and includes tools to help tune model inputs and evaluate outputs. Both models are output to `model_runs`.

9. 5 - `run_prediction.Rmd`

Loads a model created in the previous step, creates the raster using all the model inputs created in Steps 2.X, and uses the loaded model to predict each pixel. The final model is

