Processing Notes

Folder for files in processing steps below: “DataSets\_8\_17”

**Part A: create new point shapefile in proper projection**

1. downloaded electrofishing (E) data from online LTRM fisheries database for pool 26, the Illinois river reach (ILR), and the open river reach (ORR), years 1993 – present. Saved as zipped file “ltrm\_fish\_data\_0817130929.zip” and as table “ltrm\_fish\_data.csv”
2. collapsed fish sampling data into site observations and subset columns to include
   1. stratum
   2. lcode
   3. sdate
   4. utm\_e
   5. utm\_n
   6. snag
3. split the collapsed fish sampling data into two .csv files, one for UTM zone 15N and one for UTM zone 16N.

\* The new csv files are saved as “ltrm\_fish\_sites\_EF\_z15.csv” and “ltrm\_fish\_sites\_EF\_z16.csv” and also includes a column of “RowID” – unique row identifier to link back to “ltrm\_fish\_data.csv” from step 1

1. imported csv datasets from step 2 into GIS as points and re-projected zone16 points into UTM zone 15N NAD83 to match other datalayers – this step now ensures that when you map the resulting shapefile it will overlay with other datalayers nicely.
2. Merged the two point layers together and saved as new shapefile “sites\_p26ilror.shp” with attribute columns:
   1. FID – unique identifier automatically assigned in GIS program
   2. Shape – automatically assigned in GIS
   3. RowID\_– row number from original EF dataset (result of step 2) \*can use as link back to larger csv file from step 2! (note: called “Field1” in datasets\_7\_7 files)
   4. Stratum – sampling stratum (from EF dataset)
   5. Lcode – location code (from EF dataset)
   6. Sdate – sampling date (from EF dataset)
   7. Utm\_e – UTM coordinate (from EF dataset)
   8. Utm\_n – UTM coordinate (from EF dataset)
   9. Snag – presence/absence/NA (from EF dataset)

**Part B: create point shapefile with attribute table that includes aquatic area information (2 ways)**

1. Computed spatial intersection between sites\_p26ilror.shp and level III aquatic areas shapefile (aqa\_2010\_lvl3\_011918.shp) to create a new shapefile “sites\_aquaareas2.shp”
2. Join is simple intersection – a point assumes information from an aquatic area ploygon if the point directly intersects a polygon
3. The resulting shapefile has columns from both input shapefiles
4. Observations with value of 0 in all the columns from aqa\_2010\_lvl3\_011918.shp do not intersect with the aquatic areas layer – they may be very close, and are “in” aquatic areas in real life, but due to inherent errors in the geospatial production of the aquatic areas layer and generation of the sampling points’ UTM coordinates, they might not intersect.
5. Spatial intersection between sites\_p26ilror.shp and level iii aquatic areas shapefile (aqa\_2010\_lvl3\_011918.shp) to create a new shapefile “sites\_aquaareas5m2.shp”
6. Join is a spatial intersect that assumes a 5m radius buffer – a point assumes information from aquatic area polygon if the point is directly intersecting or within 5m of a polygon from the aquatic areas shapefile.
7. The resulting shapefile has columns from both input shapefiles
8. Observations with value of 0 in all the columns from aqa\_2010\_lvl3\_011918.shp do not intersect within 5m radius of aquatic areas layer.

Note that information about the level 3 aquatic areas can be found here: https://www.sciencebase.gov/catalog/item/5a708ef0e4b06e28e9cae58f

**Part C: create point shapefile with attribute table that includes distance to terrestrial areas and terrestrial land cover type**

1. Beginning with the 2010 land cover shapefile (“lc\_20102.shp”), removed observations with aquatic habitat attributes to create a terrestrial land cover layer. (…because despite its name, the land cover dataset also includes both terrestrial and aquatic habitats!) Observations with the following attributes were removed:
   1. SAV – submerged aquatic vegetation (any %) –no SAV in these pools
   2. RFAD – rooted floating aquatics (any %)
   3. OW – open water

The resulting new shapefile of terrestrial land covers is called “Terrestrial2.shp”. Information about the attribute table columns can be found in the metadata for the 2010 landcover datasets here: https://www.sciencebase.gov/catalog/file/get/58385133e4b0d9329c801cc1?f=\_\_disk\_\_16%2F73%2F0c%2F16730c888489e35dd95dcee9321363ed7d124f93&transform=1&allowOpen=true

1. Computed a distance function in GIS to identify the nearest polygon from Terrestrial2.shp to each sampling point location, and measure the straight-line distance between the two features. Note that distance is calculated in any direction (river flow direction is not accounted for). The Resulting shapefile is “sites\_terrestrial2.shp” and has two new columns:
   1. NEAR\_FID – the unique identifier of the terrestrial polygon nearest to a given point
   2. NEAR\_DIST – the distance to the nearest terrestrial polygon in meters.

\*The attribute table of this shapefile can be joined with the attribute table of Terrestrial2.shp in R using the common field “NEAR\_FID” and “FID” (respectively) to add information about land cover type of the nearest terrestrial polygon.

**Part D: create point shapefile with attribute table that includes distance to Forested areas (similar to Part C, but only for terrestrial areas with trees)**

1. Subset Terrestial2.shp to include polygons with the following LCU\_N2 attributes listed below. The new shapefile of forested areas is called “Terrestial\_Forests2.shp”. LCU\_N2 attributes include:
   1. Conifers
   2. Floodplain forest
   3. Lowland forest
   4. Populus community (Populus = Eastern cottonwood)
   5. Salix community (Salix = willow species)
   6. Upland forest
2. Computed a distance function in GIS to identify the nearest polygon from Terrestrial\_Forests2.shp to each sampling point location, and measure the straight-line distance between the two features. Note that distance is calculated in any direction (river flow direction is not accounted for). The Resulting shapefile is “sites\_forest2.shp” and has two new columns:
   1. NEAR\_FID – the unique identifier of the forested polygon nearest to a given point
   2. NEAR\_DIST – the distance to the nearest forested polygon in meters.

\*The attribute table of this shapefile can be joined with the attribute table of Terrestrial\_Forests2.shp in R using the common field “NEAR\_FID” and “FID” (respectively) to add information about land cover type of the nearest forested polygon.

**Part E: create shapefiles for general mapping purposes**

1. Dissolved polygon boundaries of Terrestrial2.shp to create Terrestrial\_dissolve2.shp
2. Selected all areas not included in Terrestrial2.shp files to create a new shapefile of aquatic-only areas (Aquatic2.shp) and dissolved polygon boundaries. Resulting shapefile = Aquatic\_dissolve2.shp
   1. Note that steps 1 and 2 above result in shapefiles that include all 3 pools: p26, la grange reach of the Illinois River, and the open river reach.

**Part F: create .txt files from all attribute tables in Parts A-D**

1. Generated text files for shapefiles involved in Parts A-D above, each saved in subfolder Data\_8\_17\AttributeTables by the name of the parent shapefile.