Outline of GIS Analysis

Generating Sampling Points For Groundwater Monitoring Near Blueberry Fields

Curtis Bohlen

September, 24, 2020

# Introduction

The goal of the GIS analysis is to produce a polygon data layer that represents the land area of the state located in blueberry fields or within one quarter mile of blueberry fields. The final file should omit areas in BPC Inspector Regions One, Four, Five, and Six. This file provides the input area for drawing a GRTS sample.

# Source Data

The starting data for preparing the GIS were as follows. Most of these data sources are on my computer as part of my GIS data infrastructure. I do not know exactly when they were downloaded from their original sources, but most were probably downloaded about 5 years ago.

1. “cnty24p”, from Maine Office of GIS. This represents the boundary of Maine Counties, at 1:24000 scale, and is used here as a convenient base layer for depicting the land area of the state of Maine.
2. The National Hydrography Dataset (NHD) NHDWaterbody and NHDArea layers for the state of Maine. This was used to allow me to generate a “water free” map of the state of Maine, so we do not try to draw random samples from the surface of lakes, ponds, rivers, or bays.
3. The “metwp24P” data layer from the Maine Office of GIS. This represents the boundary of Maine Municipalities, at 1:24000 scale. This was used to generate the data layer depicting BPC inspector regions.
4. Data on Maine cropland, downloaded from USDA’s Agricultural Statistical Service. This is a data layer showing CROPLAND only, not all agricultural lands. Thus the sampling frame likely under samples grazing lands and forest-based products like maple syrup.

USDA National Agricultural Statistics Service Cropland Data Layer. 2019. Published crop-specific data layer [Online]. Available at https://nassgeodata.gmu.edu/CropScape/ (accessed July 27, 2020). USDA-NASS, Washington, DC.

The web interface at that URL makes it easy to download crop data for the State of Maine. I downloaded 2019 data for all commodities, statewide, in UTM coordinates. I then used ArcGIS to simplify that to only include data on location of blueberries.

# Step 1: Generate a Maine “Land Only” Polygon Layer

1. Load the cnty24p polygon layer into ArcGIS.
2. Load the National Hydrography Dataset NHDWaterbody and NHDArea layers for the state of Maine.
3. Subtract them from the cnty24p data layer sequentially using the "Erase" tool in ArcGIS.
   1. erase "NHDArea", save polygon layer as cntylessarea.shape"
   2. erase "NDHWaterbody" from that, save as "MaineLandOnly.shp"
4. Save result as “MaineLandOnly” shapefile

# Step 2: Generate a Quarter Mile Buffer around Blueberry Fields

The data downloaded from USDA consists of a raster data layer depicting cropland in 30m x 30m pixels, statewide. For our purposes, we need to simplify these data, and buffer them.

1. Select Blueberry Land Only. Blueberries are Attribute Code = 242. I uses “Extract By Attribute” from the Spatial Analyst Toolset to extract only those raster points with that value
2. Select by Inspector Region. Use Extract By Mask to extract Blueberry Lands in Inspector Regions 2 and 3. To do this, you need to make sure the projection of the blueberry data and the projection of the Inspector Regions data matches. It also appears that this can only work if all the files are contained in a geodatabase, and the output goes to a geodatabase.
3. Convert the Blueberry data from regions 2 and 3 to a polygon shapefile, Using the "Raster to Polygon" tool.
4. Generate quarter mile buffers around blueberry polygons, using the "Buffer" Tool
   1. Linear Unit: 0.25 Miles
   2. Side Type: Full (This includes the ag fields themselves in the generated polygons, avoiding producing “donuts”).
   3. Dissolve Type: All
   4. Save as Blubry23Buf
5. Simplify the resulting buffer polygons. While this is not strictly necessary, given our sampling purpose, it probably makes sense. It should make the GRTS sampling go more quickly. It simplifies geometry of buffer areas, for example where buffers around two nearby fields intersect. It rounds out corners and fills in tiny holes. It does slightly alter the buffer geometry, so a few areas more than a quarter mile from agricultural lands might fall within these “simplified” polygons, and a few areas within a quarter mile won’t. But the overall impact on the area sampled area will be very small. We do this in three steps:
   1. Run the “Simplify Polygon” tool, with Simplification Algorithm = "WEIGHTED\_AREA" and tolerance = 0.05 mile
      1. Save as Blubry23BufSimp
   2. Run the "Eliminate Polygon Part" tool with AREA = 0.01 square miles (about 1/6 of the size of the smallest polygons) and "Eliminate contained parts only". This eliminates small “holes” in the buffer area.
      1. Save as “Blubry23BufSimpFill”
6. Clip that to the MaineLandOnly and save the result as “Blubry23Clip”.

# Step 4: Generate Final Area from Which to Draw Samples

We have a couple of final “Cleanup” steps before we’re done.

This workflow generated a single, complex multi-polygon, and we want to replace that with a polygon layer. This is easy to do, using the “Explode Multipart Feature” tool on the Advanced Editing Toolbar.

1. Select the “Blubry23Clip” feature layer
2. Right click, and on the context menu, select Edit Features
3. Open the Advanced Editing Toolbar (right click on any of the tool bars and a long menu with available toolbars appears).
4. Select any part of the multi-polygon layer.
5. Click on the Explode Multipart Feature button on the Advanced Editing Toolbar.
6. On the editor toolbar, select the editor menu, and save edits, then stop editing.

I also deleted some uninformative attributes, and recalculates area (“Shape\_Area”) and perimeter (“Shape\_Length”) fields. These steps are not strictly necessary.

# Step 5: Address Geometry Problems

It turns out, the shapefile we have just produced has some geometry problems. Before spsurvey runs the GRTS algorithm, it checks the geometry of all of the polygons that form the support for the random sample. R reported a “TopologyException” with “self intersections” and would not run successfully.

While we repair that, we’ll also simplify the polygon geometry , which makes it (slightly) easier to work with in R.

1. Toolbox -> Conversions -> To Geodatabase -> Feature Class to Feature Class Tool
   1. At bottom, select Environments
   2. Under “M Values”, set “Output has M values” to “Disable”
   3. Under “Z Values” set “Output has Z values” to “Disable”
   4. Save the resulting feature class as “Blubry23Flat.shp”

I tried two approaches for solving the geometry problems. The obvious one (shown next) failed, but an alternative worked.

1. Run the “Repair Geometry” tool on the “Blubry23Flat.shp” feature layer.

At this point, I reran the R script, but it still crashed GRTS with complaints about self intersections.

Following suggestion from a post online, I Buffered the shapefile one more time. A one meter buffer makes (almost) no difference in sampled area, but resolves many geometry errors.

1. Under “Geoprocessing, select the Buffer Tool. Apply to Blubry23Flat.shp”.
   1. Buffer = one meter
   2. side-type = FULL
   3. Dissolve Type = NONE
   4. Save result as “Blubry23Final.shp”

# Step 5: Run GRTS Algorithm

The accompanying R Markdown document (and the WORD version of it) contains code for generating a simple random sample of points within the polygons defined in the shapefile “Blubry23Final.shp”.

That R NotebooK will output a point shapefile, named “Blubry23RandomSD.shp”.

# Step 6: Reproject the Results and generate WGS 1984 Coordinates

We have been operating fast and loose with spatial references. The data I pulled from the Maine office of GIS is in UTM coordinates, while the data from USDA is in an Albers equal area projection. ArcGIS handled interconversion among spatial references quietly in the background, which was accurate enough for our purposes.

But, the GRTS sample used the same spatial reference as the polygon layer, so our output was in UTM. While that’s O.K., it’s better if we can generate coordinates for direct use with GPS and Google Earth. Both GPS and Google Earth use unprojected digital latitudes and longitudes, using the WGS 1984 datum.

We want to:

1. Remove unneccesary attributes (by hand). Make sure not to remover the “panel” attribute. For this simple unstratified design, “mdcaty”, “weight”, and “stratum” contain only single values and so can be removed.
2. To project the data to WGS 1984:
   1. Toolbox -> Data Management Tools -> Projections and Transformations -> Project
   2. Select WGS 1984.
   3. I saved results back as “BlubryRandomWGS84.shp”
3. Add columns to the attribute table for latitude and longitude (in decimal degrees) to the attribute table.
4. Use “Calculate Geometry” to find the relevant X and Y coordinate values.

# Step 7: Export Text File (tab- delimited text)

“Simple\_Random\_Sample.txt”