Rebecca Ubalde

Professor Kling

GEOG-485

23 October 2023

Final Term Project

**GEOG-485: Final Term Project**

**Creating a Point Feature Class from an Excel File Table and**

**Performing Density-Based Clustering on the Output Point Feature Class**

For my final project, I customized a script tool that generates a point feature class from coordinate points stored in a file table. The point feature class output is subsequently used to perform density-based clustering. Converting tabular data into spatial points then performing spatial density-cluster analysis on them is useful because it allows for the visualization and uncovering of patterns in data that may otherwise not be obvious. In this case, the custom script tool provides insights into the spatial distribution of botanical seed collections made between 2002 and 2017. For the cluster analysis, specific parameters were chosen for the clustering method, minimum feature cluster, search distance, and search time interval.

The clustering method chosen for the analysis was the density-based spatial clustering of applications (DBSCAN) method. This clustering technique identifies both cluster groups and outliers and aids in the identification of critical regions with high or low seed collection records in the Mojave Desert. Density-based clustering is especially critical to seed collection teams, who use the information to identify priority regions for seed collections, and to wildland restoration teams, who use the data to identify seed sources for regional restoration efforts. The minimum feature cluster parameter indicates that at least two plant collection points are required to be considered a valid cluster. The search distance parameter was configured to 5 miles, a typical distance at which individuals of the same species are considered part of the same population. Finally, the search time interval parameter was configured to 15 years, encompassing the entire time frame that seed collections were made, thus capturing the entire dataset.

The final deliverables submission includes:

1. Two scripts: 1) the original test script with hard coded values (FinalProjectScript1.py) and 2) the script used for the final script tool (PointsToClusteringScriptTool.py)
2. Toolbox titled *FinalProject* containing the *PointToClustering* script tool
3. Excel file titled *Collection\_Data\_Clean* for testing the script tool
4. Geodatabase named *FinalProject.gdb* which is the original workspace containing the results from the original script and script tool
5. Final Writeup

There are nine mandatory parameters in the *PointToClustering* script tool:

1. **Workspace:** this parameter allows the user to specify the working directory where the script will execute and where the output feature classes will be saved. The data type specified for this parameter is *workspace.*
2. **Input Table:** this parameter allows the user to provide the table containing the coordinate points to be converted into a point feature class. The table must contain two fields that hold the X and Y coordinates. The data type specified for this parameter is *table.*
3. **Point Output Feature Class:** this parameter allows the user to specify the name of the point feature class created from the coordinate points in the input table. The data type specified for this parameter is *string.*
4. **Cluster Output Feature Class:** this parameter allows the user to specify the name of the output feature class that results from the spatial clustering. The data type specified for this parameter is *string.*
5. **Cluster Method:** this parameter allows the user to choose the spatial clustering method used in the analysis. The data type specified for this parameter is *string.*
6. **Minimum Features Per Cluster:** this parameter allows the user to define the minimum number of features required for a cluster to be considered valid. The data type specified for this parameter is *long.*
7. **Search Distance:** this parameter allows the user to set the maximum distance within which the script will search for nearby points to form clusters. The data type specified for this parameter is *linear unit.*
8. **Time Field:** this parameter specifies the field in the input table that contains the time and date information associated with each point. The data type specified for this parameter is *string.*
9. **Search Time Interval:** this parameter allows the user to specify the time interval within which the script will search for nearby points within the temporal range. Only points within this time range will be used for forming clusters. The data type specified for this parameter is *field.*

To test the code and perform the cluster analysis, open the script tool titled *PointToClustering* in ArcGIS Pro. First, specify the workspace or geodatabase where the script will execute and save the files. Then specify the input table as *Collect\_Points\_Clean.xlsx.* Choose a string name for the Point Output Feature and the Cluster Output Feature. Select the spatial cluster method you want to apply for the analysis. I chose the *DBSCAN* method for this project. Other options include *HDBSCAN* and *OPTICS.* Specify the minimum number of points that must be present to be considered a cluster; in this case, I indicated that there must be at least two points present. Specify the search distance. I chose 5 miles because it is common to regard plants within this range as belonging to the same population; however, it is important to note that other factors may influence this classification. Then indicate the time field by defining the field name containing the date and time information. The field name found in the table should be *COLL\_DT*. Finally, define the search time interval as a string. Because the collection points were made within 2002 and 2017, I defined the search time interval as *15 Years* which successfully captures all the points (excluding outliers) into the density-cluster analysis. When running the script with the same parameters, the results should look like the results in the figure below.

A map of a city

Description automatically generated Figure 1 Results of the Coordinates-In-Table to Point Feature Class Conversion

A map of a mountain

Description automatically generated

Figure Results of the Density-Based Cluster Analysis

In summary, developing custom scripts is a flexible and effective approach for managing and manipulating spatial data. The ability to fine-tune certain parameters is particularly advantageous due to its adaptability, allowing users to specify unique parameters based on their individual research needs. In addition to the collection points, it would be beneficial to add land administrative boundaries, critical habitat, wilderness and protected lands, and historical wildfire burn scars to the map. These additional layers would help users identify agency jurisdictions and help on-the-ground crews navigate through different types of landscapes and ecological conditions. In addition to displaying high-collection densities, the map also reveals areas where collections are sparse. This scarcity can be attributed to a myriad of reasons including historically low precipitation, limited plant diversity, poor road conditions, permit requirements, or restricted access. Regions characterized by low collection densities could also influence the decision-making of natural resource management teams, leading them to consider prioritizing seed collections in these areas on years where there is abundant precipitation.