Analyzing Species Through Data

Abstract:

Understanding the various habitat preferences of the biological community especially plants has long been an interest of researchers. A species examined by this analysis is little brown jug (*Hexastylis arifolia*), an herbaceous perennial with a unique flowering structure. The data collection took place using ESRIs Arc Collector, a mobile application for precision field work. The field observations took place on a publicly owned parcel. The habitat description for little brown jug includes a soils component and during field investigations biologists noted a landscape position component also. During the analysis we will attempt to better understand the habitat desired at this site. During the course of the evaluation over 200 data points were collected that attempted to understand the forest structure and presence or absence of little brown jug. The publicly available datasets used in this analysis are, USDA Soil Survey and the US Fish and Wildlife Service's National Wetland Inventory (NWI). After the study more research is need to better predict occupied habitat.

Keywords

#Sustainability #Plants

#FieldCollection #Soils

Introduction:

The primary research questions addressed by this analysis are there trends in the general occupied habitat areas of little brown jug (*Hexastylis arifolia*). During the analysis the following subcategories of data will be discussed, the USDA Soil Survey and the USFWS National Wetland Inventory (NWI), of these additional datasets we will investigate their general size and intersections with field collected observations points. The interest in the topic is a desire to better understand species habitat modeling. While working as a biologist I have noticed the desire, interest and funding increases in the world of species modeling. Little brown jug was chosen as the target species for this analysis because it is relatively simple to identify and is commonly found with its habitat range. There is little significance of doing research on little brown jug itself, however there is significance is understanding species modeling. Understanding the habitats of certain species, especially those listed as threatened or endangered by the Endangered Species Act, is extremely valuable.

The value in developing species models and greater understanding of ranges that species occupy is related to the Endangered Species Act. All federal actions are demanded to evaluate appropriate habitat for threatened and endangered species by the Endangered Species Act. A federal action is in many cases a form of federal development, for example highway construction funded by the Federal Highways Administration or a permit issued by the US Army Coprs of Engineers. The problem that comes with modeling rare species is they are often hard to find or not in many numbers. This evaluation will try to understand the biogeography of a more common species in order to determine if having species models for rarer species have merit in being developed.

Research Context/Background:

The ultimate goal of this project is to increase our understanding of a target species biogeography within a defined study area. Biogeography has long been a interest researchers, it seeks to answer a basic question of why there are so many living things, why these living things are distributed the way they are, and are these patterns effected by human activities (Hominick, 2002). The origins of biogeography are essential to keystone research about ecology and evolutionary biology itself. Biologists have long wondered why species are present in some areas and absent in others. At the very beings of the evolutionary theory the geographic locations of species was considered to be a key factor in evaluating their origins (Darwin, 1859). The location of a species gives evidence to where it has come from and where it may have existed previously, having our data points of presence and absence allows us to understand the current distribution of the species and attempt to understand a pre colonial range.

The target species for the study is little brown jug (*Hexastylis arifolia*), a species that receives its common name from the unique flowering structure that resembles a jug. The herbaceous perennial plant is found throughout most of the Southeastern United States and is generally agreed to be common within its known range. The species habitat description is varied between different researchers and throughout their time. For example, the habitat described in a botanical guide published in 1964 describes deciduous or pine woods and swamp forests; essentially throughout North Carolina and other southern states (Radford, 1964). This description seems that this species has a large and varied habitat range thought the landscape. In later published reviews scientists seem to lower their confidence in their understanding of the habitat preferences with little brown jug. In the later review of the species biogeography the authors reveal that more study is needed on variation in habitat of *Hexastylis arifolia* (Gaddy, 1987). Scientist changing

there understanding of a species or any studied topic is not cause for concern in itself because science is a constantly evolving subject.

The leaf shape is arrow like and distinct with generally a darker shade of green and white mottling. A photo taken during the field observation is provided for reference (Figure 1). During the field efforts GPS points were taken at the location of little brown jug locations.

Figure 1:(Photo Credit: Matt Martin 5/5/2021)

In addition to the photos taken of the species location a general description of the forest structure was provided. The forest structure data was separated into, “mixed pine hardwood forest” & “pine forest”. This data was included to understand if the species had a general canopy species preference.

Materials and Methods:

The data was gathered and processed through the data life cycle eventually becoming the DataPoints_Clean file. Below is the detailed process of the data life cycle for this project. Through the field investigation over 200 data points were collected, each point includes a statement about the target species presence or absence, the overall forest structure and a photograph.

Plan:

The plan for the data collection and project itself first involved picking a target species. The target species was chosen based on its abundance and likelihood that it would be observed during the collection phase. The second part of the planning phase included how the data will be

collected, for this project I utilized ArcCollector. The third part of the planning phase focused on what parts of data to collect, I chose to collect data on forest structure, species observation and a photo.

Collect:

The collection phase involved a field effort where a dataset of over 200 spatial referenced photos and other attributes were collected. This effort utilized the ArcCollector application which was blue tooth connected to a Trimble R1 GPS receiver. This GPS receiver was used to increase the accuracy of the data points improve the overall data quality of the field effort. Beyond the traditional advantages GPS technology adds to a project the Collector application allows users to use a state-of-the-art technology. Researchers comparing collection application reviewed ArcCollector to be a game changer for biologists (Nowak, 2020).

Assure:

The quality control section of the data life cycle is curcile to delivering a well made project. One assurance method that was used was a constraint within Collector. Within the collector application a constraint was set at 5 feet on the accuracy of the points. The collector application averages three GPS points that are below the 5-foot constraint as an additional assurance that the data is the correct accuracy. An additional constraint described in the collection method is the utilization of the Trimble R1 to insure the accuracy of the data.

Describe and Preserve :

To insure the longevity and the appropriate uses the data was described in the form of metadata. To create the metadata file the USGS Metadata wizard was used, where a accurate description of the different fields was provided and any acronyms used in the primary data set. This metadata also serves to preserve the data for long-term use. One legal issue that was

considered when collecting this data was the type of target species. I chose a fairly common species that does not have known pressures of poaching for ethical and legal reasons. Many rare and all endangered species have special legal considerations that should be thought about before publishing their locations.

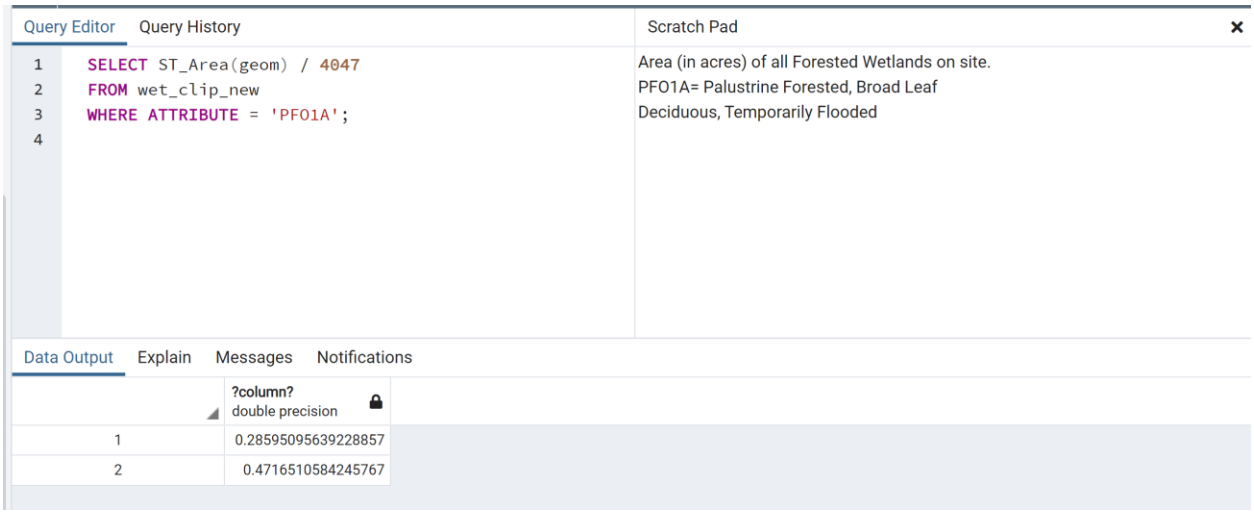
Discover, Integrate, and Analyze:

Many tools currently exist to complete these final portions of the data life cycle. In order to complete this on this project I used, PostGIS for analysis and ArcPro for analysis. For the discover and integrate portions Arcgis Online was used for its ease of share and the data is made available on the Open Science Framework (OSF). Researchers state that OSF can be an excellent tool for promoting best practices around reproducibility, transparency, and research data management (Dodge, 2021). Utilizing programs that allow accessible data review and sharing enhances the scientific communities knowledge base as a whole.

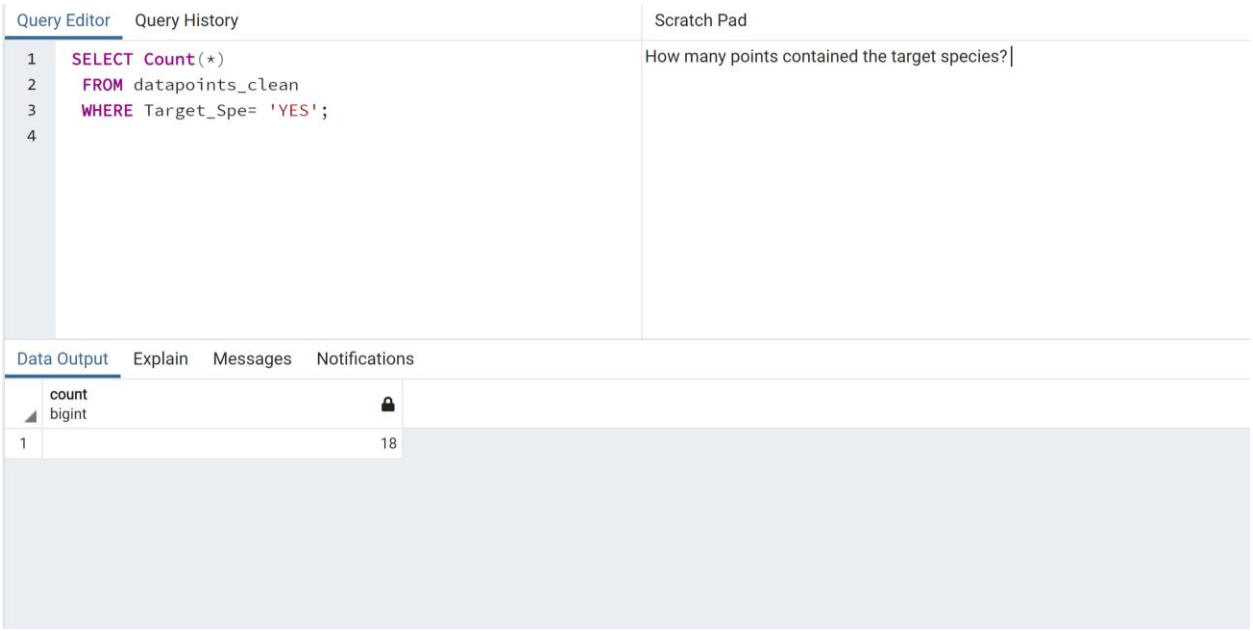
Results:

During the analysis phase of the project several basic questions were answered about the primary dataset and the general site conditions.

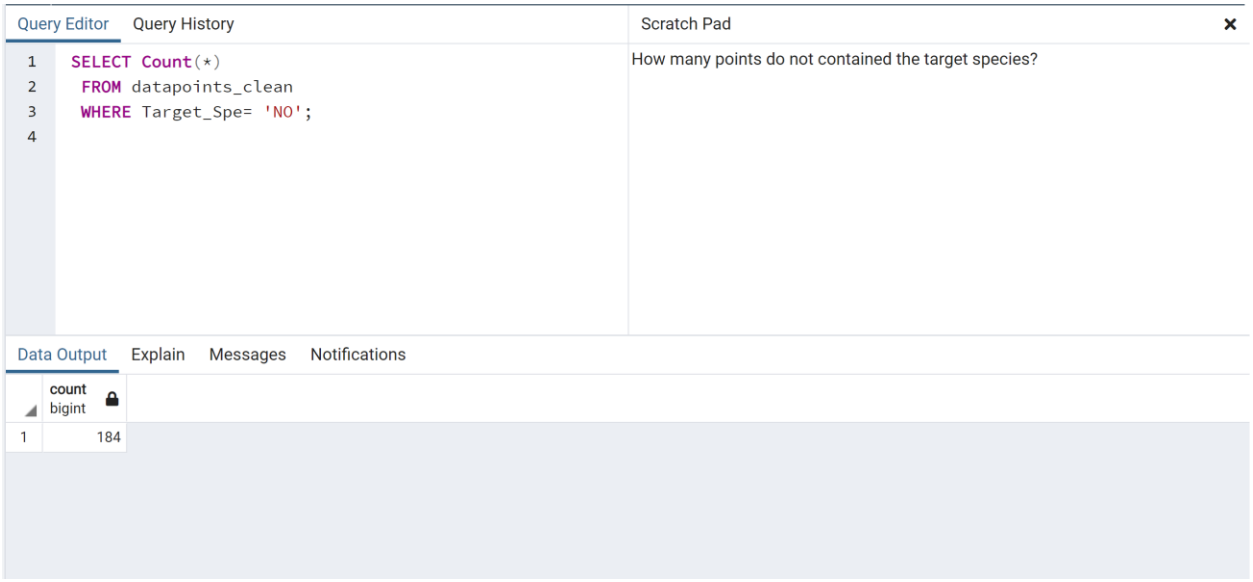
1- The area of the site that is forested wetland.



2- How many observation points contained the target species



3. How many observation points do not contain the target species



4. The locations of the target species

Query Editor | Query History

```
SELECT Target_Spe, ST_AsText(geom)
FROM datapoints_clean
WHERE Target_Spe='YES';
```

Scratch Pad: What are the locations of the "YES" species found locations

Data Output | Explain | Messages | Notifications

	target_spe character varying (254)	st_astext text
1	YES	POINT(-80.85840702999997 35.52833442000008)
2	YES	POINT(-80.85059870999999 35.52754956000007)
3	YES	POINT(-80.85065499999996 35.528091920000065)
4	YES	POINT(-80.85041239999998 35.528669440000044)
5	YES	POINT(-80.84933395999997 35.52840493000008)
6	YES	POINT(-80.84944367999998 35.528203190000056)
7	YES	POINT(-80.84976335999994 35.52781741000007)
8	YES	POINT(-80.85894893999995 35.53175868000005)
9	YES	POINT(-80.84641460999995 35.53061630000008)
10	YES	POINT(-80.84996638999996 35.529472070000054)
11	YES	POINT(-80.84995644999998 35.529789770000036)
12	YES	POINT(-80.84996925999997 35.530088730000045)
13	YES	POINT(-80.85014906999999 35.53051062000003)
14	YES	POINT(-80.85059289999998 35.530408230000035)
15	YES	POINT(-80.85073116999996 35.529718520000074)

5. The area of the soil series MeB2

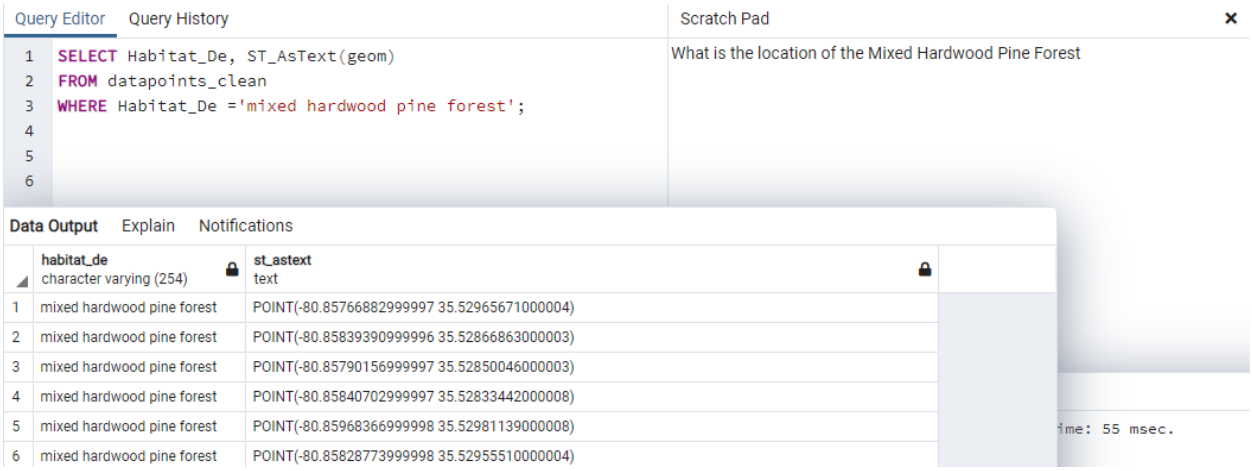
```
SELECT ST_Area(geom) / 4047
  FROM soil_clip_wgs
  WHERE MUSYM = 'MeB2';
```

How much area (acres) is MeB2?

Data Output | Explain | Notifications | Messages

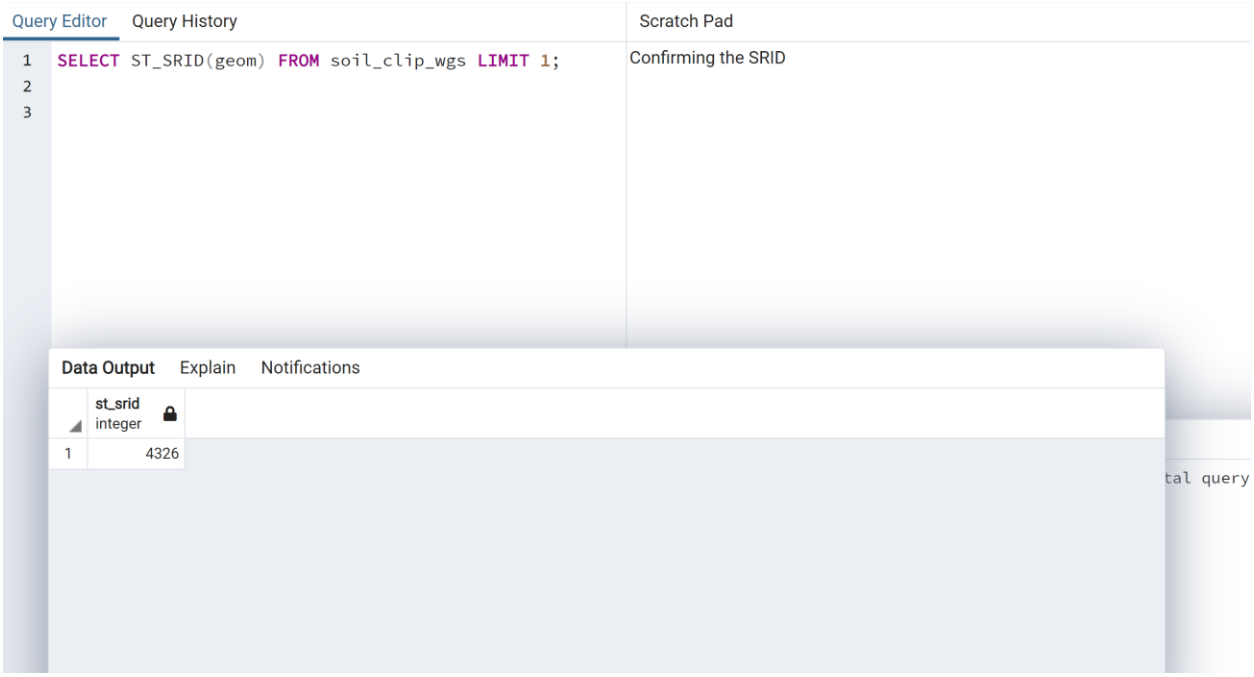
	?column? double precision
1	17.266930082179194
2	7.0742516857632785
3	7.715567619104421

6. The location of the mixed pine hardwood forest observed in the field.

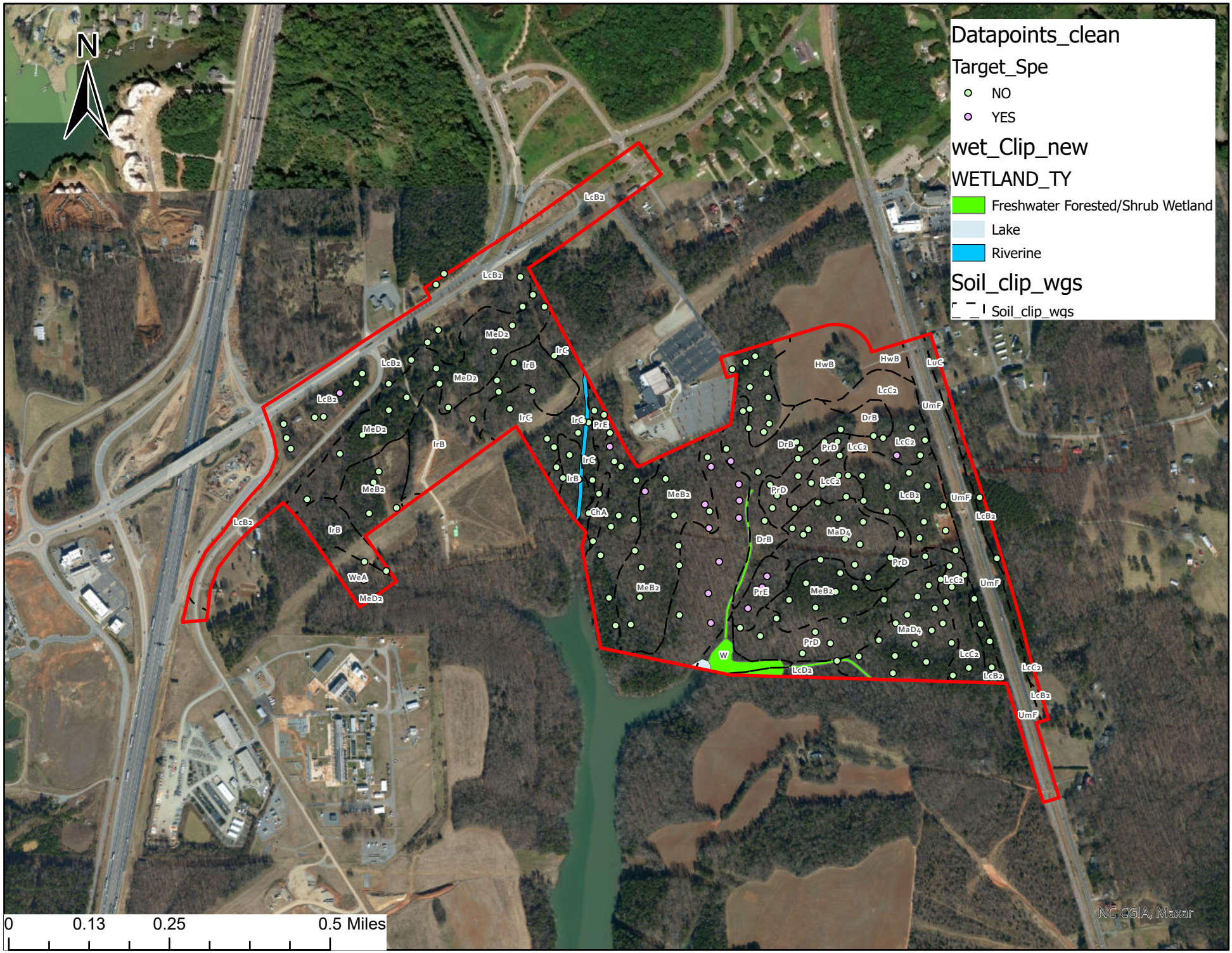


7. Confirming the SRID of the primary dataset

8. Confirming the SRID of the NWI clipped dataset



9. Figure 2: Map of the project.


N
Datapoints_clean
Target_Spe
NO
YES
wet_Clip_new
WETLAND_TY
Freshwater Forested/Shrub Wetland
Lake
Riverine
Soil_clip_wgs
Soil_clip_wgs
LcB2
MeD2
IrC
IrB
PrE
ChA
HwB
LuC
LcC2
UmF
DrB
PrD
MaD4
W
LcD2
WeA
0
0.13
0.25
0.5 Miles
NC CGIA, Maxar

Discussion:

The data was analyzed through PostGIS and ArcPro. Within PostGIS the data was reviewed for constancy and better understand for general geographies. The results section in analysis #2 and #3 was able to present data on the amount the observations of the target species and the inverse of the amount of absent observations, this analysis confirmed that the data was collected consistently and no data was missing. The analysis #1 was interesting to understand the total amount of forested wetland on the site, as seen in figure 2 none of the target species was present within the forested wetland. The area with the most occupied habitat was the soil unit MeB2, the analysis #5 revealed the total area of this soil survey is comprised on three separate polygons who ~32 acres. Another useful SQL query used was the shown in analysis #7 and #8 where we confirmed that the SRID is 4326 (WGS 84), having the various data sets in the same coordinate system is crucial to completing a accurate analysis.

The majority of the occupied habitat seen in figure 2 was located in areas described as mixed pine hardwood forests. This aligns with the literature review analysis #6 shows the location of the various pine hardwood forests. Another interesting analysis was the #4, where we could better understand the locations of the target species. In addition to completing the eight SQL queries figure 2 was also included as a map created in ArcPro. Having a visual understanding in addition to the queries helps better understand any potential trends.

Conclusion:

The conclusion of the project that the most occupied habitat on the project was on the east banks of the western most drainage feature. This population was commonly found on this slope and was entirely composed within a mixed pine hardwood forest. The soil series for this population was the MeB2. For this project site it is hard to discount areas of mixed pine

hardwood forest as potential habitat but after the field investigation we can understand that this habitat is unoccupied. There could be several reasons for this including anthropogenic activity. After the completion of this project I do not yet believe we have a large enough knowledge base to accurately predict the location of occupied habitat. More research will be needed to fully predict occupied habitat. Both the Radford and Gaddy habitat descriptions appear vague enough to still be considered correct, I believe they wrote these descriptions intentionally broad understanding that this species habits are difficult to understand.

Bibliography

Darwin, C. (1859). *The Origin of Species* .

Dodge, S. (2021). *A data science framework for movement.* Geographical Analysis.

Gaddy, L. L. (1987). *A Review of the Taxonomy and Biogeography of Hexastylis (Arisolochiaceae).* Morgantown, WV: Southern Appalachian Botanical Society .

Hominick, W. M. (2002). *Biogeography.* Entomopathogenic nematology.

Nowak, M. M. (2020). *Mobile GIS applications for environmental field surveys: A state of the art.* Global Ecology and Conservation.

Radford, A. E. (1964). *Manual of the vascular flora of the Carolinas.* Chapel Hill: University of North Carolina Press.