Retrofitting GIS to Enhance A Longstanding/Non-GIS Study

In Association With:

Utility Rights-of-Way Wildlife Research at Penn State

Grant Wills

The Pennsylvania State University

Masters of Geographic Information Systems

May 2018

Introduction:

Although it is no secret that Geographic Information Systems (GIS) can be used in a broad range of applications, virtually in any field that contains a spatial component, GIS is still a relatively new tool that has not reached its full potential. Through many different efforts, from advances in technology to the education of users and consumers alike, GIS has become an obtainable tool to be used in our everyday lives, it is no longer reserved for only the highly trained users. One area that benefits greatly from increasingly available GIS services are scientific studies. In the past GIS was used sparingly to enhance or drive a scientific study outside the field of geography, but as GIS percolates into the common place there are an increasing number of applications for its use. Since it is common practice for many scientific studies to occur over long periods of time, it is important to understand that GIS can be implemented as an ad-hoc service. Although it is generally easier to implement GIS into a new study, retrofitting GIS into an ongoing study can be done. The following sections will discuss the process of retrofitting GIS into a longstanding non-GIS study. The case study that will drive the narrative of this paper is an ecological study looking at the ecological affects that different vegetation management techniques have on utility rights-of-way (ROW).

Background:

In 1953 William Bramble & W.R. Byrnes began researching the ecological effects of utility ROW at what would later be known as the State Game Land 33 (SGL33) study site (The Pennsylvania State University, n.d.). "Today, SGL 33 is the site of the longest continuous study measuring the effects of herbicides and mechanical vegetation management practices on plant diversity, wildlife habitat, and wildlife use within a ROW" (The Pennsylvania State University, n.d.). That being said, in the 60+ years that the site has been studied the only tangible GIS data that were created prior to attempting a GIS retrofit were incomplete outlines of the individual treatment zones within the study site. On the contrary the amount of non-GIS data collected could best be described as mountainous.

To clarify the scope of this retrofit, although research has been conducted on the SGL33 site since 1953, the research can be broken down into a series of supporting but separate projects. The most current project began in the fall of 2015 and extended into the fall of 2018. This round of study marks the first time any GIS was a planned objective, albeit a very basic visualization of the study site, and respective data overlay (Wild, 2016). In the fall of 2016 I was introduced to this project, at this point the rough polygons representing the treatment zones had been created, however at nearly a year into the project there was very little to show for it in the GIS. On the contrary, there had already been lots of ecological data collected on: pollinators, birds, and plants. After a series of correspondences with the lead researcher, Dr. Carolyn Mahan, she was hopeful that I could go beyond simply mapping the study site, and rather produce some GIS products that were focused on the research data that they were actively generating. With those desires voiced the GIS retrofit was born.

To better define a GIS retrofit: it is the late addition of GIS to a current project, in such a way that supports the original objectives. It is not a separate GIS study focused on its own goals, and it must run parallel to an active study. Mapping historic data collected at the site would not be considered a GIS retrofit because the GIS work would not be conducted in parallel with an active study, it would have its own goals, and it would not have an impact on how the study was conducted.

Objectives:

The objectives of the GIS retrofit were simple. Convert all current ecological data generated from field data collection into GIS data. Produce static maps that could be used to communicate basic information about the SGL33 study site. Produce a dynamic mapping product that focuses on the ecological data collected. Create custom tools for ESRI's ArcGIS software to aid future projects involving SGL33.

Methods:

To build an understanding for what the researchers were studying a thorough literature review was conducted prior to any GIS work. At the completion of the literature review, GIS work commenced. However, before any data were manipulated a rough data structure was conceived. The data were to be stored in ESRI file geodatabase (FGDB) format, whenever possible, with at minimum three separate FGDBs. One FGDB would be to contain vector data, points, lines, polygons, a second FGDB to contain Raster data (Imagery), and a third FGDB to serve as a workspace, where data could be worked on before being archived into its respective FGDB.

Once a general data management plan was in place data was collected from all available sources. The data collected would serve primarily as foundational information to help set the scene so that the study data could be better understood. The majority of this foundational data was collected via PASDA's (Pennsylvania Spatial Data Access) website. The following is a partial list of some of the major data providers that are responsible for the foundational data:

- Centre County GIS
- National Lands Trust
- Pennsylvania Department of Conservation and Natural Resources
- Pennsylvania Department of Transportation
- Pennsylvania State Game Commission
- United States Department of Agriculture
- United States Geological Survey

The foundational data collected was not built with the scale of the SGL33 study site in mind, so to minimize the overall size of the data, and to increase the processing speed of the data a bounding box was established that fully encompassed the SGL33 study site. Most of the foundational data was clipped to the geometry of the bounding box. The overall size of the bounding box is: approximately 20,000 feet x 20,000 feet, or approximately 14.25 square miles, which is large enough to reduce the likelihood of introducing errors of analysis on account of edge effects. The bounding box's size also accounted for more cartographic freedom so that complete mapping products did not appear to be holding back any information from the viewer by displaying areas of no data. For some more generalized datasets, *i.e.* roads, both a clipped and unclipped version of the data were retained so that if a small-scale mapping product was needed there would be continuity between the large-scale and small-scale data.

In addition to the foundational data gathered, new GIS data was generated using tabularized data collected by the research team concerning the species richness of: vegetation, birds, and pollinators (primarily bees). To convert the data from tabular to ESRI feature class format the data was joined to a polygon feature dataset that represents the various treatment zones across the study site based on the key field of unique treatment zone identification codes. However, before the join was conducted the raw tabular data had to be manipulated to conform to ESRI's table standards, *i.e.* field headings could not have spaces or special characters. Additionally, to enrich the data collected, a weblink was added to every table record concerning species specific information. Species richness metrics were also added to the tabular data prior to being joined to the treatment zone feature dataset.

There was a large amount of effort placed in the procurement and generation of metadata for all gathered and generated GIS data layers. This effort involved viewing and editing the metadata file for each feature class, and populating as much information as possible to adhere to the Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) metadata model.

Through the process of collecting and creating all of the GIS data for the project, modifications were made to the data management plan. Three additional FGDBs were added as well as three standard Microsoft file folders. The resulting file structure can be viewed in *Figure 1*.

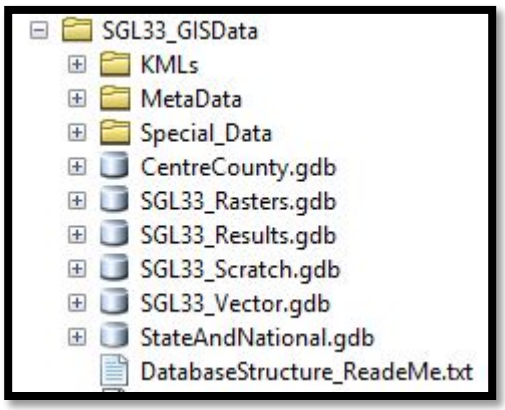


Figure 1: Image of the Data Structure for all of the SGL33 GIS Data as viewed from ESRI's ArcCatalog. For additional information on each element of the data structure see Appendix A which has the full text of "DatabaseStructure_ReadMe.txt"

With the data structure complete and populated, the remaining efforts were focused on generating mapping products. Several of the desired mapping products were static maps. All of these maps were produced using ESRI's ArcGIS for desktop 10.X. The static maps were formatted to be printed on a standard letter sized sheet of paper to increase the printability of the products. A dynamic web mapping application was also produced; however, it was produced using ESRI's ArcGIS Pro. ArcGIS Pro was used because its increased functionality and integration with ESRI's ArcGIS Online web mapping venue. By using ArcGIS Pro, the entire web map was constructed as a single mapping document and then published directly from ArcGIS Pro to ArcGIS Online. Once the web map was published to ArcGIS Online, by means of ESRI's Web App Builder interface offered through ArcGIS Online a web mapping application was formatted. The web map was reviewed several times by the research team. Each review helped refine the functionality of the web map.

With the completion of all of the requested mapping products the desired GIS tasks were completed for the project. The final process involved in the retrofit was passing the data forward. This was done using the file transfer site: https://box.com. All data and associated mapping products and documentation were uploaded to the research team's account on the file transfer site for final delivery.

Results:

At the completion of the project all stated objectives were met apart from the creation of custom tools for ESRI's ArcGIS. In total there were 8 static maps created for the project, and one dynamic map created by means of ESRI's ArcGIS Online. The static maps are presented in *Appendix B*. The web map can be viewed by viewing Penn State's Transmission Line Ecology website, within the Maps section: http://sites.psu.edu/transmissionlineecology/

Discussion:

From the surface this GIS retrofit seems very simple, gather a bunch of data, generate a few layers, and make a few maps, and then you are done. However, the devil is in the details. There are certain difficulties intrinsic to a retrofit that would not be present if the project was driven by GIS, most notably the way in which field data was collected. Had GIS been involved in the design of the project, field data collection could have been conducted so that location information of each observation could have been recorded. The only location data that existed for the field data was that the observation fell within a specific treatment zone. A treatment zone is several meters wide and could be over 100 meters long, meaning any spatial influences that may affect an observed species could not be analyzed due to the lack of adequate and standardized spatial resolution of the results. It is unrealistic to gather GPS coordinates for each observation of species such as birds, because they are often observed from a distance, however if the study site was divided into a grid and a color-coded flag or stake had been placed in the center of each grid as it exists on the ground, then the observers in the field could reference the grid the species is in at the time of observation by making a judgement call on which grid-marker the species is closest to. The data collected in this manner would have a standardized spatial resolution and a much higher level of spatial accuracy, which would allow for a greater amount of GIS analysis of the data.

An additional level of difficulty faced in retrofitting the collected data was the lack of consistency in the tabular formats of the data. This does not reflect poorly on the research done, as the collection was sufficient for the purposes of the ecological study as originally designed, however a GIS thrives on consistent data schemas, and thus posed a time-consuming hurdle to reformat the various tabular datasets generated from the field data collections. This is another example of, had a GIS professional been involved in the design of

the data collection procedures a standardized collection and table schema could have been implemented greatly reducing the amount of work prepping the data to be converted into a GIS format.

At the point when all of the data was collected and generated the project began to feel like a standard GIS project, and the generation of the mapping products went smoothly. However, the amount of time spent in data conversion pushed the schedule of the project so far to the right that the objective of creating custom GIS tools had to be taken off of the table. This was an unfortunate concession; however, the priority was to establish the use of GIS for future SGL33 studies. This could only be done if the proper amount of time was given to building a solid foundation of data, accompanied with documentation on how to use the data and where the data came from.

At the conclusion of the current round of study the SGL33 study site is now covered by 28 different raster datasets that provide historical imagery dating back to 1938, a digital elevation model, and various ecological and surface analyses. There are also 58 dedicated GIS vector layers covering the study site that range from highway shields to field data collection results. In addition to the SGL33 specific data there are more than 75 additional datasets that cover data at the County, State and National levels. This stock of over 160 data layers should allow for greater GIS exploitation of future SGL33 studies. As well as the conversion and potential GIS analysis of historic SGL33 data. In addition to setting a foundation for further GIS development at SGL33, the efforts of the GIS retrofit have had a positive impact on the current round of study. Specifically, with the development of an interactive web map that allows the research data to be displayed, for the first time, in a simple interactive graphical format.

References:

The Pennsylvania State University. (n.d.). Rights-of-Way Ecology at Penn State, History. Retrieved from http://sites.psu.edu/transmissionlineecology/history/

Wild, K. (2016). Update on Game Lands 33 and Green Lane Research. *The Asplundh TREE*, Autumn 2016.

Appendix A

Six File Geodatabases makeup the main data structure for all of SGL33's geodata. However, there are 3 additional file folders to help manage auxiliary data.

--Listed in order of most commonly used to least likely to use--

1.SGL33_Vector:

This database is the foundational database for all SGL33 core mapping and analysis needs. It houses all of the point, line, and polygon feature classes for the SGL33 study site. This database should store all finalized data, both schema and projection should be established before being added to this database. Editing of data should not occur within this database, instead a copy of the data should be placed within the Scratch database until it is finalized.

2.SGL33_Raster:

Like the Vector database this is a finalized geodatabase used for the analysis and visualization of raster datasets. Only finalized raster datasets with the extent of the study area should be stored within this geodatabase.

3.SGL33_Results:

This database is the location where any analysis data should be stored until it can be vetted and moved to the SGL33_Vector.gdb All SGL33 GIS applications/tools will push results to this database. If a new tool has been created then it should be set to push the output into this geodatabase. Once the data is finalized it should be moved accordingly, and there should be no copy of said data within the SGL33_Results.gdb. DO NOT use this geodatabase to store finalized data. Currently there is a Feature dataset named 'Clippings' this dataset must not be deleted as several tools within the SGL33_Toolbox are set to place data into this Feature dataset.

4.SGL33_Scratch:

This database is used to process data during any geoprocessing event. This database should be empty, except while a geoprocessing event is processing. All of the temporary files should be populated within the scratch database. DO NOT use this geodatabase to store finalized data.

5.CentreCounty:

This database contains data from Centre County PA at the scale of the County. All data within it originated from Centre County and can be retrieved on their Open Data access site. Use this data to supplement/visualize the surrounding area of the SGL33 study site.

6.StateAndNational:

This database holds data that has at least State-level coverage and can extend out to National-level coverage. This data is also used to supplement the SGL33 data.

Appendix A

__*__*__*__*__*__*__* FILE FOLDERS*__*__*__*__*__*__*__*__

1. KMLs:

This folder is where any KML files related to the study should be placed.

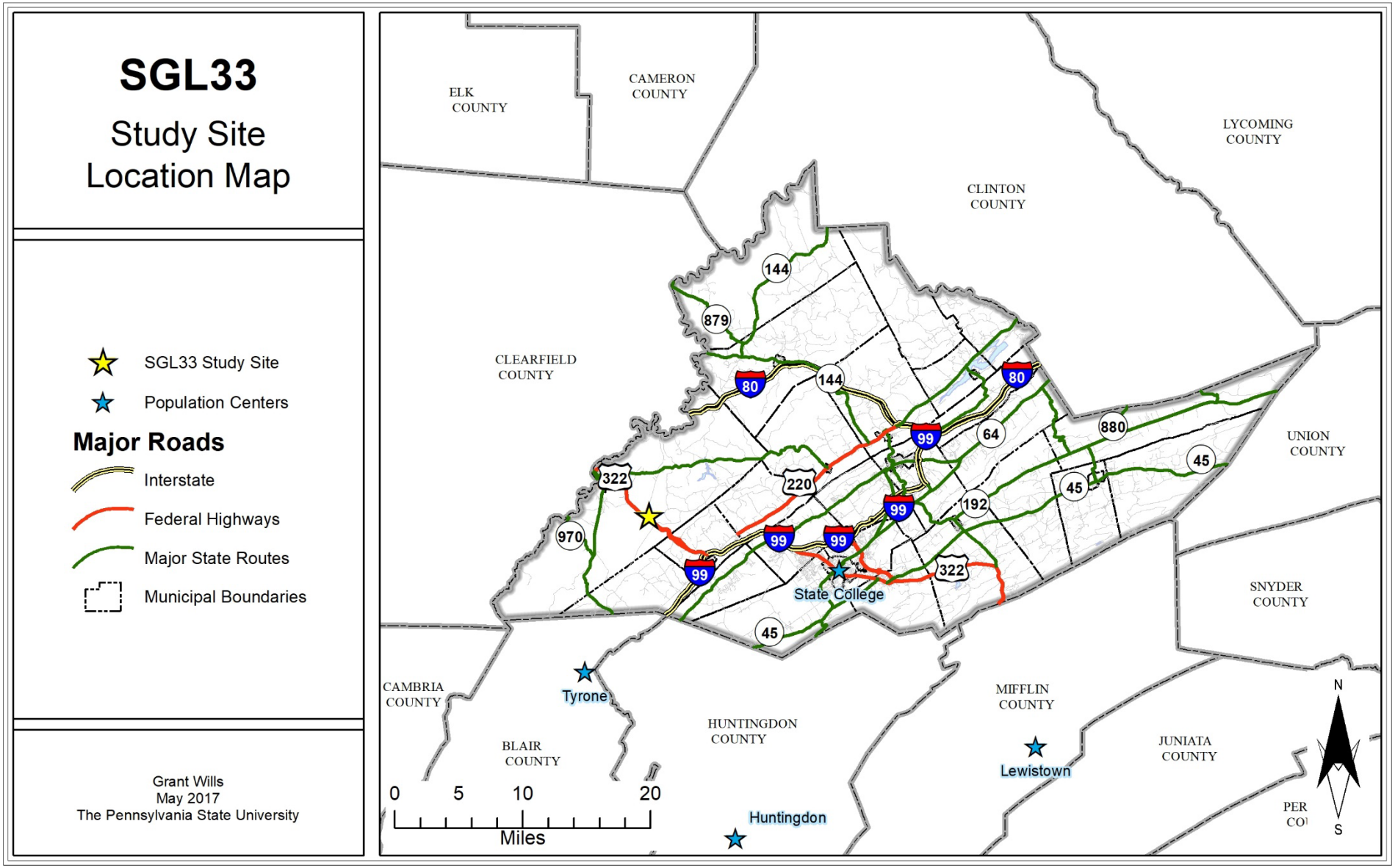
2. MetaData:

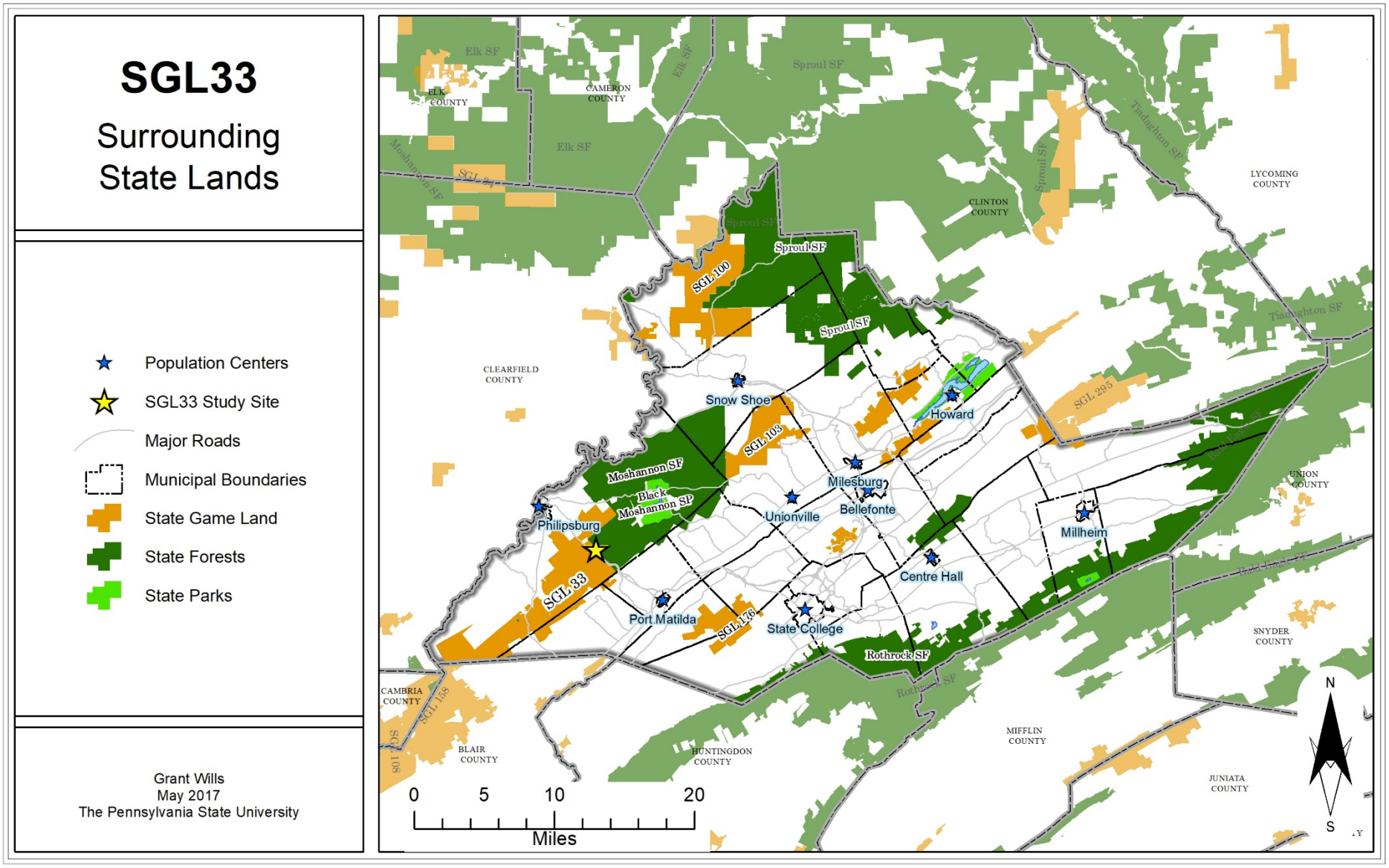
Any extra copies of metadata should be placed into this folder. Make sure to maintain good file naming standards so that the metadata can be referenced easily.

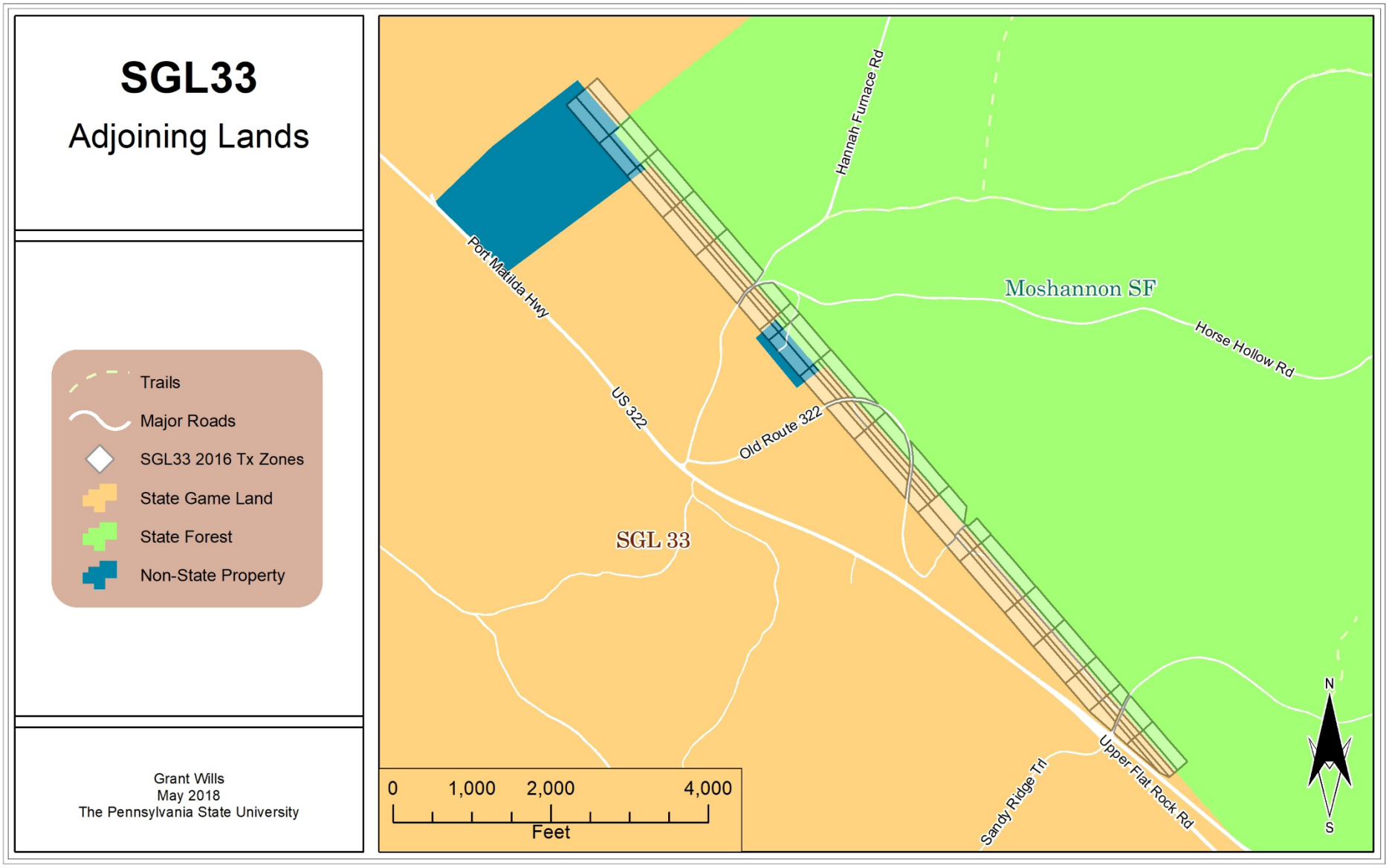
3. Special_Data:

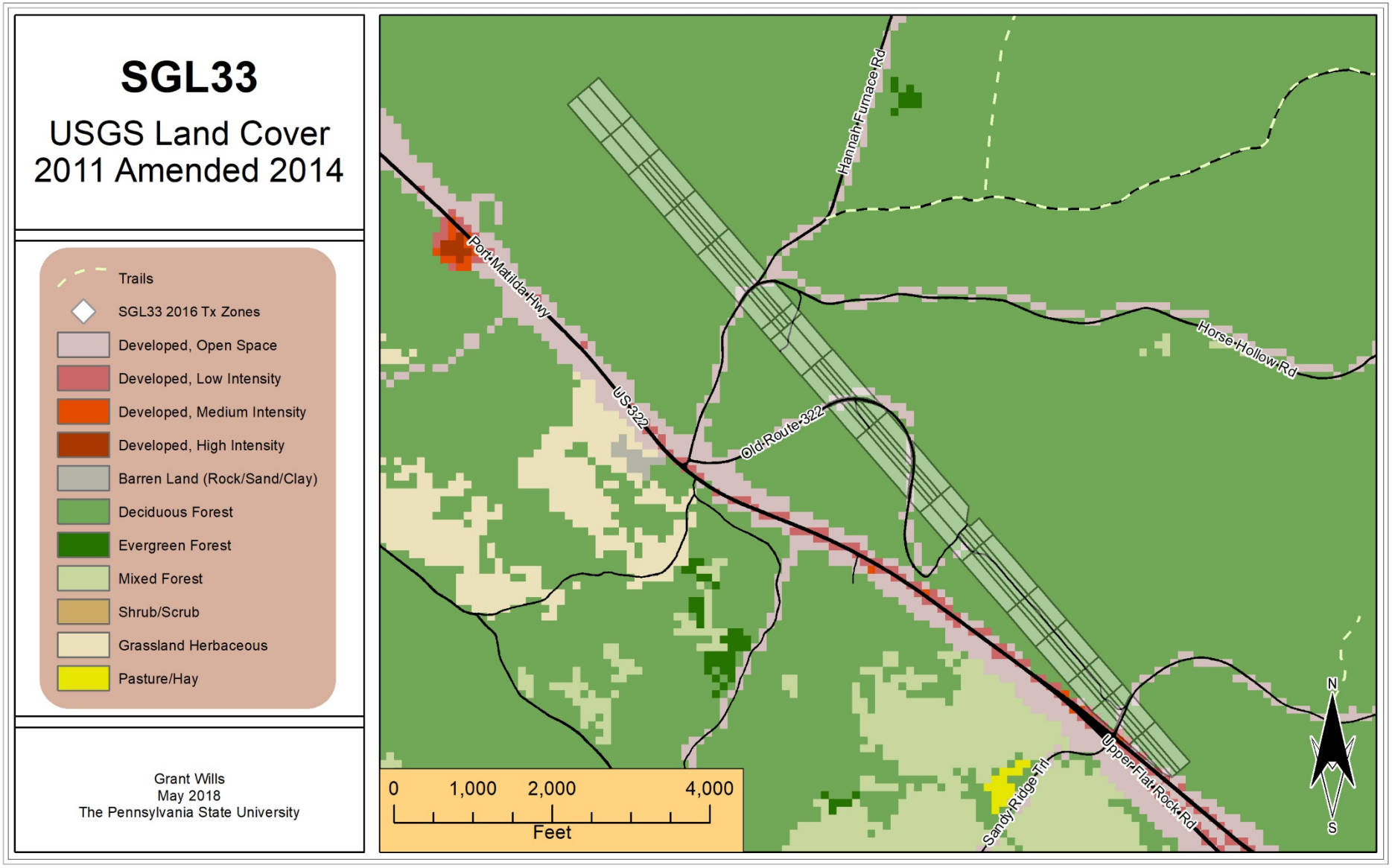
This is a file folder dedicated to obscure data that is not directly compatible with a file geodatabase, or KML. This folder currently contains LAS data but can be expanded with other misfit data. As always, make sure that any new data added to this folder follows appropriate naming standards.

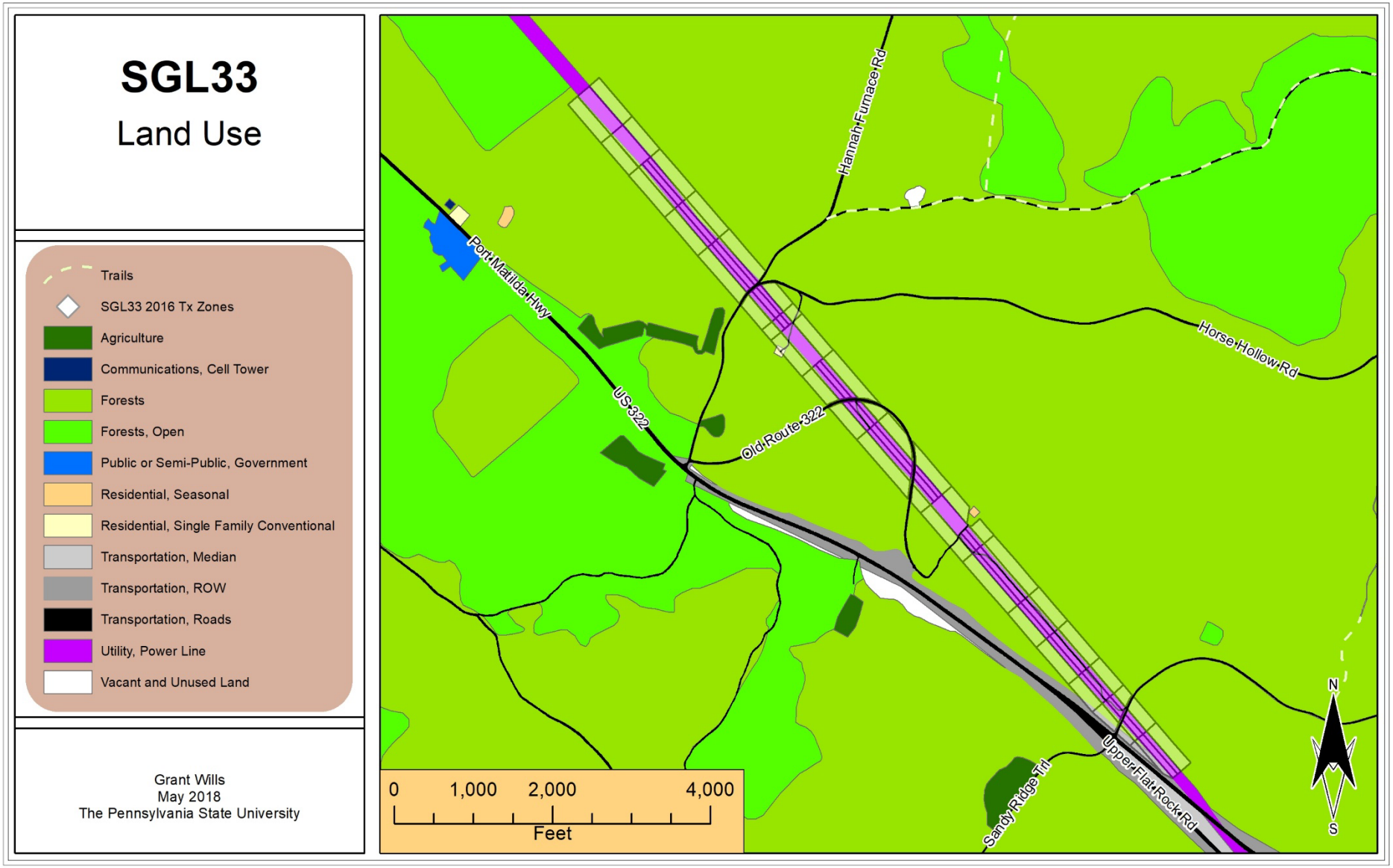
For any questions relating to the file structure, please contact Grant Wills:
gtw5016@gmail.com

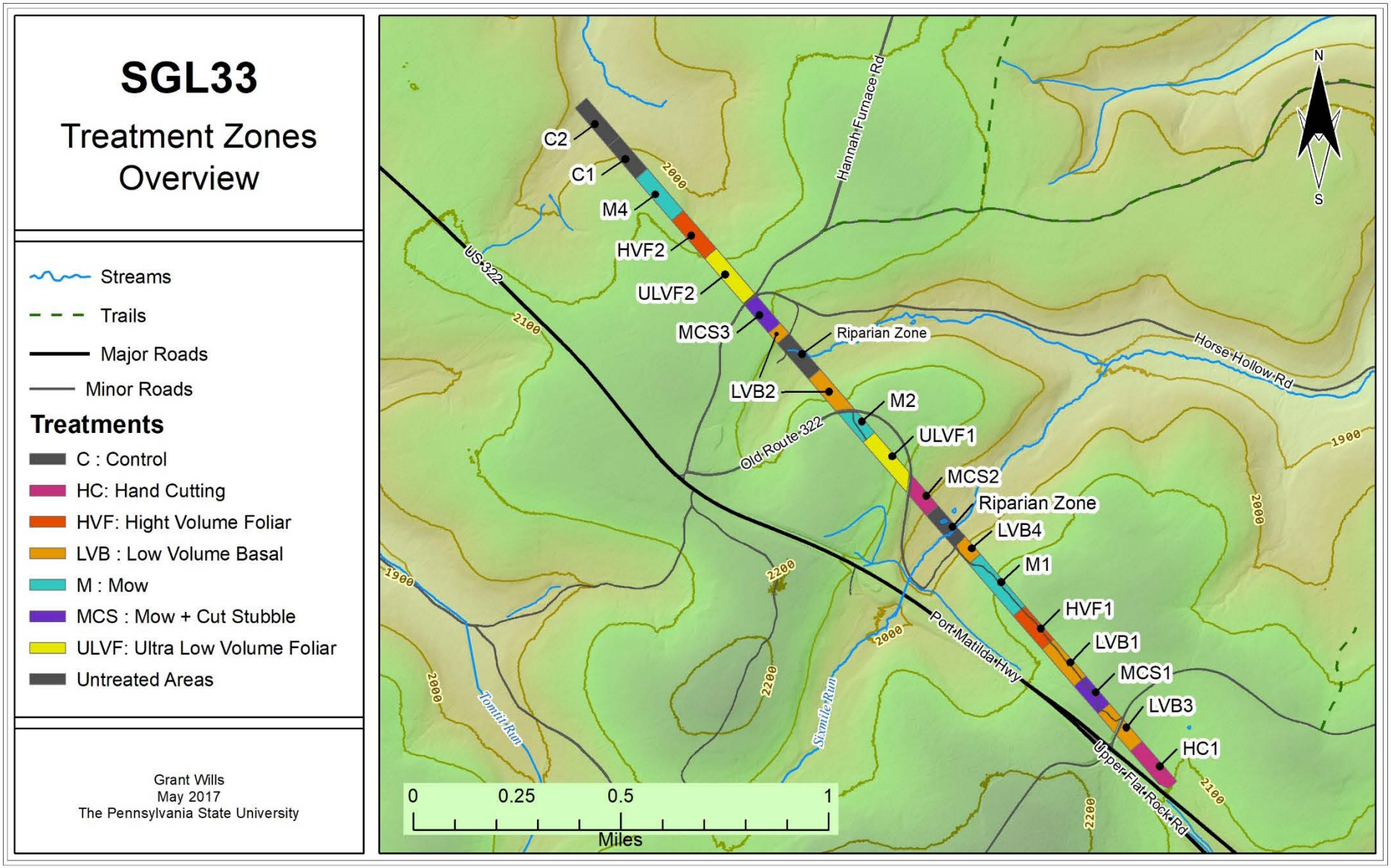

SGL33
Study Site
Location Map
SGL33 Study Site
Population Centers
Major Roads
Interstate
Federal Highways
Major State Routes
Municipal Boundaries
Grant Wills
May 2017
The Pennsylvania State University
ELK COUNTY
CAMERON COUNTY
LYCOMING COUNTY
CLINTON COUNTY
CLEARFIELD COUNTY
UNION COUNTY
SNYDER COUNTY
CAMBRIA COUNTY
BLAIR COUNTY
HUNTINGDON COUNTY
MIFFLIN COUNTY
JUNIATA COUNTY
State College
Tyrone
Huntingdon
Lewistown
144
879
80
64
880
45
322
220
192
970
99
0 5 10 20
Miles
N
S


SGL33
Surrounding State Lands
Population Centers
SGL33 Study Site
Major Roads
Municipal Boundaries
State Game Land
State Forests
State Parks
Grant Wills
May 2017
The Pennsylvania State University
Elk SF
Sproul SF
Tiadaghton SF
Moshannon SF
SGL 24
SGL 100
SGL 103
SGL 295
SGL 33
SGL 176
SGL 158
SGL 108
Black Moshannon SP
Rothrock SF
ELK COUNTY
CAMERON COUNTY
CLINTON COUNTY
LYCOMING COUNTY
CLEARFIELD COUNTY
UNION COUNTY
SNYDER COUNTY
MIFFLIN COUNTY
JUNIATA COUNTY
HUNTINGDON COUNTY
BLAIR COUNTY
CAMBRIA COUNTY
Snow Shoe
Howard
Milesburg
Bellefonte
Unionville
Philipsburg
Millheim
Centre Hall
Port Matilda
State College
0 5 10 20
Miles
N
S


SGL33
Adjoining Lands
Trails
Major Roads
SGL33 2016 Tx Zones
State Game Land
State Forest
Non-State Property
Grant Wills
May 2018
The Pennsylvania State University
Hannah Furnace Rd
Port Matilda Hwy
Moshannon SF
Horse Hollow Rd
US 322
Old Route 322
SGL 33
Sandy Ridge Trl
Upper Flat Rock Rd
0
1,000
2,000
4,000
Feet
N
S


SGL33
USGS Land Cover
2011 Amended 2014
Trails
SGL33 2016 Tx Zones
Developed, Open Space
Developed, Low Intensity
Developed, Medium Intensity
Developed, High Intensity
Barren Land (Rock/Sand/Clay)
Deciduous Forest
Evergreen Forest
Mixed Forest
Shrub/Scrub
Grassland Herbaceous
Pasture/Hay
Grant Wills
May 2018
The Pennsylvania State University
Hannah Furnace Rd
Port Matilda Hwy
US 322
Old Route 322
Horse Hollow Rd
Upper Flat Rock Rd
Sandy Ridge Trl
0
1,000
2,000
4,000
Feet
N
S


SGL33
Land Use
Trails
SGL33 2016 Tx Zones
Agriculture
Communications, Cell Tower
Forests
Forests, Open
Public or Semi-Public, Government
Residential, Seasonal
Residential, Single Family Conventional
Transportation, Median
Transportation, ROW
Transportation, Roads
Utility, Power Line
Vacant and Unused Land
Grant Wills
May 2018
The Pennsylvania State University
Hannah Furnace Rd
Port Matilda Hwy
US 322
Old Route 322
Horse Hollow Rd
Sandy Ridge Rd
Upper Flat Rock Rd
0
1,000
2,000
4,000
Feet
N
S


SGL33
Treatment Zones Overview
Streams
Trails
Major Roads
Minor Roads
Treatments
C : Control
HC: Hand Cutting
HVF: Hight Volume Foliar
LVB : Low Volume Basal
M : Mow
MCS : Mow + Cut Stubble
ULVF: Ultra Low Volume Foliar
Untreated Areas
Grant Wills
May 2017
The Pennsylvania State University
C2
C1
M4
HVF2
ULVF2
MCS3
Riparian Zone
LVB2
M2
ULVF1
MCS2
Riparian Zone
LVB4
M1
HVF1
LVB1
MCS1
LVB3
HC1
US-322
Old Route 322
Hannah Furnace Rd
Horse Hollow Rd
Port Matilda Hwy
Upper Flat Rock Rd
Tomtit Run
Sixmile Run
2000
2100
1900
2200
N
S
0
0.25
0.5
1
Miles

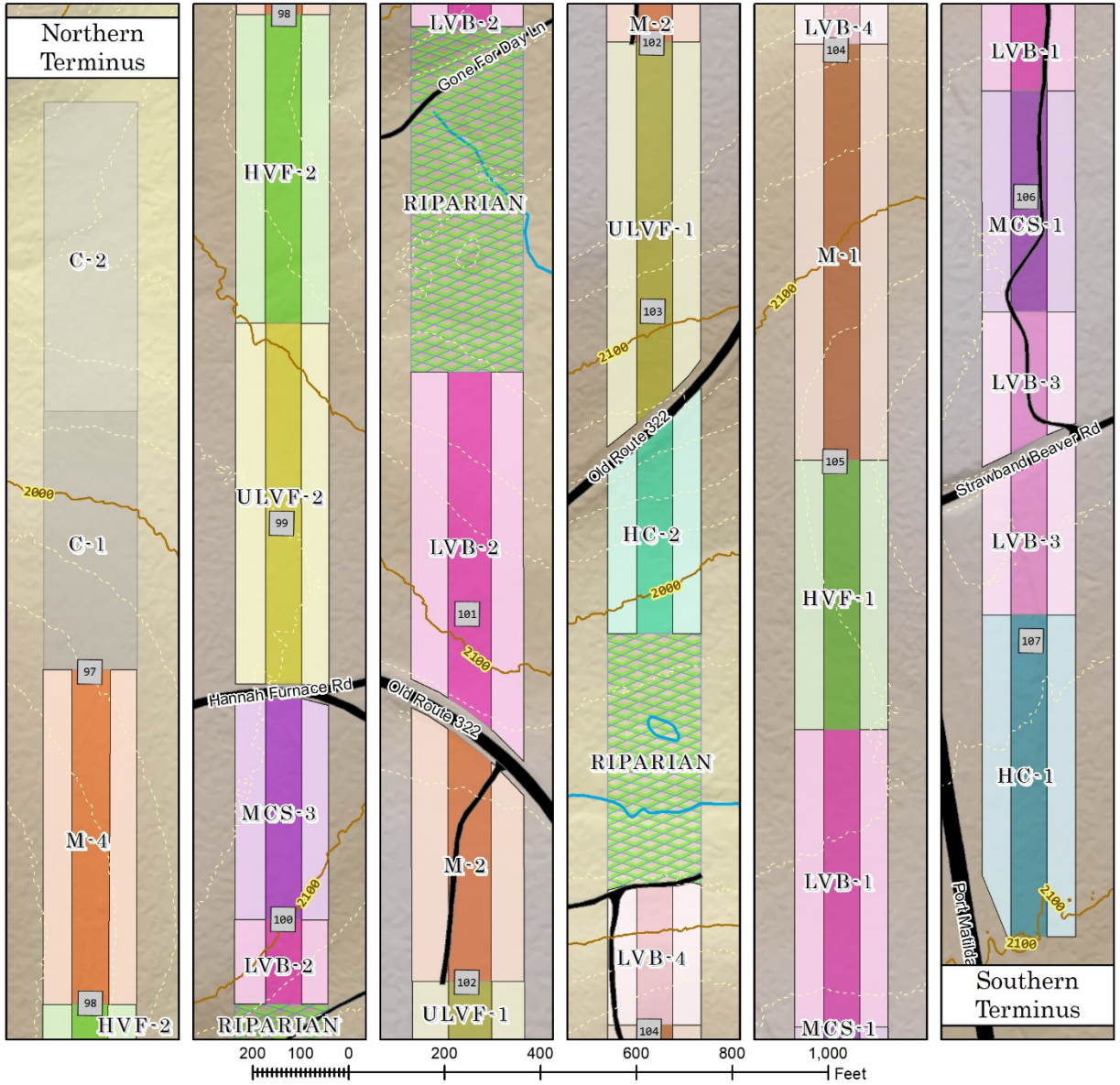
SGL33

2016 Treatment Map

Treatment (TX) Types

C = Control
HVF = High Volume Foliar
HC = Hand Cutting
LVB = Low Volume Basal
M = Mowing
MCS = Mowing+Cut Stubble
ULVF = Ultra Low Volume Foliar

TX Zone	Acres	2016 TX
HC-1	2.93	HC
LVB-3	2.76	LVB
MCS-1	2.01	ULVF
LVB-1	2.7	LVB
HVF-1	2.45	ULVF
M-1	2.01	ULVF
LVB-4	1.71	LVB
HC-2	2.26	HC
ULVF-1	2.93	ULVF
M-2	1.81	ULVF
LVB-2	3.78	LVB
MCS-3	2.01	ULVF
ULVF-2	3.31	HVF
HVF-2	2.81	HVF
M-4	3.04	M



Grant Wills
September 2017
The Pennsylvania State University

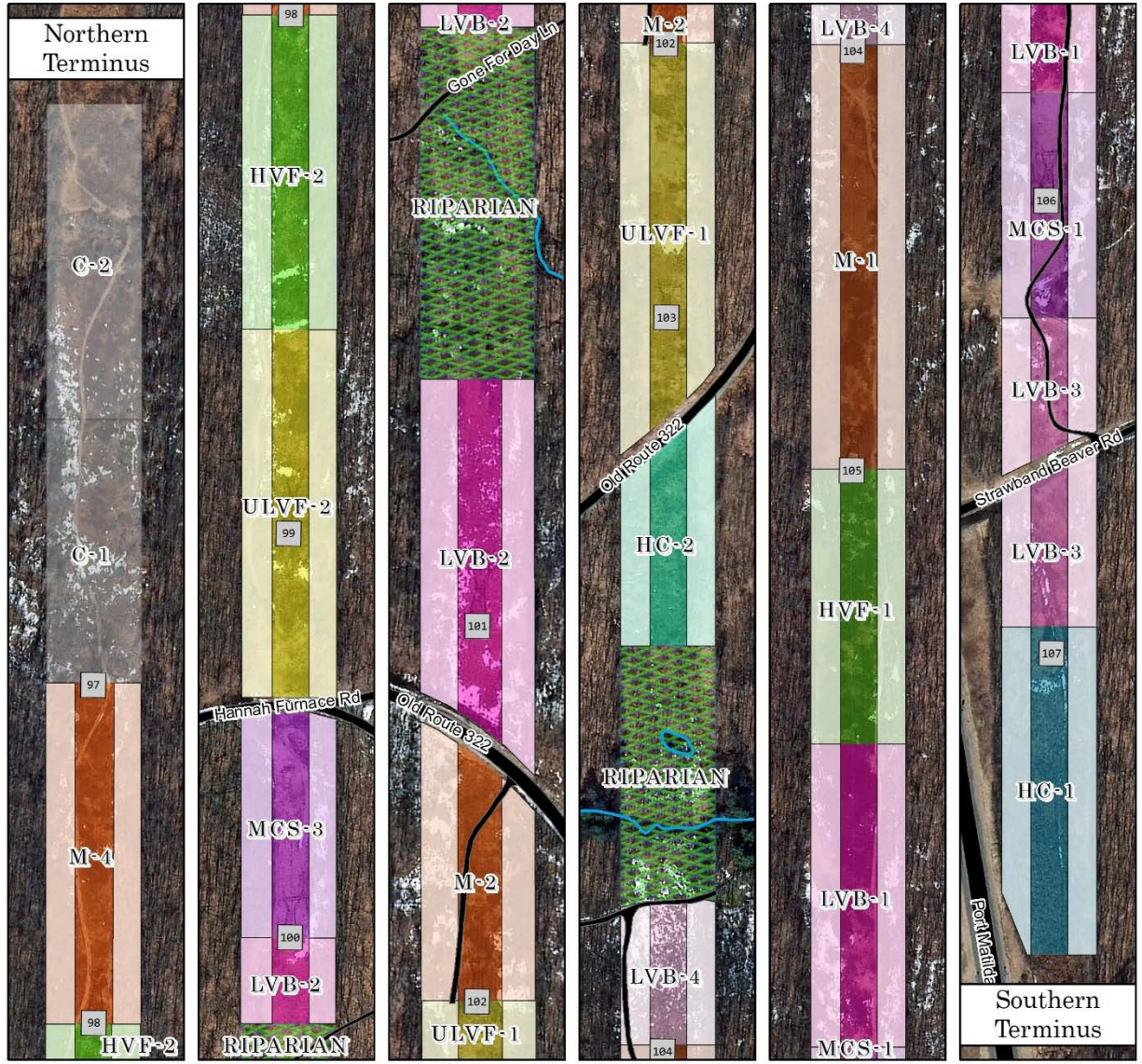
SGL33

2016 Treatment Map

Treatment (TX) Types

C = Control
HVF = High Volume Foliar
HC = Hand Cutting
LVB = Low Volume Basal
M = Mowing
MCS = Mowing+Cut Stubble
ULVF = Ultra Low Volume Foliar

TX Zone	Acres	2016 TX
HC-1	2.93	HC
LVB-3	2.76	LVB
MCS-1	2.01	ULVF
LVB-1	2.7	LVB
HVF-1	2.45	ULVF
M-1	2.01	ULVF
LVB-4	1.71	LVB
HC-2	2.26	HC
ULVF-1	2.93	ULVF
M-2	1.81	ULVF
LVB-2	3.78	LVB
MCS-3	2.01	ULVF
ULVF-2	3.31	HVF
HVF-2	2.81	HVF
M-4	3.04	M



Grant Wills
September 2017
The Pennsylvania State University