Lab Report

Title: Raster ETL Operations and Spatiotemporal Visualization in the ArcGIS Ecosystem
Notice: Dr. Bryan Runck
Author: Luke Zaruba
Date: 2 November 2022

Project Repository: [Link Here]
Time Spent: 12.0 hours

Abstract
In this project, the goal was to further develop skills for performing ETL processes, by diving into exploring funcitonalities within the ArcGIS ecosystem for working with LiDAR data and spatiotemporal multidimensional rasters. The specific processes that were explored include creating DEMs and TINs, as well as creating a space-time cube from a multidimensional raster. ETL pipelines were developed using open source Python libraries and ArcPy, Esri's primary Python library supporting ArcGIS Pro/Desktop, to programatically perform the operations necessary and develop a platform for which future projects can use the products of this project to easily reproduce any of the operations that were performed.

Problem Statement
The problem at hand is two-fold, first, a pipeline will be created to streamline the extraction and transformation of LAZ files to DEMs and TINs. Second, a pipeline will be created for extracting PRISM data and processing the data into a space-time cube. Broadly speaking, ETL operations for non-vector (raster and point cloud) datasets will be explored by extracting data via API calls, transforming data into a variety of different formats, before creating various visualizations of the transformed datasets.

Table 1. Analysis Requirements for Part 1

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	LiDAR Data	Raw .LAZ/.LAS files	Point Cloud	N/A	[MN DNR]	Uncompress and convert to LASD
2	PRISM Data	Multidimensional raster dataset containing precipitation averages for the contiguous U.S.	Raster	Avg. Precipitation	[PRISM]	Convert BIL to MD raster, to STC

Input Data
The first dataset used in this initial part of the lab is LiDAR data that was retrieved from the Minnesota DNR. The files were originally stored as .LAZ files on the DNR's FTP server, which are simply compressed .LAS files. The second dataset used was PRISM data, which was extracted from the site's FTP page. The file was initially zipped, but after unzipping, the file contained a .BIL file. This type of file can be used to represent rasters, wit hvalues stored in

binary format. Additionally, the file has 12 different bands, which was used to represent the temporal component of the dataset.

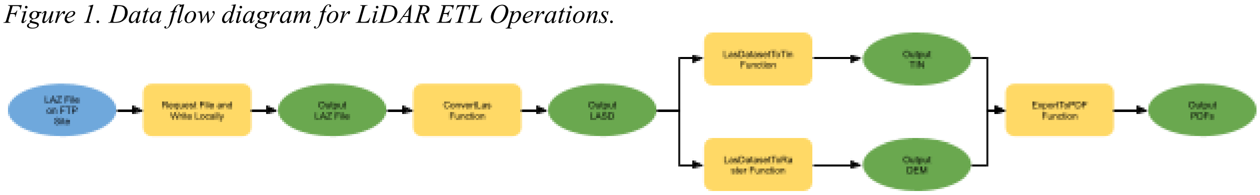
Table 2. Input Datasets for Part 1

#	Title	Purpose in Analysis	Link to Source
1	MN DNR LiDAR	Raw input LiDAR point clouds	[MN DNR]
2	PRISM Data	Average precipitation measurements for the U.S.	[PRISM]

Methods

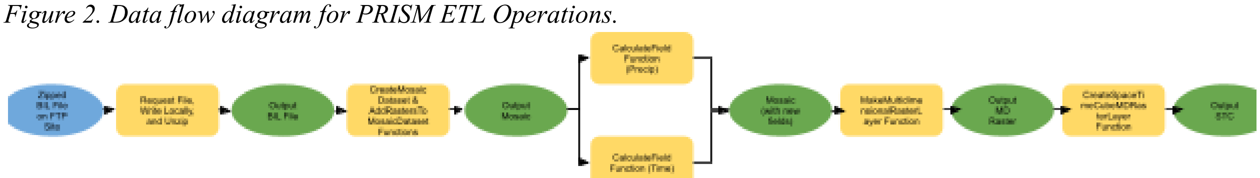
First, I started by extracting the compressed LiDAR files from the FTP site and manually (through Python's file methods) writing them on to local files. After this, the files needed to be decompressed and turned into an LASD. Luckily, the operation necessary for decompressing the files actually has an optional LASD output that can be created as well, which saves the step of creating an LASD. After this, the LASD can easily be converted to both a TIN and a raster (a DEM in this case). With the DEM and TIN, the datasets need to be loaded into map frames in layouts, which is done manually. The final step was to programatically export the layouts to PDFs.

Figure 1. Data flow diagram for LiDAR ETL Operations.



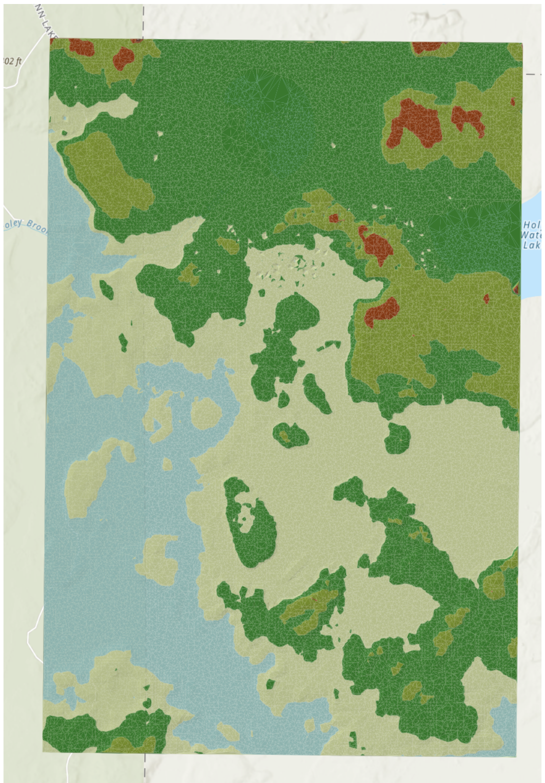
With the PRISM analysis, the sequence of operations started very similar to that of the LiDAR pipeline's, with extraction being necessary from the PRISM FTP site. The difference, though, was that a ZIP file was downloaded from PRISM, so this needed to be unzipped before continuing on with the regular extraction steps. After this, the file then need to be converted into a mosaic, which allows us to leverage some functionalities that are only available in mosaic datasets. With the mosaic created, two fields were added to store the precipitation values and the temporal aspect of the applicable band that the value was associated with. The mosaic could then be converted to a multidimensional raster, after the two fields were created, which simply serves as an intermediary between the mosaic and the space time cube, since there is not direct conversion between the two types of layers.

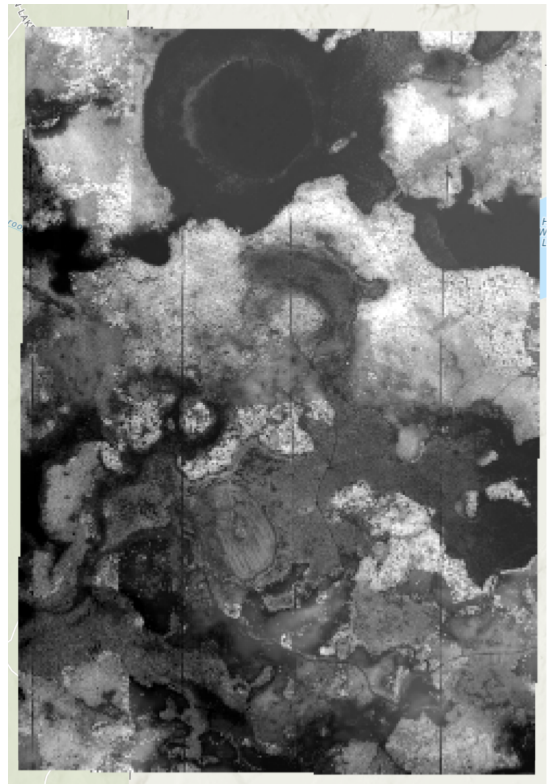
Figure 2. Data flow diagram for PRISM ETL Operations.



Results

In the end, I was able to produce two maps showing a TIN and DEM representation of a LiDAR tile in Aitkin County, Minnesota, and an animated visualization of PRISM data over time, across the United States. A brief glimpse of the two maps can be found below, and the actual PDF maps and the GIF animation can be found at the project GitHub repository, which is linked at the top of this document.





Results Verification

In order to verify the accuracy of my operations, I performed manual inspections of the intermediary data produced by any operations I performed, as well a more thorough manual inspection of my final results. For the purposes of this lab, no quantitative methods were used to determine the accuracy of the analysis.

Discussion and Conclusion

I think that I learned a lot about the tools that are available in the ArcGIS ecosystem for dealing with LiDAR and spatiotemporal/multidimensional rasters. Furthermore, the project gave me an opportunity to improve my skills related to ETL operations and processes, which is crucial and relevant to any analysis. I also went ahead and expanded on this by adding some additional tools to the LiDAR ETL pipeline, which can be seen at the start of the notebook for that respective component. I created functionality to better organize downloads into subdirectories, as well as prevent unnecessary downloads if the data already exists. Another cool feature I made was downloading the tile map for a specified county and then displaying the map in the notebook prior to any download so that you can see which tiles you want to download.

References

N/A

Self-score

Category	Description	Points Possible	Score
Structural Elements	All elements of a lab report are included **(2 points each)**: Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score	28	**28**
Clarity of Content	Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**.	24	**22**
Reproducibility	Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified.	28	**27**
Verification	Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**.	20	**20**
		100	**97**