GeoProcessing Batch Script: Feature and Raster Analysis

Amanda L. Riggs

GEOG-6400

East Carolina University

April 30, 2024

Introduction

I created the *GeoProcessing Batch Script: Feature and Raster Analysis* tool out of a need to preprocess a significant amount of data at once. My thesis study utilizes the InVEST® (Integrated Valuation of Ecosystem Services and Tradeoffs) software suite, which models are used to assess natural capital provided by ecosystem services. To utilize the InVEST tools, significant amounts of different types of data need to be collected and processed, which vary depending on the specific model requirements. ArcGIS Pro can use the outputs generated by the InVEST models to visualize changes over time and help me answer questions related to urban development in the area. Such significant amounts of preprocessing require a script tool that will help streamline the data preprocessing that software suits use such as InVEST. Therefore, I decided to create the script as part of my final project for Geography 6400, Seminar in Geographic Information and Analysis, GIS Applications Programming Edition course.

Most of the data preprocessing needs consist of various data types such as rasters, feature classes, and CVS files, which need clipping, masking and projecting to a study area with the same spatial references and deposited in a default geodatabase. Therefore, the tool I created can use a feature layer boundary layer to clip and mask multiple feature and raster layers simultaneously, project them to the desired coordinate system, and deposit them into the project default geodatabase for the user's convenience.

Code, as such, requires the use of many tools and proves to be challenging. The main challenge was getting the tool to deposit the final outputs to the default geodatabase. However, after many trials and errors, the script tool processed the data successfully. The tool clips, masks, sets the spatial references, and projects large amounts of data types simultaneously and converts the outputs to the user's default geodatabase. Such a tool would be helpful for others who are using similar software modeling suites that need large amounts of preprocessing, as well as convenient for users who are geoprocessing spatial data solely in ArcGIS Pro.

Methodology

InVEST models prove to be powerful tools for simulating the movement of nutrients and sediment within a watershed and assessing potential flooding in urban areas. These models aid in decision-making for conservation and land management and help identify areas prone to flooding while mitigating the risks of water pollution.

My study will employ a variety of InVEST models, such as nutrient and sediment export models, stormwater runoff, and urban flooding models. Through spatial analysis and modeling, the study will evaluate the impacts of land use changes on ecosystem services and urban water management and attempt to place a monetary value on ecosystem services and tradeoffs in Pitt County over twenty years, requiring much data.

Input Data Requirements

Effective utilization of the InVEST models requires careful preprocessing of spatial data inputs. Key data inputs include land use/land cover maps, soil hydrologic group raster data, precipitation data, and biophysical tables. Additionally, optional inputs such as road centerlines and built infrastructure maps may enhance model accuracy. Figure 1 shows the various data types I need to run InVEST's models according to the various InVEST (2024) models I plan to use for analysis. For this tool, I created it to process all the feature and raster data for now with plans to continue working on the tool to process all the required data needs for my thesis project.

Data Types	InVEST Models					
	Urban Stormwater	RouteDEM	DelineateIt	Urban Flood Risk	Sediment Delivery	Nutrient Delivery
Area of Interest	X			X		X
Band Index		X				
Biophysical Table	X			X	X	
Borselli IC0					X	X
Borselli K					X	
Built				X		X
Damage Loss				X		
Digital Elevation		X	X		X	
Drainages (Raster)					X	X
Erosivity (Raster)					X	
Land Use/Land	X			X	X	
(Raster)						X
Maximum L Value					X	
Maximum SDR					X	
Water						X
Precipitation	X					
Rainfall Depth				X		
Replacement Cost	X					
Snap Distance			X			
Soil Erodibility					X	
Soil Hydrologic	X			X		
Subsurface						
Maximum						X
(Vector) optional	X					
Table (CSV)	X					
Watershed Outlets			X			
Watersheds					X	
(number)	X					

Figure 1. This table includes the various data required to run the InVEST models. The model types are listed on the X axis, and the data needs are listed on the Y axis.

Although InVEST provides two preprocessing tools I can use to delineate watersheds (DelineateIt) and fill sinks (RouteDEM), there is not a tool that clips or sets coordinate systems for any of the dataset type. Considering that I plan to run six models for my thesis project, it would be time-consuming to run projections, clips, masks, and geodatabase conversions on every single data set I will need, which is a lot since I need to model twenty years' worth of data. Therefore, this project would be a good opportunity to design a script tool that will do most of the preprocessing for me without having to utilize all the available ArcGIS tools separately and repeatedly.

Script Development for Preprocessing

When creating the script, the challenging part is using a feature layer as a study boundary to mask the rasters. Therefore, I had to adjust the script to utilize different tools available in ArcGIS Pro and combine them in logical flow that would process different types of data. To create the tool, I had to decide on parameters that would work efficiently and work with different types of data. Originally, I began with different parameters for each type of data, such as a parameter for the land use rasters, precipitation rasters, study boundary feature layers, as well as built environment layers. However, after running the tool several times trying to get the logic correct, I realized that I only need four parameters.

The script requires a feature layer input parameter for the clipping and masking, a coordinate system parameter that enables users to choose the coordinate system they desired, and a parameter for feature layers that allows multiple inputs, and a raster layer input that allows multiple inputs, saving a significant amount of time entering in data layers because it provides a chevron that a user can check that will check all feature layer files at once and add them to the parameters.

There are seven key geoprocessing tools that the script needs to process the data which includes Project, Clip, Feature Class to Geodatabase, Feature to Raster, Project Raster, Extract by Mask, and Raster to Geodatabase. The main body of the workflow consists of setting the workspace and parameters, setting the coordinate system, projecting feature layers to the desired coordinate system, then clipping the feature class layers. Next, the script moves on to process the raster layers which require using the Feature to Raster tool to convert the boundary feature layer into a raster to use for processing the rasters. I have not found a tool that can mask or clip data that are different types. Therefore, my tool will be convenient for others as well. After processing the rasters, all data is then deposited into the default geodatabase. Figure 2 depicts a flow chart of the basic workflow that includes fifteen steps required to process all the different data types.

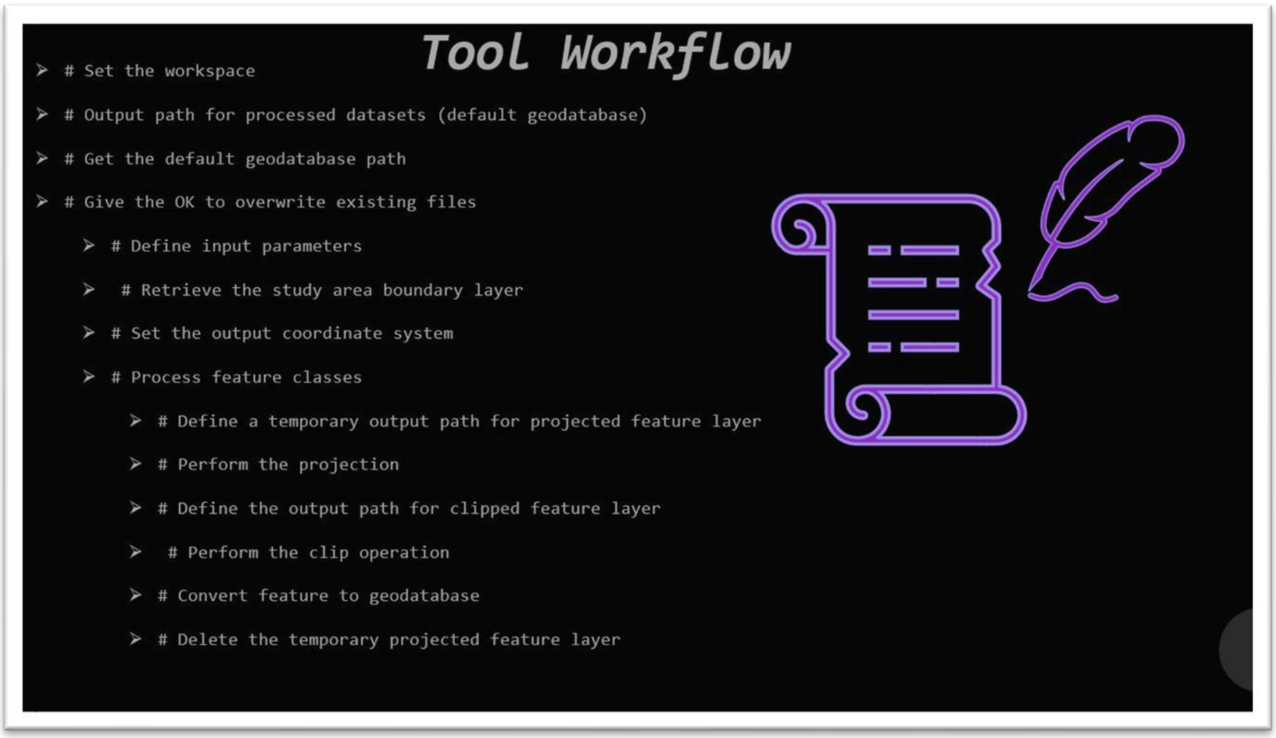

Figure 2. A flow chart of the basic workflow that includes the fifteen steps needed to process different data types.

Results

Scenario Demonstration

The sample run of the GeoProcessing Batch Script included nine datasets that feature classes and rasters. Figure 3 depicts a map of the raw data layers used for demonstration, which includes a railroad feature class, districts feature layer, buildings feature layer, two land use/land cover rasters, an annual precipitation raster, a hydrology index raster, roads feature class, and a waterway feature class. All these datasets have different spatial references and extents and need the tool to process them down to a study area.

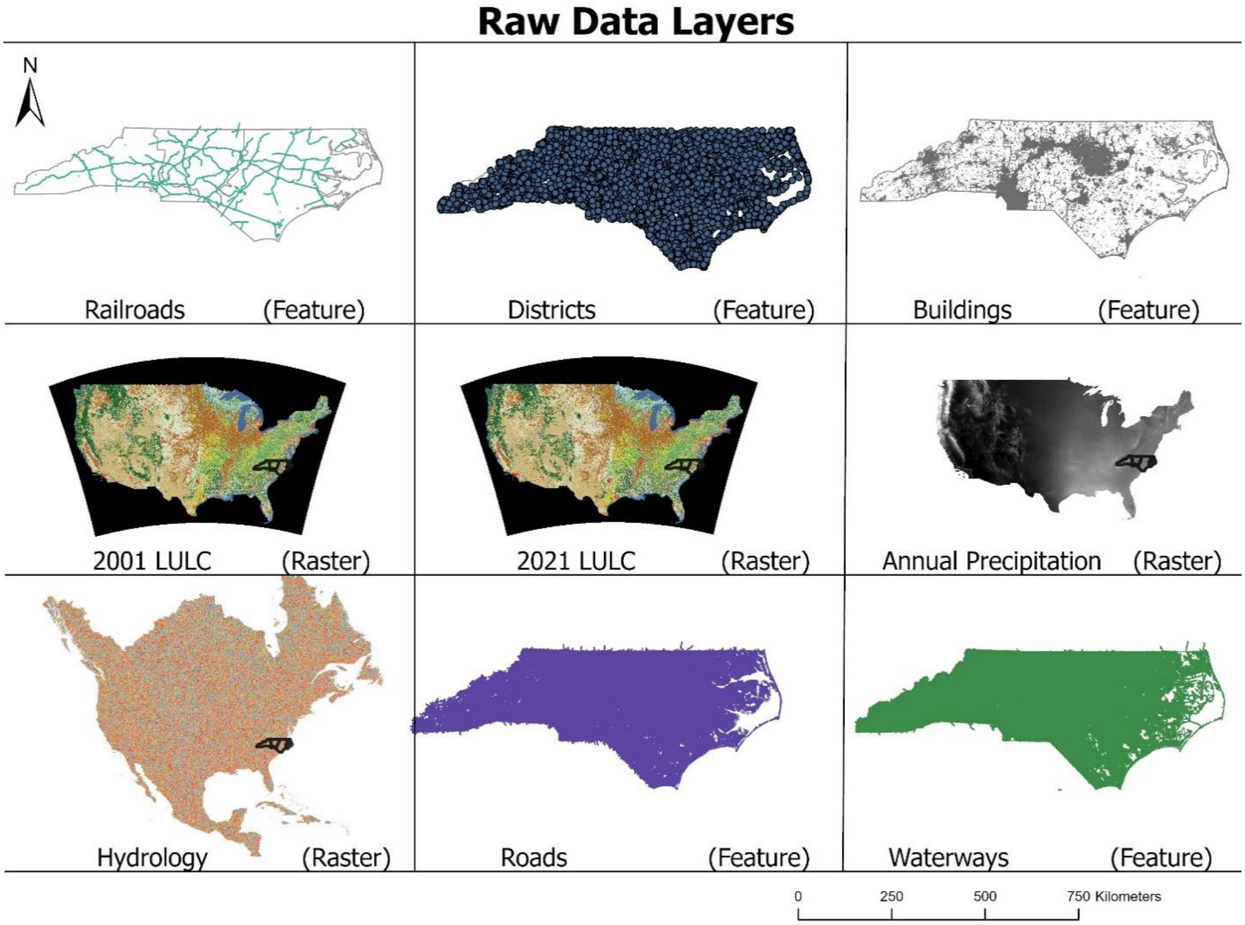

Figure 3. A map of nine different raw data types prior to processing.

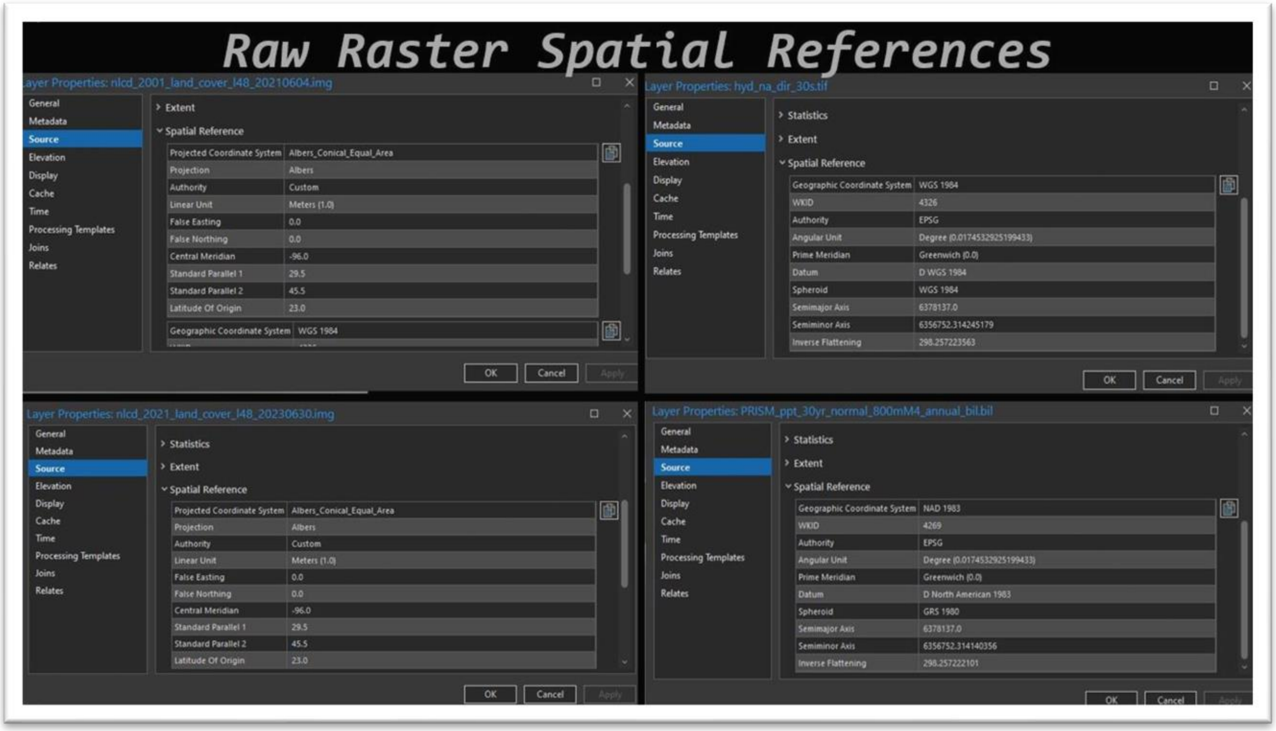

Figure 4. This is an image of the property's windows of each of the four raw raster data layers, depicting the spatial references prior to processing.

Figure 4 depicts the spatial references of all the raw raster layers prior to processing. Because the feature data already had the same spatial references because they came from the same source and I used their reference to project the raster data, I did not include a figure of the feature layer spatial references.

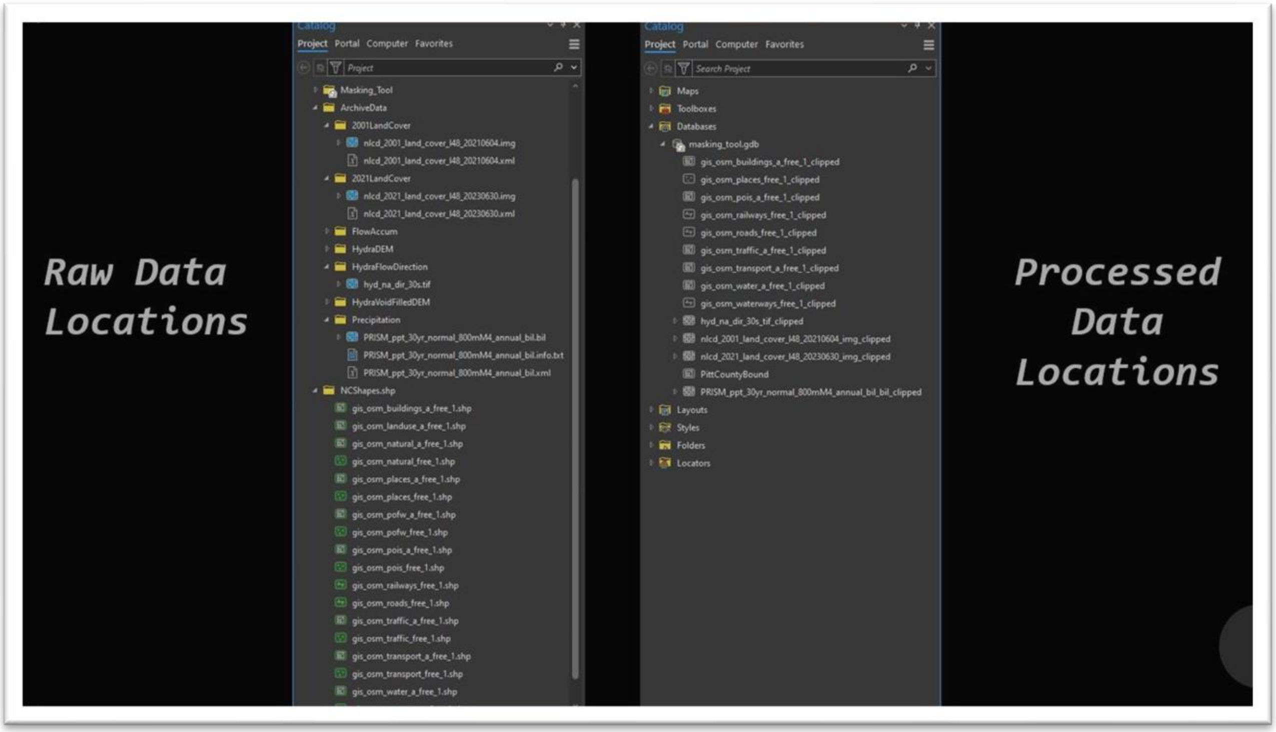

Figure 5. This is an image of the locations of the demonstration data before and after processing.

Figure 5 depicts the location of the data used for the demonstration. The left image shows the raw data locations in their respective folders prior to processing. The right image shows the location of the processed data, showing it as the default geodatabase in the demonstration project.

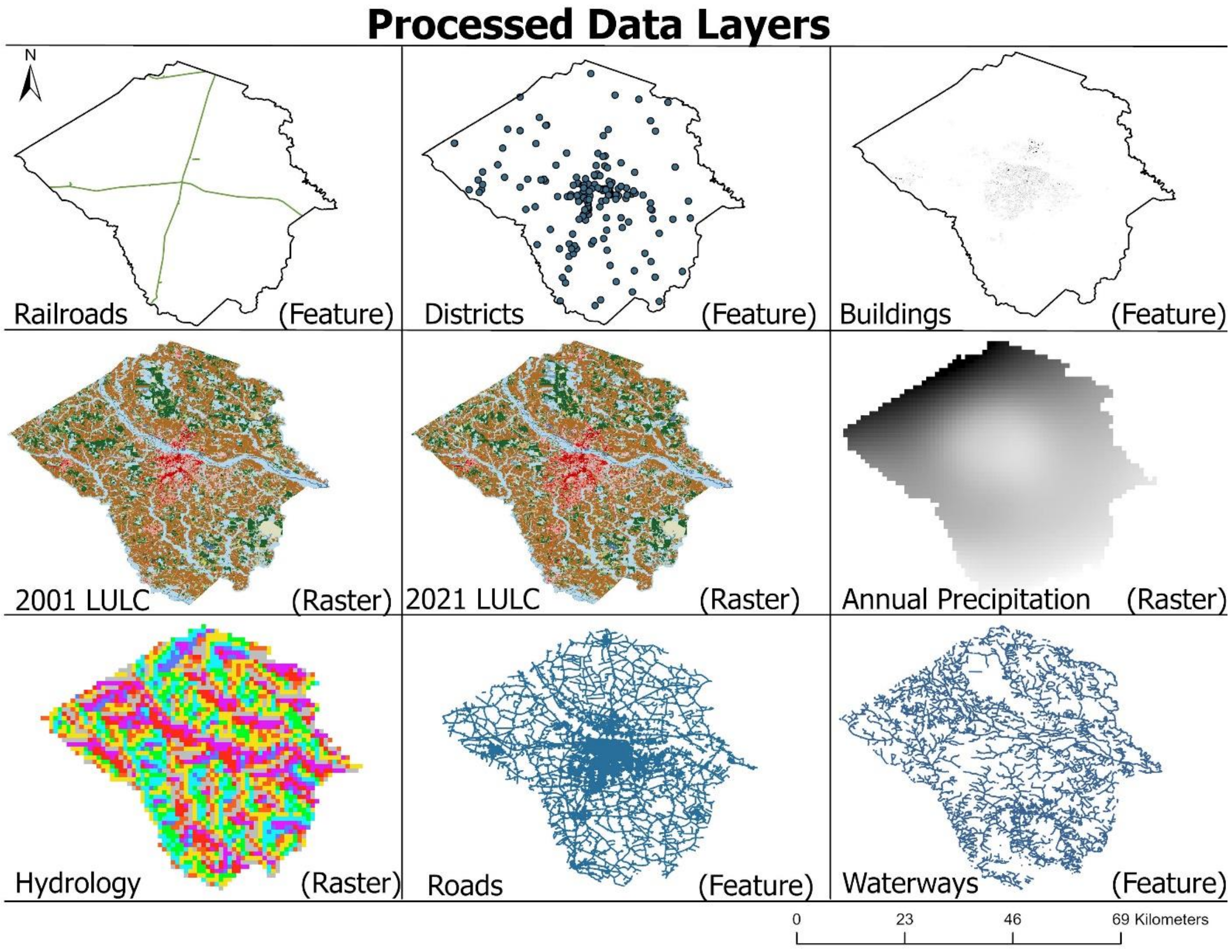

Figure 6. A map of nine different outputs generated by the script.

After several trials and errors, the script completed a successful run and generated nine outputs, all with the same coordinate system and extent. Figure 6 depicts a map of the different data outputs generated by the tool including a railroad feature class, districts feature layer, buildings feature layer, two land use/land cover rasters, an annual precipitation raster, a hydrology index raster, roads feature class, and a waterway feature class. The rasters all have different file extensions such as img, tif, and bil, which all were processed successfully using the script tool.

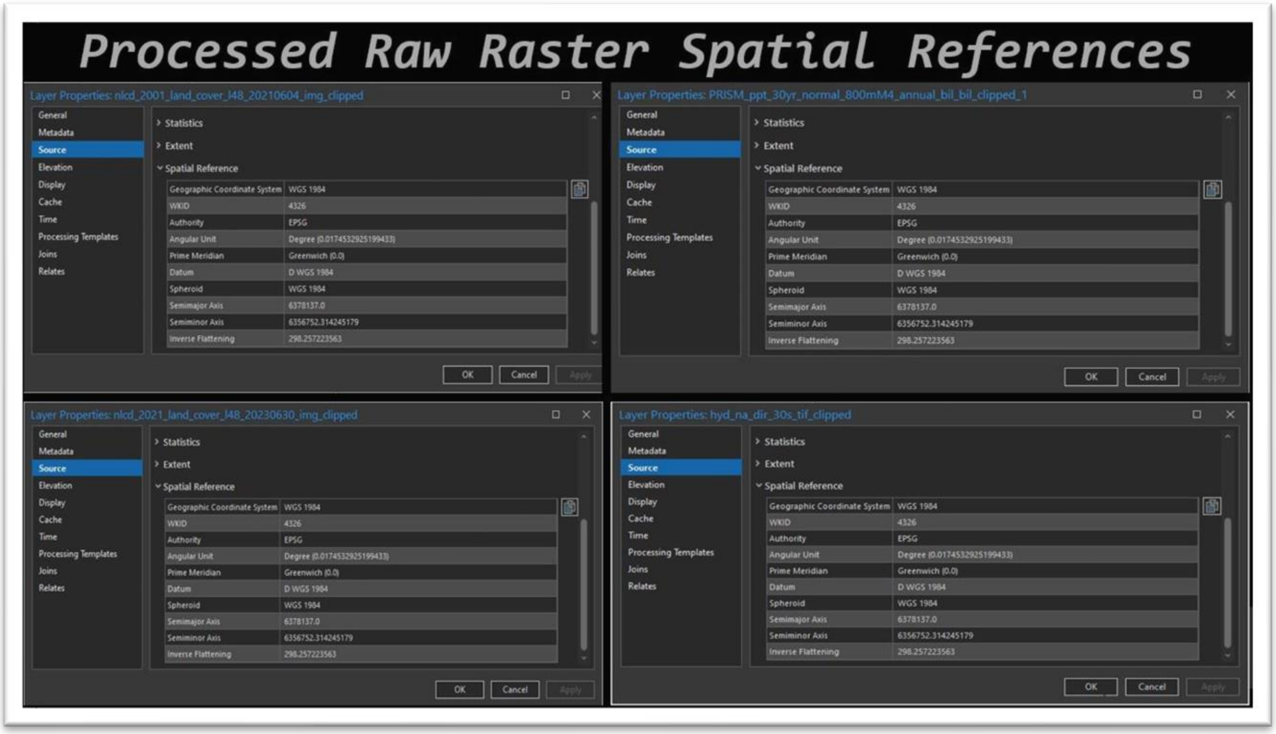

Figure 7. This is an image of each of the data's raster's properties after processing, showing that they all are projected to the same spatial references.

Figure 7 depicts the spatial references of all the processed raster layers after running the script showing that they are all projected to the same coordinate system.

Summary and Conclusions

Overall, the script performed as expected and will be a substantial help with my thesis project for future preprocessing data to use in the InVEST models I have chosen to use for analysis as well as many other future uses involving projects that require both feature and rasters processed to the same extent. The script took nine raw data layers of different file extensions, clipped, masked, and projected them all to the default geodatabase of the ArcGIS Pro project. The script will benefit other users that have similar needs to process a substantial number of features and raster data types simultaneously. However, there are some limitations that users should consider before utilizing the script.

Use Limitations

The script requires a boundary feature layer for parameter (0) that includes an osm_id field in the attribute table. The tool may require specific data formats or data structures, limiting its compatibility with certain datasets or data sources. Also, large datasets or complex processing tasks may lead to longer processing times, affecting overall efficiency, especially when processing multiple layers in batch mode. Another limitation is the tool's memory usage may increase significantly when processing large datasets, potentially leading to memory-related errors or slowdowns on systems with limited resources. Also, depending on the input data and processing operations, the tool may generate large output files, which could consume storage space and impact file management.

Future Improvements

Although the tool works as expected to process raster and data layers, I have other processing needs that involve other data types such as CSV files and text files that I would like to also include in the script. Also, although the processed data included all the attributes appended to them, there is not an actual snippet of any appending tools. Other data that some users may attempt to process could need an append tool included in the script. Other tools I would like to include involve delineating watersheds and filling sinks, as my project involves the flow of watercourses and the dispersion of nutrients and sediments. Moreover, others that are using the same types of software or simply working on a hydrology project in ArcGIS Pro, could also use the script to streamline processing times. Other future analysis using the script is to utilize it in other software suits to enhance its versatility.

The GeoProcessing Batch Script: Feature and Raster Analysis is a critical tool for me to process data that will allow me to assess the impacts of urbanization on ecosystem services by modeling my processed data in the InVEST software suite. By utilizing the script in combination with InVEST my thesis study aims to provide valuable insights into urban planning and environmental management.

References

Natural Capital Project [NCP], (2024). InVEST 3.14.1. *Stanford University, University of Minnesota, Chinese Academy of Sciences, The Nature Conservancy, World Wildlife Fund, Stockholm Resilience Centre, and the Royal Swedish Academy of Sciences*. https://naturalcapitalproject.stanford.edu/software/invest

OpenStreetMap. [Data]. *Open Data Commons Open Database License (ODbL0 by the OpenStreetMap Foundation*. Copyright Creative Commons Attribution-ShareAlike 2.0 . https://www.openstreetmap.org/#map=5/38.007/-95.844

USGS. (n.d.). National land cover database. *U.S. Geological Survey U.S. Department of the Interior*. https://www.usgs.gov/centers/eros/science/national-land-cover-database

USAFacts. (2024). Our changing population: Pitt County, North Carolina. *USAFacts*. Pitt County, NC population by year, race, & more | USAFacts