

Vector Data Analysis – Overlay Techniques

Vector Data Analysis – Overlay Techniques

Environment Settings

Environment Settings. Before we can discuss vector data analysis, we must 1st take a short digression into environment settings.

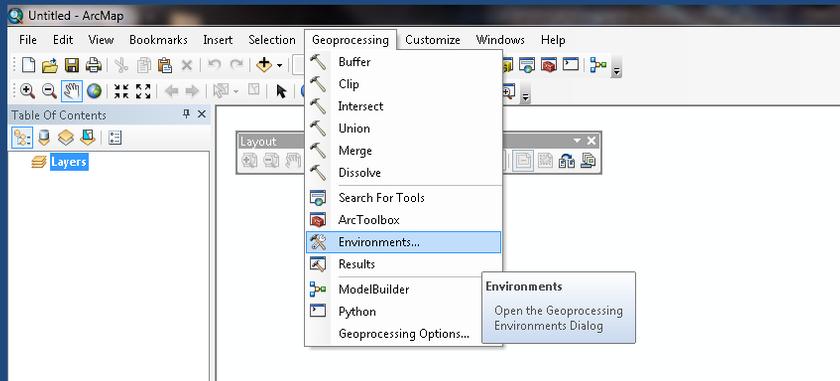
What are environment settings?

- Each tool uses a set of parameters to execute an operation
- Environment settings allow the user to set defaults for these common parameters (ie. output location, extents, cell sizes, etc.)
- Allows the user to remain organized as well as make tasks quicker and easier

And environment setting is a set of parameters that a user will set prior to executing a tool or operation. Environment settings allow the user to determine settings for the input, processing, and output of the tool or operation. Environment settings allow the user to set defaults for parameters such as output location, analysis extents, cell sizes, and so on. Environment settings allow the user to remain organized as well as make tasks quicker and easier as environment settings can be shared between tools and operations.

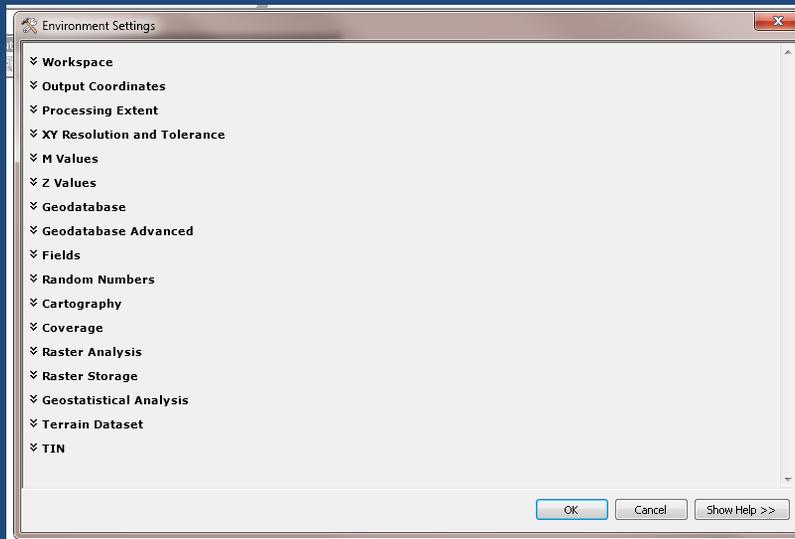
Location

- Geoprocessing -> Environments
- Can be set From ArcCatalog and ArcMap



In ArcMap, and ArcCatalog, you can access the environment settings dialog box by clicking Geoprocessing from the menu bar followed by environments.

Opens a list of parameters



This is the environment settings dialog box as opened in either ArcMap or ArcCatalog. You can click the double down chevrons to expand each environment setting and change its values.

Examples

- **Workspace** – set current and scratch workspace
- **Output Coordinates** – Set a geographic transformation for all output coordinates

Let's review a few environment settings. The workspace environment setting sets the current and scratch workspace. The workspace is where tool will look for input values by default, and where will place intermediate and output files by default. The output coordinates environment setting sets a geographic transformation for all output coordinates when the tool or operation is run. You can find more information about other environment settings in ArcGIS desktop help.

Coverage Data Format

Coverage Data Format

Coverage

- A data model that stores vector data
- The data contains the spatial and attribute data for geographic features
- Stores feature classes

The coverage data format stores vector data, such as points, lines, and polygons. A coverage file contains the spatial and attribute data for the geographic features that it represents. A coverage is similar to a shapefile, however, the file format is quite different, and the data is stored in separate feature classes.

Storing Coverages

- Stored as a directory
- Naming:
 - Can't be longer than 13 characters
 - Can't contain spaces
 - Can't start with a number
 - Doesn't have an extension
 - Must be in all lowercase letters

To store a coverage, there are a few restrictions to keep in mind. 1st, a coverage is stored as a directory containing multiple files. The name of the coverage cannot be longer than 13 characters, a cannot contain spaces, they cannot start with a number, it does not have a filename extension, and it must be named using all lowercase letters. The reason for these restrictions, is at a coverage is considered a legacy format, and was created during a time when computers and operating systems did not support longer and more complex filenames.

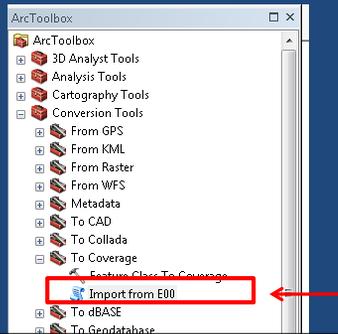
Coverage Composition

- Composed of a set of files
- Each file contains information concerning the feature classes
- Contents appear differently in Windows Explorer v. ArcCatalog

A single coverage is composed of multiple files, similar in the way that a shapefile is composed of multiple files. Each file contains information concerning one part of the entire feature class. The contents of a coverage will appear differently in Windows Explorer versus ArcCatalog, and it is not recommended that you ever manually move or change any of the files inside of the coverage folder. It is recommended that you manage the coverages using ArcCatalog as it simplifies the coverage files into a single entry in the ArcCatalog file viewer.

Converting Coverages

- Uses script tool to convert to a feature dataset
- Located in ArcToolbox:
 - Conversion Tools -> To Coverage -> Import from E00 (E00 being the interchange format)



The reason coverages still need to be learned is that quite a bit of data is still provided in this legacy format. Therefore, while it is not recommended that you save newly created data into the coverage format, you must know how to convert the coverages into a more contemporary GIS data format. To make things slightly more complicated, coverage files are often delivered in an interchange format, known as the E00 format. In ArcGIS, a script tool is provided to convert coverages in the interchange format to a feature data set. This tool is located arc toolbox under the conversion tools toolbox, to coverage toolset, and is named import from E00.

Vector Overlays

Vector Overlays

Overlay Operations

- Overlay Operations involve combining spatial and attribute data from two or more spatial data layers.

Overlay operations involve combining spatial and attribute data from two or more spatial data layers. In other words, we are stacking the data, and looking for where the layers overlap. While the overlay operations may seem simple when used alone, they can actually be quite powerful operations when combined in series. An example of an overlay operation, would be the clip operation. For example, let's say I have a data set that contains all the streets in the United States of America. However, I only want the streets in Arkansas. I can use the clip operation to cut out all the streets that are only in Arkansas by providing the state outline of Arkansas as the clipping feature. One thing to note about overlay operations, are that overlays require that the data be in the same coordinate system. This is so that the software can perform the highest accuracy of overlay. Some software packages automatically re-project on-the-fly if the data is not in the same coordinate system, but it is preferable if you convert all data sets to the same coordinate system before performing the overlay operation.

Vector Overlay

- Vector Overlay involves combining point, line, polygon geometry and associated attributes
 - Overlay creates new geometry
- Large attribute tables may result if overlay operations combine many layers
 - Duplicate attribute fields may also exist.

A vector overlay involves combining point, line, or polygon geometry and their associated attributes. All overlay operations create new geometry and a new output geospatial data set. You should be cautioned that with certain overlay operations, very large attribute tables may result if the overlay operations combine many layers, and each layer has a very large attribute table. Additionally, it might be possible that the combined attribute tables would cause duplicate attribute fields to exist. In these cases, you should consider reducing the number of transferred attributes to the minimum required, and renaming duplicate fields so that there is no ambiguity.

Basic Cases of Overlay - Clip

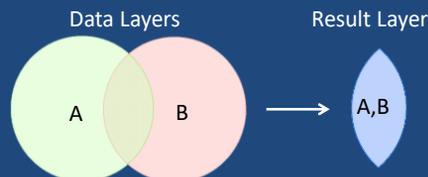
- Clip defines the areas for which features will be output based on a 'clipping' polygon
- Attributes of clipping layer not transferred to clipped data



The first basic case of overlays is the clip function. The clip function defines the area for which features will be output based on a "clipping" polygon. I like to think of it in terms of cookie dough, and a cookie-cutter. The cookie dough is the data layer that we want to reduce in size based on our cookie-cutter. The clipping layer is the cookie-cutter that will determine how we cut out the part of the cookie dough we are interested in. It is important to note that in the clip operation, only the geometry and attributes of the data layer are transferred to the result layer. Neither the geometry or attribute of the clipping layer are included in the output, just as you would not make your cookies in the oven with the cookie-cutter still surrounding the cookie dough.

Basic Cases of Overlay - Intersection

- Intersection combines data from at least two layers but only for region where all layers contain data
- Similar to Clip, but the attributes are transferred to the output data

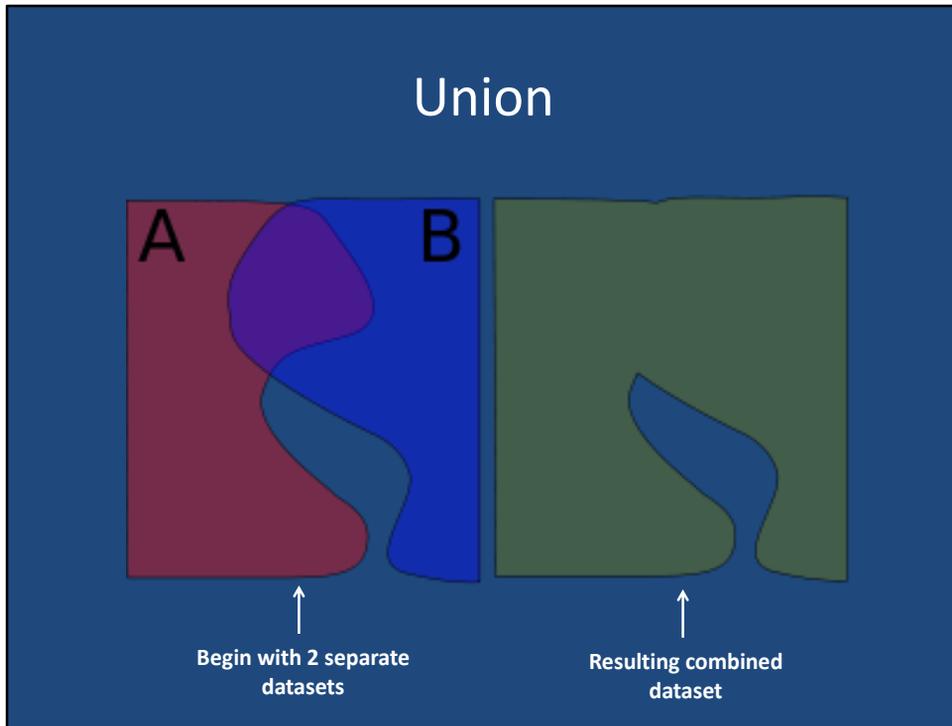


The next basic case of overlay is the intersection operation. The intersection operation combines data from two or more input layers, but the combination only exists for the regions where all input layers contain data. This is similar to the clip function, however the attributes and geometry of all input data layers are transferred to the result layer.

Basic Cases of Overlay - Union

- Union is an overlay that includes all data from all of the input data layers
- No geographic data are discarded in the union operation and corresponding attribute data are saved for all regions.

The third basic case of overlay is the union overlay operation. The union operation is an overlay that includes all data from all of the input data layers. Since the union operation combines all attributes, and all geometry, no geographic data are discarded in the union operation, and all corresponding attribute data are saved for all regions.

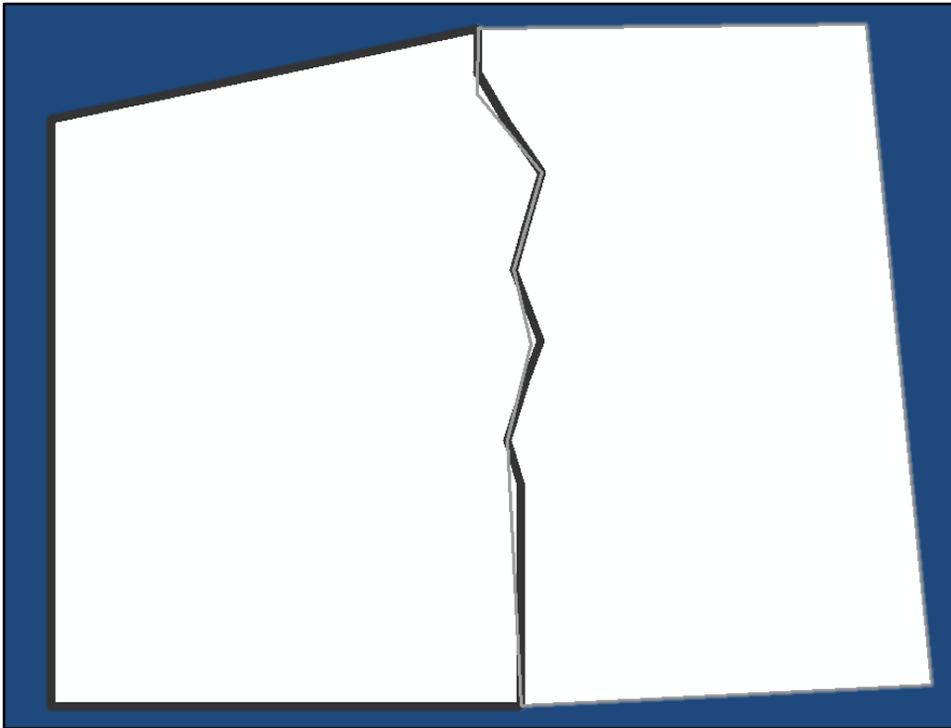


There is an illustration of a union overlay where we initially had to separate data sets, a and B, that, when union, result in a single, combined data set.

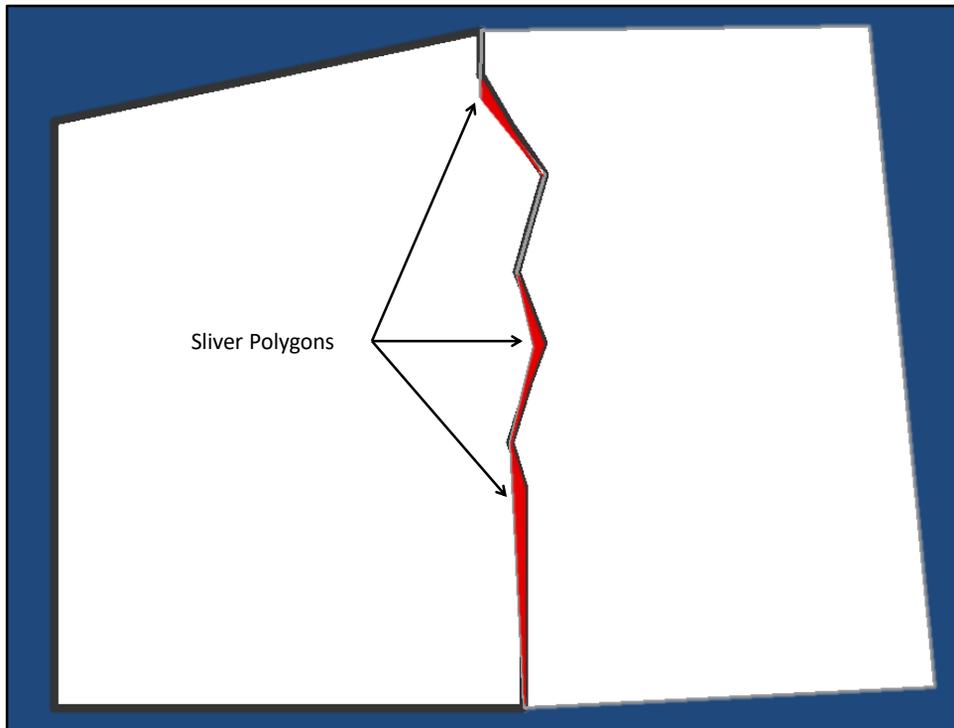
Vector Overlay Problems

- Common Features represented in both layers, but with slightly different geometry
- This creates sliver polygons when the overlay operations are performed
- Several methods exist to reduce the occurrence of sliver polygons.

When performing vector overlays, there are a few problems that you should be aware of. In the case where the input features represent common features that are represented in both layers, they may have slightly different geometry. This creates sliver polygons when the overlay operations are performed. For instance, imagine we have one input data set that represents the national boundary of Canada created at a 1 to 5,000,000 scale. And we have a second input data set of the national boundary of the United States of America created at 12 1 million scale. If we stack these two layers on top of each other, it is probable that the boundary lines will not overlay exactly, as they were created at different scales, and therefore the Canadian boundary line has probably been simplified more than the United States boundary line. The areas where the two input data sets do not match, or result in very small, typically insignificant sliver polygons that need to be dealt with. There are several methods that exist to reduce the occurrence of sliver polygons, such as manually snapping the boundaries together, or introducing a fuzziness to the edges of the overlay operation, or simply manually deleting the sliver polygons.



To illustrate the idea of a sliver polygon, here are two different data sets, one outlined in a bold black line, and the other outlined in a thinner great line. Where the two data sets overlay, they do not match exactly for whatever reason. If we perform in intersect, clip, or union operation on these two input data sets, we are going to create three sliver polygons.



The sliver polygons created between these two input data sets are highlighted in red. The sliver polygons are probably insignificant, and are the results of data sets being produced by different agencies, at different times, and probably at different scales. It is up to you to determine how you would like to solve for the sliver polygons.