Spatial Analysis Tip Sheet – Spatial Joins and Overlays, Density, and Proximity Analysis

Barbara Parmenter
Tufts University
October/Nov.1, 2007

Requirements for spatial analysis in GIS

All data and the data frame must be in the same coordinate system:
Before doing any analysis functions in ArcGIS, make sure that all the data layers you will be using are in the same map projection/coordinate system, and that the data frame is also in that coordinate system. For ANY analysis in ArcGIS, the layers you are using in your analysis MUST be in the same coordinate system, and this must match the coordinate system of the data frame. If you don't do this, you'll either get errors, or it will appear to run, but nothing will happen. If your data is in a different coordinate system, the easiest way to convert it to set the data frame to the correct one (e.g., Mass State Plane, Mainland, NAD 83, meters), then right-click on the layer that is not in that coordinate system and choose Data-Export. In the dialog box, say you want the new file to have the same as the data frame's coordinate system, and proceed from there. This creates a copy of the original shape file in the correct coordinate system. See additional tip sheets concerning projections if you need more help.

Set the Geoprocessing Environment Settings
For most or all of these analysis functions, it will also help things if you set the Environment Settings. I recommend you set these before you do anything by going to the main toolbar in ArcGIS and clicking on Tools - Options, then clicking on the Geoprocessing tab, and finally clicking on the "Environments" button there. The critical things for you to set are:

1. Under the *General Settings* area;
    a. Set the *current workspace* to be where you are pulling data from (just click on a folder, not a file - the folder within which most of your GIS data is located)
    b. Set the *scratch workspace* to a temp folder on your local computer or your H: drive that you have write access to. If you don't have write access, this will cause errors. Create a temp folder on the local computer under your user name if you have to (putting the scratch folder on the C: drive will save space on your H: drive).
    c. Most important, set the Output Coordinate System to be the one you want all your analysis to be in
    d. Set the *Output Extent* to be the largest extent of your study area – use an existing file that covers the study area or create a new one. This is critical for raster analysis – you must use the same extent for all your raster work The Extent will always be rectangular even if you use an irregular polygon area to set it
2. Under the *Raster Analysis Settings* area:
    a. Set the cell size (use this for all your raster analysis, so think about the cell size carefully)
    b. Set a mask if you want (a polygon layer that represents the study area – the area outside this layer will not be included in the analysis – if you use an irregular polygon layer to set this, the mask will reflect that

Spatial Joins and Overlays

Spatial joins and overlays are methods for combining information between GIS layers. Most of these tools will result in the attribute information from one layer being combined with the attribute table of another layer. This then makes further analysis possible. For example, you could perform a spatial join between facilities that release pollutants and town boundaries. You could then calculate how many facilities are in each town and the quantity of pollutants released in each town.

All these tools are found in *ArcToolbox – Analysis Tools – Overlay*.

The *Spatial Join function* can also be accessed by right-clicking on the layer to which you want to join the information from another layer – look for Joins and Relates – Join, and make sure to choose the method to *Join data from another layer based on spatial location*.

A major difference between the *Spatial Join* and the other *Overlay* tools is that the *Spatial Join* does not alter the original geography in any way. It simply joins the attribute tables of two layers together based on common spatial locations. In that way it is similar to a table join (based on a common attribute field value). The other *Overlay* tools will result in altered geography – for example, if you perform an Intersect between town boundaries and water bodies, the result will be a water bodies layer where water polygons are split at town boundaries instead of crossing them. This would then allow you to calculate the total surface are of water bodies per town.

Important note: when using the spatial overlay tools, ArcGIS **DOES NOT recalculate area, length, or perimeter, or any of the other attributes** in the new shape file that results. You must recalculate area, length, and perimeter using the Calculate Geometry tool in the attribute table.

For help on these tools, go to the online ArcGIS 9.2 Help for Geoprocessing Tools - http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Geoprocessing_tools – then go to *Fundamental Tool Concepts*, then look for *Overlay Analysis*

Below are some types of analyses that can be performed with spatial joins and overlays. Note that a few of these can be achieved by using the Select by Location tools.

Overlay Analysis

Questions to be answered	Example	Data/tool set
How many of features from one layer are inside or near features from another layer	How many low income housing tax credit properties are within each ward in Boston? What is the total number of low income units by ward?	*Spatial Join* – works best for points in polygons or points near lines

Tag features in one layer with an important attribute from an overlaying layer	Attach the number of floors from a parcel layer to building footprints to create a 3-D building view; tag each state-regulated facility emitting water pollutants with the watershed name in which it lies.	*Spatial Join* – works best for points in polygons or points near lines *Intersect* overlay works better for polygons in polygons or lines in polygons – the result is a new shape file of the intersection of the two layers, with attributes from both tables
What combination of properties from two or more layers occur in a given area? Where are certain combinations of properties found? How much of one layer is found in another layer?	What is the soil type and land cover for any given area? Where can you find residential zoning on currently forested land? Summarize the total acres of each soil type by watershed	*Union* overlay – the result a new shape file with the combined polygons of each layer splitting each other attached to a table combining the attributes of each original layer
Exclude areas of one layer based on features in another layer	What is the buildable area of a town, excluding current buildings and floodplains?	*Erase* tool (needs ArcInfo license). Use one layer (e.g., building footprints and selected "special flood hazard areas" from floodplains) to erase through another layer (e.g., parcels)

Density Analysis

There are two basic approaches to exploring the density of values in GIS. You map simple calculations of the value of a given attribute by area, or you can create density surfaces for points or line vector data using the *Spatial Analyst* extension. Note that to use the *Spatial Analyst* extension, you must first enable it. To do this in ArcMap, click on the *Tools* menu, go to *Extensions*, and check mark *Spatial Analyst*. You can then load the *Spatial Analyst* toolbar (*View – Toolbars*) or open *ArcToolbox – Spatial Analyst* and access the tools from there.

Note: if you need to create a new field for area or length, you can open the attribute table (polygons or lines only), add a new field (type=double), right-click on the field name, and use the *Calculate Geometry* function to calculate area or length in the units of your choice. If you want a

permanent density field for further analysis, create a new field and use the *Field Calculator* to add the results of a calculation dividing your value by area.

Questions to be answered	Example	Data/tool set
What is the density of a value in a defined area	Show population density (population/area) by census tract	Symbology properties for mapping – normalize one value by the area
Visualize density patterns across an area	Show crime, population, disease, or pollution patterns	Spatial Analyst – Density tools
Be able to make a numeric estimate of a density value for any point in an area	What is the development pressure as represented by the density of new building permits on environmentally sensitive features? What is the surrounding street connectivity index for every school in a city?	Spatial Analyst – Density tools Spatial Analyst – Extraction – Extract Values to Points

Density surface notes:
For help, see ArcGIS 9.2 online help – Overview of Density Tools - http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=1860&pid=1859&topicname=An_overview_of_the_Density_toolset

You can create density surfaces for points or lines.

You can create a density surface that simply counts the points or lines per area, or you can use an attribute field within the point or line field (e.g., population count, dollar value, total toxics released, number of employees).

Make sure your cell size is consistent with both your purpose and with other grid layers in your analysis. Think about cell size, extent and mask in advance.

The smaller the search radius you choose, the more detailed the density surface is. The larger the search radius, the broader the pattern. Large is good for seeing broad patterns over a region. Small is good for estimating what is going on in a specific area.

Kernel density tools smooth the surface by assigning a larger weight to the point or line and dropping away more quickly from those points. The regular point or line density tools simply count what is within your search radius and divide by area without any smoothing.

The default map that ArcGIS creates for a density surface is not necessary the best way of visualizing it. Once it is created, go into the Symbology properties and explore different ways of mapping, different classes, excluding 0 values, possibly excluding outlier values, etc.

Proximity Analysis

Questions to be answered	Example	Data/tool set	Functions
What's within a set distance?	Characterize the demographics of ¼ mile around parks	Vector – Selection *Vector – Analysis Tools - Proximity*	Select by location Buffer (then overlay or spatial join)
What's the closest facility in terms of straight line distance (crow flies)? On average, how far are one set of points from the closest point or line in another layer?	Distance between parcel points and the nearest grocery store as the crow flies (one result per parcel point) Distance between schools and the closest Toxic Release Inventory reporting facility. On average, how close are Boston area schools to a TRI facility? Which schools are within ¼ mile of a facility?	*Vector – Analysis Tools-Proximity*	*Near* tool (ArcInfo only) – works on points to points or points to line - (for polygons, need to convert polygons to centroid points first[1]) The result is that two new fields are added to the input feature data layer, a NEAR_FID recording the feature ID of the nearest feature, and a NEAR_DIST field containing the distance to that facility. You can create a table join between the input layer (e.g., parcel centroids) and the "near features" (e.g., grocery stores) based on the Near_FID and original FID of the near features. You could also get statistics from the Near_Dist field to know max, min, mean, and standard deviation for distances between these points.
What's the closest facility along a street or other set path? What is the closest facility along a path/street in terms of time or other cost?	Distance between parcel points and the nearest grocery store along a street network (one result per parcel point)	*Vector – Network Analyst* (you must first set up a network first using existing street or other line data[2])	Closest Facility function You can specify only distance as the cost of travel, or time or other cost.

[1] To convert a polygon or line to centroid points, use ArcToolbox – Data Management Tools – Features – Feature to Point tool (ArcInfo only). This creates a new point data layer with all the attributes of the original polygon or line file. The point is the centroid of the polygon, or the center point of the line feature.

[2] To create a network from an existing street centerline file, open ArcCatalog, right-click on the street centerline file, and choose New

What is the straight line distance between all points in one layer to all points in another layer?	Distance between parcel points and *all* grocery stores in an area (you can specify a maximum search radius) Distance between schools and all nearby TRI reporting facilities? How many Boston metro schools are withn ¼ mile of more than one TRI facility?	*Vector – Analysis Tools-Proximity* Hawths Tools	*Point Distance* tool – output is a table with the input layer FID and the Near_ID ("near" layer FID) and Near_Dist for every point to every point. Alternatively, use Hawths Tools – Distances Between Points (Between Layers) – with this you can specify a unique ID other than FID. Output is a table.
What is the straight line distance between all or selected points *within* a layer?	How far is each school in a layer from every other school?	Hawths Tools – Analysis Tools	*Distances Between Points (Within a Layer)* – output is a matrix of distances between points, or summary stats
What is the distance at any given location on a map to the nearest feature type of interest?	Visualize distance from roads or parks or pollutant sources across a surface. Create a distance preference grid for use in another model (e.g., for affordable housing, the shorter distance to a transit station the better)	*Raster – Spatial Analyst - Distance*	Euclidean Distance is the tool for straight line distance. Cost Distance can incorporate variables that make travel more difficult (e.g., slope, crime, forest)
How far are points from the edge of the nearest polygon, point, or line?		*Raster – Spatial Analyst – Distance, then Extraction – Extract Values to Points*	Use one of the distance tools, then extract the values of that grid to points. This is a variation of some of the earlier questions. Using a raster distance grid and extracting point values will be a rougher approximation of distance than doing a true point-to-point distance. But this function works for distance to polygons where as the point functions do not.