Topic Series 19

GIS Spatial Operators, Overlay Operations, Models and Modeling

Spatial Operators

We have briefly touched on concepts on spatial arrangement before and will now examine arrangement of entities we describe by **point, line and polygon** vector constructs. One type of assessment GIS can help us with is in determining **density** and **pattern** among features.

In **point** represented entities, the density may be of importance (for example, bird counts or stems in a stand). The **pattern** can be characterized as **uniform** if the number of points per unit area for a small area is the same as the number of points per unit area over a different area or a larger region. If spread as equally spaced, we further describe the pattern as **regular** as opposed to **random** (scattered). And as before, points can be also occur in a **clustered pattern**. **(overhead)**

Polygons can be treated in some respects in similar fashion as points. We can look at **pattern and density of polygons**. To get **density**, we must also consider the **area** of each class of polygons and the **number of polygons** represented. For some applications, it may be important to know the **number of polygons** per unit area (say in **housing assessments**) while in others the **percentage area** covered by a class and the **contiguity** may be important (**wildlife habitat**). There are a number of measures of contiguity that are used in wildlife sciences that are covered in the textbook.

Line density is measured as **length of lines per unit area**. This measure may be important again in wildlife applications. For instance, we may be interested in road density as a limiting factor in black bear populations. **Distributions** of lines can be assessed by various forms of **intercept methods**. **Directionality and orientation** of linear features is also something that can be observed in GIS. Examples of these types of assessments can be drawn from remote sensing. There are numerous studies that deal with **lineaments** as interpreted from images as possible indicators of **geologic faults**. **(overhead)**

Proximity Analysis

Proximity analysis we'll define as some of the **basic operators** in GIS that **act on spatial arrangement** of point, line and polygon entities. It relies on the measurement capabilities of GIS to examine neighborhoods for closeness of features. **Distance between two features** can be determined roughly by use of interactive measurement capabilities (for instance the **measure tool** in ArcView). More robust approaches would utilize the data search capabilities to select the features of interest and then **calculate exact distances** between them. **What is another form of proximity analysis we've touched on? (Ans. Buffers)**

The example illustrates some of the more complex things we would do with proximity searches. The scenario here is disaster response. Here, an oil spill has occurred. To mobilize efforts we need to know where the closest response team is. This could involve other facilities located near a city. Speed of boats that can arrive on the scene can be a factor entered into the GIS. Therefore, response time is also a component of the closeness determination.

Aggregation - is a form of proximity analysis that utilizes nearness of feature to each other to form groups. One reason for use of this type of analysis is to generalize data (remove extraneous or confusing small polygons).

Proximity Operations

Conflation - is the opposite process to development of buffer zones around areas. This is done to represent an area with the smallest possible dimension feature. For instance, the T- shaped irrigation area in the example, or the centerline of a road.

Centroids - Polygons can be represented by centroids. These can be either determined geometrically or, in the case of Arc/Info by placement of label points to represent a single graphic representation of the polygon location.

Thiessen Polygons - these are the opposite of centroids. They define equal areas around points. Although these do not have a lot of practical application in natural resource management, they can be used to visualize areas of influence such as wildlife habitat ranges or perhaps areas of influence of individual tree canopies. They are also useful as a tool in some **market analysis** work such as **areas of influence** for shopping malls, etc.

Network Operations

Network operations rely on the aspect of topology that involves **connectivity** of linear entities. They also rely on attributes of the lines that dictate how they can be traversed. One of the basic operations is to determine the shortest path between two points along a network. This process is called routing.

A road network has multiple roads that are one way. The problem also must consider features such as the roads-under-construction. As we have stated before, these processes are under greater use in emergency services and in delivery services in urban areas.

Cars with GIS and GPS systems have already been developed that demonstrate the utility of routing to get you from city to city with guiding maps and your present location. Larger trucking firms are now putting GPS into their vehicles with network capabilities. You can now buy cars with this same capability.

Routing can involve both constraints on the roads (**one-way**, **speed limits**, etc.) or constraints due to **type of road** (travel slower on gravel). Additional barriers may pop up and have to be added that change older routing plans. GIS lets us efficiently look at these problems and added constraints as they come up. **Network analysis** allows us to plan bus routes to pick up kids and deliver them to school. But the network must be optimized so that no child has to walk further than one block to a bus stop.

Overlay Operations

The concept of map overlays has been around almost as long as maps (well maybe the ancient Romans didn't have **velum or acetate**, but map comparisons have been around a long time). In modern times, we've used transparent drafting materials to superimpose different information on cartographic bases to evaluate spatial relationships between mapped variables. Now, computers make this process infinitely easier and more accurate (assuming we have **correct projections/datums**). The big downside to computer overlay is that we can too easily combine data sets that should not be combined (i.e. wrong datum, great differences in scale, etc.) and as a result, get wrong answers to our questions.

This topic will examine the many facets of overlay operations available with GIS that include simple cartographic overlay to complex data extraction procedures.

Thematics to General Reference or Cartographic Overlay

Map separates. In data collection and entry, it is better to keep **different themes** as separate layers so they can be **recombined** in different ways. **Cartographic overlay** is the first and most frequently used form of overlay you used in the laboratory exercises in ArcView. You've spend a lot of time bringing in coverages (themes) and graphically superimposing them to look at the relative spatial relationships among features. This operation is reserved for **cartographic output** while the next, **overlay**, gives us both the combination of theme graphics and their attributes.

Overlays

In **overlay operations** the merger of the attribute tables is as important as the graphics. The combined attributes give us more information about each area to work with to either extract more specific information or make decisions. We can now ask, “where are trees located on clay soils” or “how many hectares are occupied by brush on loam soils.” These are both examples of basic polygon overlay capabilities in GIS. Note that although the result polygons seem extremely complex, the combined attribute tables and associated polygons from the overlay operation give us the flexibility to spatially query the attributes in a number of ways to produce the desired information.

Applications

On the top we have a three coverage example to give a composite data set based on **overlays**. The combination of **soils, crop, and farm practice** could, in theory, be used to **predict yield**. This same sort of example can be used **in forestry: timber type, soil (SI), silvicultural practice; in wildlife management: habitat, edge, management practices; in fisheries: water quality, structure, forage populations.**

The second example deals not with recombination of polygons, but with finding **point or line features** that fall **within polygons**.

Raster Overlay (Map Algebra)

Here we are talking about **overlay of raster data** structures to determine new information by mathematical **combination of the numeric values (raster example)**. These operations can utilize binary representations of files to find coincidence of features. This is a simple example of a **multiply operation**.

Map algebra can also be used to determine the **maximum or minimum** value in coincident cells. Here we are interested in what the maximum rainfall is for two or more dates, each as a separate layer in the database.

Overlay operations can be combined with a form of **recode** operation. The predetermined outcomes can be listed in a **look-up table** to consult in order to code the final map.

Weighting involves use of a operation that multiplies the cell values in each input coverage by a constant. In an example, the outcome is determined by **weighted inputs** of elevation, soils, and slope to determine the sites suited for agriculture. In this case soils are more important so they carry twice the weight of the other layers.

A **matrix add by recode** is something like a lookup table mentioned before. Here we define a matrix of all possible combinations with each input represented on the axes of the matrix. In this way, we can ensure that there are no ambiguous codes (same code for two different combinations).

Vector Overlay

In GIS based on vector topologies, overlay operations involve recombining the vectors and creating output attribute tables that reflect the combination of **all input attributes referenced to the new vector attributes**.

With vector overlay functions, we use **Boolean-base operators** (answer yes/no condition questions) to combine areas of interest.

Consider the examples: **(overhead)**
To combine A and B:

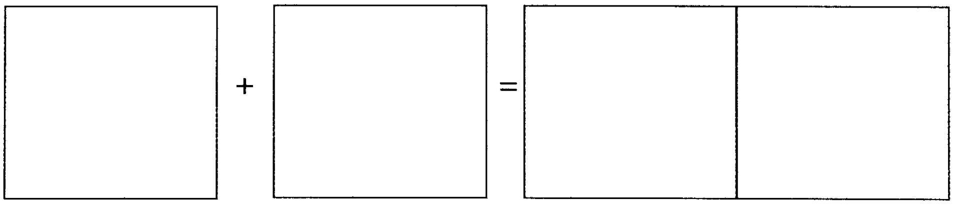
Intersect - if we ask question where is A AND B (yes) you get the area on left. In Arc/Info, the intersect operation that accomplishes this is called **Identity (information from both coverages is retained in the output)**. The **Intersect** command in Arc/Info extracts the information from one coverage that falls within another. If we ask what area by A XOR B (exclusive OR) but not both (yes) we get the right hand side dark area. The area can be A or B but not both.

Union - is to combine both into one common area (all features in both coverages are retained). A OR B we get the right side answer. The opposite answer is to show all area that is common to neither (NOT A OR B).

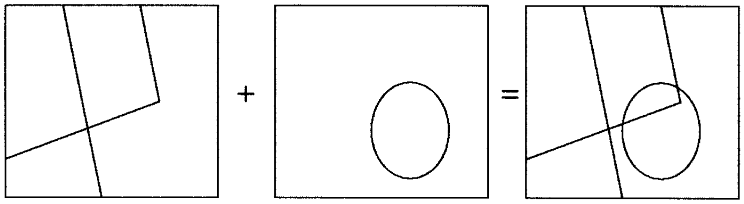
NOT - this is used to exclude area from the coverage of interest based on overlap from another coverage (bottom right).

APPENDS, UNION, INTERSECT, IDENTITY functions: **(overhead)**

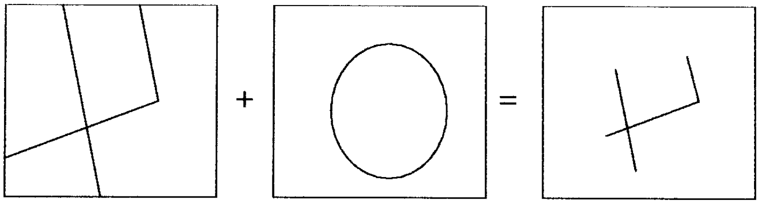
Append



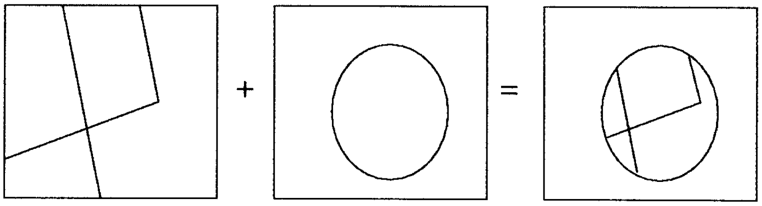
Union



Intersect



Identity



Clip and Mask

These are fairly standard operations to GIS.

Clipping involves extraction of features in one coverage based on their overlap with features of another (often referred to as "**cookie cutter**"). This is accomplished by the **Intersect and Clip** commands in Arc/Info. Clipping is an **extraction operation** while **masking in an exclusion operation**.

Masking involves using a coverage that excludes everything but the information in the area of interest (simple example is a window). **Non-destructive** erase function.

Erase operations use the features of one coverage to define what to eliminate in another. This is a **destructive** function.

Replace

This is a special type of overlay operation where you ask that information in one coverage takes precedence over the other.

Models and Modeling

Models

Two definitions:

1) Recall that a model is a **mathematical representation** of real world conditions.

2) A model can also be a set of procedures that can be repeated for different areas with similar conditions.

We try to simplify the world into a logical ordered set of expressions. **Every GIS coverage is a model.** Complexity in modeling operations comes into play when we use one or more coverages to try to **simulate or predict conditions** of the landscape (present, future or past). So a model can use models to model things!!!???

In general, we can think of modeling processes in GIS as use of a sequence of **commands** and **procedures** to accomplish a complex task. Simple models will subject data sets to some logical or mathematical progression to derive new information for use in a decision process. More complex models may use many data sets and produce different levels of both intermediate and final products.

You may hear the term **decision support system** used to reflect the capabilities of complex modeling systems that provide information on land/resource management issues. For example, many large forestry companies use GIS for their timber management and harvest scheduling. The GIS-based decision support systems they use are complex models that often combine GIS and linear programming to determine optimal scenarios for management of wood supplies.

A simple modeling concept is the **overlay process**. We can combine or overlay GIS coverages to look at the co-existence of two or more features on the landscape (say hardwoods and bottomland soils). Here we are interested in how well we can predict timber type based on soils. The result of this spatial overlay helps us predict (model) where hardwoods occur in other areas given soils data.

Some aspects of models include:

- Representation of reality (top left)
- Show interrelationships among features (top right)
- Can be a set of procedures (bottom left)
- Can predict condition (past, present, future) of an area based on other areas (bottom right)
-

Generalization Model This simple example demonstrates how we take the original data and reduce the complexity to produce a more generalized coverage. This may be done as a simple **recode** of data based on criteria that define new categories (recall classification processes discussed earlier).

Another modeling process is where we have **statistical surfaces**; the model is derived from the field contour data by interpolation of contours from the original point elevation data. The contour map is a model developed by the interpolation modeling process.

Site Suitability Models

A type of model that we often employ in GIS is for site suitability determination. **Site Suitability** analysis usually involves use of combined procedures to define a location or area that is best suited to some activity or use. The same processes can also be combined to define areas to be avoided for specific purposes.

More frequently, suitability analysis involves combination of the graphics and attributes of several coverages**.** **Degree of suitability** can be determined in order to rate sites based on different desirable characteristics. **Sensitivity analysis** is especially useful in environmental applications.

The **database approach** to site selection makes use of the multiple phase query operations that can narrow down the selections. The final selections can then be displayed and also subjected to other geographic analysis (proximity measures, etc.) to select the site.

An example that utilizes both database and geographic measurement capabilities of GIS can be fairly complex. For instance, a **flow diagram** of all steps to select a site. Flow diagrams are very useful planning tools in the model development process and should be utilized to organize the flow of data. A well thought out flow chart will save an enormous amount of time and frustration in developing a complex model. Modern software now has the ability to graphically define spatial models (e.g. ERDAS Spatial Modeler).

Overlays

Overlay capability is frequently the main focus of site suitability and other modeling activities. This process provides outputs that can be evaluated based on the combined attribute inputs of the input coverages.

Binary Overlays The original attributes of the input coverages can be **re-coded** in a preliminary step to produce the suitable(1) / non-suitable (0) categories for recombination. The **multiply operation** is used to determine if an area is suitable based on two coverages. The **additive model** would rank the sites. This process can also work for multiple coverages given as binary inputs. If all factors are equally rated, then the highest sum for the location is the best site. Weighting can be used to place more importance on some of the inputs.

Matrix overlay operations utilize a matrix (**look up table**) to predetermine ratings based on different overlay outcomes.

Consider the multiplicative and additive overlay combinations of three coverages (soils, slope, elevation). The multiplicative function yields an output that indicates if the site is suitable. The additive model can be used to rank the sites as Best, Acceptable, or unsuitable. The overlay results can be shown in topological format for polygons 1-6. In this type of combination, the original attributes were preserved and additional attributes were created to show the results of the recombination.

Environmental Modeling

GIS is used heavily in **environmental models.** Information on complex landscapes can be used to determine the impacts dues to pollution. The top of the page indicates two point sources of pollution and the **proximity to wetlands**. A more plausible scenario would be that there is information known about the **nature of spread of the hazard**. (down stream, down wind). This information can be generated by modeling the flow patterns through analysis of **statistical surfaces** (DEM for flows) or **network (stream beds) operations**.

GIS can also be used to look at **3-D relationships** (atmospheric effects, or ground water contamination at different levels.)

GIS can be used to determine potential impact sensitivity to a particular hazard.

A simple **predictive model** allows inputs of topography (DEM), hydrology, land cover, dam size and location to determine the possible extent of a reservoir and what will be impacted by the impoundment.

Statistical Models

GIS data can be used to develop statistical models. These are models that are mathematically descriptive of **trends, averages, deviations**. Examples of statistical models are in utilization of tabular data to predict trends or future conditions (e.g. regression) then express these in the resulting geographic output graphics.

Examples of statistical models:

1) prediction of future volume of timber given age class, site quality, etc.
2) prediction of wildlife population based on present populations, habitat quality and distribution, predicted changes in land cover.
3) Agriculture yields based on past yields, nutrient status, weather patterns
4) Population distribution trends based on historic changes in distribution of land-cover.

Time-Series Models

These models may be either **historic** (document past changes) or **predictive** (urban sprawl based on present land-use, topography, transportation, past spread)

Process Models

These are a set of steps that can be repeated by substitution of different data sets.

Model Validation

A short note on model validation. Computerized mapping and data delivery has often failed to meet expectations of the users due to a lack of understanding of the veracity and utility of data sets used in the decision process. We've discussed on prior occasions the problems associated with mixing incompatible data. These can be further exacerbated by using the wrong data in modeling to develop new information. Hence there is a need for establishment of the credibility of the data and models. Data credibility is best determined by recording the origins and accuracy of input data sets through close ties to metadata. Model credibility can be assessed by testing to independent sets of knowns.

Question: **(overhead)**

1) are the data **representative** of the area?
2) **Model combines data correctly**?
3) **Model produces acceptable results**?

Key assurances can be incorporated by: **(overhead)**

1) small **pilot** to test data inputs
2) use of a **control** or test database in the model
3) **backtracking** procedures to reproduce inputs from the outputs.