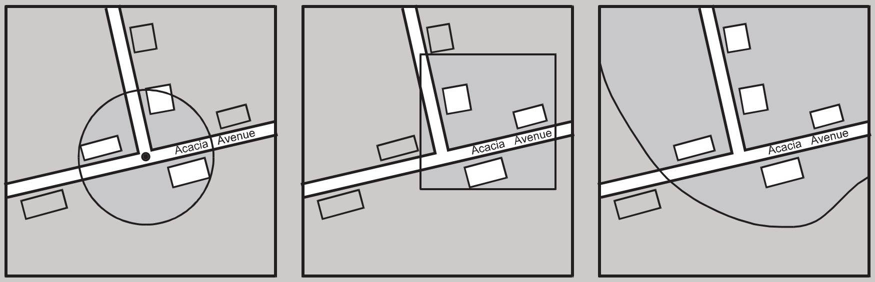
CHAPTER 9

GEOGRAPHIC QUERY AND ANALYSIS



How many houses are within 50 m of this junction?

How many children live in this 100 m grid square?

Which households fall within the floodplain?

Spatial analysis is in many ways the crux of GIS, because it includes all of the transformations, manipulations, and methods that can be applied to geographic data to add value to them, to support decisions, and to reveal patterns and anomalies that are not immediately obvious – in other words, spatial analysis is the process by which we turn raw data into useful information. If GIS is a method of communicating information about the Earth's surface from one person to another, then the transformations of spatial analysis are ways in which the sender tries to inform the receiver, by adding greater informative content and value, and by revealing things that the receiver might not otherwise see. Some methods of spatial analysis were developed long before the advent of GIS, and carried out by hand, or by the use of measuring devices like the ruler. The term analytical cartography is sometimes used to refer to methods of analysis that can be applied to maps to make them more useful and informative, and spatial analysis using GIS is in many ways its logical successor.

Spatial analysis is the crux of GIS. Spatial analysts can reveal things that might otherwise be invisible – it can make what is implicit explicit.

Here, we will look first at some definitions and basic concepts of spatial analysis. Further, we look at spatial analysis grouped into six distinct categories – queries and reasoning, measurements, transformations, descriptive summaries, optimization, and hypothesis testing.

Methods of spatial analysis can be very sophisticated, but they can also be very simple. A large body of methods of spatial analysis has been developed over the past century or so, and some methods are highly mathematical – so much so, that it might sometimes seem that mathematical complexity is an indicator of the importance of a technique. But the human eye and brain are also very sophisticated processors of geographic data, and excellent detectors of patterns and anomalies in maps and images. So the approach taken here is to regard spatial analysis as spread out along a continuum of sophistication, ranging from the simplest types that occur very quickly and intuitively when the eye and brain focus on a map, to the types that require complex software and sophisticated mathematical understanding. Spatial analysis is best seen as collaboration between the computer and the human, in which both play vital roles. Effective spatial analysis requires an intelligent user, not just a powerful computer. Spatial analysis helps us in situations when our eyes might otherwise deceive us.

There are many possible ways of defining spatial analysis, but all in one way or another express the basic idea that information on locations is essential – that analysis carried out without knowledge of locations is not spatial analysis. One fairly formal statement of this idea is: 'Spatial analysis is a set of methods whose results are not invariant under changes in the locations of the objects being analyzed'. The double negative in this statement follows convention in mathematics, but for our purposes we can remove it: 'Spatial analysis is a set of methods whose results change when the locations of the objects being analyzed change'.

On this test the calculation of an average income for a group of people is not spatial analysis, because it; in no way depends on the locations of the people. But the calculation of the center of the New Delhi's population is spatial analysis, because the results depend on knowing where all Delhi residents are located. GIS is an ideal platform for spatial analysis because its data structures accommodate the storage of object locations.

Spatial analysis can be used to further the aims of science, by revealing patterns that were not previously recognized, and that hint at undiscovered generalities and laws. Patterns in the occurrence of a disease may hint at the mechanisms that cause the disease, and some of the most famous examples of spatial analysis are of this nature, including the work of Dr. John Snow in unraveling the causes of cholera (Figure 9.1 and Box 18).

It is interesting to speculate on what would have happened today, if early epidemiologists like Snow had access to a GIS. The rules governing research today would not have allowed Snow to remove the pump handle, except after lengthy review, because the removal constituted an experiment on human subjects. To get approval, he would have to shown persuasive evidence in favour of his hypothesis, and it is doubtful that the map would have been sufficient, because several other hypotheses might have explained the pattern equally well. First, it is conceivable that the population of Soho was inherently at risk of cholera, perhaps by being comparatively elderly, or because of poor housing conditions. The map would have been more convincing if it had shown the rate of incidence, relative to the population at risk. For example, if cholera was highest among the elderly, the map could have shown the number of cases as a proportion of the population over 50 years. Second, it is still conceivable that the hypothesis of transmission through the air between carriers could have produced the same observed pattern, if the first carrier lived in the center of the outbreak. Snow could have eliminated this alternative if he had been able to produce a sequence of maps, showing the locations of cases as the outbreak developed. Both of these options involve simple spatial analysis of the kind that is readily available today in GIS. Spatial analysis in GIS provides tools that are far more powerful than the map at suggesting causes of disease.

Today the causal mechanisms of diseases like cholera, which results in short, concentrated out breaks, have long since been worked out. Much more problematic are the causal mechanisms of diseases that are rare and not sharply concentrated in space and time. This example is of inductive use of spatial analysis, to examine empirical evidence in the search for patterns that might support new theories or general principles. Other uses of spatial analysis are deductive, focusing on the testing of known theories or principles against data. A third type of application is normative, using spatial analysis to develop or prescribe new or better designs, for the locations of new retail stores, or new roads, or new manufacturing plant.

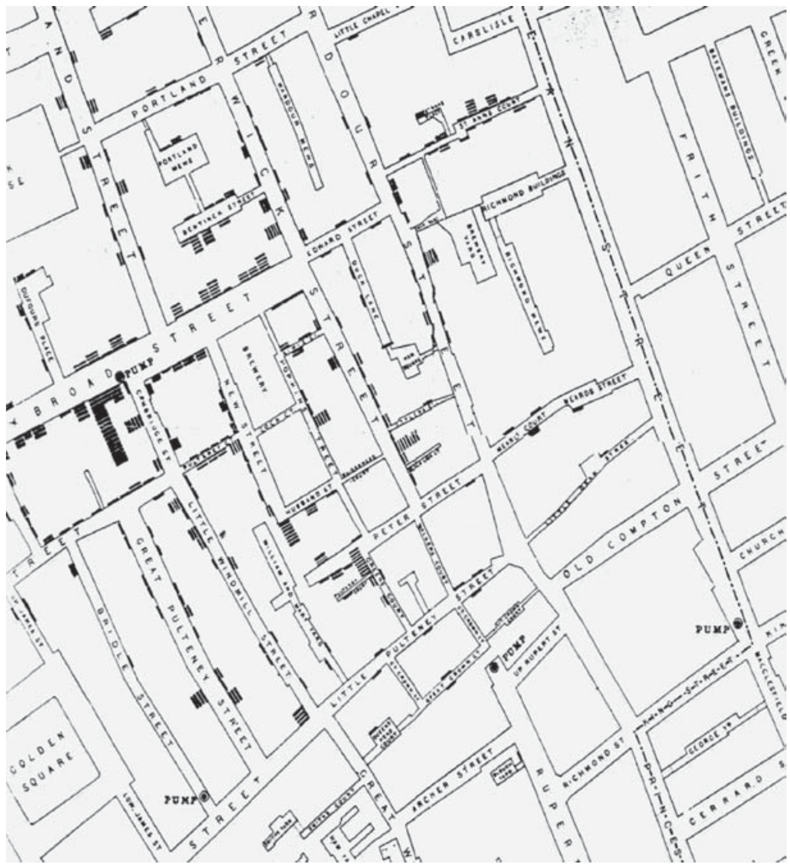


Figure 9.1: Dr. John Snow and the causes of cholera in London.

Box 18: Dr. John Snow and the causes of cholera

In the 1850s cholera was very poorly understood, and massive outbreaks were a common occurrence in major industrial cities. An outbreak in London in 1854 in the Soho district was typical of the time, and the deaths it caused are mapped in Figure 9.1. The map was made by Dr. John Snow, who had conceived the hypothesis that cholera was transmitted through the drinking of polluted water, rather than through the air, as was commonly believed. He noticed that the outbreak appeared to be centered on a public drinking water pump in Broad Street, and if his hypothesis was correct, the pattern shown on the map would reflect the locations of people who drank the pump's water. There appeared to be anomalies, in the sense that deaths had occurred in households that were located closer to other sources of water, but he was able to confirm that these households also drew their water from the Broad Street pump. Snow had the handle of the pump removed, and the outbreak subsided, providing direct causal evidence in favour of his hypothesis. This was perhaps the first use of cartographic techniques for solving a real world problem.

Box 19: Spatial relations and analysis on geometric objects

There are nine methods for testing spatial relations between geometric objects. Each takes as input two geometries and evaluates whether the relation is true or not.

Equals – are the geometries the same.

Disjoint – do the geometries share a common point

Intersects – do the geometries intersect

Touches – do the geometries intersect at their boundaries

Crosses – do the geometries overlap

Within – do the geometries within another

Contains – does one geometry completely contain another

Overlaps – do the geometries overlap

Relate – are the intersections between the interior, boundary or exterior of the geometries.

Seven methods support spatial analysis on these geometries:

Distance – determines the shortest distance between any two points in two geometries.

Buffer – returns a geometry that represents all the points whose distance from the geometry is less than or equal to a user defined distance.

Convex hull – returns a geometry representing the convex hull of a geometry (convex hull is the smallest polygon that can enclose another geometry without any concave areas).

Intersection – returns a geometry that contains just the points common to both input geometries.

Union – returns a geometry that contains all the points in both input geometries.

Difference – returns a geometry containing the points that are different between the two geometries.

SymDifference – returns a geometry containing the points that are in either of the input geometries, but not both.

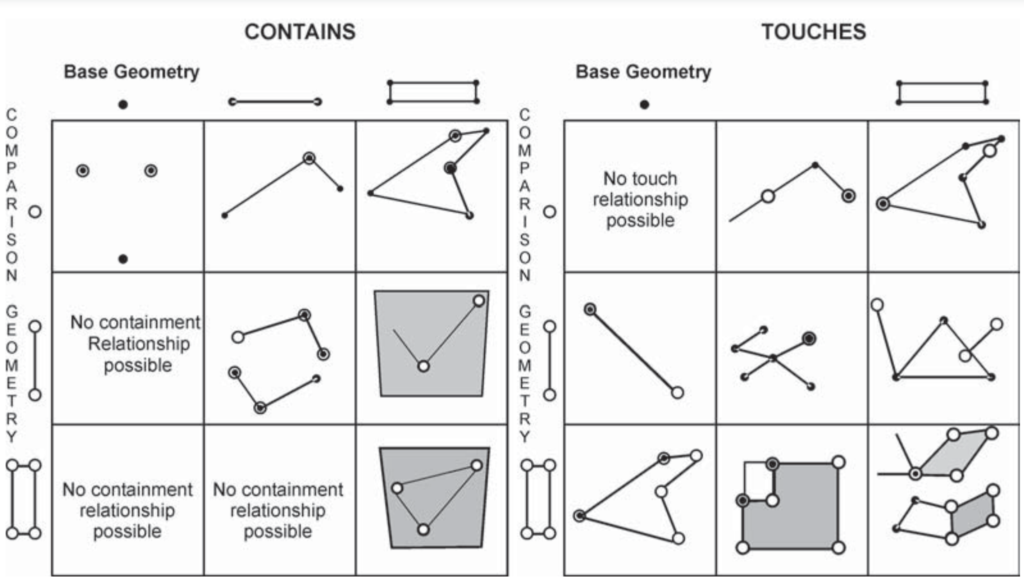


Figure 9.2: Examples of possible relations for two geographic database.

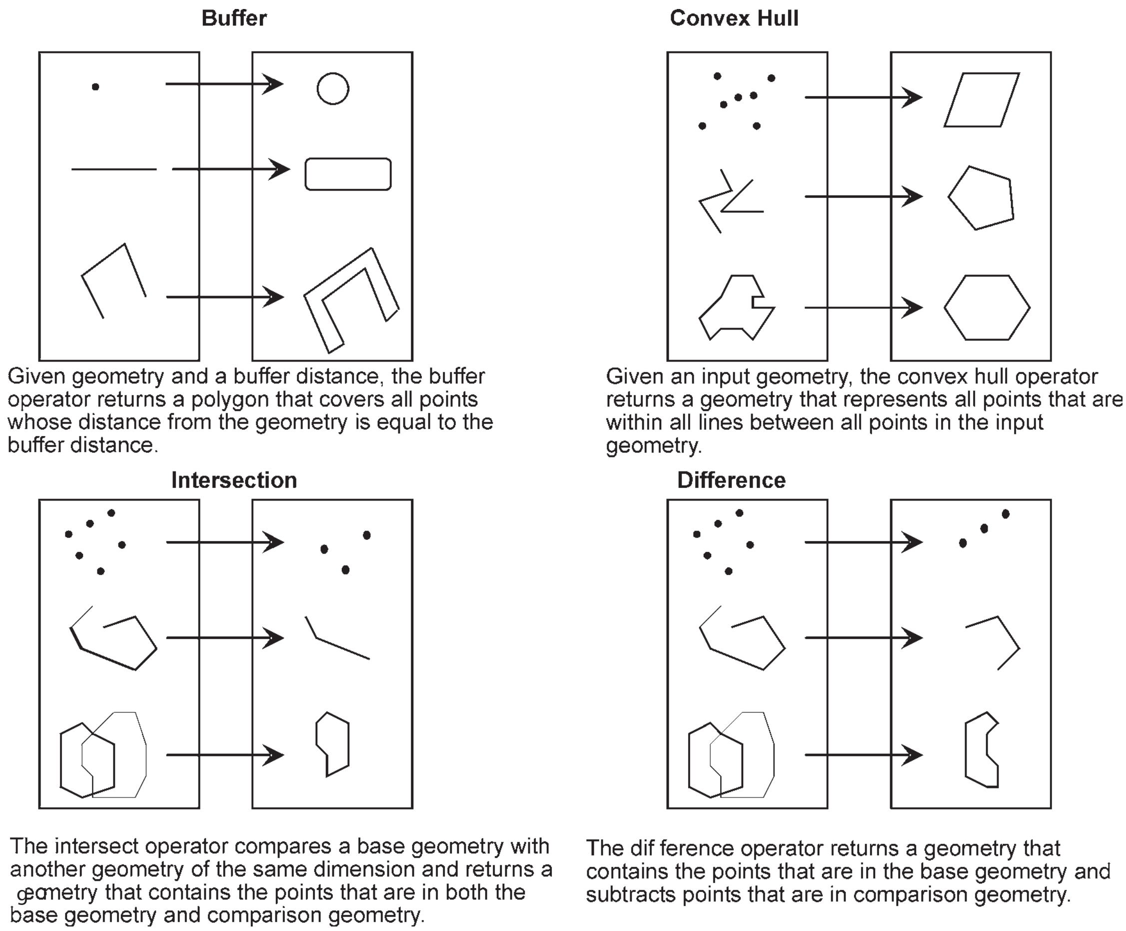


Figure 9.3: Examples of spatial analysis methods on geometries.

TYPES OF SPATIAL ANALYSIS

We would focus on methods of spatial analysis using six general headings:

i. *QUERIES AND REASONING* are the most basic of analysis operations, in which the GIS is used to answer simple questions posed by the user. No changes occur in the database, and no new data are produced. The operations vary from simple and well-defined queries like 'how many houses are found within 1 km of this point', to vaguer questions like 'which is the closest city to New Delhi going east', where the response may depend on the system's ability to understand what the user means by 'going east'.

ii. *MEASUREMENTS* are simple numerical values that describe aspects of geographic data. They include measurement of simple properties of objects, like length, area, or shape, and of the relationships between pairs of objects, like distance or direction.

iii. *TRANSFORMATIONS* are simple methods of spatial analysis that change datasets, combining them or comparing them to obtain new datasets, and eventually new insights.

Transformations use simple geometric, arithmetic, or logical rules, and they include operations that convert raster data into vector data, or vice versa. They may also create fields from collections of objects, or detect collections of objects in fields.

iv. *DESCRIPTIVE SUMMARIES* attempt to capture the essence of a dataset in one or two numbers. They are the spatial equivalent of the descriptive statistics commonly used in statistical analysis, including the mean and standard deviation.

v. *OPTIMIZATION TECHNIQUES* are normative in nature, designed to select ideal locations for objects given certain well-defined criteria. They are widely used in market research, in the package delivery industry, and in a host of other applications.

vi. *HYPOTHESIS TESTING* focuses on the process of reasoning from the results of a limited sample to make generalizations about an entire population. It allows us, for example, to determine whether a pattern of points could have arisen by chance, based on the information from a sample. Hypothesis testing is the basis of inferential statistics and lies at the core of statistical analysis, but its use with spatial data is much more problematic.

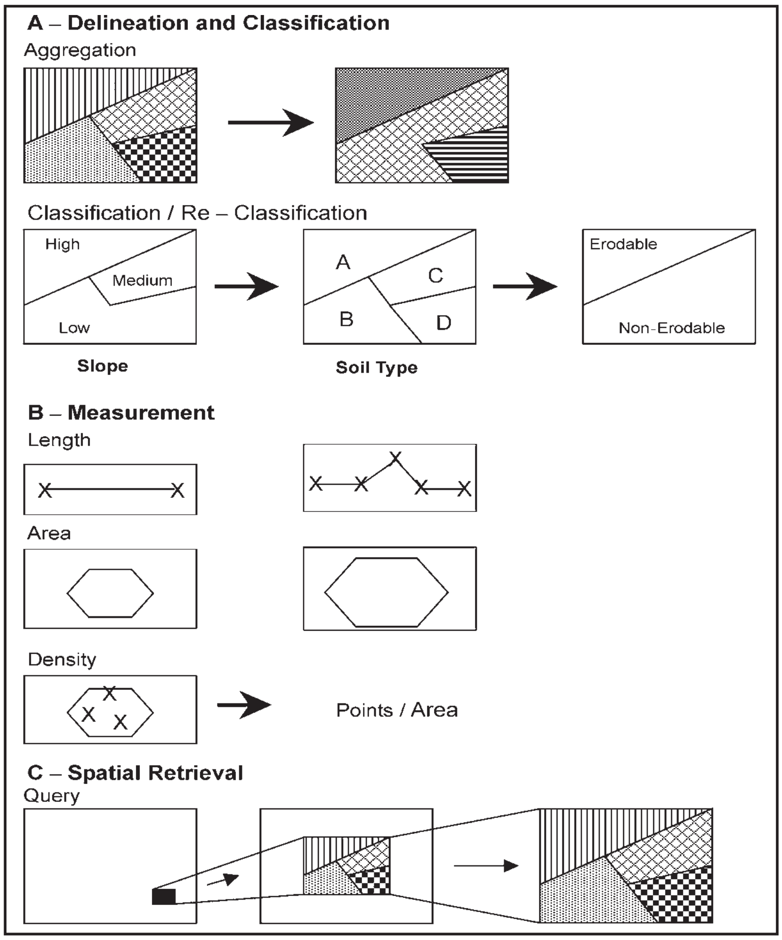
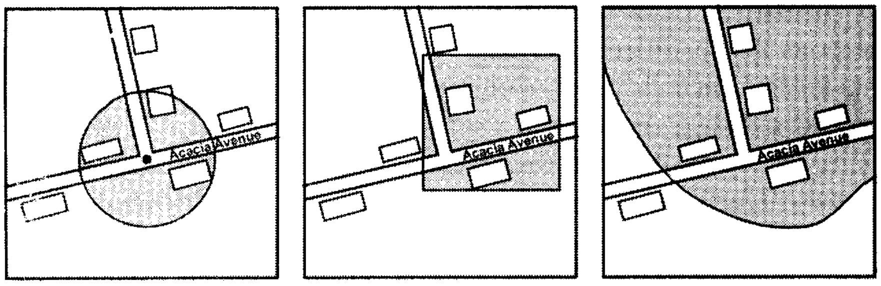


Figure 9.4: Spatial retrieval, delineation and classification, and measurement are separate functions, but are commonly used together.



How many houses are within 50 m of this junction? How many children live in this 100 m grid square? Which households fall within the floodplain?

Figure 9.5: Some examples of spatial query.

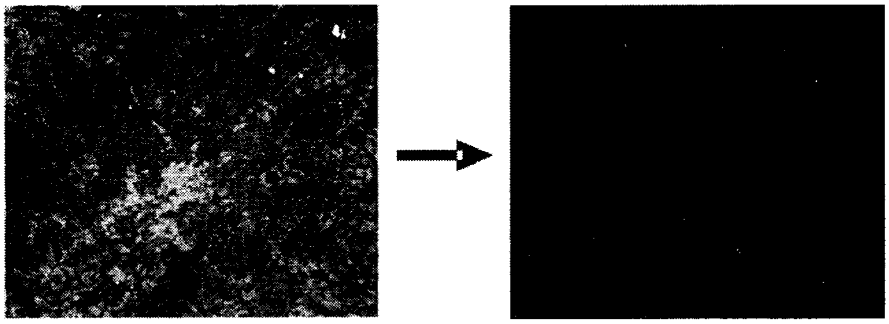


Figure 9.6: Example of *re-classification* where the modification in attribute values are made to produce new object data sets.

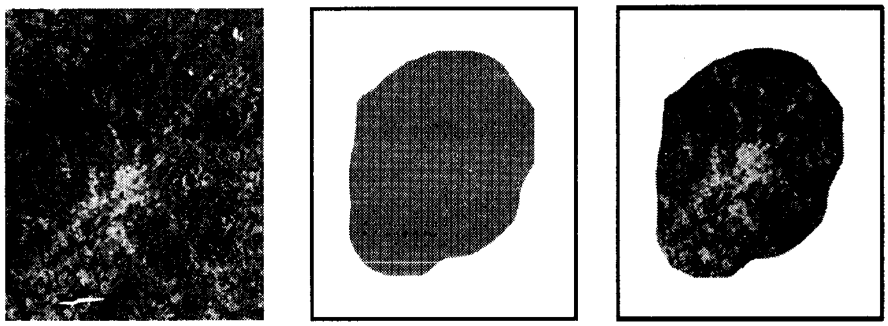


Figure 9.7: Example of *Cookie Cutting* where overlaying of datasets is made, using one dataset as a sieve or cookie cutter to select a subset of the other dataset.

QUERIES AND REASONING

In the ideal GIS it should be possible for the user to interrogate the system about any aspect of its contents, and obtain an immediate answer. Interrogation might involve pointing at a map, or typing a question, or pulling down a menu and clicking on some buttons, or sending a formal SQL request to a database. Today's user interfaces are very versatile, and have very nearly reached the point where it will be possible to interrogate the system by speaking to it – this would be extremely valuable in vehicles, where the use of more conventional ways of interrogating the system through keyboards or pointing devices can be too distracting for the driver.

The very simplest kinds of queries involve interactions between the user and the various views that a GIS is capable of presenting. A Catalog view shows the contents of a database, in the form of storage devices (hard drives, Internet sites, floppies, CDs, or ZIP disks) with their associated folders, and the datasets contained in those folders. The Catalog will likely be arranged in a hierarchy, and the user is able to expose or hide various branches of the hierarchy by clicking at appropriate points. Different types of datasets are symbolized using different icons, so the user can tell at a glance which files contain grids, polygons, points, etc.

In contemporary software environments, such as Microsoft's Windows or Windows NT, or the Macintosh or Unix environments, many kinds of interrogation are available through simple pointing and clicking. For example, in ArcCatalog simply pointing at a dataset icon and clicking the right mouse button exposes basic statistics on the dataset when the Properties option is selected. The metadata option exposes the metadata stored with the dataset, including its projection and datum details, the names of each of its attributes, and its date of creation.

The map view of a dataset shows its contents in visual form, and opens many more possibilities for querying. When the user points to any location on the screen the GIS display the pointer's coordinates, using the units appropriate to the dataset's projection and coordinate system. Today's GIS supports much more sophisticated forms of query than these. Suppose both the map view and the table views are displayed on the screen simultaneously. Linkage allows the user to select objects in one view, perhaps by pointing and clicking, and to see the selected objects highlighted in both views. Linkage is often possible between other views, including the histogram and scatterplot views. For example, by linking a scatterplot with a map view, it is possible to select points in the scatterplot and see the corresponding objects highlighted on the map. This kind of linkage is very useful in examining residuals, or cases that deviate substantially from the trend shown by a scatterplot. The term exploratory spatial data analysis is sometimes used to describe these forms of interrogation, which allow the user to explore data in interesting and potentially insightful ways. Exploratory spatial data analysis allows its users to gain insight by interacting with dynamically linked views.

Second, many methods are commonly available for interrogating the contents of tables, such as SQL. SQL is a standard language for querying tables and relational databases. The language becomes much more powerful when tables are linked, using common keys, and much more complex and sophisticated queries, involving multiple tables, are possible with the full language. More complex methods of table interrogation include the ability to average the values of an attribute across selected records, and to create new attributes through arithmetic operations on existing ones (*e.g.,* create a new attribute equal to the ratio of two selected attributes).

The term reasoning encompasses a collection of methods designed to respond to more complex forms of query and interrogation. Humans have sophisticated abilities to reason with spatial data, often learned in early childhood, and if computers could be designed to emulate these abilities then many useful applications would follow. One is in the area of navigation. Humans are very skilled at direction giving, and computer emulation of these skills would be useful in the design of in-vehicle navigation systems. The difference between the directions given orally to a person and in GIS is obvious that the human's are given in familiar terms, and they use many more landmarks and hints designed to make the driver's task less error-prone and to allow the driver to recover from mistakes. They also use gestures such as pointing that cannot be easily represented in digital form.

One major difference between the two sets is in the use of vague terms. Computers are generally uncomfortable with vagueness, preferring the precise terms (like, start from Geography department, turn right from V.C. lodge, stay straight up to University circle, turn left on Dodhpur road, turn left on to Medical road and stay straight up to Zakaria market). But the world of human communication is inherently vague and full of terms and phrases like 'near', 'north', 'too far', or 'watch out for' that defy precise definition. Very often the meaning of human terms depends on the context in which they are used. For example, Agra may be 'near' New Delhi in a conversation in Chennai, but not in a conversation in Aligarh.

MEASUREMENTS

Many types of interrogation ask for measurements – we might want to know the total area of a parcel of land, or the distance between two points, or the length of a stretch of road – and in principle all of these measurements are obtainable by simple calculations inside a GIS. Comparable measurements by hand from maps can be very tedious and error prone. In fact it was the ability of the computer to make accurate evaluations of area quickly that led the Canadian government to fund the development of the world's first GIS, the Canada Geographic Information System, in the mid-1960s, despite the primitive state and high costs of computing at that time. Evaluation of area by hand is a messy and tedious job. The dot-counting method uses transparent sheets on which randomly located dots have been printed – all area on the map is estimated by counting the number of dots falling within it. In the planimeter method a mechanical device is used to trace the area's boundary, and the required measure accumulates on a dial on the machine.

Box 20: *What is an algorithm?*

> Algorithm is a procedure consisting of a set of unambiguous rules which specify a finite sequence of operations that provides the solution to a problem, or to a specific class of problems. Each step of an algorithm needs to be unambiguous and precisely defined and the actions to be carried out must be rigorously specified for each case. An algorithm always arrives at a problem solution after a finite and reasonable number of steps. An algorithm that satisfies these requirements can be programmed as software for a computer.

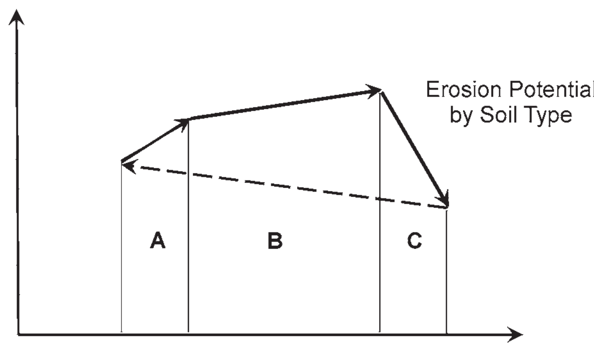


Figure 9.8: The algorithm for calculation of the area of a polygon given the coordinates of the polygon's vertices. The polygon consists of the three arrows and one arrow with dashed line forming the fourth side. Trapezia are dropped from each edge to the x axis and their areas are calculated as (difference in x) times average of y. The trapezia for the first three edges, shown in 'A' 'B' and 'C', are summed. When the fourth trapezia is formed from the dashed arrow its area is negative because its start point has a larger x than its end point. When this area is subtracted from the total, the result is the correct area of the polygon.

By comparison, measurement of the area of a digitally represented polygon is trivial and totally reliable. The common algorithm (Box 20) calculates and sums the areas of a series of trapezia, formed by dropping perpendiculars to the x axis as shown in Figure 9.8. By making a simple change to the algorithm it is also possible to use it to compute a polygon's centroid.

DISTANCE AND LENGTH

A metric is a rule for the determination of distance between points in a space. Several kinds of metrics are used in GIS, depending on the application. The simplest is the rule for determining the shortest distance between two points in a flat plane, called the pythagorean or straight-line metric. If the two points are defined by the coordinates (X1, Y1) and (X2, Y2), then the distance D between them is the length of the hypotenuse of a right-angled triangle (Figure 9.9), and pythagoras's theorem tells us that the square of this length is equal to the sum of the squares of the lengths of the other two sides. So a simple formula results:

$$D = \sqrt{(x_2\ \ x_1)2\ (y_2\ \ y_1)^2}$$

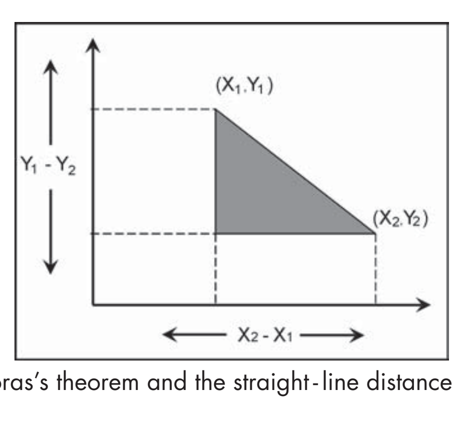


Figure 9.9: Pythagoras's theorem and the straight-line distance between two points.

The Pythagorean metric gives a simple and straightforward solution for a plane, if the coordinates X and Y are comparable, as they are in any coordinate system based on a projection, such as the UTM. But the metric will not work for latitude and longitude, reflecting a common source of problems in GIS – the temptation to treat latitude and longitude as if they were equivalent to plane coordinates.

Distance between two points on a curved surface such as that of the Earth requires a more elaborate approach. The shortest distance between two points is the length of a taut string stretched between them, and if the surface is spherical that is the length of the arc of the great circle between them (the circle formed by slicing the sphere through the center and through the two points).

Given latitude and longitude for two points, the length of this arc is:

$$D = R \cos^{-1} [\sin Ø1 \sin Ø2 + \cos Ø1 \cos Ø2 \cos(\lambda_1 - \lambda_2)]$$

where R is the radius of the Earth (6378 km to the nearest km and assuming a spherical Earth). In some cases it may be necessary to use the ellipsoid model of the Earth, in which case the calculation of distance is more complex.

In many applications the simple rules – the Pythagorean and great circle equations–are not sufficiently accurate estimates of actual travel distance, and we are forced to resort to summing the actual lengths of travel routes. In GIS this normally means summing the lengths of links in a network representation, and many forms of GIS analysis use this approach. If a line is represented as a polyline, or a series of straight segments, then its length is simply the sum of the lengths of each segment, and each segment length can be calculated using the pythagorean formula and the coordinates of its endpoints. But here two problems arise with this simple approach.

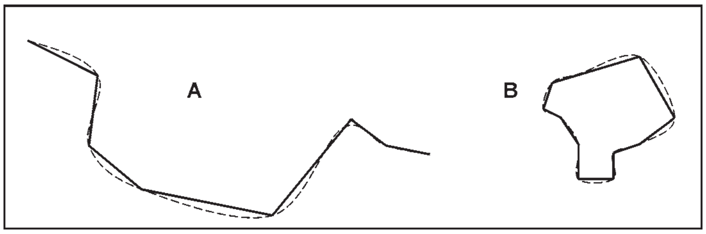


Figure 9.10: (A) – The polyline representations of smooth curves tend to be shorter in length.
(B) – But estimates of area tend not to show systematic bias because of the effects of overshoots and undershoots cancel out.

First, a polyline is often only a rough version of the true object's geometry. A river, for example, never makes sudden changes of direction, and Figure 9.10 shows, how smooth curves have to be approximated by the sharp corners of a polyline. Because there is a tendency for polylines to short-cut corners, the length of a polyline tends to be shorter than the length of the object it represents. There are some exceptions, of course – surveyed boundaries are often truly straight between corner points, and streets are often truly straight between intersections. But in general the lengths of linear objects estimated in a GIS, and this includes the lengths of the perimeters of areas represented as polygons, are often substantially shorter than their counterparts on the ground. Note that this is not similarly true of area estimates, because shortcutting corners tends to produce both underestimates and overestimates of area, and these tend to cancel out (Figure 9.10)

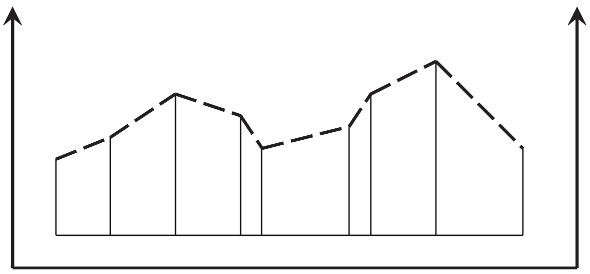


Figure 9.11: The length of a path on earth's surface (dashed line) remains longer than the length of its horizontal projection.

Second, the length of a line in a two-dimensional GIS representation will always be the length of the line's planar projection, not its true length in three dimensions, and the difference can be substantial if the line is steep (Figure 9.11). In most jurisdictions the area of a parcel of land is the area of its horizontal projection, not its true surface area. A GIS that stores the third dimension for every point is able to calculate both versions of length and area, but not a GIS that stores only the two horizontal dimensions.

Shape

GIS are also used to calculate the shapes of objects, particularly area objects. In many countries the system of political representation is based on the concept of constituencies, which are used to define who will vote for each place in the legislature. In the USA and also in India, and in many other countries that derived their system of representation from the UK, there is one place in the legislature for each district. It is expected that districts will be compact in shape, and the manipulation of a district's shape to achieve certain overt or covert objectives is termed Gerrymandering, after an early governor of Massachusetts, Elbridge Gerry (the shape of one of the state's districts was thought to resemble a salamander, with the implication that it had been manipulated to achieve a certain outcome in the voting; The construction of voting districts is an example of the principles of aggregation and zone design.

Anomalous shape is the primary means of detecting gerrymanders of political districts. An easy way to define shape is by comparing the perimeter length of an area to its area measure. Normally the square root of area is used, to ensure that the numerator and denominator are both measured in the same units. A common measure of shape or compactness is:

$$S = P/3.54\sqrt{A}$$

where P is the perimeter length and A is the area.

The factor 3.54 (twice the square root of π) ensures that the most compact shape, a circle, returns a shape of 1.0, and the most distended and contorted shapes return much higher values.

Slope and Aspect

The most versatile and useful representation of terrain in GIS is the digital elevation model, or DEM. This is a raster representation, in which each grid cell records the elevation of the Earth's surface, and reflects a view of terrain as a field of elevation values. The elevation recorded is often the elevation of the cell's central point, but sometimes it is the mean elevation of the cell, and other rules have been used to define the cell's elevation (the rules used to define elevation in each cell of the US Geological Survey's GTOPO30 DEM, which covers the entire Earth's surface, vary depending on the source of data.

Knowing the exact elevation of a point above sea level is important for some applications, including prediction of the effects of global warming and rising sea levels on coastal cities, but for many applications the value of a DEM lies in its ability to produce derivative measures through transformation, specifically measures of slope and aspect, both of which are also conceptualized as fields. Imagine taking a large sheet of plywood and laying it on the Earth's surface so that it touches at the point of interest. The magnitude of steepest tilt of the sheet defines the slope at that point, and the direction of steepest tilt defines the aspect. This sounds straightforward, but it is complicated by a number of issues. First, what if the plywood fails to sit firmly on the surface, but instead pivots, because the point of interest happens to be a peak, or a ridge? In mathematical terms, we say that the surface at this point lacks a well-defined tangent, or that the surface at this point is not differentiable, meaning that it fails to obey the normal rules of continuous mathematical functions and differential calculus. The surface of the Earth has numerous instances of sharp breaks of slope, rocky outcrops, cliffs, canyons, and deep gullies that defy this simple mathematical approach to slope, and this is one of the issues that led Benoit Mandelbrot to develop his theory of fractals, or mathematical functions that display behaviours of this nature. Mandelbrot argues in his books (Mandelbrot 1977, 1983) that many natural phenomena are fundamentally incompatible with traditional mathematics, and need a different approach.

A simple and satisfactory alternative is to take the view that slope must be measured at a particular resolution. To measure slope at a 30 meters resolution, for example, we evaluate elevation at points 30 meters apart and compute slope by comparing them. The value this gives is specific to the 30 meters spacing, and a different spacing would have given a different result. In other words, slope is a function of resolution, and it makes no sense to talk about slope without at the same time talking about a specific resolution or level of detail. This is convenient, because slope is easily computed in this way from a DEM with the appropriate resolution.

Second, there are several alternative measures of slope, and it is important to know which one is used in a particular software package and application. Slope can be measured as an angle, varying from 0 to 90 degrees as the surface ranges from horizontal to vertical. But it can also be measured as a percentage or ratio, defined as rise over run, and unfortunately there are two different ways of defining run. Figure 9.12 shows the two options, depending on whether run means the horizontal distance covered between two points, or the diagonal distance (the adjacent or the hypotenuse of the right-angled triangle respectively). In the first case (opposite over adjacent) slope as a ratio is equal to the tangent of the angle of slope, and ranges from zero (horizontal) through 1 (45 degrees) to infinity (vertical). In the second case (opposite over hypotenuse) slope as a ratio is equal to the sine of the angle of slope, and ranges from zero (horizontal) through 0.707 (45 degrees) to 1 (vertical). To avoid confusion we will use the term slope only to refer to the measurement in degrees, and call the other options tan(slope) and sin(slope) respectively.

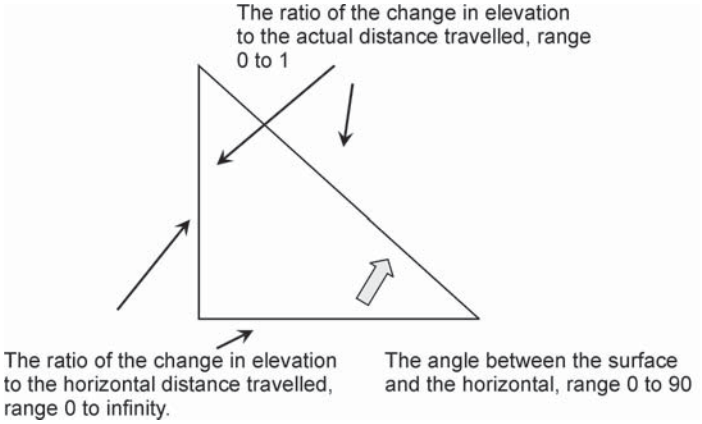


Figure 9.12: Three alternative definitions of slope.

When a GIS calculates slope and aspect from a DEM, it does so by estimating slope at each of the data points of the DEM, by comparing the elevation at that point to the elevations of surrounding points. But the number of surrounding points used in the calculation varies, as do the weights given to each of the surrounding points in the calculation.

Slope and aspect are the basis for many interesting and useful forms of analysis. Slope is an input to many models of the soil erosion and runoff that result from heavy storms. Slope is also an important input to analyses that find the most suitable routes across terrain for power lines, highways etc.

TRANSFORMATIONS

In this section, we look at methods that transform GIS objects and databases into more useful products, using simple rules. These operations form the basis for many applications, because they are capable of revealing aspects that are not immediately visible or obvious.

BUFFERING

One of the most important transformations available to the GIS user is the buffer operation. Given any set of objects, which may include points, lines, or areas, a buffer operation builds a new object or objects by identifying all areas that are within a certain specified distance of the original objects. Figure 9.13 shows instances of a point, a line, and an area, and the results of buffering. Buffers have many uses, and they are among the most popular of GIS functions:

- A builder wishes to develop a residential colony, but is concerned of flooding in rainy season. He is required to avoid construction within l00 meters of streams – the builder could build buffers 100 meters wide around all streams to identify these flooding areas.

- A retailer is considering developing a new store on a site, of a type that is able to draw consumers from up to 4 km away from its stores the retailer could build a buffer around the site to identify the number of consumers living within 4 km of the site, in order to estimate the new store's potential sales.

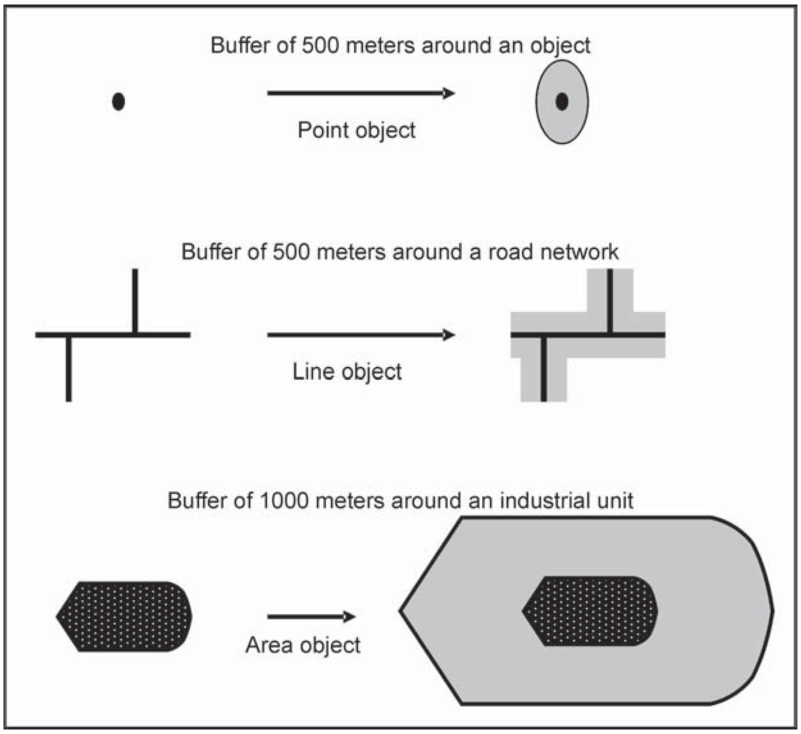


Figure 9.13: Buffers of constant width drawn around a point, line and a polygon.

Buffering is possible in both raster and vector GIS, in the raster case, the result is the classification of cells according to whether they lie inside or outside the buffer, while the result in the vector case is a new set of objects. But there is an additional possibility in the raster case that makes buffering more useful in some situations. Figure 9.15 shows a city; average travel speeds vary in each cell of the raster outside the city. Rather than buffer according to distance from the city, we can ask a raster GIS to spread outwards from the city at rates determined by the travel speed values in each cell. Where travel speeds are high the spread will extend further, so we can compute how far it is possible to go from the city in a given period of time. This idea of spreading over a variable surface is easily implemented in raster representations, but impossible in vector representations.

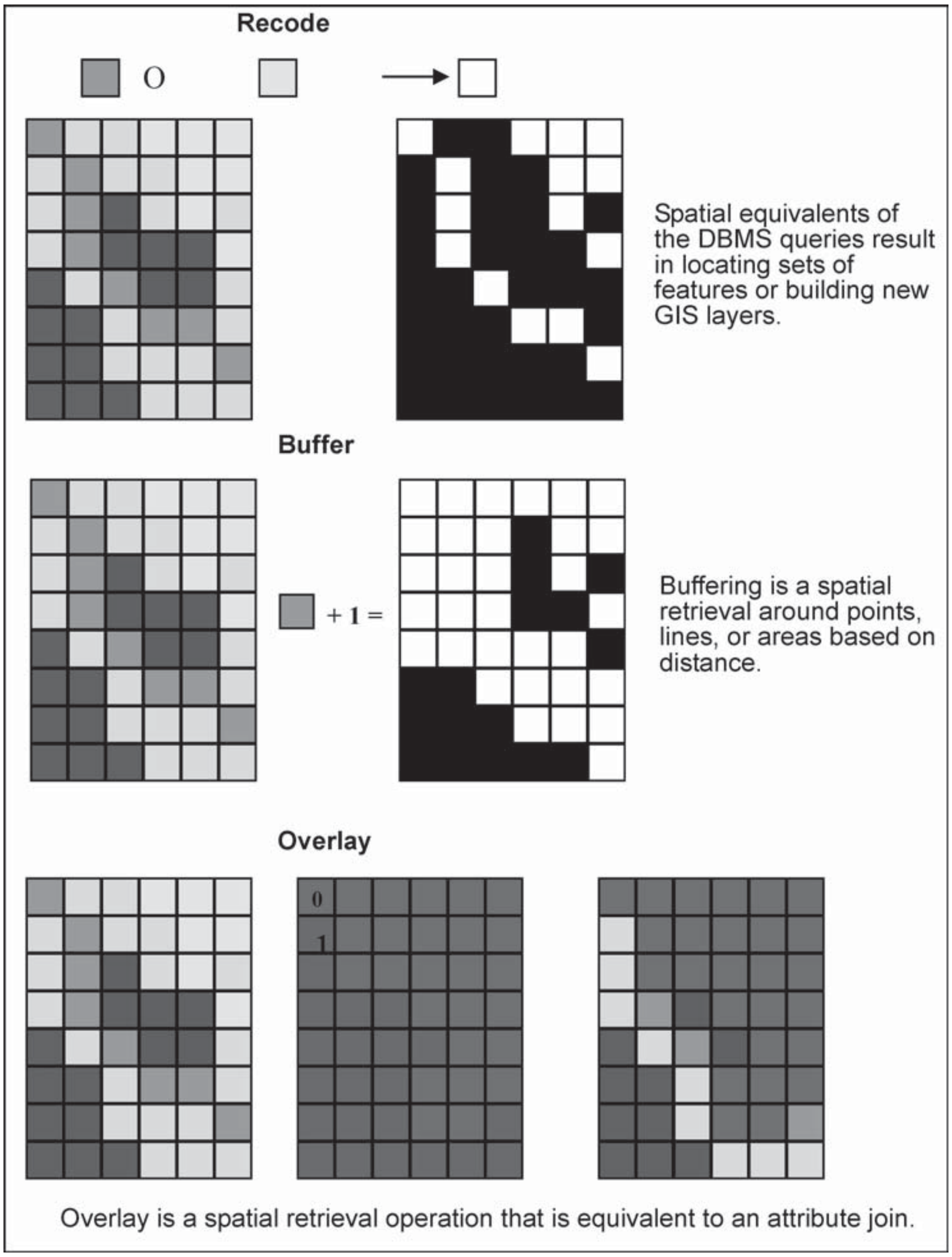


Figure 9.14: Spatial retrieval operations.

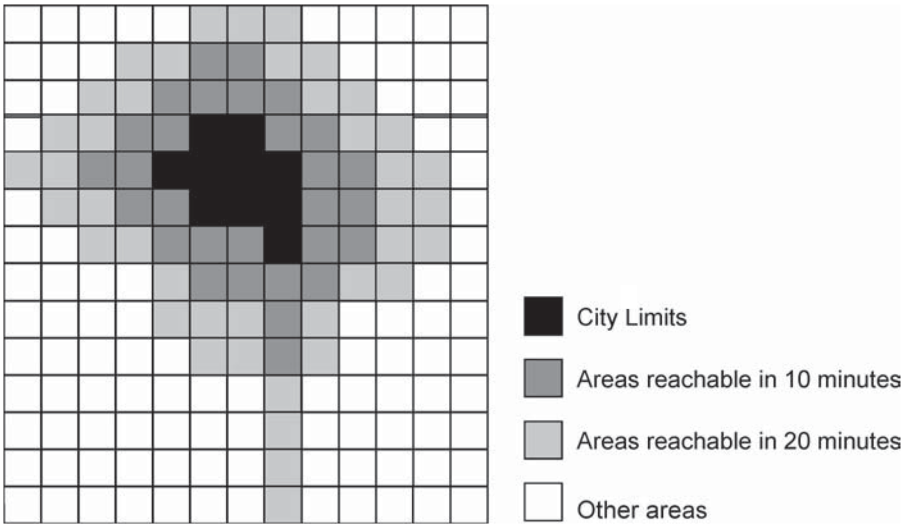


Figure 9.15: A raster generalization of the buffer function where changes may be controlled by some variable (example is of travel speed, whose value is recorded in every raster cell)

POINT IN POLYGON

In its simplest form, the point in polygon operation determines whether a given point lies inside or outside a given polygon. In more elaborate forms there may be many polygons, and many points, and the task is to assign points to polygons. If the polygons overlap, it is possible that a given point lies in one, many, or no polygons, depending on its location. Figure 9.17 illustrates the task. The operation is popular in GIS analysis because it is the basis for answering many simple queries:

- The points represent instances of a disease in a population, and the polygons represent reporting zones such as wards-the task is to determine how many instances of the disease occurred in each ward (in this case the ward should not overlap and each point should fall into exactly one polygon).
- The points represent the locations of a tube-well owned by a person, and the polygons are parcels of land-the task is to determine the owner of the land where tube-well lies has necessary permission and owner of the land which is irrigated by tube-well has paid the necessary fees.
- The points represent the residential locations of an industry, and the polygons represent the entire settlement-the task is to ensure that each worker of the industry receives the invitation for a function by mail.

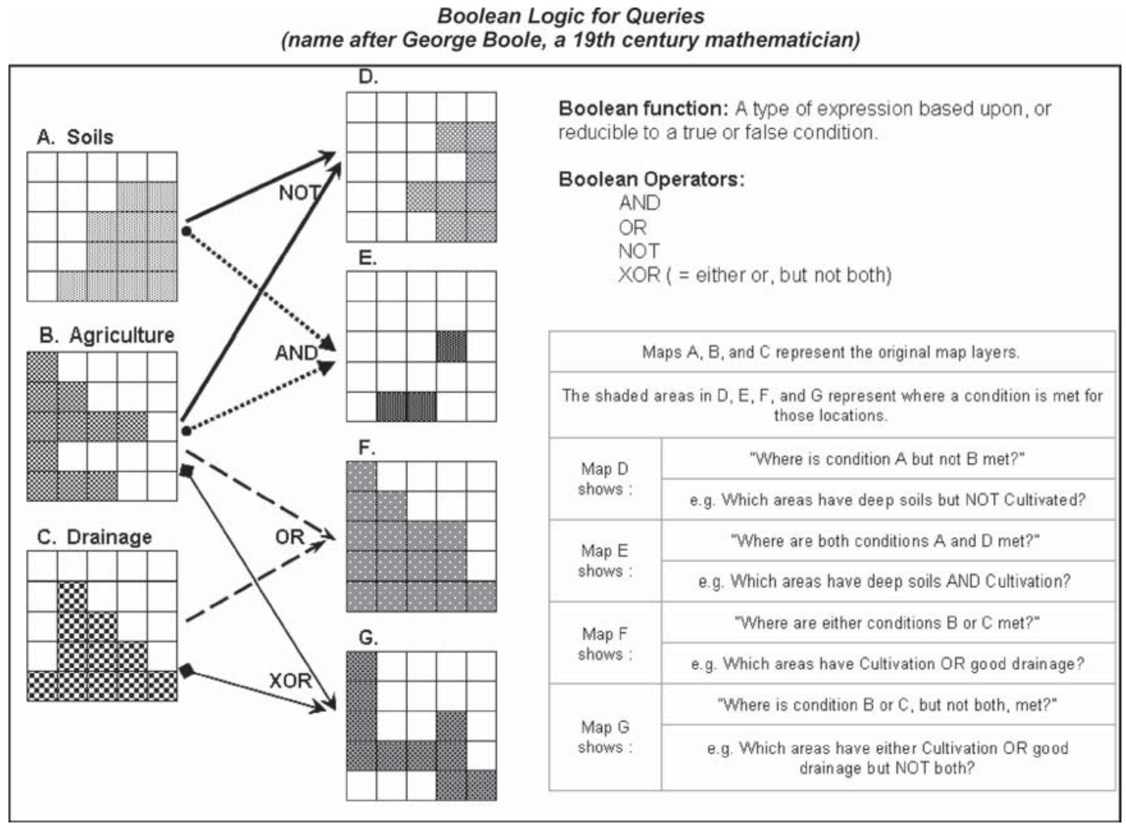


Figure 9.16: Examples of Boolean logic using Boolean operators.

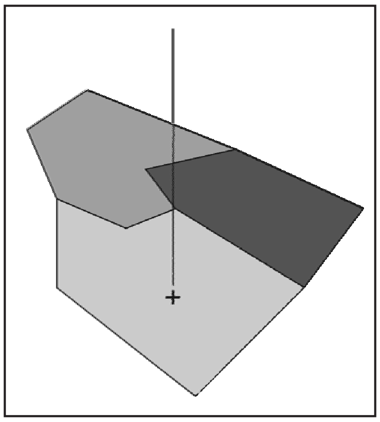


Figure 9.17: The point in polygon problem.

The point in polygon operation makes sense from both the discrete object and the field perspectives. From a discrete object perspective both points and polygons are objects, and the task is simply to determine enclosure. From a field perspective, polygons representing a variable such as land ownership cannot overlap, since each polygon represents the land owned by one owner, and overlap would imply that a point is owned simultaneously by two owners. Similarly from a field perspective there can be no gaps between polygons. Consequently, the result of a point in polygon operation from a field perspective must assign each point to exactly one polygon.

The standard algorithm for the point in polygon operation is shown in Figure 9.14. In essence, it consists of drawing a line vertically upwards from the point, and determining the number of intersections between the line and the polygon's boundary. If the number is odd the point is inside the polygon, and if it is even the point is outside. The algorithm must deal successfully with special cases, for example, if the point lies directly below a corner point of the polygon. Some algorithms extend the task to include a third option, when the point lies exactly on the boundary. But others ignore this, on the grounds that it is never possible to determine location with perfect accuracy, and so never possible to determine if an infinitely small point lies on an infinitely thin boundary line.

Polygon Overlay

Polygon overlay is similar to point in polygon transformation in the sense that two sets of objects are involved, but in this case both are polygons. It exists in two forms, depending on whether a field or discrete object perspective is taken. From the discrete object perspective, the task is to determine whether two area objects overlap, to determine the area of overlap, and to define the area formed by the overlap as one or more new area objects (the overlay of two polygons can produce a large number of distinct area objects, see Figure 9.18). This operation is useful to determine answers to such queries as:

- How much area lies in the shaded zone?
- How much of this land parcel is shaded but not the white polygon?
- What proportion of the land area outside the shaded but inside the white polygon?

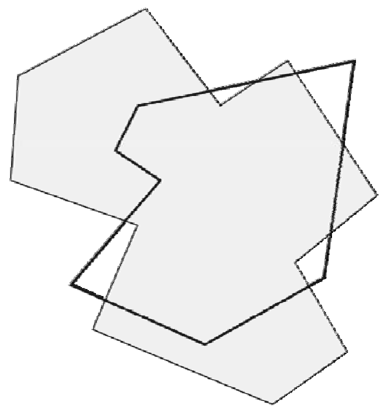


Figure 9.18: An example of polygon overlay, in the discrete object case. Here the overlay of two polygons produces nine polygons. One has the property of both, four have the properties of shaded but not the white polygon and four are outside the shaded but inside the white polygon.

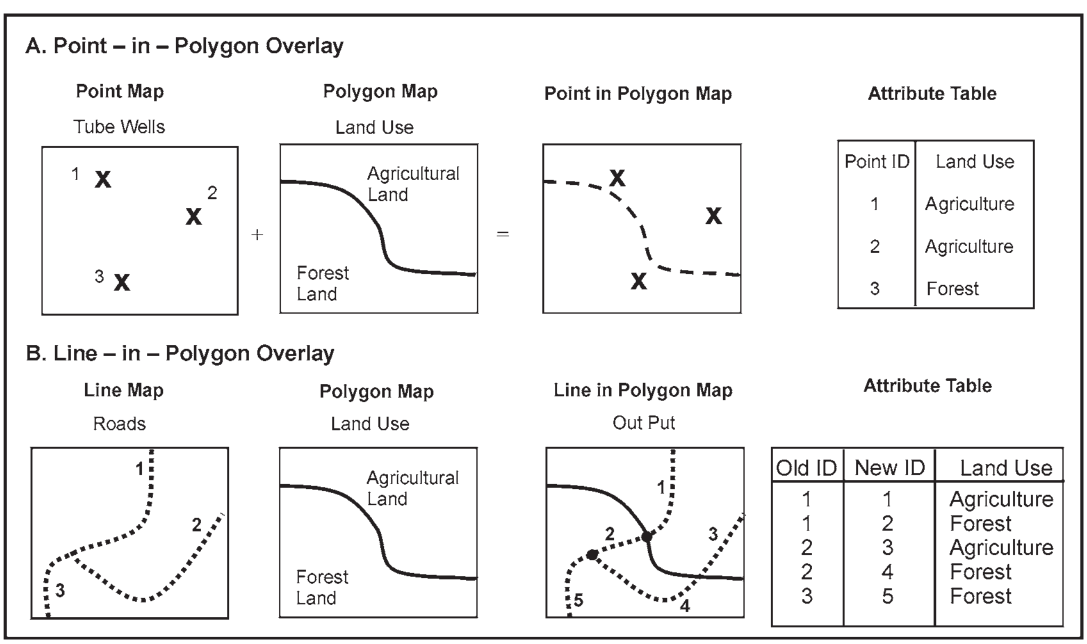


Figure 9.19: Vector overlays (point in polygon and line in polygon).

C. Polygon – on – Polygon Overlay

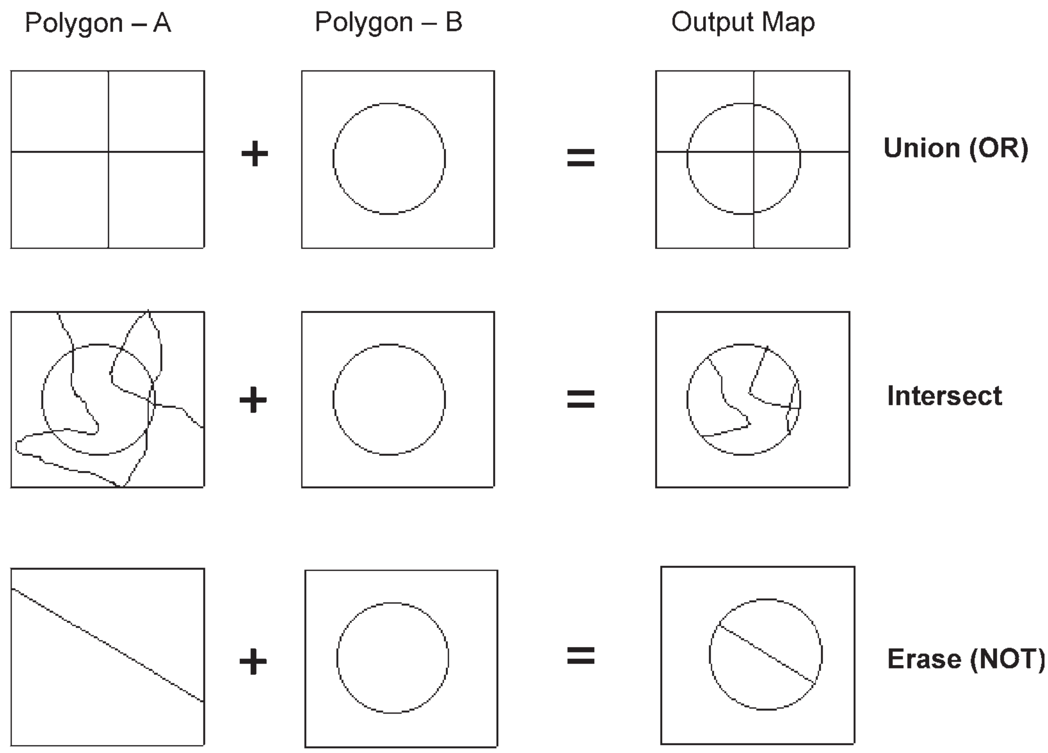


Figure 9.20: Vector overlays (polygon on polygon)

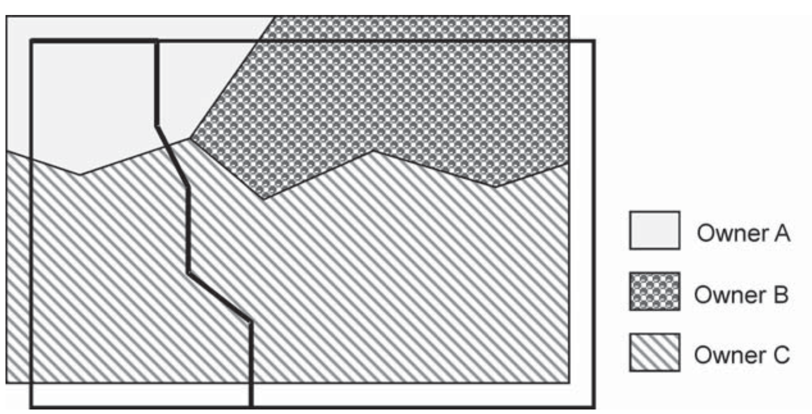


Figure 9.21: Polygon overlay in the field case. Where a dataset representing two types of land cover. (one on the left, say X and another in right, y). It is overlaid on a dataset showing three land parcels owned by three different persons. The overlay result will be a single dataset in which every point is identified with one land cover type and one ownership type. There will be five polygons, as land over X intersects two ownership types and land cover y intersects with three.

From the field perspective the task is somewhat different. Figure 9.21 shows two datasets, both representations of fields–one differentiates areas according to land ownership, and the other differentiates the same region according to land cover class. In the terminology of ESRI's Arc Info, both datasets are instances of area coverages, or fields of nominal variables represented by non-overlapping polygons. The methods discussed earlier in this chapter could be used to interrogate either dataset separately, but there are numerous queries that require simultaneous access to both datasets, for example:

⇒ What is the total area of land owned by A and with land cover class X?
⇒ Where are the areas that is owned by C and have land cover class Y?
⇒ What is the land cover class and who is the owner of the point indicated by the user?

None of these queries can be answered by interrogating one of the datasets alone the data sets must somehow be combined so that interrogation can be directed simultaneously at both of them. The field version of polygon overlay does this by first computing a new dataset in which the region is partitioned into smaller areas that have uniform characteristics on both field variables. Each area in the new dataset will have two sets of attributes – those obtained from one of the input datasets, together with those obtained from the other. All of the boundaries will be retained, but they will be broken into shorter fragments by the intersections that occur between boundaries in one input data set and boundaries in the other. Unlike the two input datasets, where boundaries meet in a junction of three lines, the new map contains a new junction of four lines, formed by the new intersection discovered during the overlay process. Because the results of overlay are distinct in this way it is almost always possible to discover whether a GIS dataset was formed by overlaying two earlier datasets.

With a single dataset that combines both inputs, it is an easy matter to answer all of the queries listed above through simple interrogation. It is also easy to reverse the overlay process-if neighbouring areas that share the same land cover class are merged. Polygon overlay is a computationally complex operation, and much work has gone into developing algorithms that function efficiently for large datasets. One of the issues that must be tackled by a practically useful algorithm is known as the spurious polygon or coastline weave problem. It is almost inevitable that there will be instances in any practical application where the same line on the ground occurs in both datasets.

Rivers and roads often form boundaries in many different datasets – a river may function both as a land cover class boundary and as a land ownership boundary, for example. But although the same line is represented in both datasets, its representations will almost certainly not be the same- They may have been digitized from different maps, subjected to different manipulations, obtained from entirely different sources (an air photograph and a topographic map), and subjected to different measurement errors. When overlaid, the result is a series of small slivers. Paradoxically, the more care one takes in digitizing or processing, the worse the problem becomes, as the result is simply more slivers, albeit smaller in size.

Today, a GIS offers various methods for dealing with the problem, the most common of which is the specification of a tolerance. If two lines fall within this distance of each other, the GIS will treat them as a single line, and not create slivers. The resulting overlay contains just one version of the line, not two. But at least one of the input lines has been moved, and if the tolerance is set too high the movement can be substantial, and can lead to problems later.

Overlay in raster is an altogether simpler operation, and this has often been cited as a good reason to adopt raster rather than vector structures. When two raster layers are overlaid, the attributes of each cell are combined according to a set of rules. For example, suppose the task is to find all areas that belong to owner A and have land use class X. Areas with these characteristics would be assigned a value, say 1, and all other areas would be assigned a value of 0. The important difference between raster and vector overlay in vector overlay there is no rule for combination, and instead the result of overlay contains all of the input information, rearranged and combined so that it can be used to respond to queries and can be subjected to analysis.

Spatial Interpolation

Spatial interpolation is a pervasive operation in GIS. Although it is often used explicitly in analysis, it is also used implicitly, in various operations such as the preparation of a contour map display, where spatial interpolation is invoked without the user's direct involvement. Spatial interpolation is a process of intelligent guesswork, in which the investigator attempts to make a reasonable estimate of the value of a field at places where the field has not actually been measured. Spatial interpolation is an operation that makes sense only from the field perspective. Spatial interpolation finds applications in many areas:

- In contouring, when it is necessary to guess where to place contours in between measured locations.
- In estimating the elevation of the surface in between the measured locations of a DEM.
- In estimating rainfall, temperature, and other attributes at places that are not weather stations, and where no direct measurements of these variables are available.
- In resampling rasters, the operation that must take place whenever raster data must be transformed to another grid.

In all of these instances spatial interpolation calls for intelligent guesswork, and the one principle that underlies all spatial interpolation is the Tobler Law-'all places are related but nearby places are more related than distant places'. In other words, the best guess as to the value of a field at some point is the value measured at the closest observation points - the rainfall here is likely to be more similar to the rainfall recorded at the nearest weather stations than to the rainfall recorded at more distant weather stations. A corollary of this same principle is that in the absence of better information, it is reasonable to assume that any field exhibits relatively smooth variation-fields tend to vary slowly, and to exhibit strong positive spatial autocorrelation, a property of geographic data.

Here we discuss two commonly used methods of spatial interpolation: inverse distance weighting (IDW), which is the simplest method; and Kriging, a popular statistical method that is grounded in the theory of regionalized variables and falls within the field of geostatistics.

INVERSE DISTANCE WEIGHTING (IDW): IDW is the workhorse of spatial interpolation, the method that is most often used by GIS analysts. It employs the Tobler law by estimating unknown measurements as weighted averages over the known measurements at nearby points, giving the greatest weight to the nearest points. IDW provides a simple way of guessing the values of a field at locations where no measurement is available.

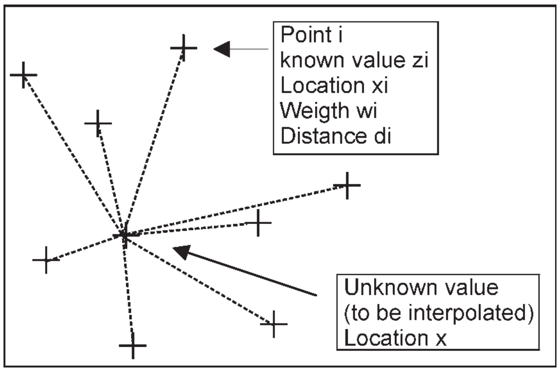


Figure 9.22: Notation used in the equations defining spatial interpolation.

IDW achieves the desired objective of creating a smooth surface whose value at any point is more like the values at nearby points than the values at distant points. If it is used to determine z at a location where z has already been measured it will return the measured value, because the weight assigned to a point at zero distance is infinite, and for this reason IDW is described as an exact method of interpolation because its interpolated results honour the data points exactly (an approximate method is allowed to deviate from the measured values in the interests of greater smoothness, a property which is often useful if deviations are interpreted as indicating possible errors of measurement, or local deviations that are to be separated from the general trend of the surface.

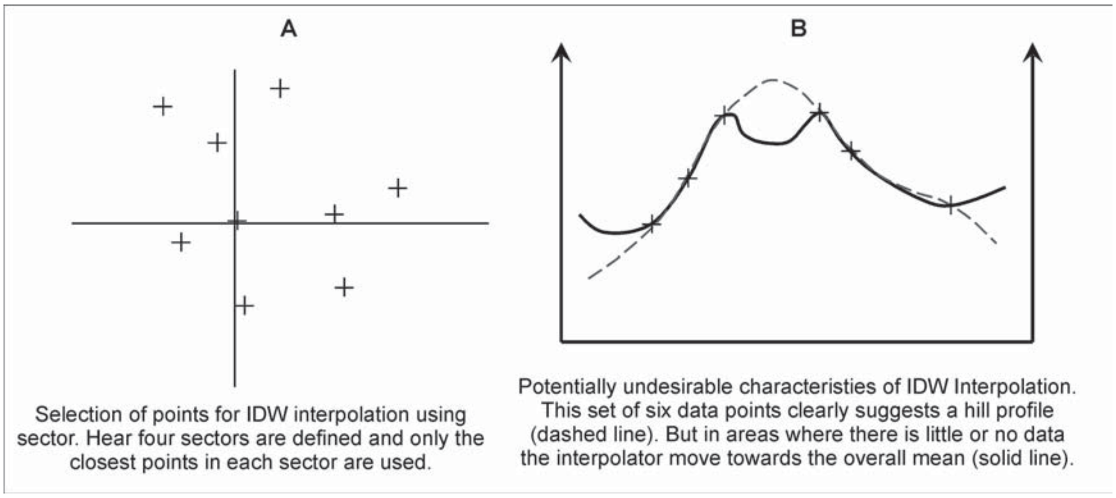


Figure 9.23: IDW interpolation.

But because IDW is an average it suffers from certain specific characteristics that are generally undesirable. A weighted average that uses weights that are never negative must always return a value that is between the limits of the measured values-no point on the interpolated surface can have an interpolated z that is more than the largest measured z, or less than the smallest measured z. IDW interpolation may produce counterintuitive results in the areas of peaks and pits, and outside the area covered by the data points.

In short, the results of IDW are not always what one would want. There are many better methods of spatial interpolation that address the problems that were just identified, but the ease of programming of IDW and its conceptual simplicity make it among the most popular.

Kriging: Of all of the common methods of spatial interpolation it is Kriging that makes the most convincing claim to be grounded in good theoretical principles. The basic idea is to discover something about the general properties of the surface, as revealed by the measured values, and then to apply these properties in estimating the missing parts of the surface.

Smoothness is the most important property, and it is operationalized in Kriging in a statistically meaningful way. There are many forms of Kriging, but all are firmly grounded in theory. Suppose we take a point x as a reference, and start comparing the values of the field there with the values at other locations at increasing distances from the reference point. If the field is smooth (if the Tobler law is true, that is, if there is positive spatial autocorrelation) the values nearby will not – very different-z(x) will not be very different from z(xi). To measure the amount, we take the difference and square it, since the sign of the difference is not important:

$$(z(x) - z(xi))^2$$

We could do this with any pair of points in the area.

As distance increases, this measure will likely increase also, and in general a monotonic (consistent) increase in squared difference with distance is observed for most geographic fields (z must be measured on a scale that is at least interval, though indicator Kriging has been developed to deal with the analysis of nominal fields). In Figure 9.24, each point represents one pair of values drawn from the total set of data points at which measurements have been taken.

The vertical axis represents one half of the squared difference (one half is taken for mathematical reasons), and the graph is known as the semivariogram (or variogram for short the difference of a factor of two is often overlooked in practice, though it is important mathematically). To express its contents in summary form the distance axis is divided into a number of ranges or buckets, as shown, and points within each range are averaged to define the heavy points shown in the figure. This semivariogram has been drawn without regard to the directions between points in a pair. Kriging responds both to the proximity of sample points and to their directions.

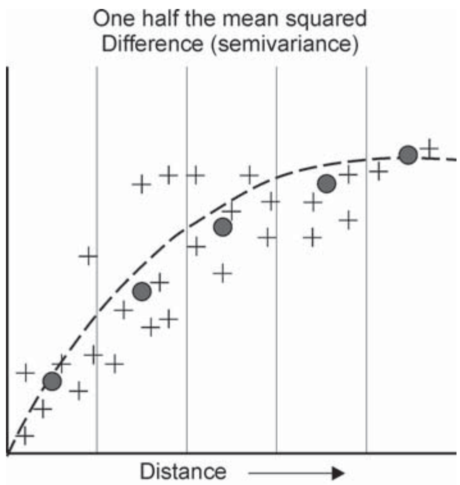


Figure 9.24: A semivariogram, here each cross represents a pair of points. The solid circles are obtained by averaging within the ranges of the distance axis. The dashed line is the best fit to the five points.

Density Estimation and Potential

Density estimation is in many ways the logical twin of spatial interpolation – it begins with points, and ends with a surface. But conceptually the two approaches could not be more different, because one seeks to estimate the missing parts of a field from samples of the field taken at data points, while the other creates a field from discrete objects.

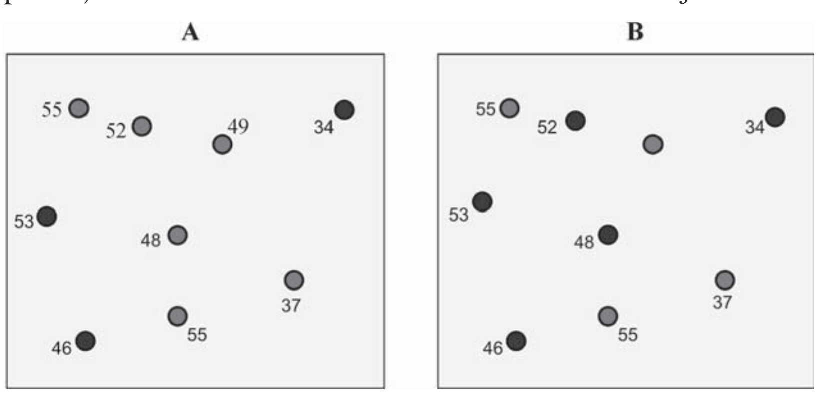


Figure 9.25: Two identical datasets but with different representations.
A – represents a field of atmospheric temperature recorded at nine sample points.
B – nine discrete objects representing population of different settlements in thousands.
Spatial interpolation is appropriate for case A, while for case B density estimation is suitable.

Figure 9.25 illustrates this difference. The two datasets in the diagram look identical from a GIS perspective – they are both sets of points, with locations and a single attribute. But one shows sample measurements from a field, and the other shows the locations of discrete objects. In the discrete object view there is nothing between the objects but empty space – no missing field to be filled in through spatial interpolation. It would make no sense at all to apply spatial interpolation to a collection of discrete objects – and no sense at all to apply density estimation to samples of a field. Density estimation makes sense only from the discrete object perspective, and spatial interpolation only from the field perspective.

Density estimation could be applied to any type of discrete spatial object, it is most often applied to the estimation of point density, and that is the focus here. The most obvious example is the estimation of population density, and but it could be equally well applied to the density of different kinds of diseases, or animals, or any other set of well-defined points.

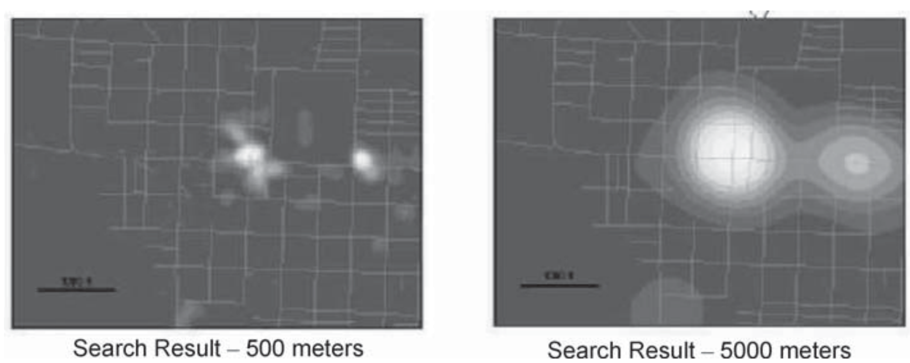


Figure 9.26: Density estimation using two different distance parameters in the respective kernel functions, displaying smoother and less peaked nature of the surface that results from the larger distance parameter.

Advanced Spatial Analysis

There are also some complex spatial analysis in GIS, which uses advanced conceptual frameworks. These spatial analysis are the outcome of advancement in technology. The advent and easy availability of large datasets and fast computing led new ways of thinking about spatial analysis. Now loads of datasets collected and archived everyday like continuous imaging of every corner of the earth or socio-economic information of population for every settlement or even the use of credit card all over the world. All this leads to thinking of interesting patterns, anomalies, truths – myths and many of these are captured in through ***data mining.*** Data mining is used to detect anomalies and patterns in vast archives of digital data. The objective of it is to find patterns that stand out from the normal in an area.

Descriptive Summaries

CENTERS: To analyze the numerical summaries generally we measure by methods of central tendency. Like **mean** is one method which is broadly citing the average of data series, similarly **median,** where the value is as such that one half of the numbers are larger and one half are smaller. Although mean can be computed only for numbers measured on interval of ratio scales, the median can be computed for ordinal data. For nominal data appropriate measure of central tendency is the **mode.**

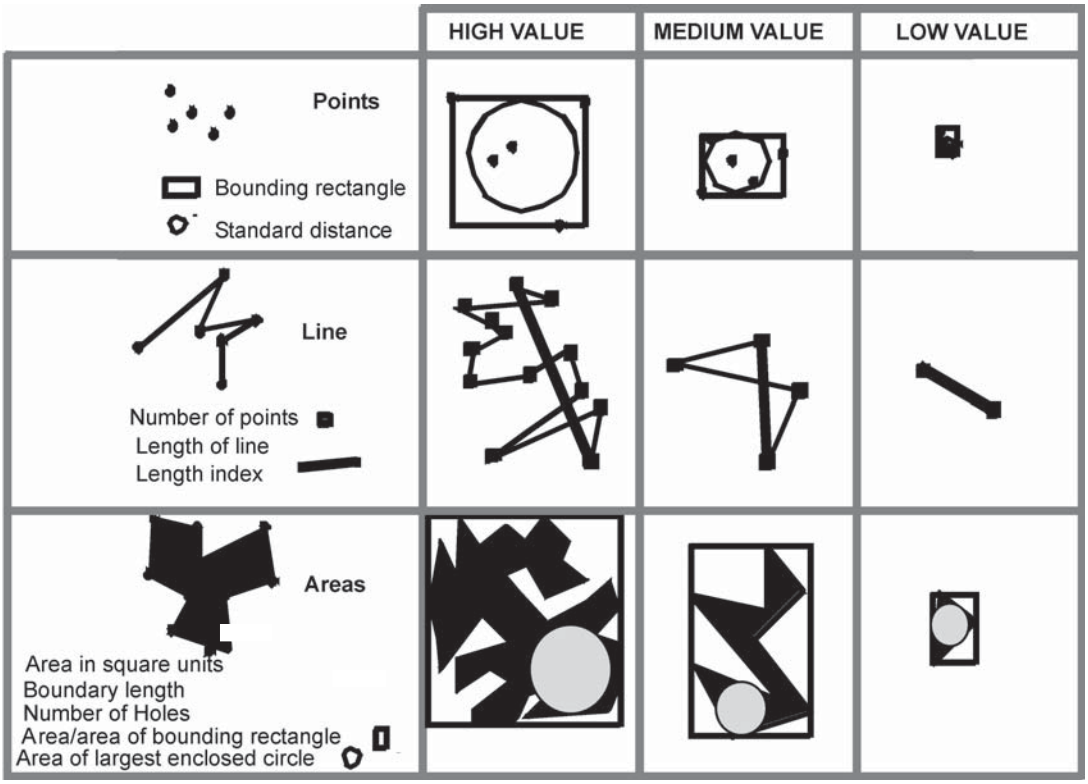


Figure 9.27: Statistics and features.

The spatial equivalent of the mean would be some kind of center, which is calculated to summarize the positions of a number of points in GIS. The center is the most convenient way of summarizing the locations of a set of points.

DISPERSION: Central tendency is the obvious choice if a set of numbers are to be summarized in a single value, but where there is opportunity for a second summary value, the measure of choice for numbers with interval or ratio properties is **standard deviation** or the **variance** is often used, which is the square of the standard deviation (the mean squared difference from the mean). But it is not convenient to measure for descriptive purpose. Standard deviation and variance are more appropriate measures of dispersion than the range because as averages they are less sensitive to the specific values of the extremes. Measures of dispersion are applied in many areas of GIS. A simple measure of dispersion in two dimensions is the mean distance from the centroid.

HISTOGRAMS AND PIE CHARTS: Histograms (bar graphs) and pie charts are two of many ways of visualizing the content of a geographic database. A histogram shows the relative frequencies of different value of an attribute by ordering them on the X axis and displaying frequency through the length of a bar parallel to the Y axis. Attributes should have interval or ratio properties, although ordinal properties are sufficient to allow the values to be ranked and histogram based on ordinal data is useful representation. A pie chart is useful for nominal data and is used to display the relative frequencies of distinct values, with no necessity for ranking. Pie charts are also useful in dealing with attributes measured on cyclic scales. Both take a single attribute and organize its values in a form that allow quick comprehension.

SCATTERPLOTS: We looked at descriptive summaries of single set of objects, further the power of GIS lies in its ability to compare sets of attributes – often thought of as the process of overlaying layers. Where we tend to explore vertical relationships (in GIS vertical refers to comparison of attributes) rather than horizontal ones. Scatterplots are useful visual summaries of relationships between attributes. It display the value of one attribute plotted against the other. If both sets of attributes belong to the same objects then the construction of a scatterplot is straight. Further if both are attributes of raster datasets, then scatterplot is built by comparing the datasets pixel by pixel. But if the attributes are from different sets of vector objects, which do not coincide in space then it is sorted by interpolating the datasets and inventing a geographic data.

SPATIAL DEPENDENCE: The fundamental problem of spatial analysis is selecting appropriate digital representations from the real world. The Tobler's first law of geography states that everything is related to everything else but near things are more related than distant things. The real world without spatial dependence is impossible to imagine. Thus, spatial dependence is crucial for GIS. It is inherently scale specific and can be measured at any spatial resolution. However, a dataset can exhibit positive spatial dependence at one scale but negative at another scale. Spatial dependence is a very useful descriptive summary of geographic data and a fundamental part of its nature. The **semivariogram** of a raster dataset elaborates how difference increases with distance and whether difference ceases to increase beyond a certain range. The computation of semivariogram in different directions, we can also determine whether a dataset displays marked anisotropy or distinct behaviours.

FRAGMENTATION AND FRACTIONAL DIMENSION: In GIS, maps may show many patches with each patch representing an area of uniform class and this may be bounded by patch of different class. For example, a soil map where we may be interested in the degree to which the landscape can be fragmented (meaning breaking in small or large patches). Fragmentation statistics provide the numerical basis for this purpose. Here we can analyze the number of patches, their shape or size etc, as a way of summarizing the geographic details. The concept of fractals is used as a way f summarizing the relationship between apparent length and level of geographic detail in the form of fractional dimensions. Smooth lines would indicate fractional dimension close to one while contorted lines would indicate towards higher values.

Optimization

Optimization is a prime example of GIS utility to support spatial decisions. It can be by many ways like optimum location of points, routing on a network, selection of optimum paths across continuous space, locating facilities etc. The methods also divide between those that are designed to locate points and routes on network and those designed to locate points and routes in continuous space without respect to the existence of roads or other links.

POINT LOCATIONS: It is an instance of location in continuous space and identifying location that minimizes total distance with respect to a number of points. The analogous problem on a network would involve finding that location on the network which minimize total distance to a number of points, also located on the network, using routes that are considered on the network using routes that are constrained to the network. Location allocation involves two types of decisions – where to locate and how to allocate demand for service.

ROUTING PROBLEMS: This is another area of optimization where routing and scheduling or decisions about he optimum tracks are considered. At the root of all routing problem is the shortest path, the path through the network between a defined origin and destination that minimizes distance or travel time. Attributes such as length, travel speed, restrictions on travel direction and level of congestion are taken into account. A GIS can be very effective at solving routing problems because it is able to examine vast numbers of possible solutions quickly.

OPTIMUM PATHS: Here the concern is for finding optimum path across continuous space for linear facilities like highways, pipelines or even airline path etc. Again emphasis would be on shortest route may be to save fuel, time or avoiding the restrictions if there are any. These are normally sorted in raster, where each cell may be assigned a friction value, equal to the cost or time associated with moving across the cell in the horizontal or vertical directions.

Hypothesis Testing

Another kind of complex spatial analysis deals with the testing of hypotheses and drawing of inference and its relationship to GIS. It is about methods of inference drawn from information about a sample to a more general information for a larger population. Hypothesis testing is based on two concepts – confidence limits and inferential tests, which are basically statistical testing. The focus here is on the issue of using these approaches with geographic data in a GIS context.

HYPOTHESIS TESTS ON GEOGRAPHIC DATA: Although inferential tests are standard practice in much of science, they are very problematic for geographic data. In GIS, we analyze all the data that are there in a given area rather than sample. The example can be of sampling topographic elevation. The ability to estimate is the base of spatial interpolation. So here on one side banking on spatial interpolation, we can not believe in independence of geographic samples (basic assumption of statistical tests). Another important issue in this context is about the earth's surface which is heterogeneous, making it difficult to take samples that are truly representative for any large region. So what an investigator do, when inferential tests on geographical data are unacceptable, certainly investigator cannot discard spatial data. Here rather investigator may abandon inferential approach. The results obtained from the data are descriptive of the study area but it need not to be generalized. This approach, using local statistics observes the differences in the results of analysis over space. It represents a compromise between nomothetic and idiographic positions. Generalization is very tempting but the heterogeneous nature of the earth's surface makes it difficult. If generalization is necessary, then it can be accomplished by appropriate experimental design, replicating the study in a sufficient number of distinct areas to ensure confidence.

This page intentionally left blank