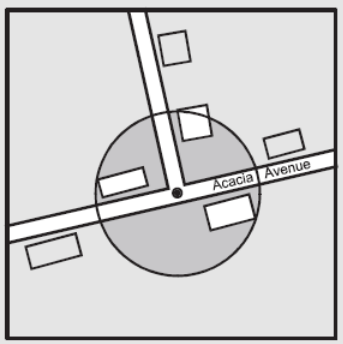
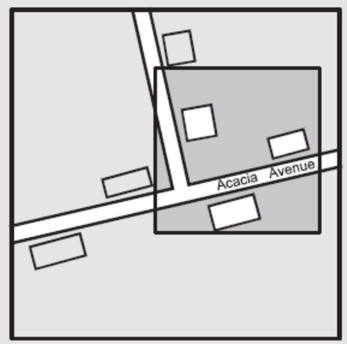
Chapter 9

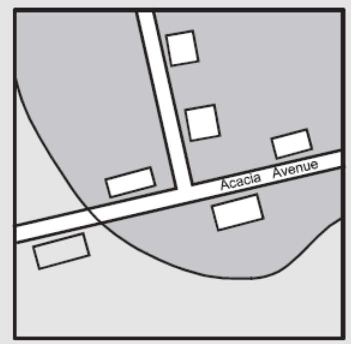
Geographic Query and Analysis



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.
  - *Spatial analysis is the crux of GIS. Spatial analysts can reveal things that might otherwise be invisible – it can make what is implicit explicit.*
- We look at spatial analysis grouped into six distinct categories – queries and reasoning, measurements, transformations, descriptive summaries, optimization, and hypothesis testing.

- 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.
  - 'Spatial analysis is a set of methods whose results change when the locations of the objects being analyzed change'.
- 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. (Figure 9.1 and Box 18).
  - 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

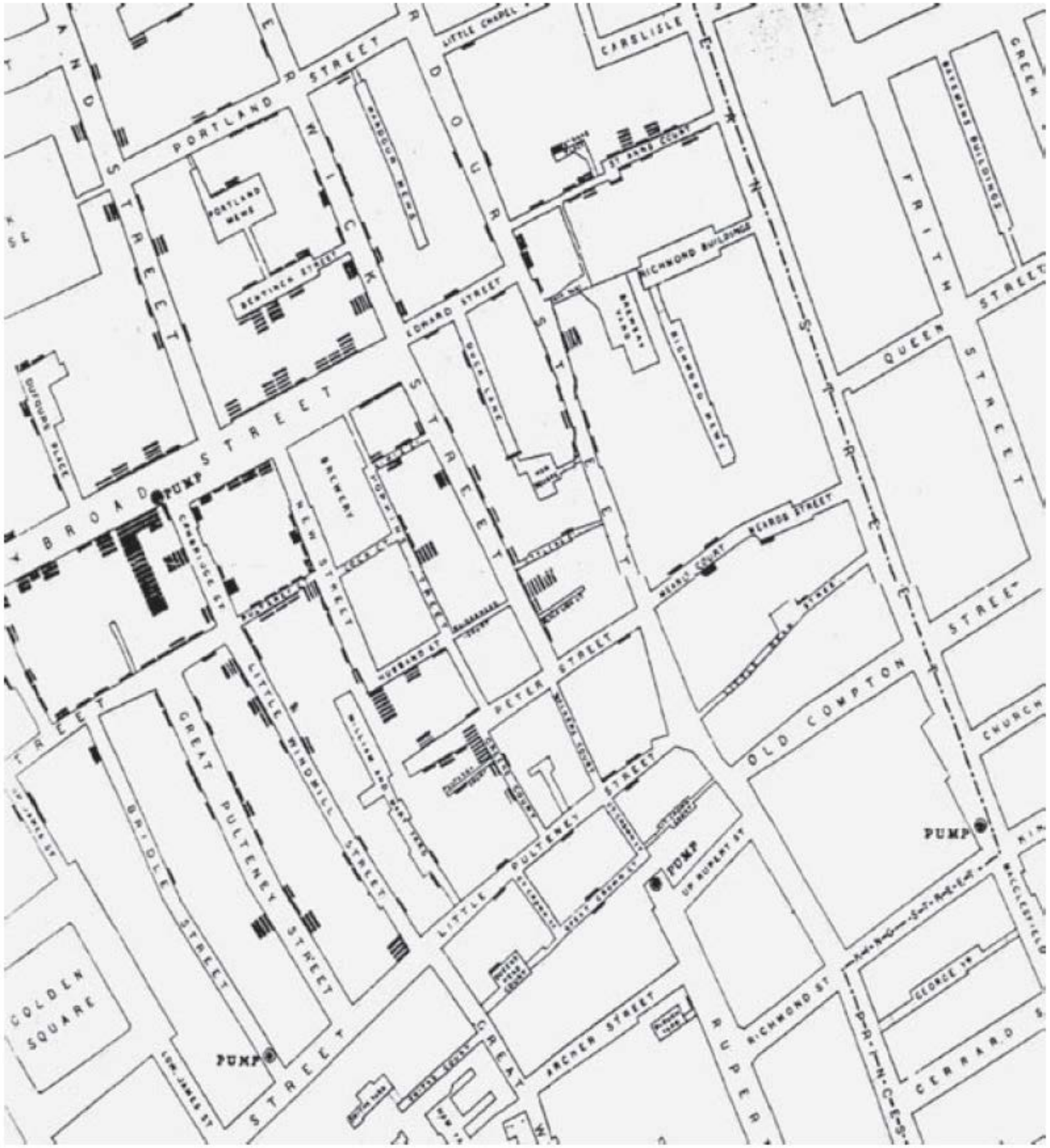


Figure 9.1: Dr. John Snow and the causes of cholera in London.

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.

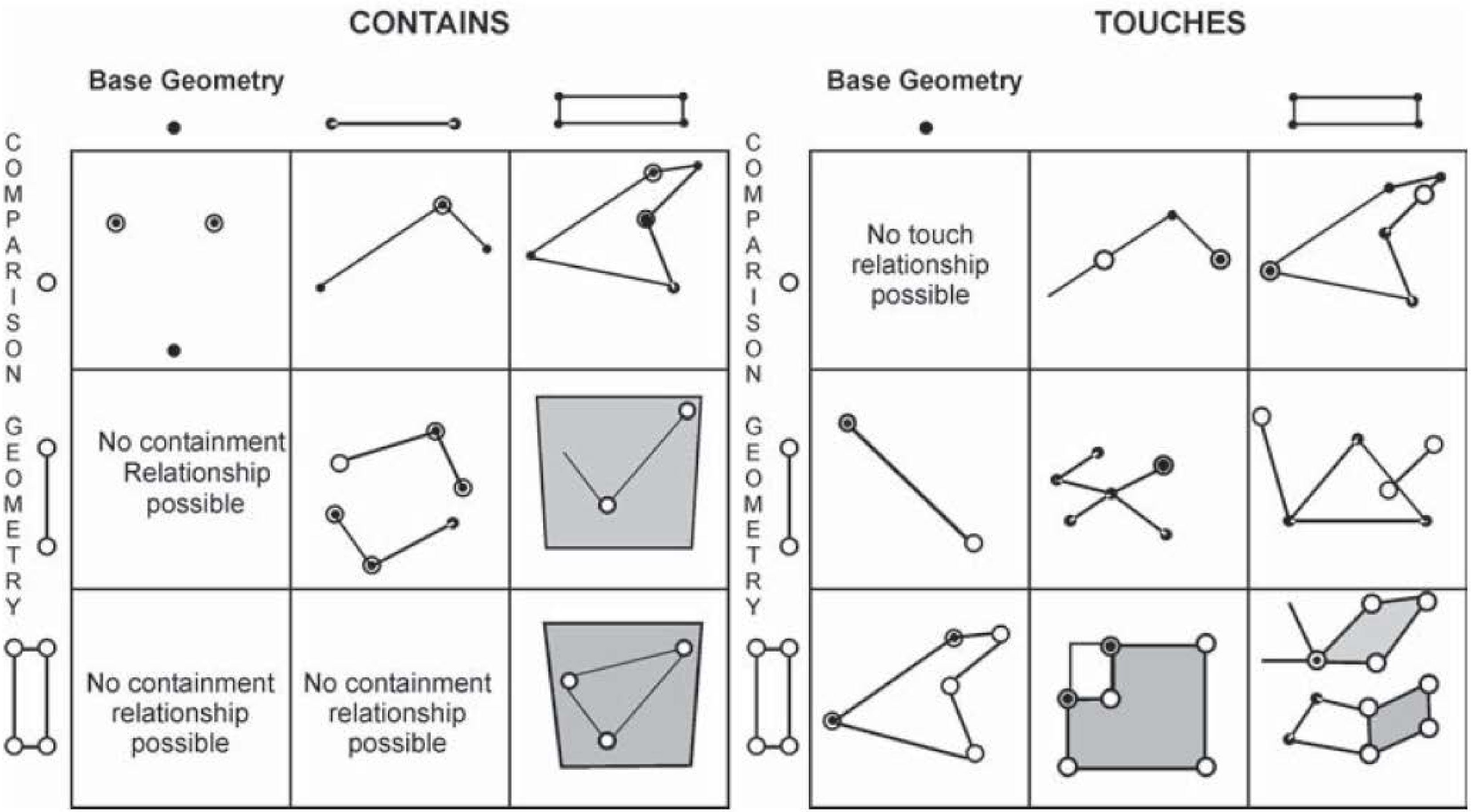


Figure 9.2: Examples of possible relations for two geographic database.

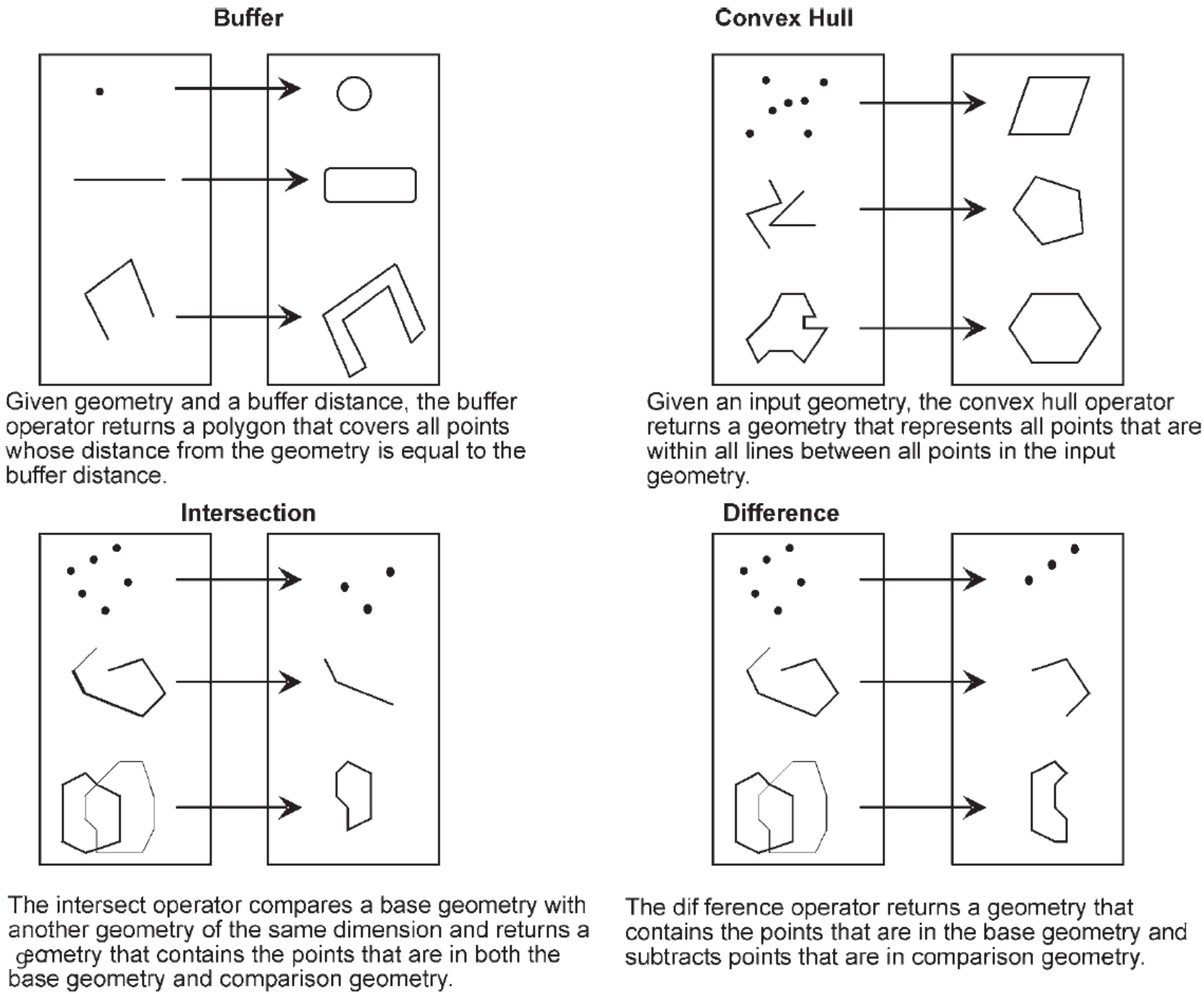


Figure 9.3: Examples of spatial analysis methods on geometries.

TYPES OF SPATIAL ANALYSIS

i. Queries and reasoning

- These are the most basic of analysis operations, in which the GIS is used to answer simple questions posed by the user.
- 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’.
  - The response may depend on the system’s ability to understand what the user means by ‘going east’.

ii. Measurements

- 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

- 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

- 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

- 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

- 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.
- But, its use with spatial data is much more problematic.

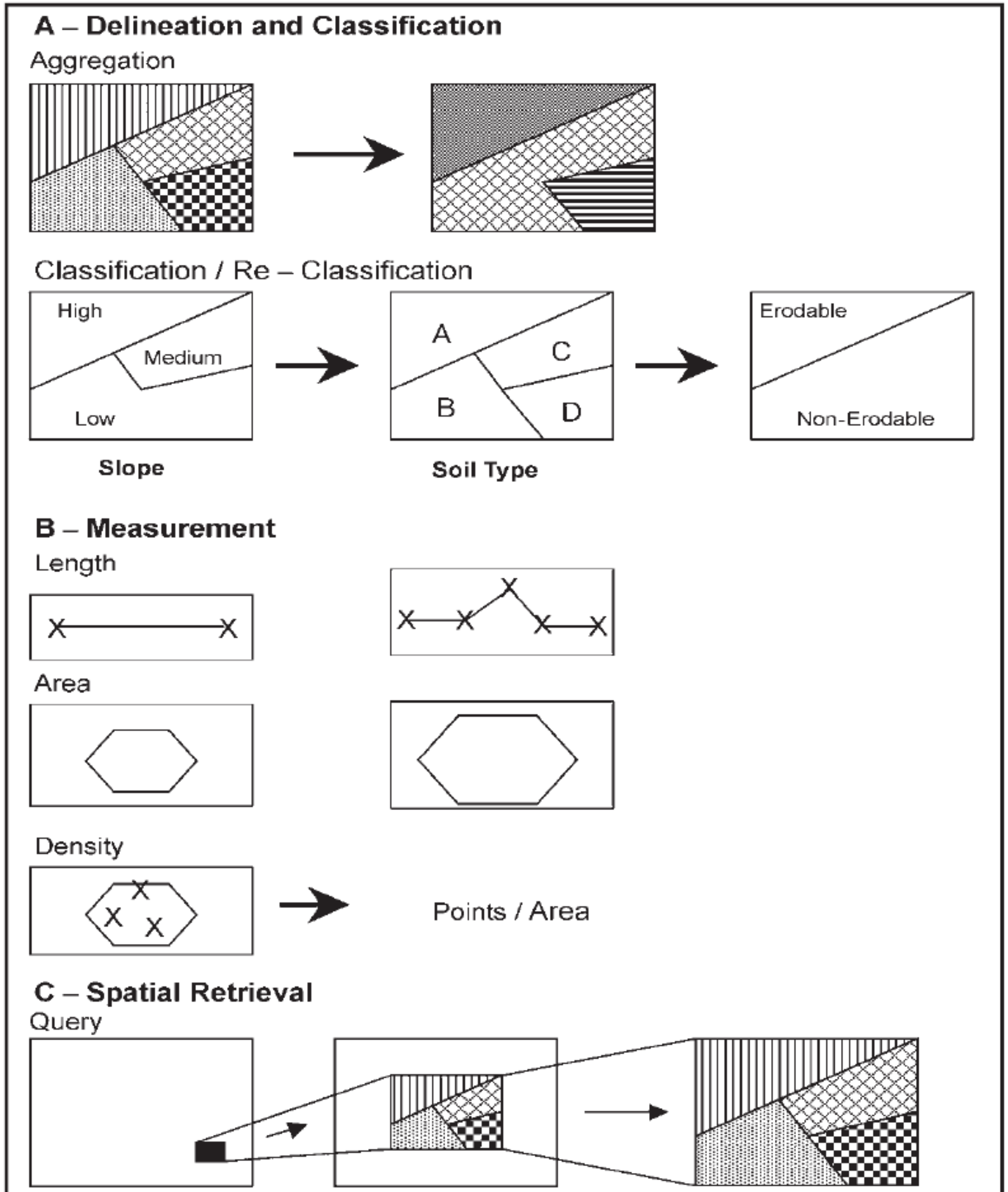
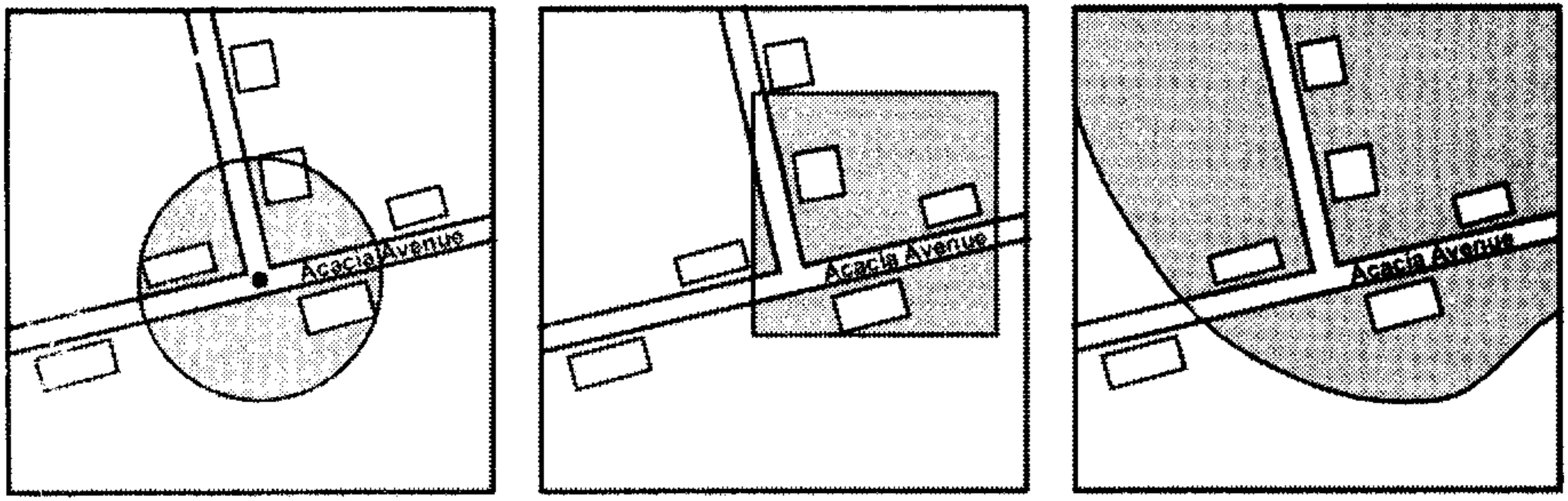


Figure 9.4: Spatial retrieval, delineation and classification, and measurement are separate functions, but are commonly used together.



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.

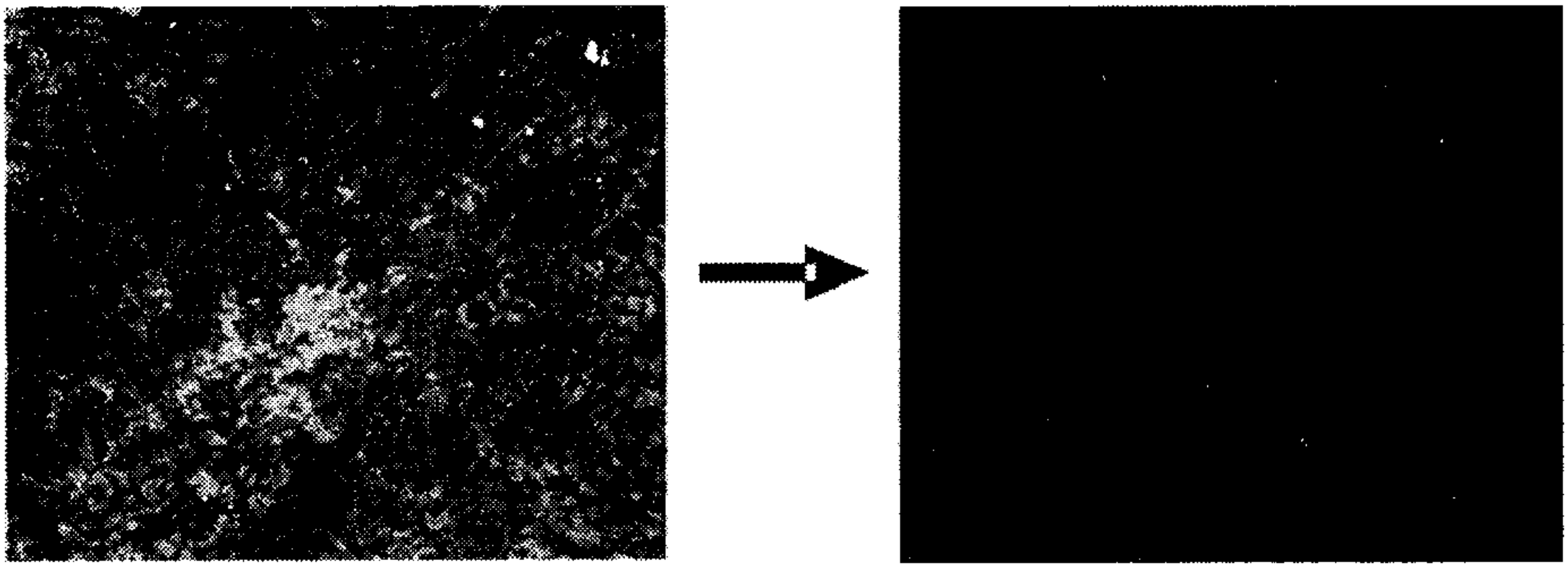


Figure 9.6: Example of *re-classification* where the modification in attribute values are made to produce new object data sets.

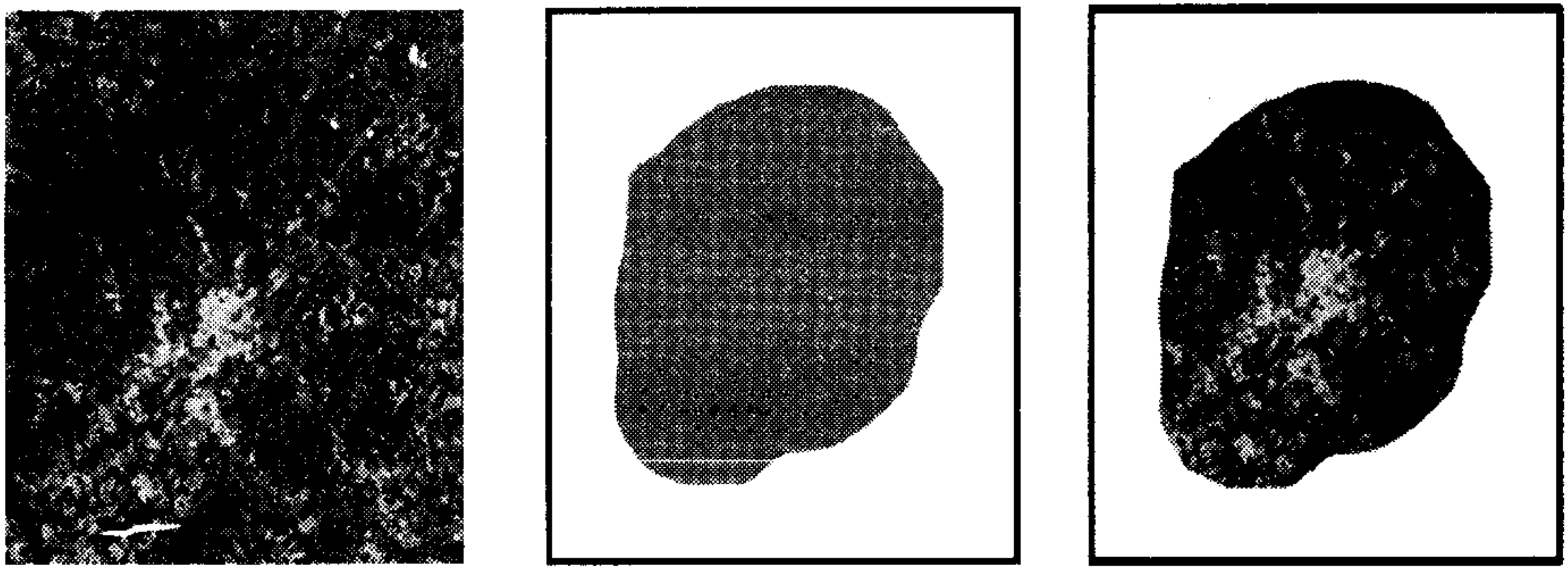


Figure 9.7: Example of *Cookie Cutting* where overlaying of datasets is made, using one dataset as a sieve or cookie cutter to select a subset of the other dataset.

QUERIES AND REASONING

- Many kinds of interrogation are available through 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.
- 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.
  - The map view of a dataset shows its contents in visual form, and opens many more possibilities for querying.
  - Exploratory spatial data analysis allows its users to gain insight by interacting with dynamically linked views.
- 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.
- The term reasoning encompasses a collection of methods designed to respond to more complex forms of query and interrogation.
  - One useful application of reasoning is in the area of navigation.

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, *etc*.
- In principle all of these measurements are obtainable by simple calculations inside a GIS.
- 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.

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.

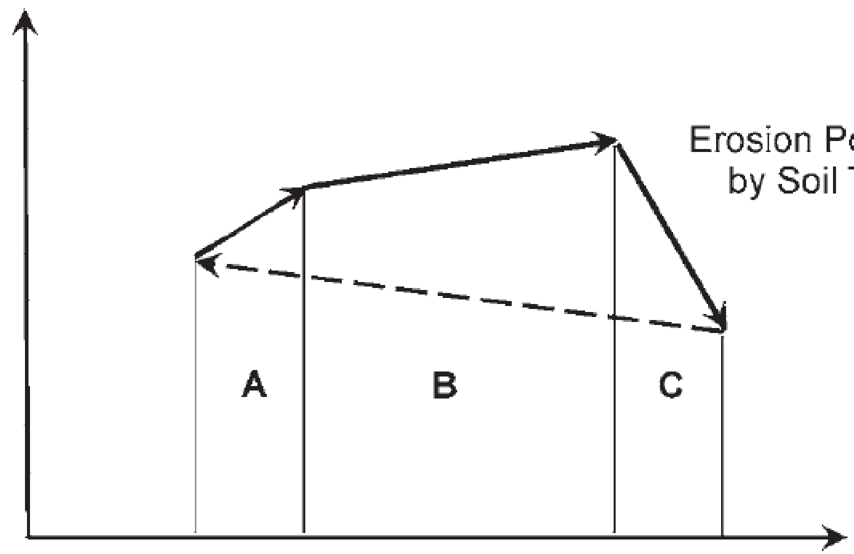


Figure 9.8: The algorithm for calculation of the area of a polygon given the coordinates of the polygon's vertices. The polygon consists of the three arrows and one arrow with dashed line forming the fourth side. Trapezia are dropped from each edge to the x axis and their areas are calculated as (difference 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.

DISTANCE AND LENGTH

- The simplest is the rule for determining the shortest distance between two points in a flat plane, called the pythagorean or straight-line metric.

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

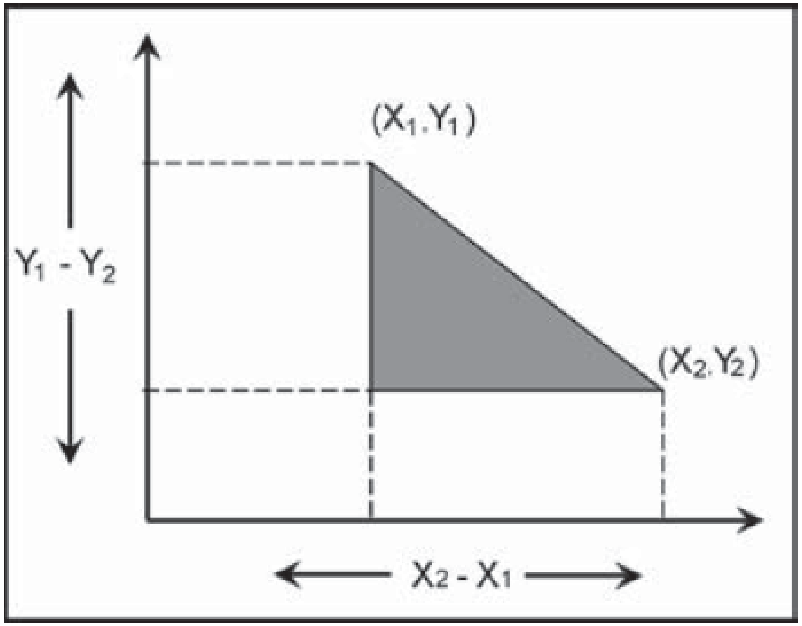


Figure 9.9: Pythagoras's theorem and the straight-line distance between two points.

- Distance between two points on a curved surface such as that of the Earth is
  - The shortest distance between two points is the length of a taut string stretched between them.
  - 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:

$$D = R\cos^{-1}\left[\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2)\right]$$

    – 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.
  - First, a polyline is often only a rough version of the true object's geometry.
    - 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.
    - But in area estimates, shortcutting corners tend to be canceled out.

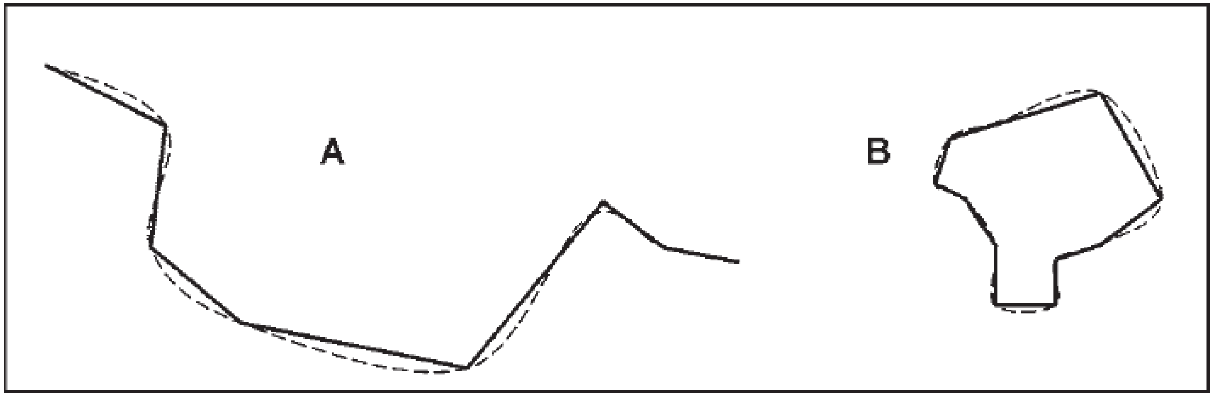


Figure 9.10: (A) – The polyline representations of smooth curves tend to be shorter in length. (B) – But estimates of area tend not to show systematic bias because of the effects of overshoots and undershoots cancel out.

- 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).

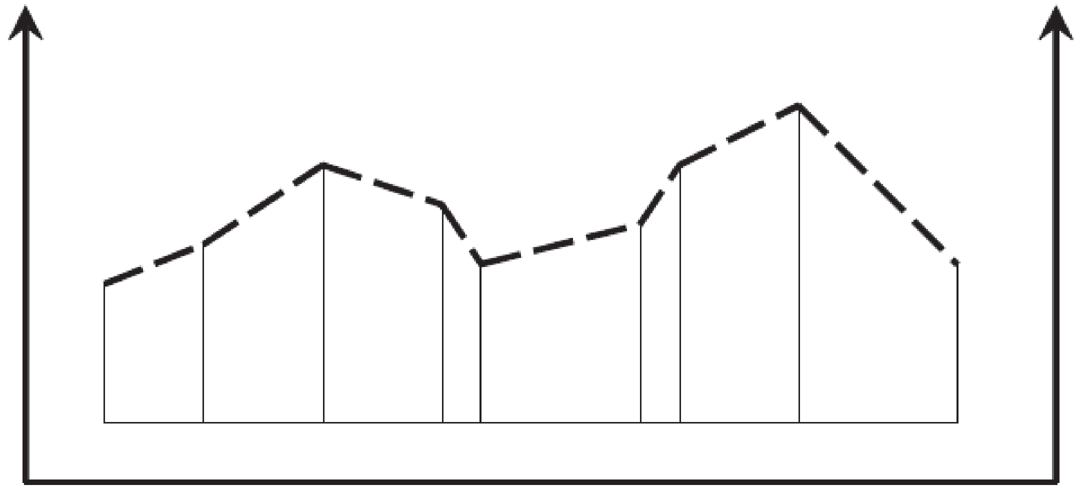


Figure 9.11: The length of a path on earth's surface (dashed line) remains longer than the length of its horizontal projection.

SHAPE

- GIS are also used to calculate the shapes of objects, particularly area objects.
- An easy way to define shape is by comparing the perimeter length of an area to its area measure.
- 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 magnitude of steepest tilt of the surface defines the slope at that point, and the direction of steepest tilt defines the aspect.
- Slope can be measured as an angle, varying from 0 to 90 degrees as the surface ranges from horizontal to vertical.
- What if the point of interest happens to be a peak, or a ridge?
  - A simple and satisfactory alternative is to take the view that slope must be measured at a particular resolution.
    - Slope is specific to the resolution, and a different resolution would have given a different result.
  - Second, there are several alternative measures of slope and can also be measured as a percentage or ratio, defined as rise over run.

- Figure 9.12 shows the two different ways of defining run.
  - Run means the horizontal distance covered between two points.
    - Slope as a ratio is equal to the tangent of the angle of slope.
  - Run means the diagonal distance.
    - Slope as a ratio is equal to the sine of the angle of slope.
- 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.
- Use of slope;
  - Input to many models of the soil erosion and runoff that result from heavy storms.
  - Input to analyses that find the most suitable routes across terrain for power lines, highways etc.

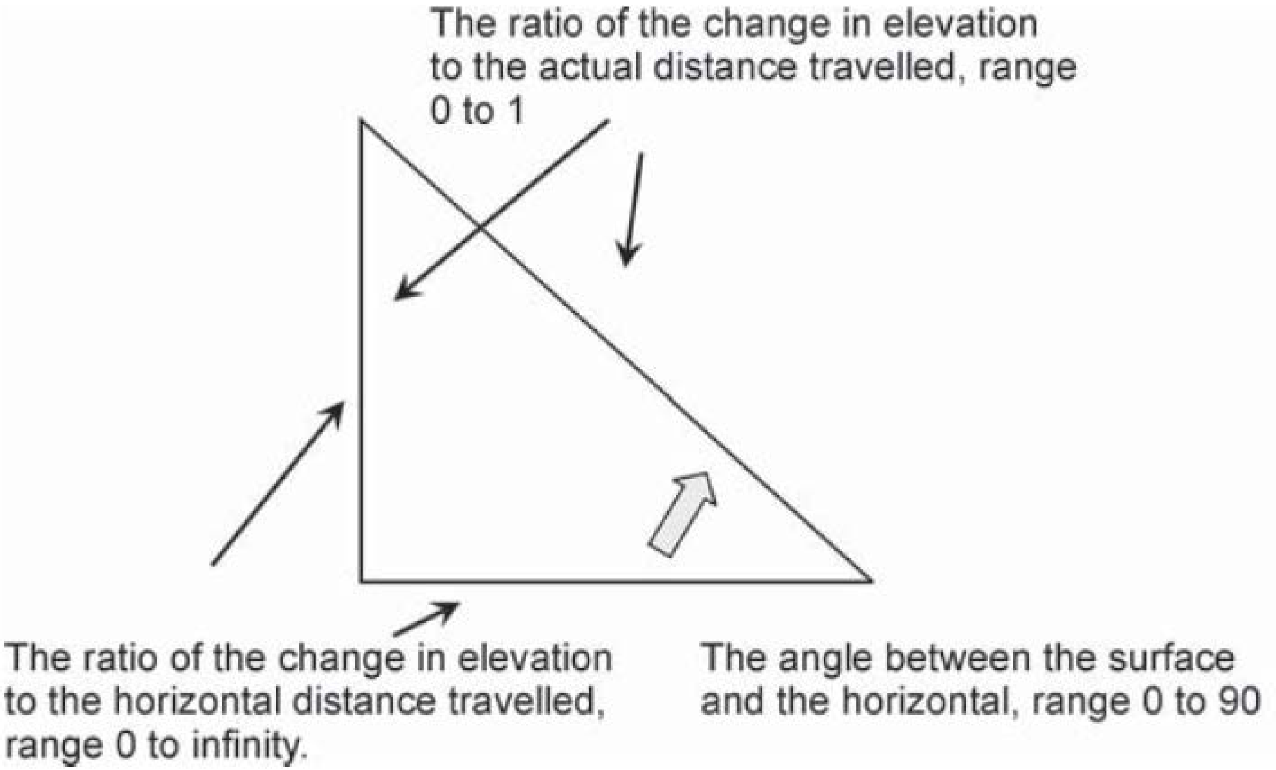


Figure 9.12: Three alternative definitions of slope.

BUFFERING

- 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.
- 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.
  - In the vector case, the result is a new set of objects.
- There is an additional possibility in the raster case that makes buffering more useful in some situations.
  - Figure 9.15 shows an example; buffering to spread outwards from the city at rates determined by the travel speed values in each cell.

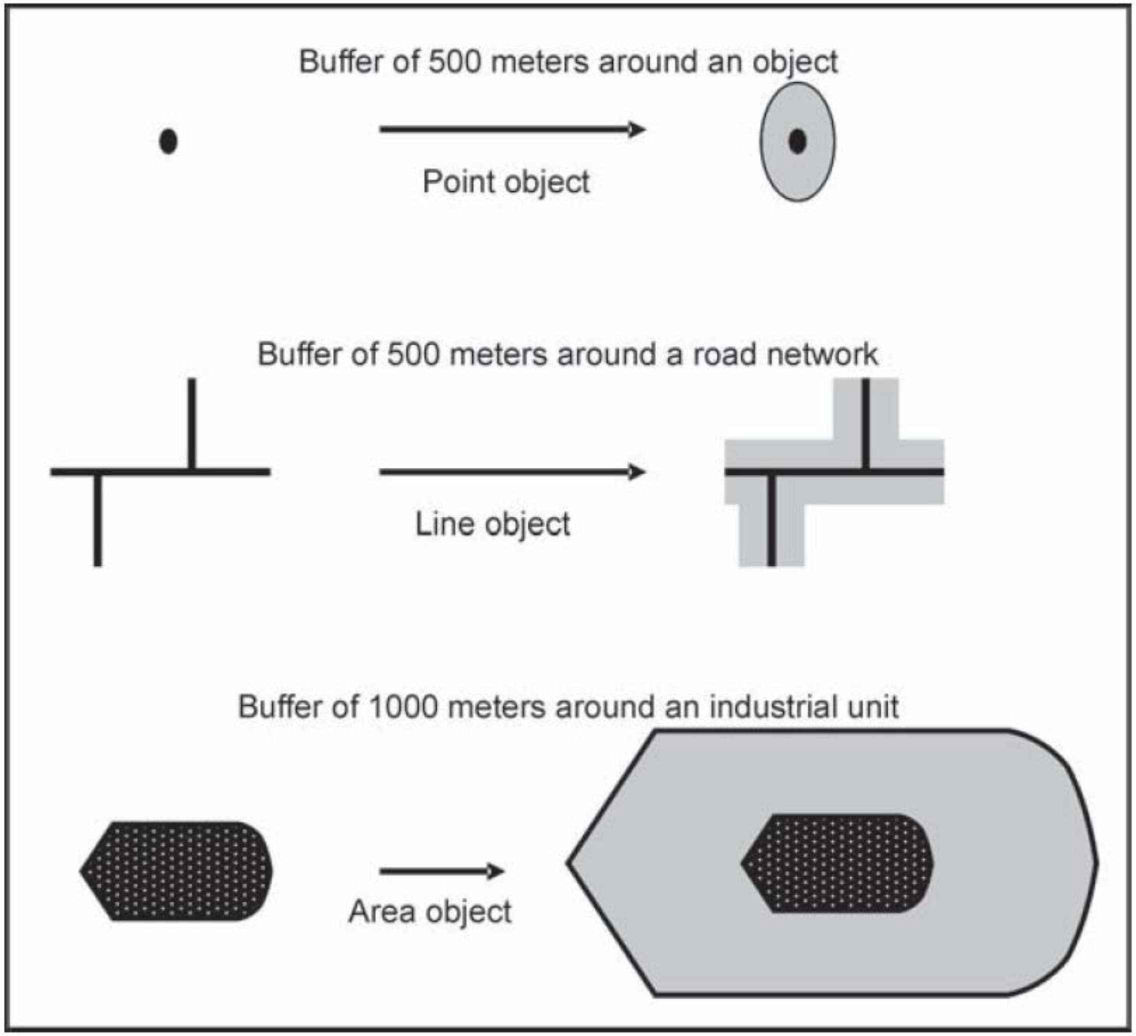


Figure 9.13: Buffers of constant width drawn around a point, line and a polygon.

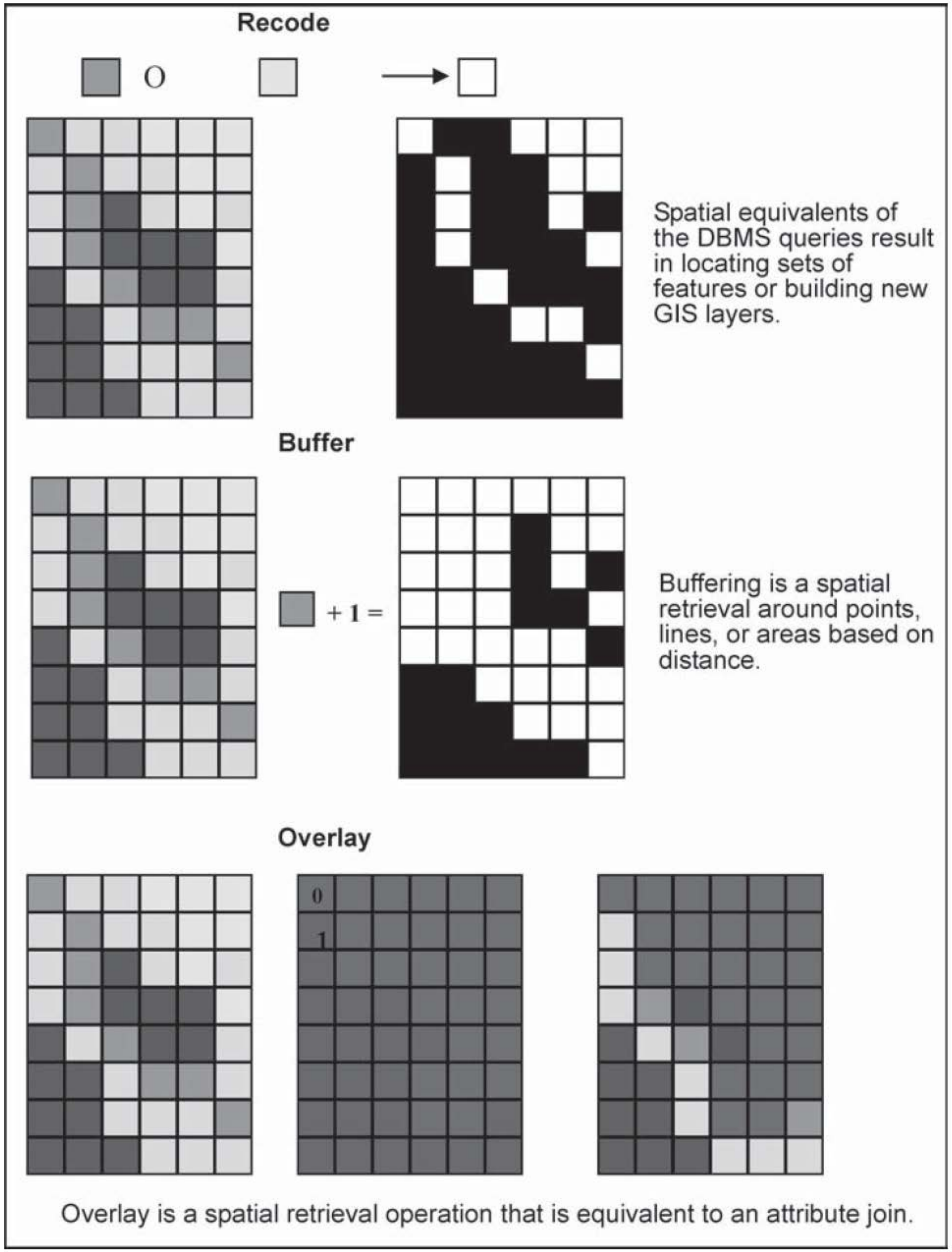


Figure 9.14: Spatial retrieval operations.

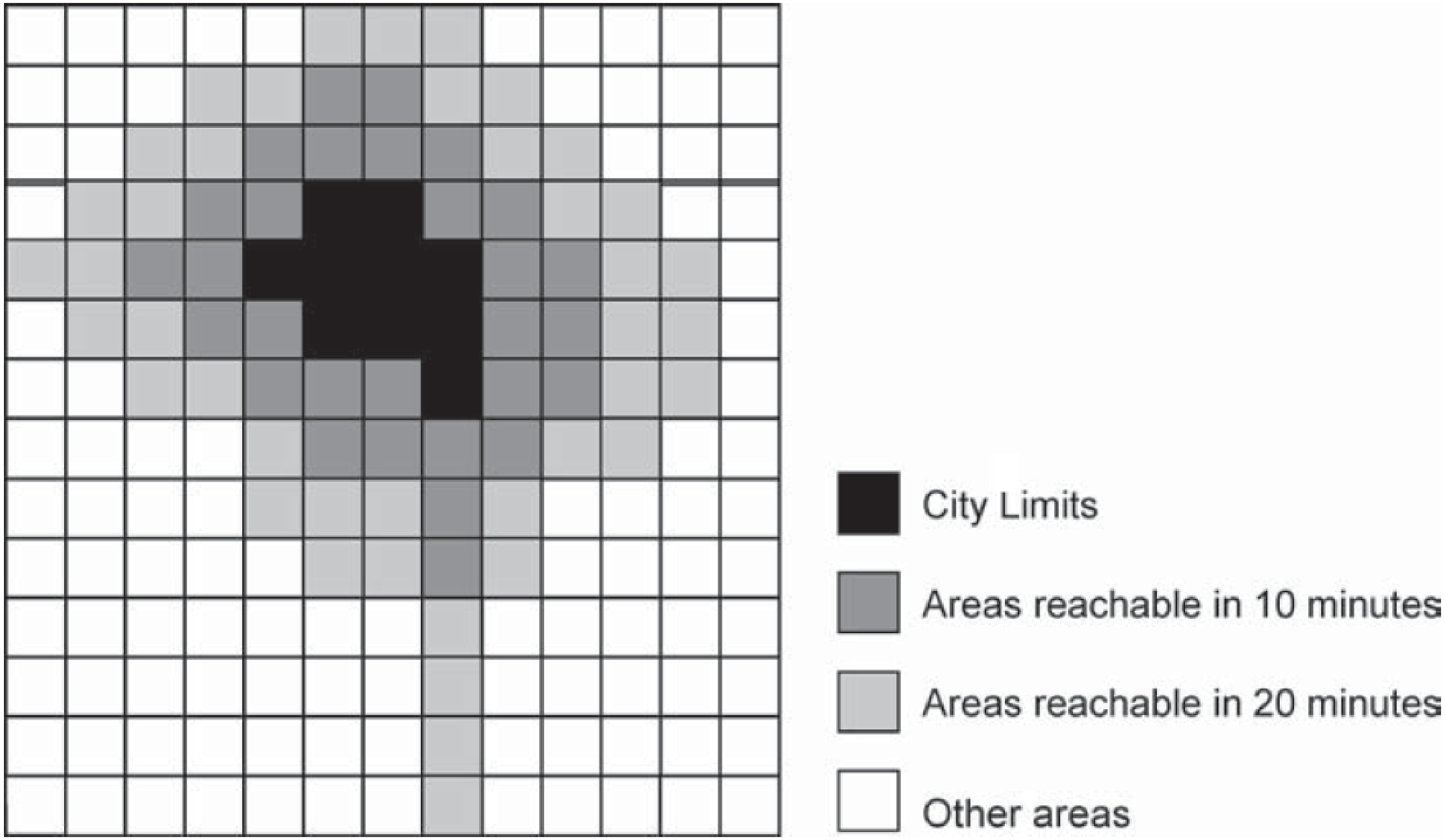
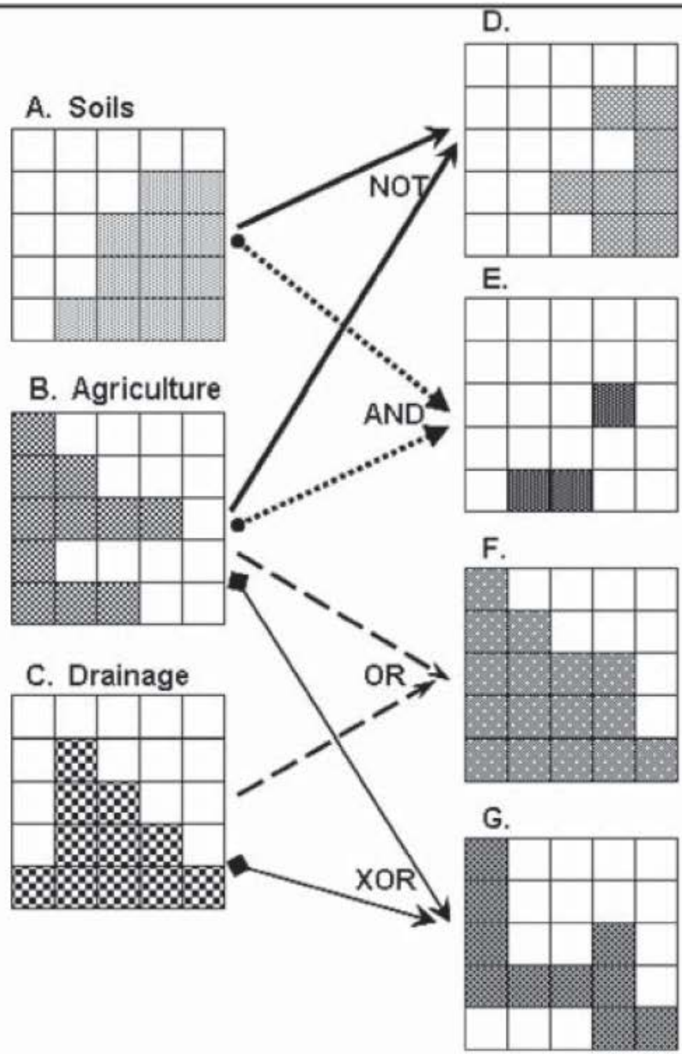


Figure 9.15: A raster generalization of the buffer function where changes may be controlled by some variable (example is of travel speed, whose value is recorded in every raster cell)

Boolean Logic for Queries
(name after George Boole, a 19th century mathematician)



Boolean function: A type of expression based upon, or reducible to a true or false condition.

Boolean Operators:

- AND
- OR
- NOT
- XOR (= either or, but not both)

Maps A, B, and C represent the original map layers.	
The shaded areas in D, E, F, and G represent where a condition is met for those locations.	
Map D shows :	"Where is condition A but not B met?"
	e.g. Which areas have deep soils but NOT Cultivated?
Map E shows :	"Where are both conditions A and D met?"
	e.g. Which areas have deep soils AND Cultivation?
Map F shows :	"Where are either conditions B or C met?"
	e.g. Which areas have Cultivation OR good drainage?
Map G shows :	"Where is condition B or C, but not both, met?"
	e.g. Which areas have either Cultivation OR good drainage but NOT both?

Figure 9.16: Examples of Boolean logic using Boolean operators.

Point in Polygon

- 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.
- Figure 9.17 illustrates the task.

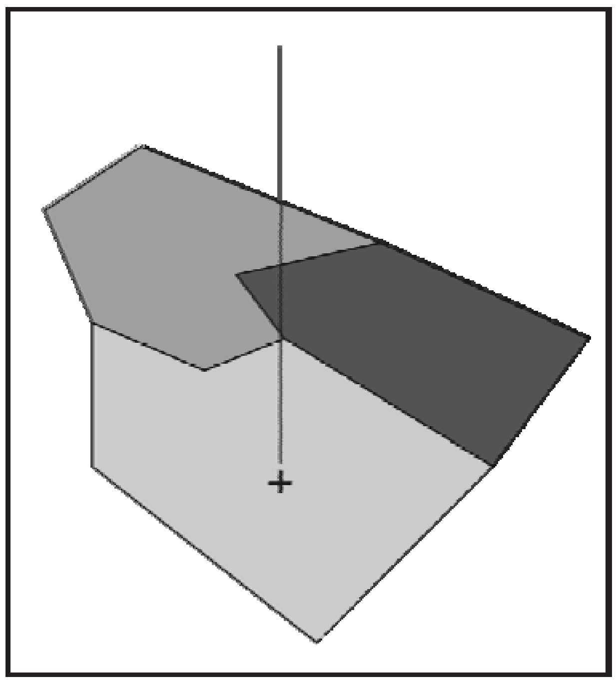


Figure 9.17: The point in polygon problem.

- The operation is popular in GIS analysis because it is the basis for answering many simple queries.
  - For example, to determine how many instances of the disease occurred in each ward.
    - The points represent instances of a disease in a population.
    - The polygons represent reporting zones such as wards
- 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 and overlap would imply that a point is owned simultaneously by two owners.
    - There can be no gaps between polygons and each point is assigned to exactly one polygon.

- The standard algorithm for the point in polygon operation is shown in Figure 9.17.
  - 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.

POLYGON OVERLAY

- Polygon overlay 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?

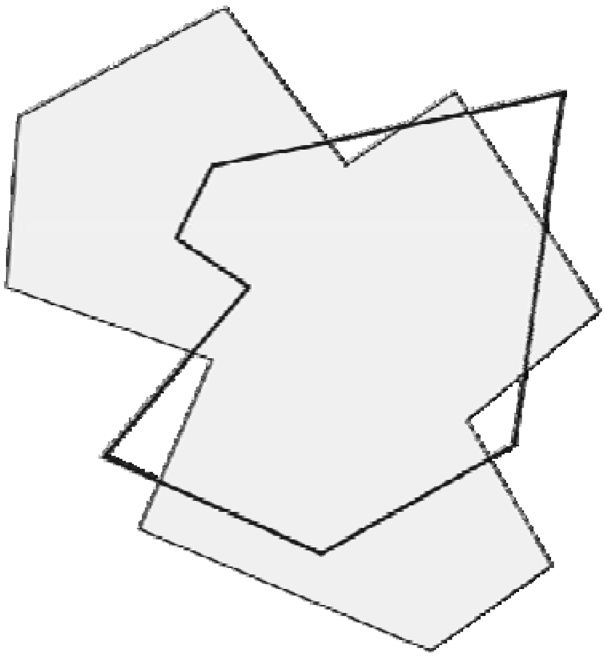
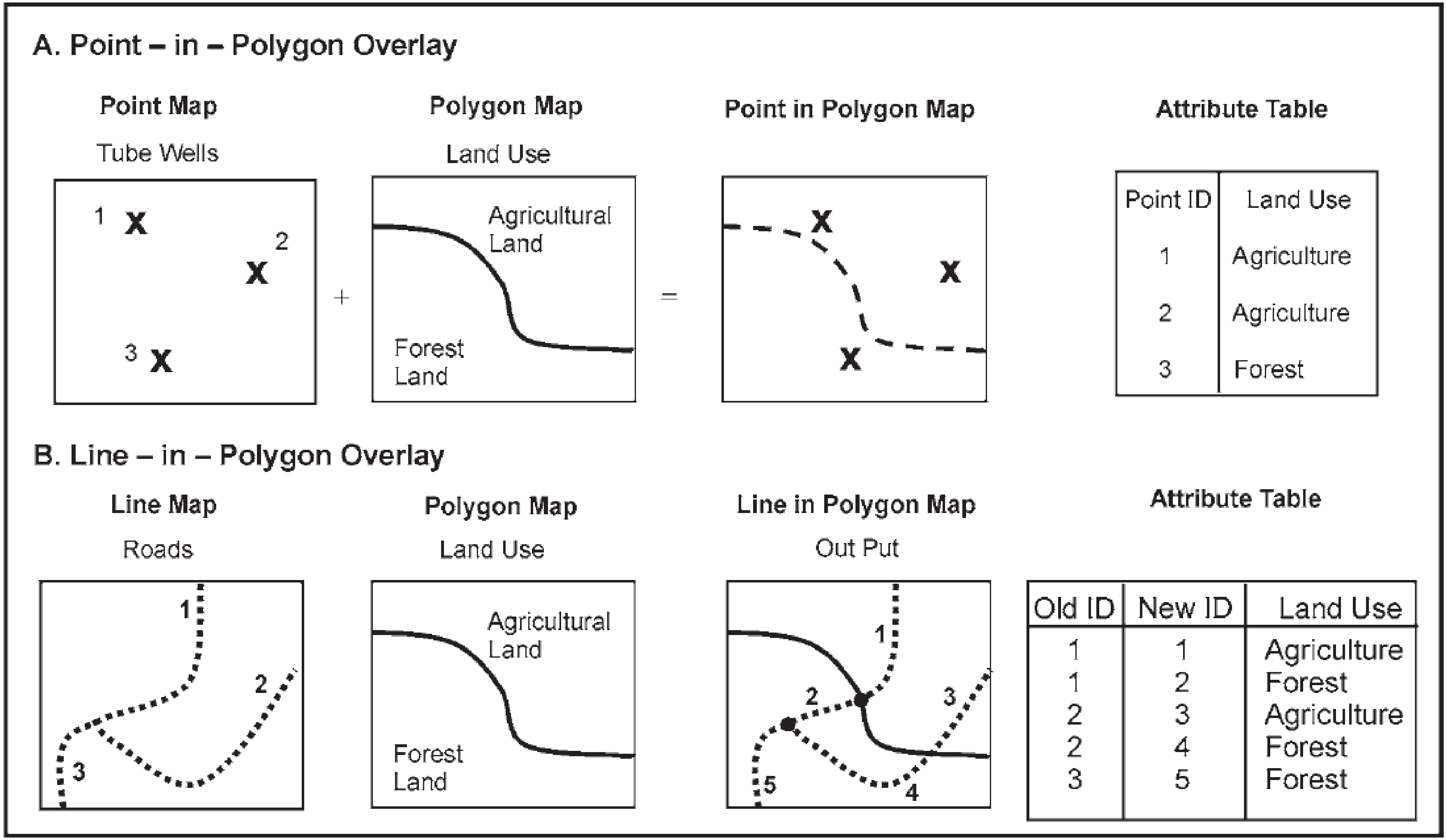


Figure 9.18: An example of polygon overlay, in the discrete object case. Here the overlay of two polygons produces nine polygons. One has the property of both, four have the properties of shaded but not the white polygon and four are outside the shaded but inside the white polygon.



Point ID	Land Use
1	Agriculture
2	Agriculture
3	Forest

Old ID	New ID	Land Use
1	1	Agriculture
1	2	Forest
2	3	Agriculture
2	4	Forest
3	5	Forest

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

C. Polygon – on – Polygon Overlay

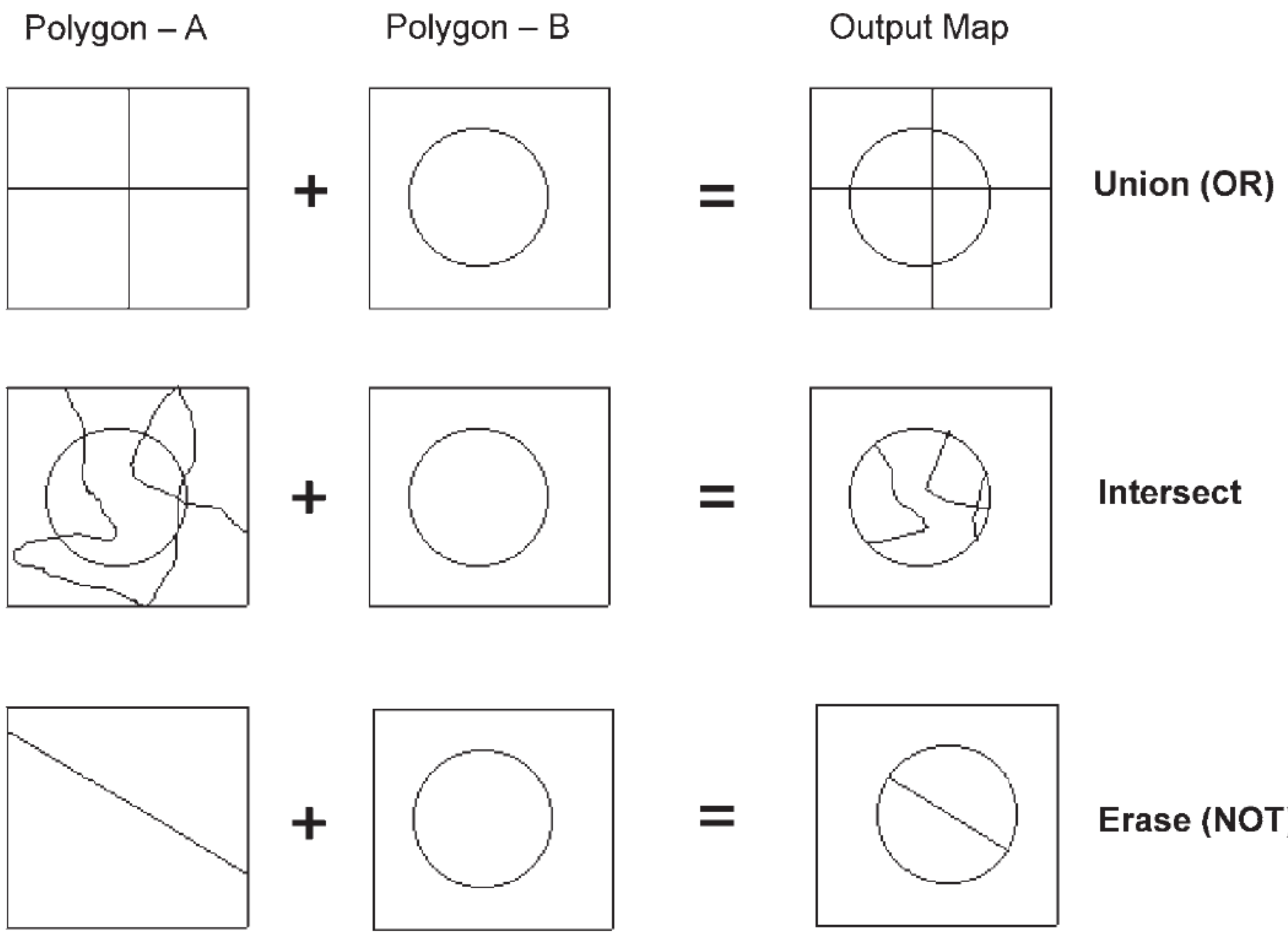


Figure 9.20: Vector overlays (polygon on polygon)

- From the field perspective,
  - There are numerous queries that require simultaneous access to both datasets (Figure 9.21), 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?
  - 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.

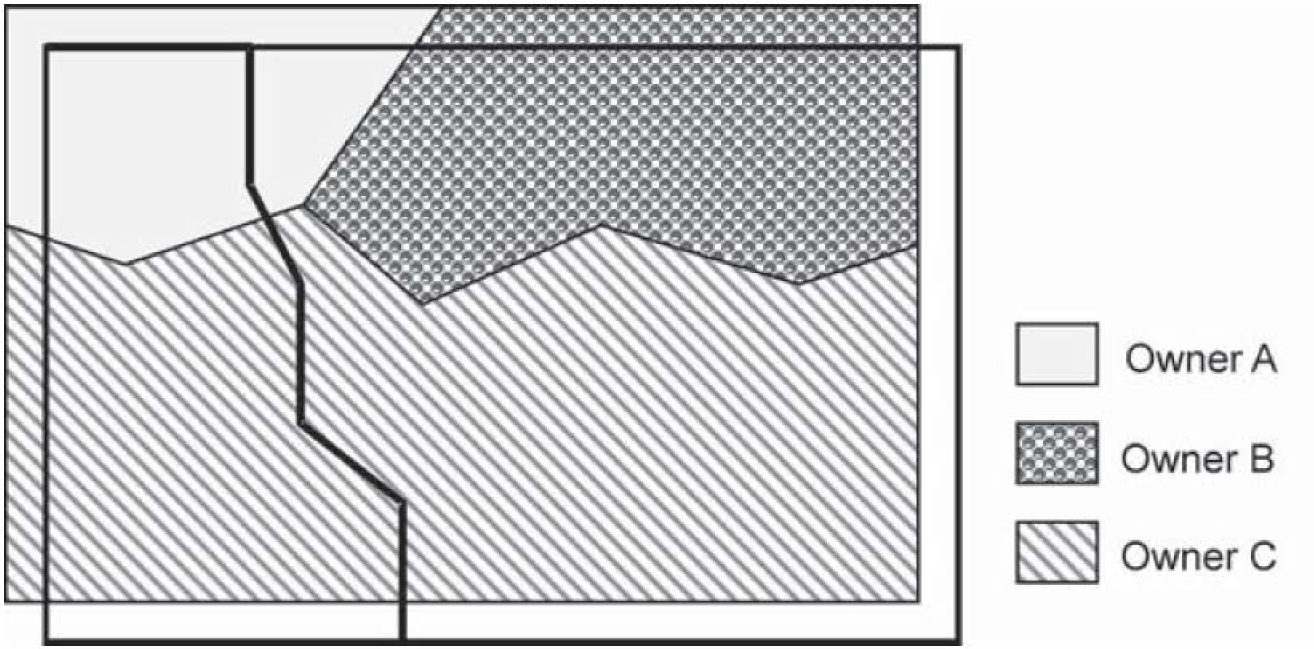


Figure 9.21: Polygon overlay in the field case. Where a dataset representing two types of land cover. (one on the left, say X and another in right, y). It is overlaid on a dataset showing three land parcels owned by three different persons. The overlay result will be a single dataset in which every point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

- The spurious polygon or coastline weave problem
  - There will be instances in any practical application where the same line on the ground occurs in both datasets.
  - When overlaid, the result is a series of small slivers.
  - 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.

- 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.
  - In raster overlay, the attributes of each cell are combined according to a set of rules.
  - In vector overlay, there is no rule for combination, and 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 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.

- 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 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.
- Two commonly used methods of spatial interpolation:
  - Inverse distance weighting (IDW), which is the simplest method.
  - 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 which is most often used by GIS analysts.
- IDW employ 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.

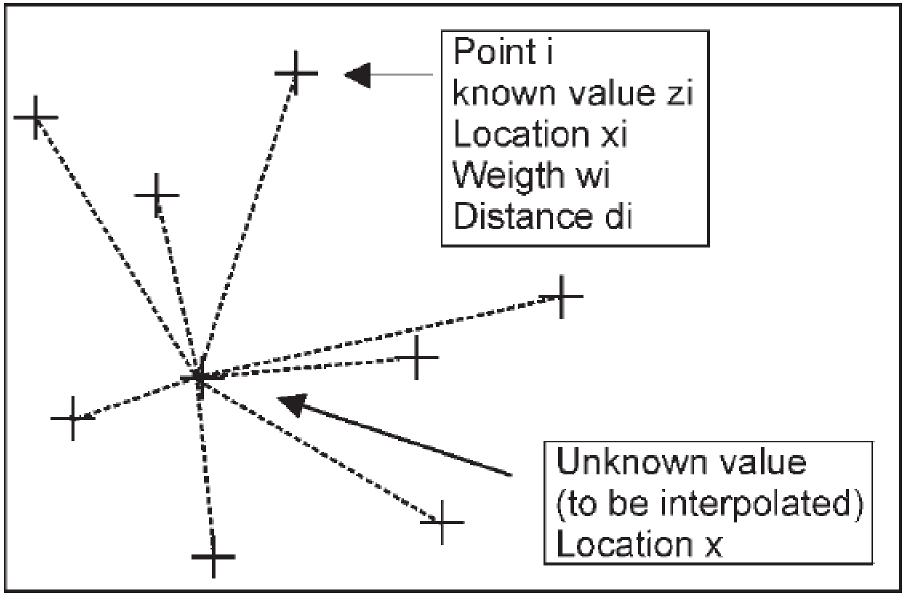
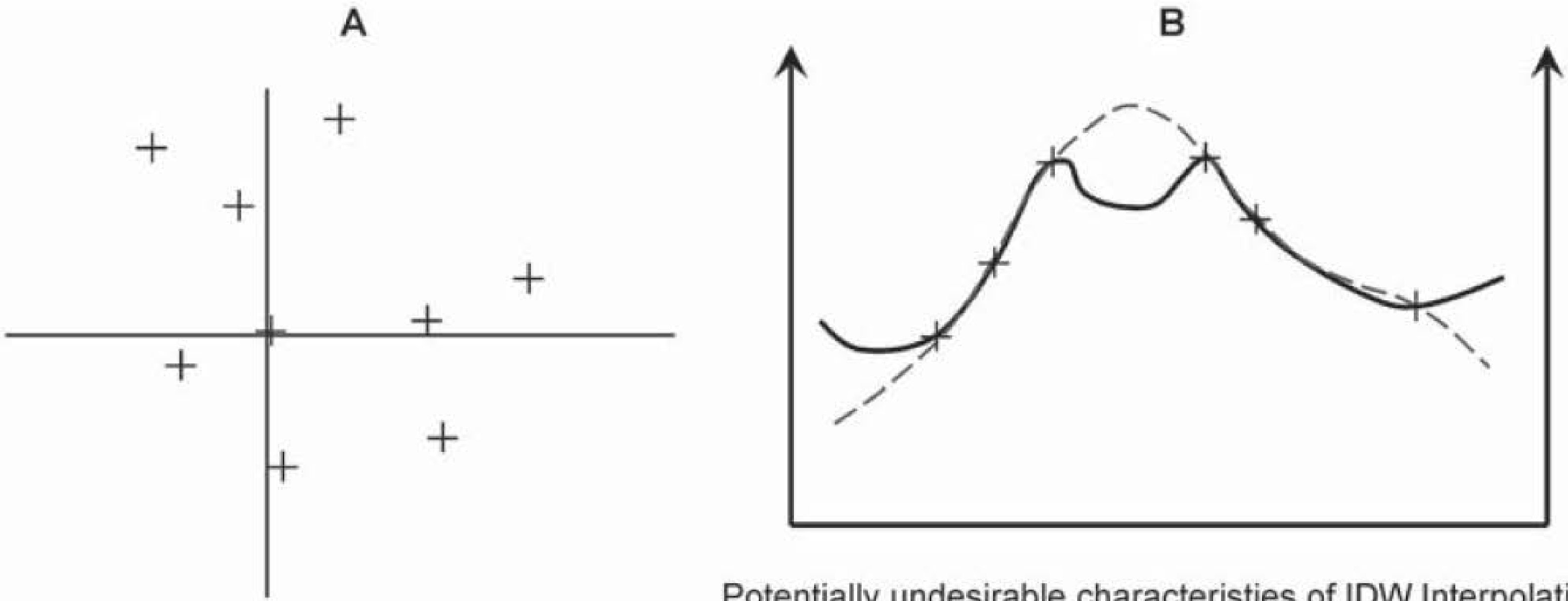


Figure 9.22: Notation used in the equations defining spatial interpolation.

- 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.



Selection of points for IDW interpolation using sector. Hear four sectors are defined and only the closest points in each sector are used.

Potentially undesirable characteristies of IDW Interpolation. This set of six data points clearly suggests a hill profile (dashed line). But in areas where there is little or no data the interpolator move towards the overall mean (solid line).

Figure 9.23: IDW interpolation.

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.
- To measure the amount, we take the difference and square it.

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

- 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).

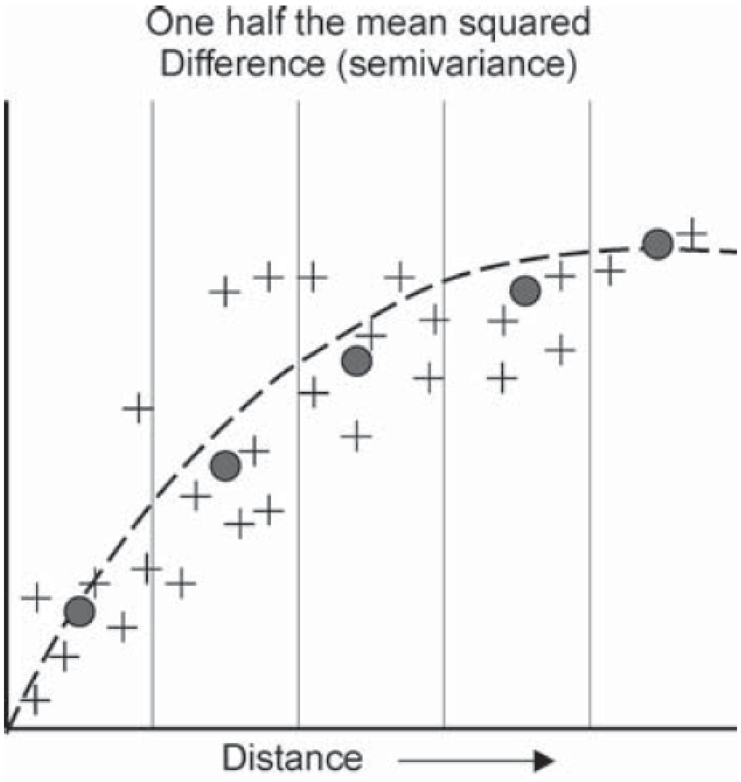


Figure 9.24: A semivariogram, here each cross represents a pair of points. The solid circles are obtained by averaging within the ranges of the distance axis. The dashed line is the best fit to the five points.

- 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).
- 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.
- Kriging responds both to the proximity of sample points and to their directions.

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, spatial interpolation seeks to estimate the missing parts of a field from samples of the field taken at data points, while density estimation creates a field from discrete objects.
  - Figure 9.25 illustrates this difference.
  - In the discrete object view there is nothing between the objects but empty space – no missing field to be filled in through spatial interpolation.
  - 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.
  - For example, estimation of population density, density of different kinds of diseases, or animals, or any other set of well-defined points.

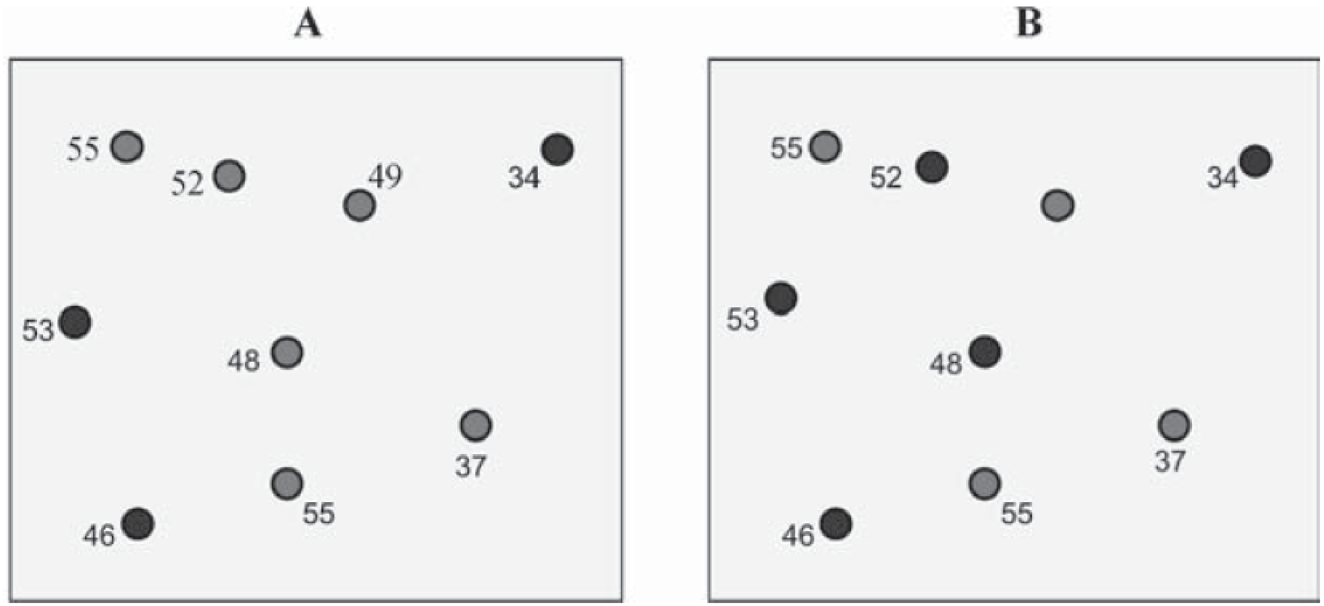
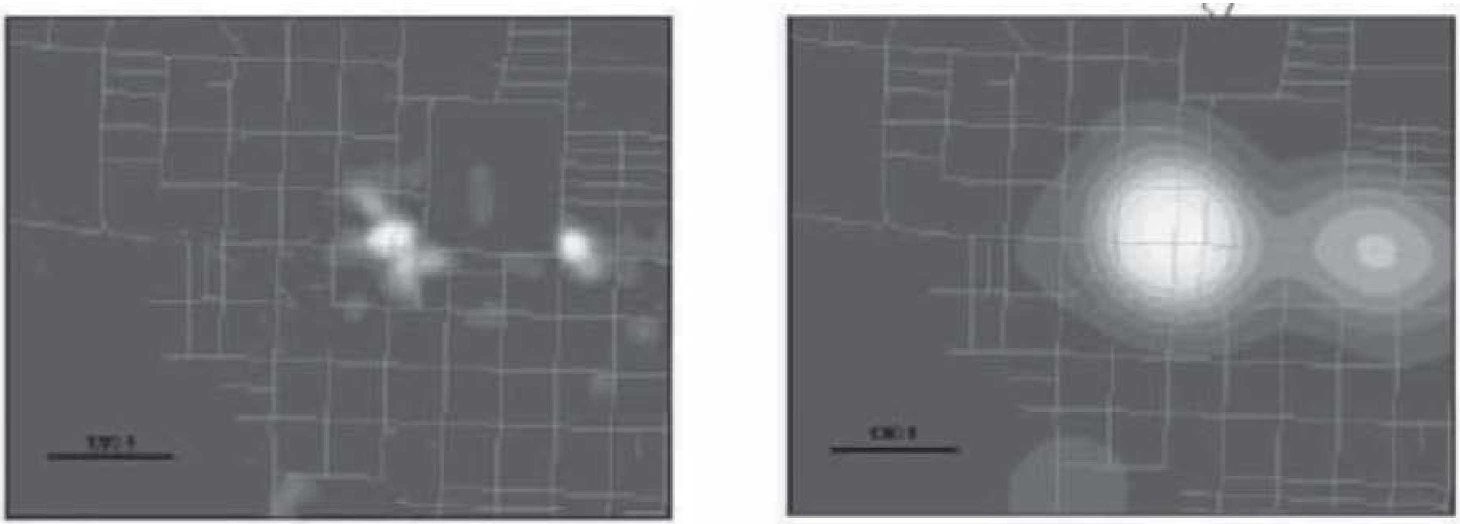


Figure 9.25: Two identical datasets but with different representations.
A – represents a field of atmospheric temperature recorded at nine sample points.
B – nine discrete objects representing population of different settlements in thousands.
Spatial interpolation is appropriate for case A, while for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



Search Result – 500 meters   Search Result – 5000 meters

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.
- The advent and easy availability of large datasets and fast computing led new ways of thinking about spatial analysis.
  - 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

- **Mean** is one method which is broadly citing the average of data series.
- **Median** is the value where 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**.
- 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.

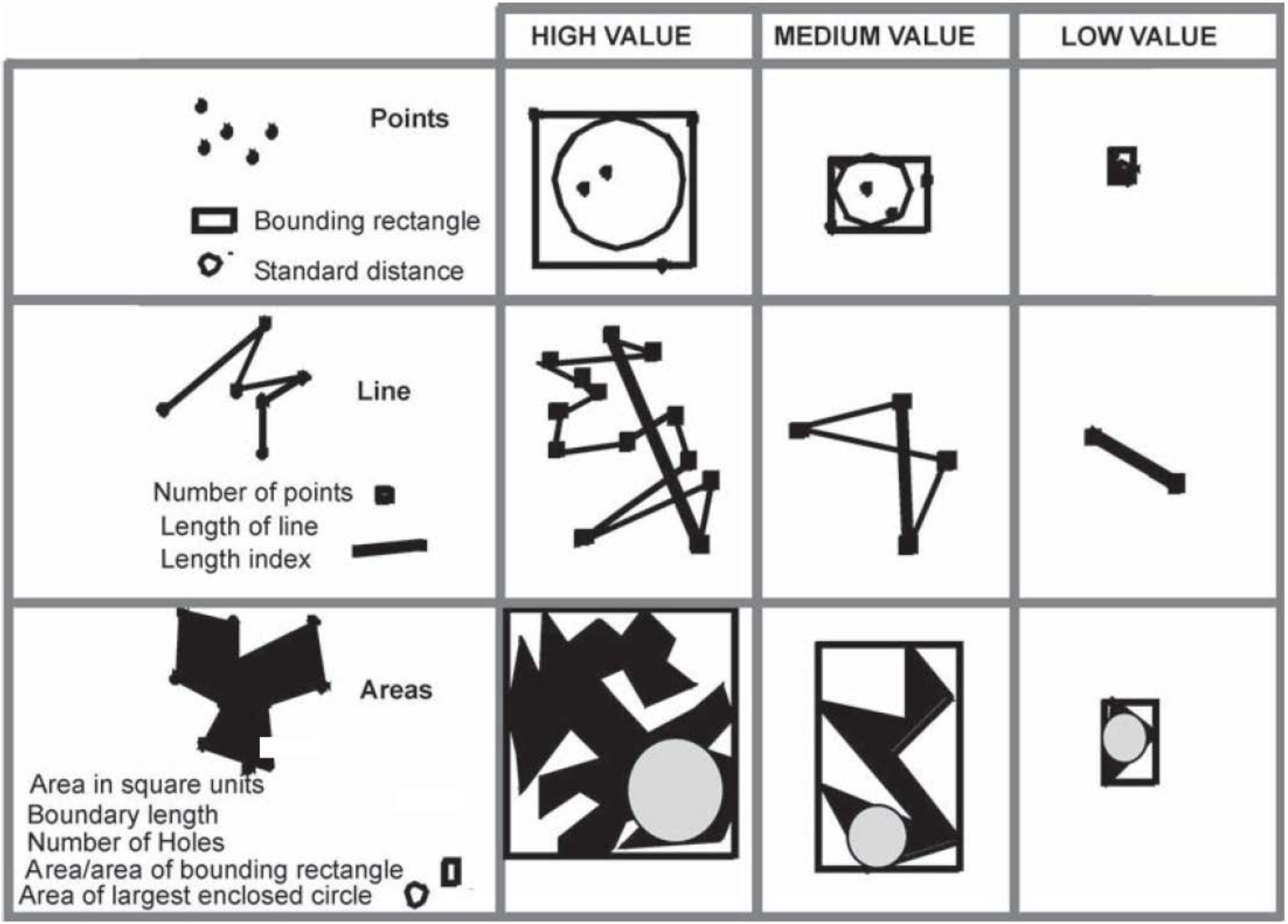


Figure 9.27: Statistics and features.

Dispersion

- 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.
  - Standard deviation is the mean squared difference from the mean and Variance is the square of the standard deviation.
- 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.
- 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.

Scatterplots

- 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.
- If both are attributes of raster datasets, then scatterplot is built by comparing the datasets pixel by pixel.
- 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 Tobler's first law of geography states that everything is related to everything else but near things are more related than distant things.
- Spatial dependence is inherently scale specific and can be measured at any spatial resolution.
- 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

- Using fragmentation statistics we can analyze the number of patches, their shape or size etc, as a way of summarizing the geographic details.
  - In maps, a patch represents an area of uniform class.
- The concept of fractals is used as a way of 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.

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.
- 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.

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 to save fuel, time or avoiding the restrictions if there are any.

HYPOTHESIS TESTING

- Hypothesis testing 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.

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.
  - Another important issue is about the earth's surface which is heterogeneous, making it difficult to take samples that are truly representative for any large region.
- 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.