GEO 465/565 - Lecture 4 - "The Nature of Geographic Data"

(from Longley et al., GI Systems and Science, 2001)

3.2

What are geographic data?

Geographic data link place, time, and attributes.

Place

Place, or location, is essential in a geographic information system. Locations are the basis for many of the benefits of geographic information systems: the ability to map, to link different kinds of information because they refer to the same place, and to measure distances and areas. Without locations, data are said to be "aspatial" and have no value at all within a geographic information system.

Time

Time is an optional element. Many aspects of the earth's surface are slow to change and can be thought of as unchanging. Height above sea level changes slowly because of erosion and movements of the earth's crust, but these processes operate on scales of hundreds or thousands of years, and for most applications (except geophysics) we can safely omit time from the representation of elevation. On the other hand, atmospheric temperature changes daily, and dramatic changes sometimes occur in minutes with the passage of a cold front or thunderstorm, so time is distinctly important.

Attributes

Attributes refer to descriptive information. The range of attributes in geographic information is vast. Some attributes are physical or environmental in nature (e.g., atmospheric temperature or elevation), while others are social or economic (e.g., population or income). There are five main types of attributes: nominal, ordinal, interval, ratio, and cyclic.

3.3
How are geographic data represented?

This course and your textbooks focus on one particular form of representation that is becoming increasingly important in our society—representation in digital form.

Digital representations of geography hold enormous advantages over previous types, such as paper maps, written reports from explorers, and spoken accounts. We can use the same digital devices—PCs, the Internet, or mass storage devices—to handle every type of information, independent of its meaning. Digital data are easy to copy, they can be transmitted at very high speeds, they can be stored at high density in very small spaces, and they are less subject to the physical deterioration that affects paper and other physical media.

Perhaps more importantly, data in digital form are easy to transform, process, and analyze. Geographic information systems allow us to do things with digital representations that we were never able to do with paper maps: to measure accurately and quickly, to overlay and combine, and to change scale, zoom, and pan without respect to map sheet boundaries. Digital representations open up a vast array of processing possibilities that will be explored in more depth later in the course.

3.3
Problems with representing geographic data

Any application of GIS requires clear attention to questions of what should be represented and how it should be represented. There are many possible ways of representing the geographic world in digital form, none of which is perfect and none of which is ideal for all applications.

One of the most important criteria for the usefulness of a representation is its accuracy. Because the geographic world is almost infinitely complex, there are always choices to be made in building any representation. You must decide what to include and what to leave out.

In principle, if we collected enough atoms of geographic information, we would be able to build a complete representation of the world. The idea of integrating all available geographic information into a single digital representation underlies the idea of "Digital Earth," a concept that originated with former U.S. Vice President Al Gore in his book Earth in the Balance (Gore 1992), and which was explored further in one of his speeches.

Fascinating though this scenario is, it glosses over the fundamental problem, which is that the world is in effect infinitely complex. The closer we look at the world, the more detail it reveals. The shoreline of Maine in the United States appears complex on a map, but even more complex when examined in greater detail.

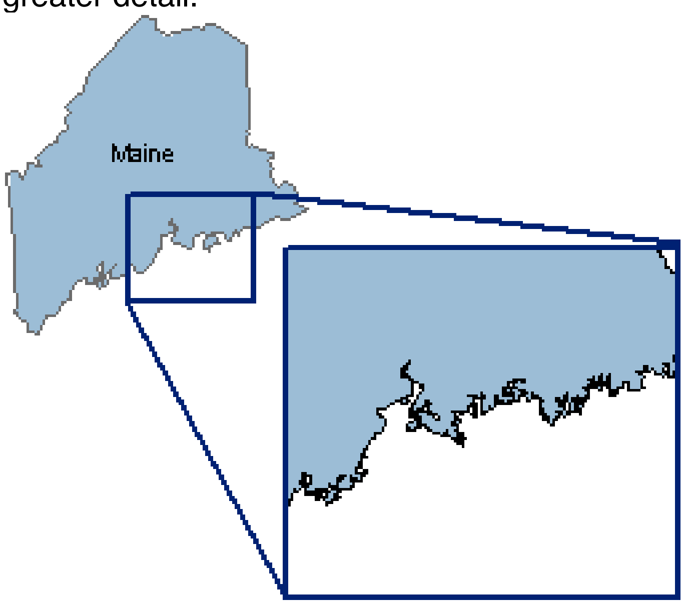


Fig. 1.1

To characterize the world completely, we would have to specify the location of every person, every blade of grass, every grain of sand—in fact, every subatomic particle, clearly an impossible task. So, in practice, any representation must be partial—that is, it must limit the level of detail provided, ignore change through time, ignore certain attributes, or simplify in some other way.

One very common way of limiting detail is by throwing away or ignoring information that applies only to small areas; in other words, not looking too closely. The image you see on a computer screen is composed of a million or so basic elements or pixels and, if the whole earth were displayed at once, each pixel would cover an area roughly 10km on a side, or about 100 sq km. At this level of detail, the island of Manhattan occupies roughly 10 pixels and virtually everything on it except Central Park is a blur. We would say that such an image has a spatial resolution of about 10km and know that anything much less than 10km across is virtually invisible.

Another strategy for limiting detail is to observe that many properties remain constant over large areas. For example, in describing the elevation of the earth's surface, we could take advantage of the fact that roughly two-thirds of the surface is covered by water, with its surface at sea level. Of the 5 million pieces of information needed to describe elevation at 10km resolution, approximately 3.4 million will be recorded as zero, a colossal waste. If we could find an efficient way of identifying the area covered by water, we would need only 1.6 million real pieces of information.

Humans have found many ingenious ways of describing the earth's surface. This ingenuity is itself the source of a substantial problem for GIS: there are many ways of representing the earth's surface and users of GIS thus face difficult and at times confusing choices.

3.5 Discrete objects and fields

The two fundamental ways of representing geography are discrete objects and fields.

Discrete objects

The discrete object view represents the world as objects with well-defined boundaries in empty space. Just as the desktop may be littered with books, pencils, or computers, the geographic world is littered with cars, houses, forest stands, and other discrete objects. One characteristic of the discrete object view is that objects can be counted. For example, there may be 49 houses in a particular subdivision.

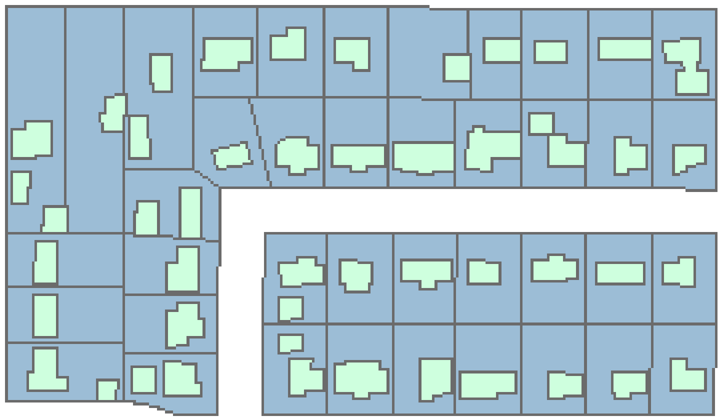


Fig. 1.2

Geographic objects are identified by their dimensionality. Objects that occupy area, including lakes, parcels, and forest stands, are termed two-dimensional and generally referred to as areas or polygons. Other objects that are linear, including roads, railways, and rivers, are termed one-dimensional and generally referred to as lines. Objects that are single locations, including individual animals and buildings, are termed zero-dimensional and generally referred to as points.

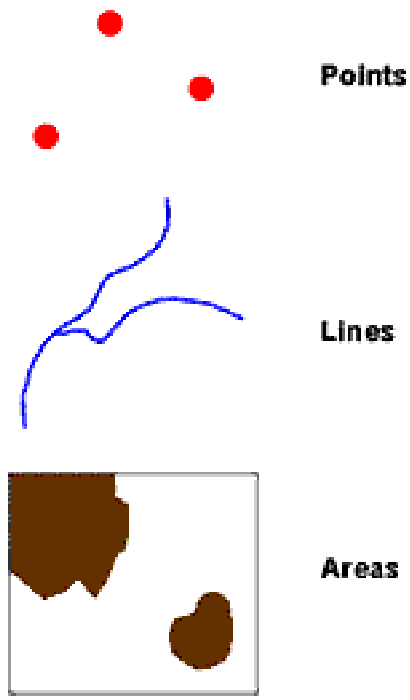


Fig. 1.3

The discrete object view leads to a powerful way of representing geographic information about objects. Consider a class of objects of the same dimensionality—for example, all the grizzly bears in the Kenai Peninsula of Alaska. We would naturally think of these objects as points. We might want to know the sex of each bear and its date of birth if our interests were in monitoring the bear population. We might also have a collar on each bear that transmitted the bear's location at regular intervals.

All of this information could be expressed in a table, like the one shown below, with each row corresponding to a different discrete object and each column to an attribute of the object. To reinforce a point made earlier, this is a very efficient way of capturing raw geographic information on grizzly bears.

Bear ID	Sex	Estimated year of birth	Date of collar installation	Location, noon on 31 July 2000
001	M	1996	02241999	-150.6432, 60.0567
002	F	1994	03311999	-149.9979, 59.9665
003	F	1991	04211999	-150.4639, 60.1245
004	F	1992	04211999	-150.4692, 60.1152

Fig. 1.4 Example of representation of geographic information as a table. The locations and attributes are for each of four grizzly bears in the Kenai Peninsula of Alaska. Locations, in degrees of longitude and latitude, have been obtained from radio collars. Only one location is shown for each bear, at noon on July 31, 2000.

Fields

While we might think of terrain as composed of discrete mountain peaks, valleys, ridges, slopes, etc., and think of listing them in tables and counting them, in practice there are unresolvable problems of definition for all of these objects. Instead, it is much more useful to think of terrain as a continuous surface in which elevation can be defined rigorously at every point. Such continuous surfaces form the basis of the other common view of geographic phenomena, known as the field view. The field view represents the real world as a finite number of variables, each one defined at every possible position.

Discrete objects are distinguished by their dimensions and naturally fall into categories of points, lines, and areas. Fields, on the other hand, can be distinguished by what varies and how smoothly. A field of elevation, for example, varies much more smoothly in a landscape that has been worn down by glaciation or flattened by blowing sand than one recently created by cooling lava. Cliffs are places in fields where elevation changes suddenly rather than smoothly.

Fields can also be created from classifications of land, into categories of land use or soil type. Such fields change suddenly at the boundaries between different classes. Other types of fields can be defined by continuous variation along lines rather than across space. Traffic density, for example, can be defined everywhere on a road network and flow volume can be defined everywhere on a river. Below is an example of field-like phenomena.

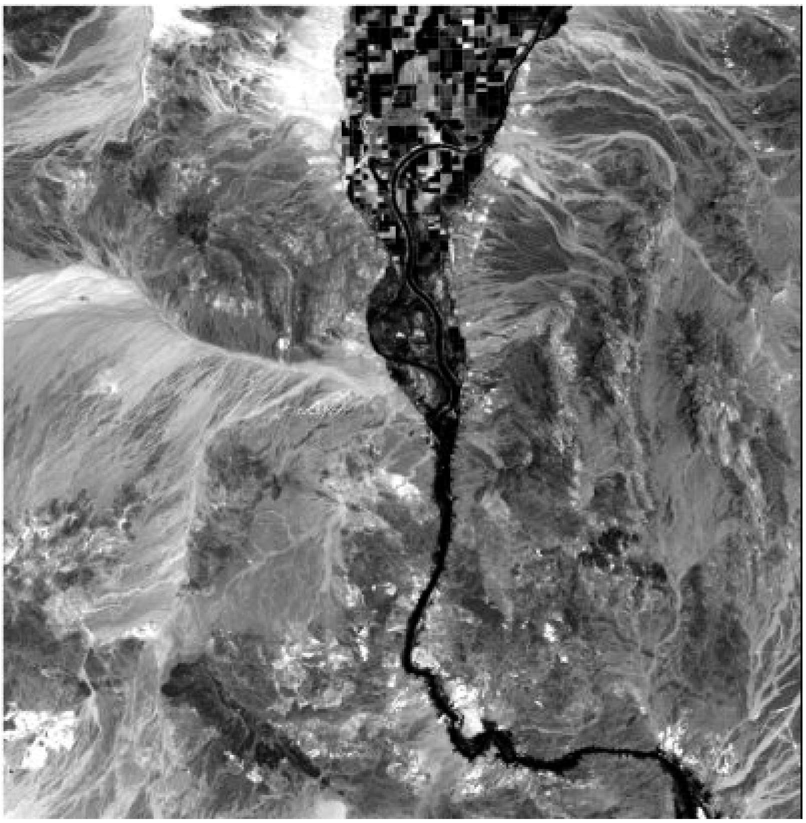


Fig 1.5 An image of part of the lower Colorado River in the southwestern USA. The lightness of the image at any point measures the amount of radiation captured by the satellite's imaging system. Image derived from a public domain SPOT image, courtesy of Alexandria Digital Library, University of California, Santa Barbara.

3.6
Rasters and Vectors

Fields and discrete objects define two different conceptual views of geographic phenomena, but neither solves every problem of digital representation. A field view potentially contains an infinite amount of information if it defines the value of the variable at every point, since there is an infinite number of points in any defined geographic area. Discrete objects can also require an infinite amount of information for full description—for example, a coastline contains an infinite amount of information if it is mapped in infinite detail. Thus fields and objects are no more than conceptualizations, or ways in which we think about geographic phenomena. They are not designed to deal with the limitations of computers.

Two methods of representing geographic data in digital form are raster and vector. In principle, both can be used to code fields and discrete objects, but in practice there is a strong association between raster and fields and between vector and discrete objects.

Raster data

In a raster representation, geographic space is divided into a rectangular array of cells, each of which is usually square. All geographic variation is expressed by assigning properties or attributes to these cells. The cells are sometimes called pixels (short for picture elements).

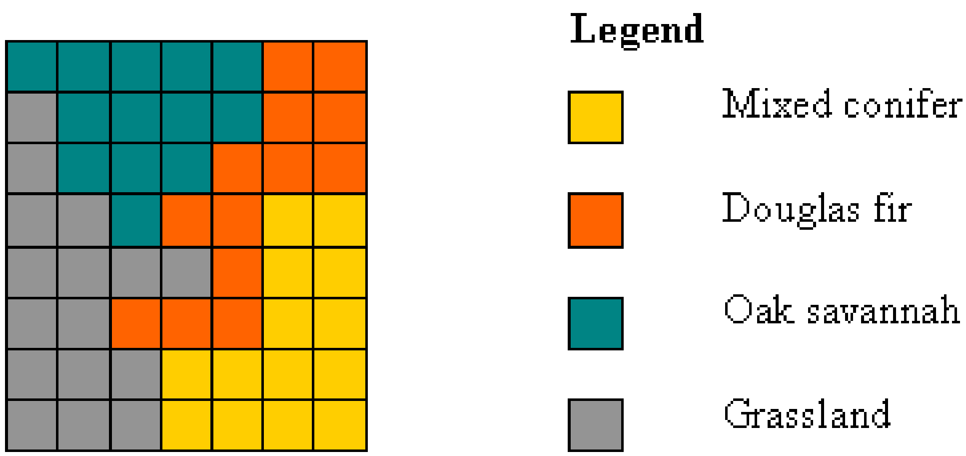


Fig. 1.6

One common form of raster data comes from remote sensing satellites. Data from the Landsat satellite, for example, which is commonly used in GIS applications, come in cells that are 30 meters on a side on the ground, or approximately one-tenth of a hectare in area. Remote sensing is a complex topic and further reading is available in Chapter 9 of the Longley et al. textbook, in the course GEO 444/544, and at web sites such as NASA's (www.nasa.gov).

When information is represented in raster form, all detail about variation within cells is lost and instead the cell is given a single value. Suppose we wanted to represent the map of the counties of Texas as a raster. Each cell would be given a single value to identify a county, and we would have to decide on a rule to apply when a cell falls in more than one county. Often, the rule is a simple plurality: the county with the largest share of the cell's area gets the cell. Sometimes the rule is based on the central point of the cell, and the county at that point is assigned to the whole cell. The graphic below shows these two rules in operation. The largest share rule is almost always preferred, but the central point rule is sometimes used in the interests of faster computing.

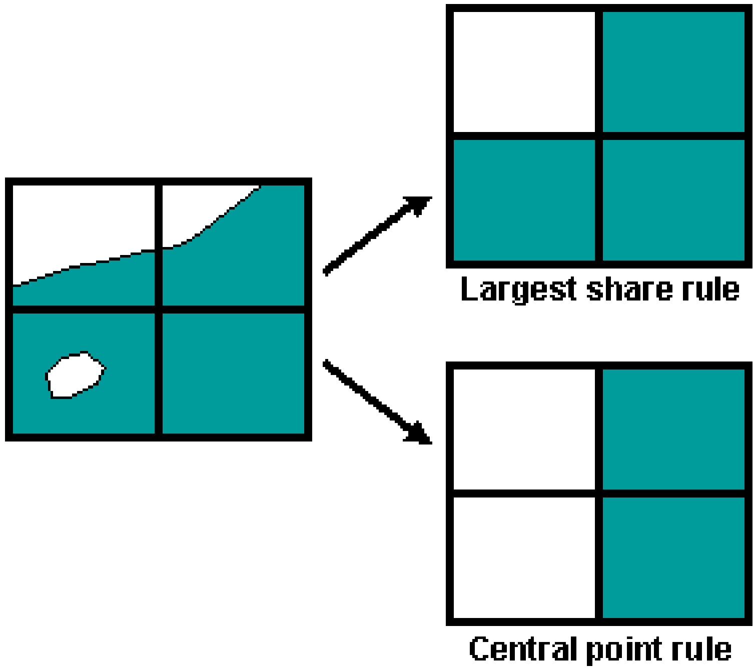


Fig. 1.7 Examples of the largest share rule, where a cell's value is on the value that occupies the largest share of the cell's area, and the central point rule, where a cell's value is based on the value that occupies the central point of the cell.

Vector data

In a vector representation, all lines are captured as points connected by straight lines (some GIS software allows points to be connected by curves rather than straight lines, but in most cases curves have to be approximated by straight lines). An area is captured as a series of points or vertices connected by straight lines as shown below. The straight edges between vertices explain why areas in vector representation are often called polygons, and the terms polygon and area are often used interchangeably. Lines are captured in the same way, and the term "polyline" has been coined to describe a curved line represented by a series of straight segments connecting vertices.

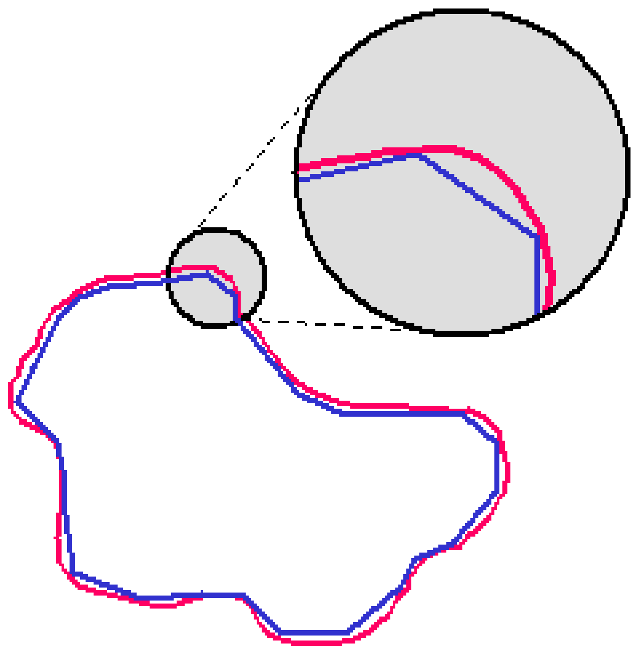


Fig. 1.8 An area (red line) and its approximation by a polygon (blue line).

To capture an area object in vector form, only the locations of the points that form the vertices of a polygon must be captured. This seems simple and also much more efficient than a raster representation, which would require us to list all the cells that form the area. To create a precise approximation to an area in raster, it would be necessary to resort to using very small cells and the number of cells would rise proportionately. But things are not quite as simple as they seem. The apparent precision of vector is often unreasonable, because many geographic phenomena simply cannot be located with high accuracy.

So, although raster data may look less attractive, they may be more honest to the inherent quality of the data. Also, various methods exist for compressing raster data that can greatly reduce the capacity needed to store a given data set. The choice between raster and vector is often complex, as shown in the table below.

Issue	Raster	Vector
Volume of data	Depends on cell size	Depends on density of vertices
Sources of data	Remote sensing, imagery	Social and environmental data
Applications	Resources, environmental	Social, economic, administrative
Software	Raster GIS, image processing	Vector GIS, automated cartography
Resolution	Fixed	Variable

Fig. 1.9 - Relative advantages of raster and vector

The nature of geographic data

5.2 Spatial Autocorrelation

The First Law of Geography, formulated by Waldo Tobler, states that everything is related to everything else, but near things are more related than distant things. Spatial autocorrelation is the formal property that measures the degree to which near and distant things are related.

Positive spatial autocorrelation occurs when features that are similar in location are also similar in attributes. Negative spatial autocorrelation occurs when features that are close together in space are dissimilar in attributes. Zero autocorrelation occurs when attributes are independent of location.

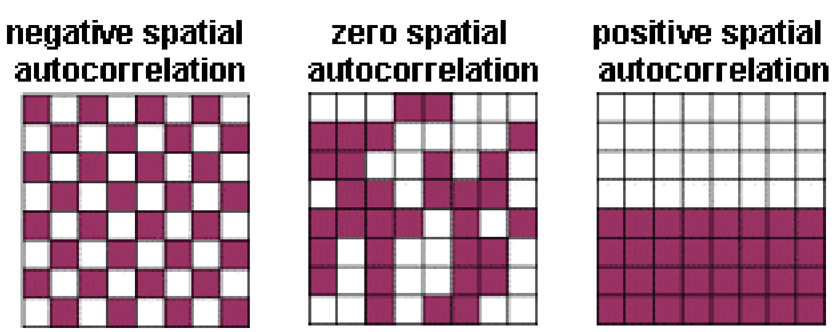


Fig. 1.11 Arrangements of dark and light colored cells exhibiting negative, zero, and positive spatial autocorrelation.

5.4 Spatial Sampling

The quest to represent the complex real world requires us to abstract, or sample, events and occurrences. For many purposes, geographic data are only as good as the sampling scheme used to create them.

You can think of sampling as the process of selecting points from a continuous field or, if the field has been digitized as a mosaic of objects, of selecting some of these objects while discarding others.

Classical statistics often emphasizes the importance of randomness in sound sample design. The purest form, simple random sampling, is well known: each element is assigned a unique number, and a specified number of elements are selected using a random number generator. In the case of a spatial sample from continuous space, x,y coordinates might be randomly sampled within the range of x and y values. Because each randomly selected element has a known probability of selection, it is possible to make robust and defensible generalizations to the population from which the sample was drawn.

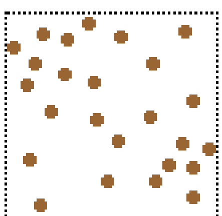


Fig. 1.12 - a spatially random sample

Randomly drawn elements, however, can be disproportionately concentrated among some parts of the population at the expense of others, particularly when the sample size is small relative to the population from which it was drawn. For example, a survey of household incomes might happen to select households with unusually low incomes. Spatially systematic (or stratified) sampling attempts to deal with this problem and ensure greater evenness of coverage across the sample area by identifying a regular sampling interval, which results in a regularly spaced grid.

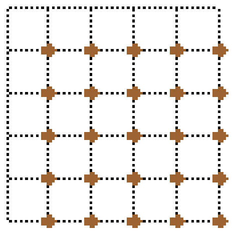


Fig. 1.13 - a spatially systematic (stratified) sample

Spatially systematic sampling also has weaknesses. Imagine that the grid pattern above were to coincide with the grid plan of a city. In a survey of urban land use, it is extremely unlikely that the attributes of street intersections would be representative of land uses elsewhere in the block structure. A number of hybrid sample designs have been devised to get around the vulnerability of spatially systematic and random sampling. These include stratified random sampling to ensure evenness of coverage and periodic random changes in the grid width of a spatially systematic sample, perhaps subject to minimum spacing intervals.

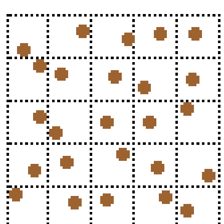


Fig. 1.14 - a stratified random sample

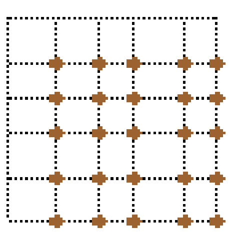


Fig. 1.15 - a sampling scheme with periodic random changes in the grid width of a spatially systematic sample

Simple random or spatially systematic sampling presumes that each observation is of equal importance in building a representation. Many times this is not the case, and it may be more efficient and necessary to devise application-specific sample designs to improve quality of representation, while minimizing resource costs of collecting data.

5.5 Spatial Interpolation

Spatial interpolation is the process of filling in the gaps between sample observations. It requires an understanding of the attenuating effect of distance between sample observations and selection of an appropriate interpolation function. This concept focuses on principles that are used to describe effects over distance. For an introduction to some of the mathematical functions used to describe these effects, refer to Section 5.5 in the Longley et al. text.

A literal interpretation of Tobler's Law implies a continuous, smooth, attenuating effect of distance upon the attribute values of adjacent or contiguous spatial objects, or incremental variation in attribute values as we traverse a field. For example, the polluting effect of a chemical or oil spill decreases in a predictable (and in still waters, uniform) fashion with distance from the source of the spill, aircraft noise pollution decreases with distance from the flight path, and the number of visits to a national park might decrease at a regular rate with distance from the park.

The literal interpretation of Tobler's Law may be appropriate for many applications. The notion of smooth and continuous variation underpins many of the representational traditions in cartography, such as the creation of maps showing contour lines. Contour lines connect points with equal elevation above sea level.

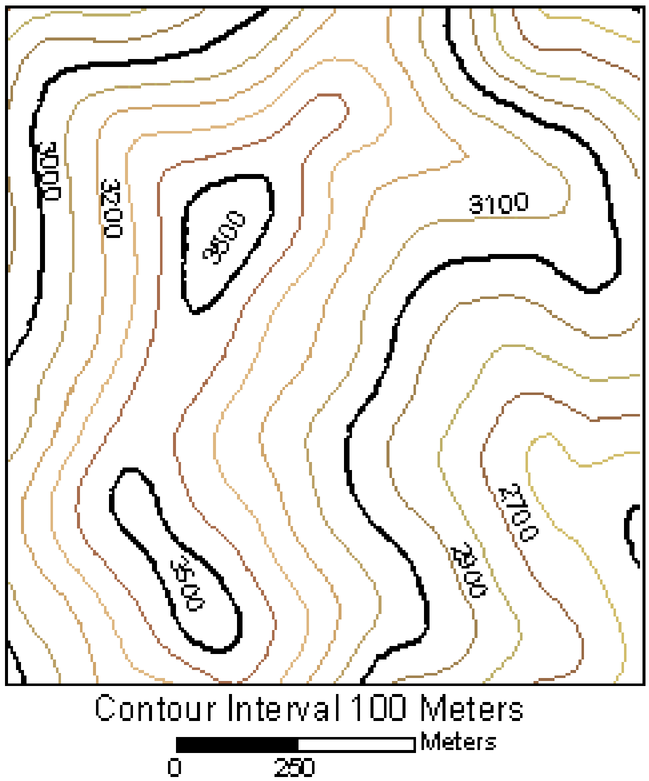


Fig. 1.16 Each contour line in the graphic represents the same elevation. Notice that the lines are relatively smooth.

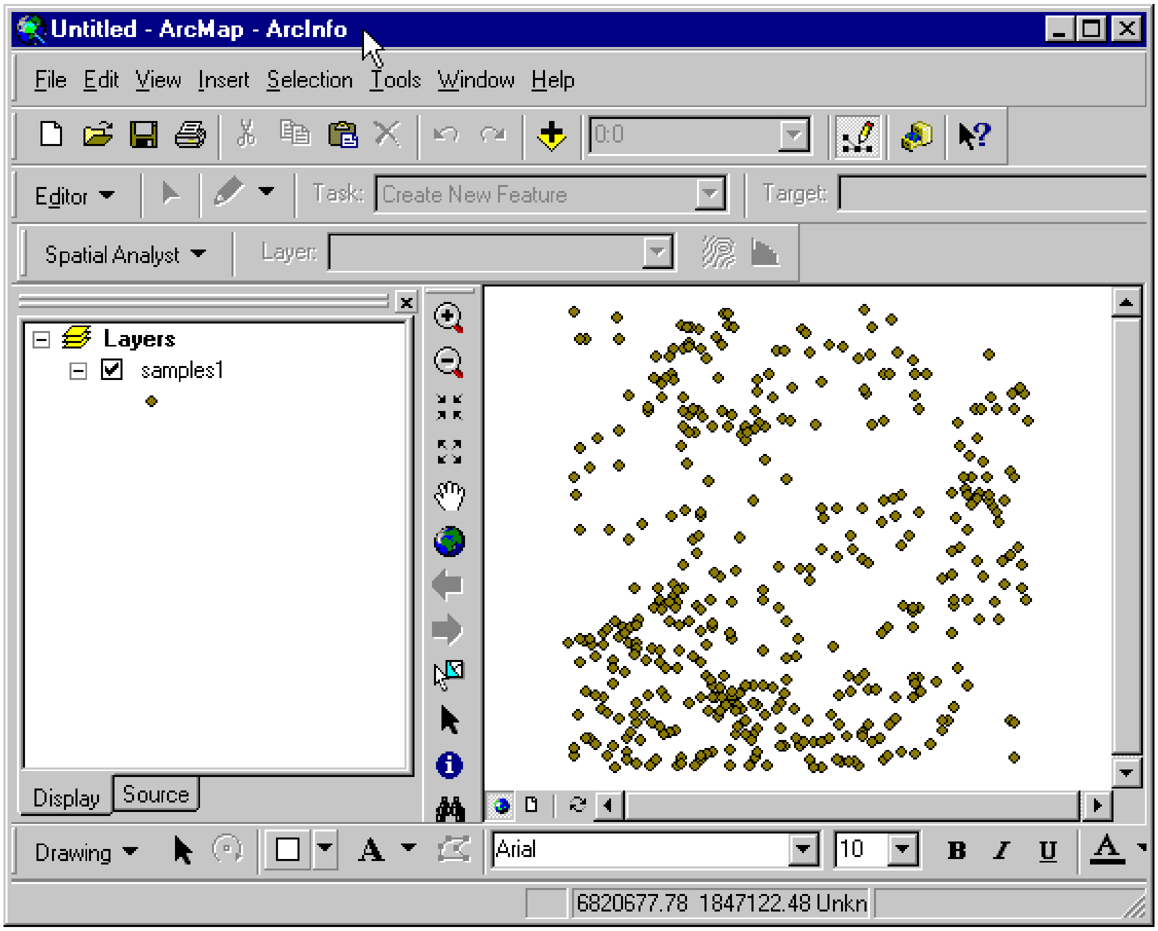
Our understanding of spatial structure, however, often tells us that variation is not smooth and continuous. In such circumstances, the true nature of geographic data might be better represented using other interpolation methods and functions. Thus, we need to adapt our thinking to accommodate discontinuities in physical (e.g., fault lines or cliffs) and socioeconomic distributions (e.g., shifts in household income distributions on crossing the U.S. - Mexico border).

EXAMPLE

Explore how sampling scheme can affect spatial interpolation:

In this exercise, you will see that representative sample points are important for interpolation. You will also experimented with spatial interpolation, using the inverse distance weighted interpolation method.

add a layer of sample points.



1.1

examine the attributes for the sample points.

Attributes of samples1

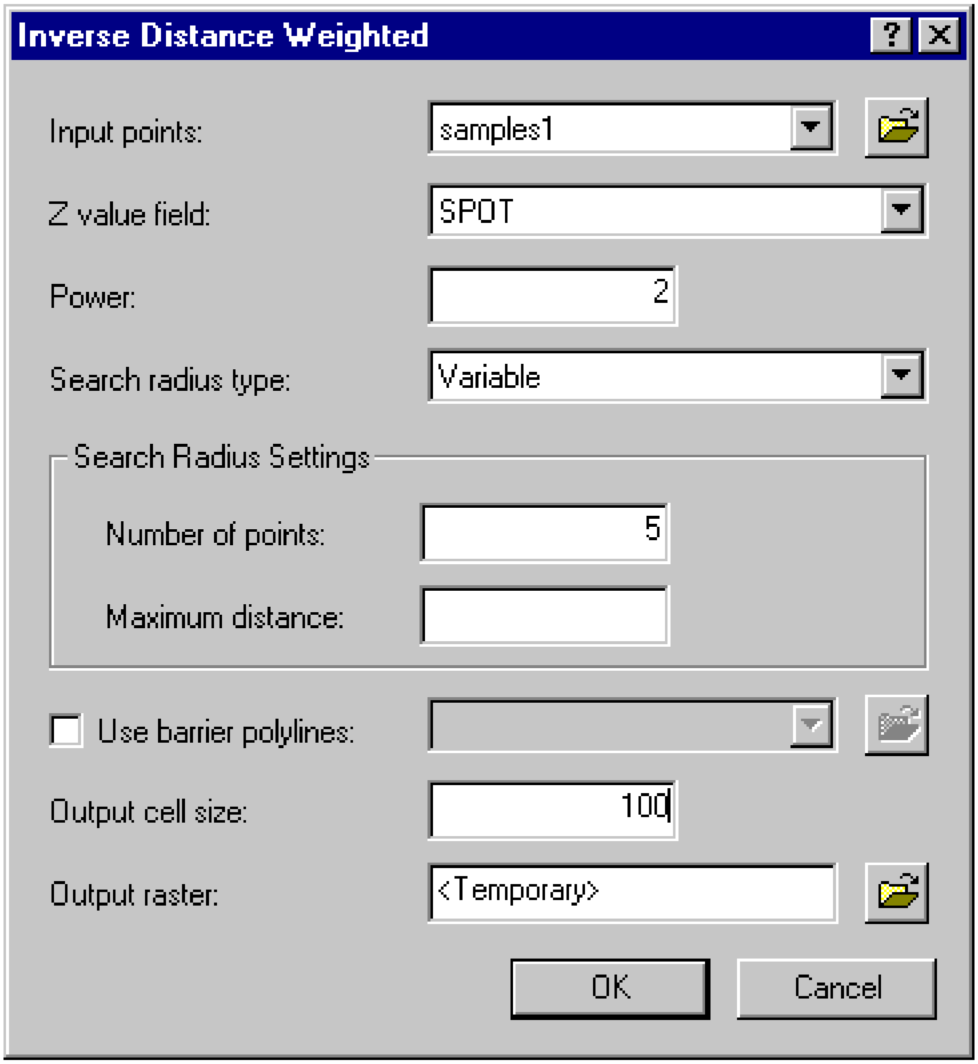
FID	Shape	VIPCOV_ID	SPOT
0	Point	1	2069.617
1	Point	2	2071.096
2	Point	3	1873.528
3	Point	4	1628.654
4	Point	5	1929.667
5	Point	6	1570.301
6	Point	7	2068.834
7	Point	8	2100.807
8	Point	9	2108.473
9	Point	10	2180.691
10	Point	11	1960.529
11	Point	12	1959.872

Record: 1 Show: All Selected Records

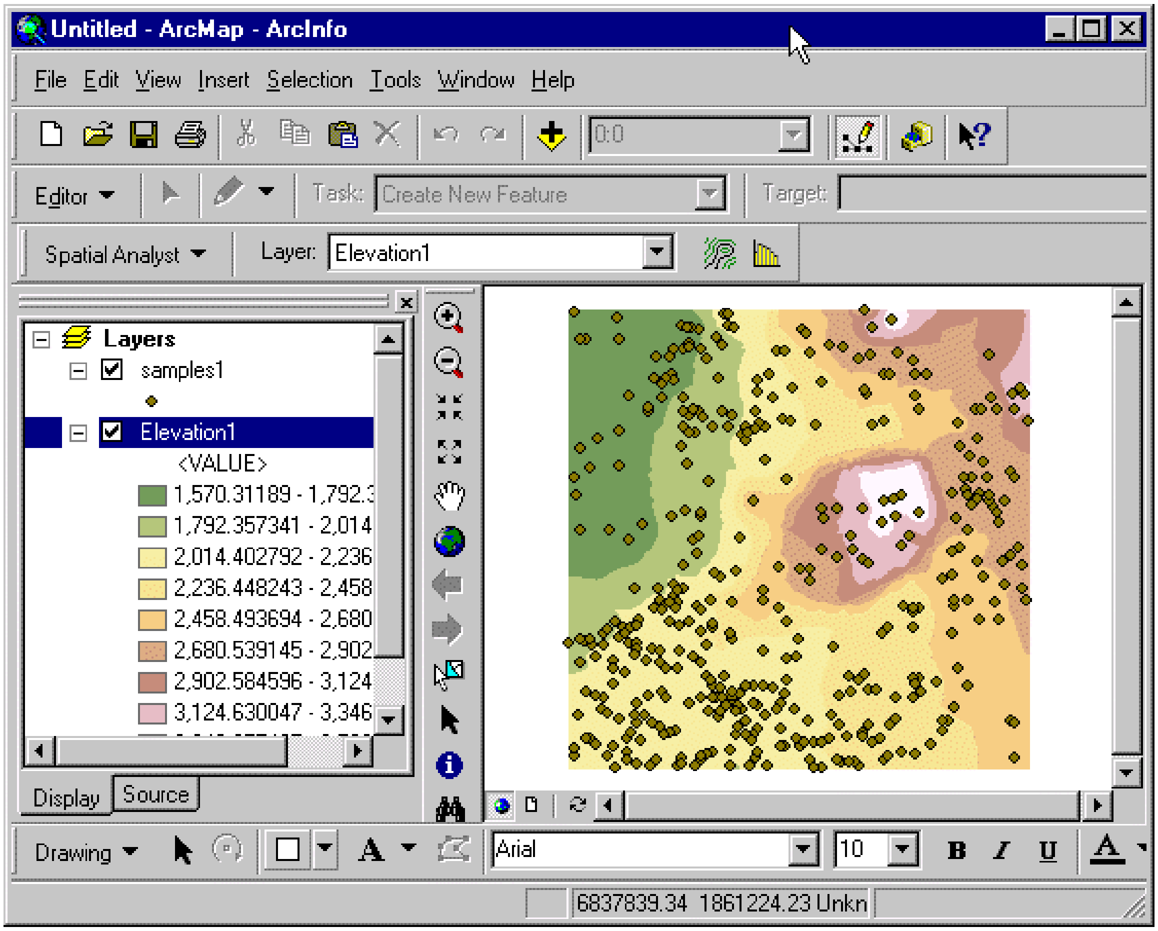
1.2

use the sample points to interpolate an elevation surface. You'll use the Inverse Distance Weighted (IDW) interpolation method. This method is a common method and follows Tobler's Law by interpolating values as weighted averages over the known measurements at nearby points, giving the greatest weight to the nearest points.

Use Number of Points = 5, Output cell size = 100


Inverse Distance Weighted
Input points:
samples1
Z value field:
SPOT
Power:
2
Search radius type:
Variable
Search Radius Settings
Number of points:
5
Maximum distance:
Use barrier polylines:
Output cell size:
100
Output raster:
<Temporary>
OK
Cancel

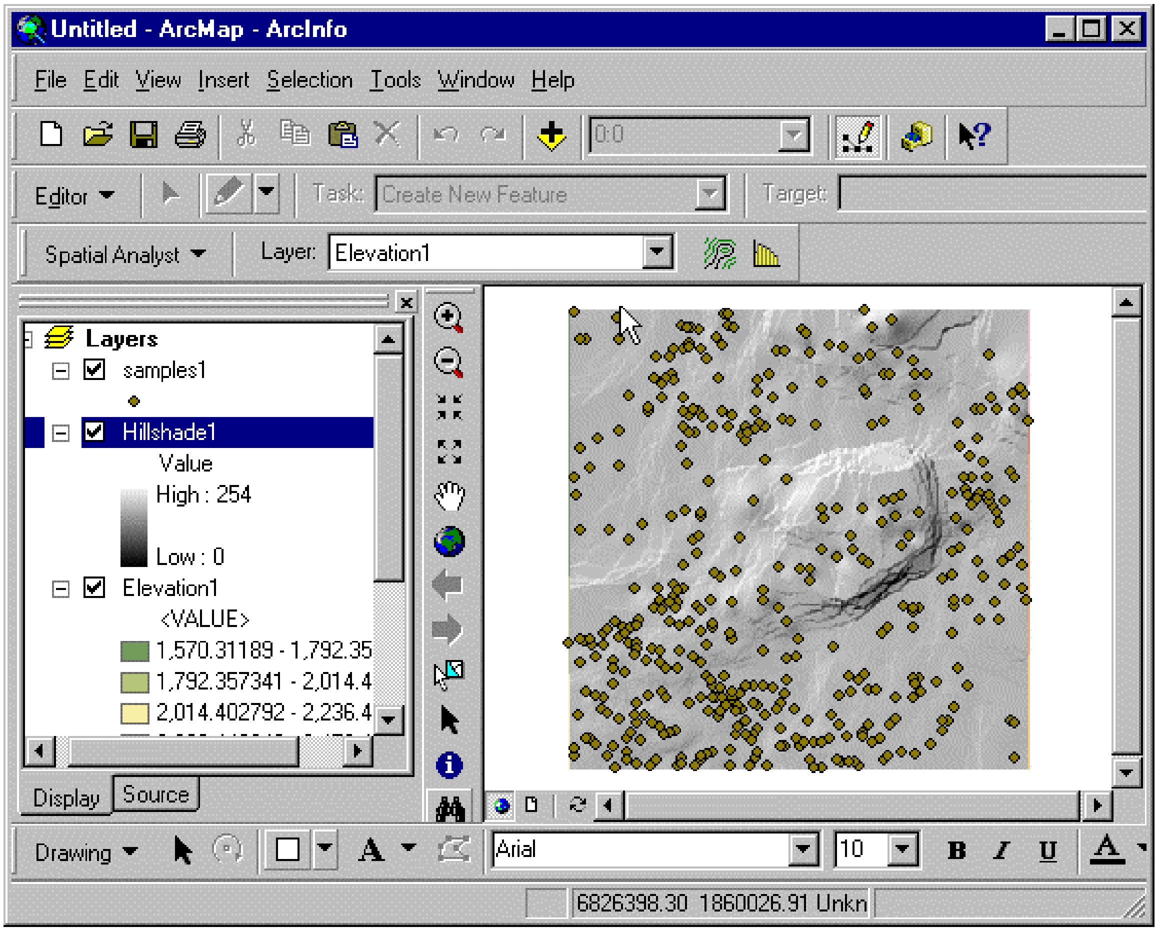
1.3



1.4

Another useful way to visualize a surface is as a hillshade. A hillshade shows hypothetical illumination of a surface.

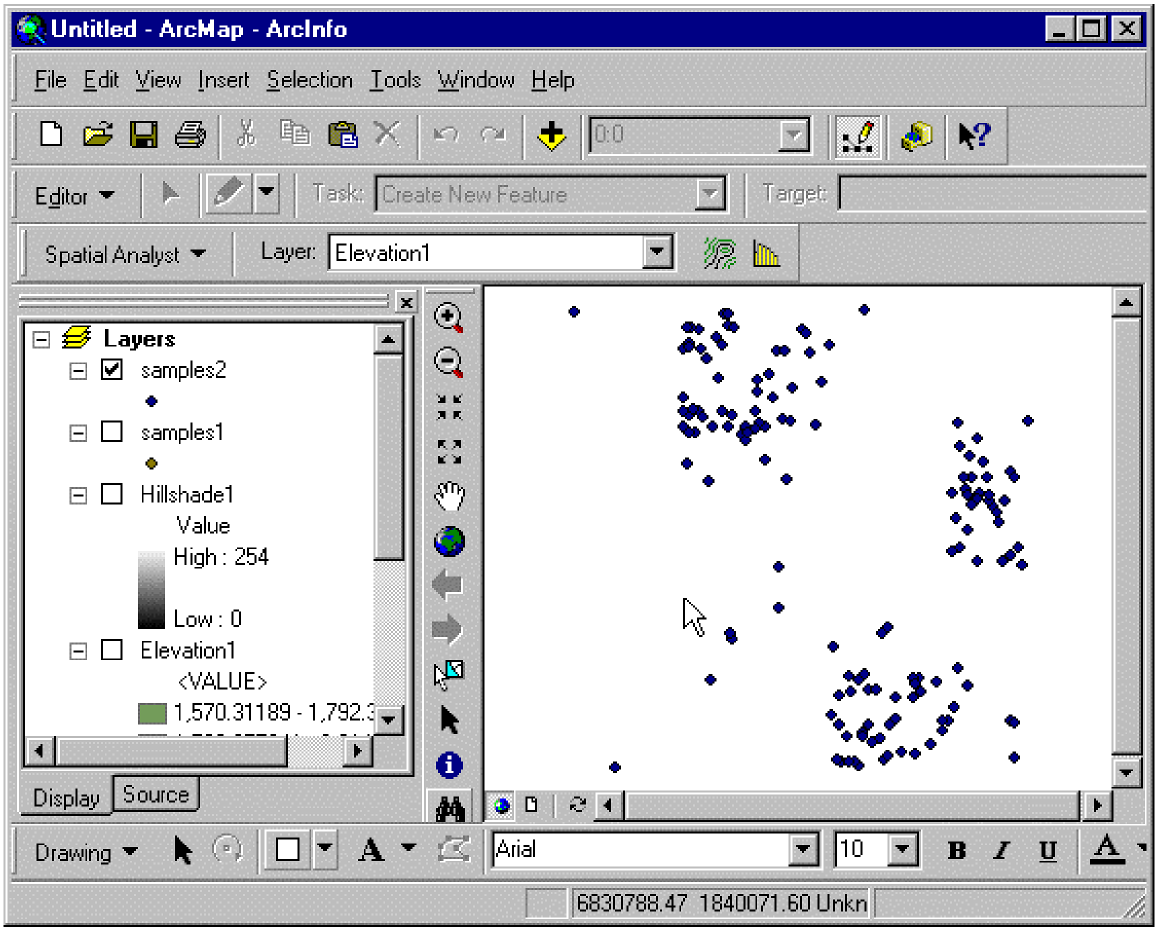
The hillshade should help you visualize the terrain. There is one big hill near the center with a smaller hill above it.



1.5

Add another set of sample points

Now you'll compare your results with those obtained using observations from a different sample survey.



1.6

Notice that the second set of sample observations is not spread evenly across the sample area. Some areas have clusters of observations, while others are represented by just a few points.

The geography of this particular sampling scheme also bears little correspondence to the local variability of topography—when it would make much more sense to sample more points in parts of the map where there are abrupt changes of slope. In fact, this set of observations has been deliberately chosen to illustrate how unrepresentative samples can present a distorted picture of geographic reality.

Next, examine the attributes for this set of sample points

Attributes of samples2

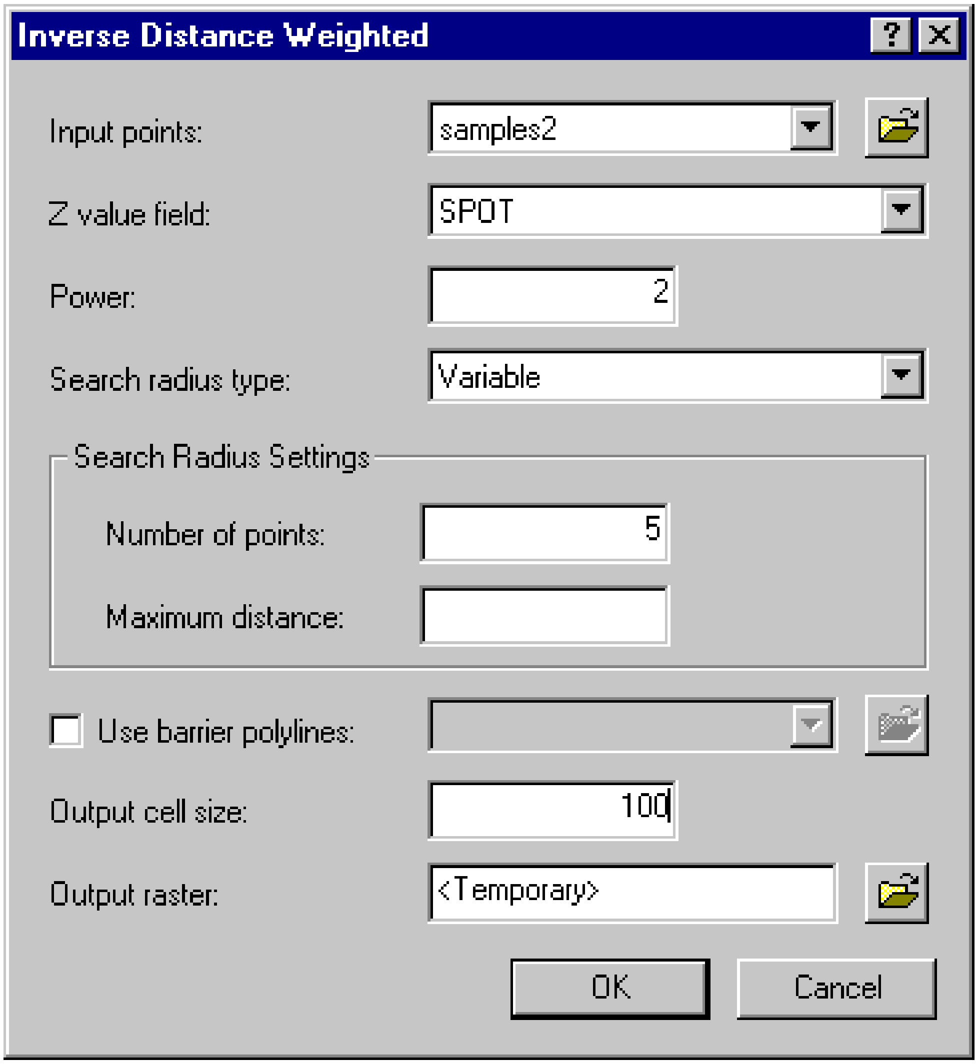
FID	Shape*	VIPCOV_ID	SPOT
0	Point	13	2442.666
1	Point	14	2562.508
2	Point	15	2411.133
3	Point	16	2731.217
4	Point	17	2796.279
5	Point	18	2743.069
6	Point	19	1915.177
7	Point	20	2019.58
8	Point	21	2029.279
9	Point	22	2378.975
10	Point	23	2344.069
11	Point	24	2354.487

Record: 1 Show: All Selected Records (0 ou

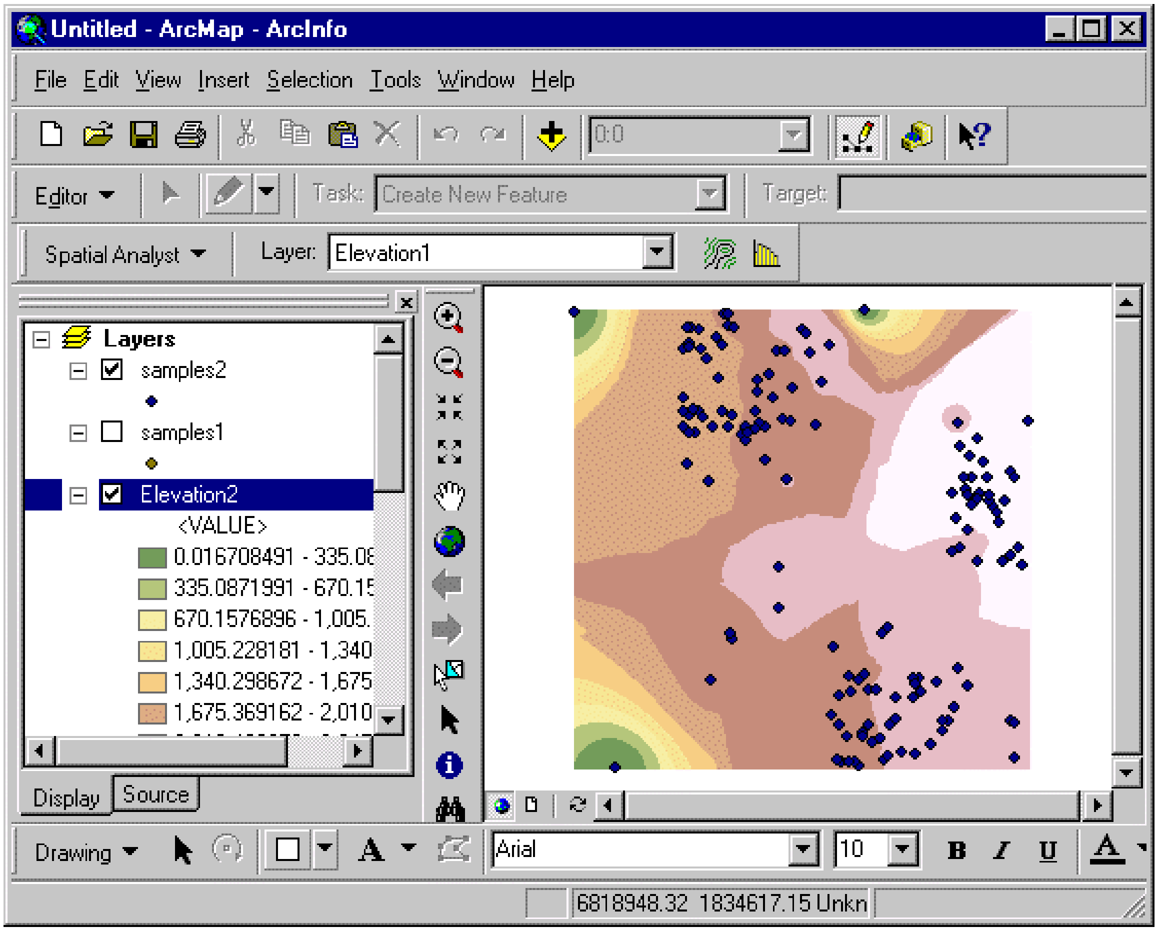
1.7

Interpolate second set of elevation values

Use IDW again with same values


Inverse Distance Weighted
Input points:
samples2
Z value field:
SPOT
Power:
2
Search radius type:
Variable
Search Radius Settings
Number of points:
5
Maximum distance:
Use barrier polylines:
Output cell size:
100
Output raster:
<Temporary>
OK
Cancel

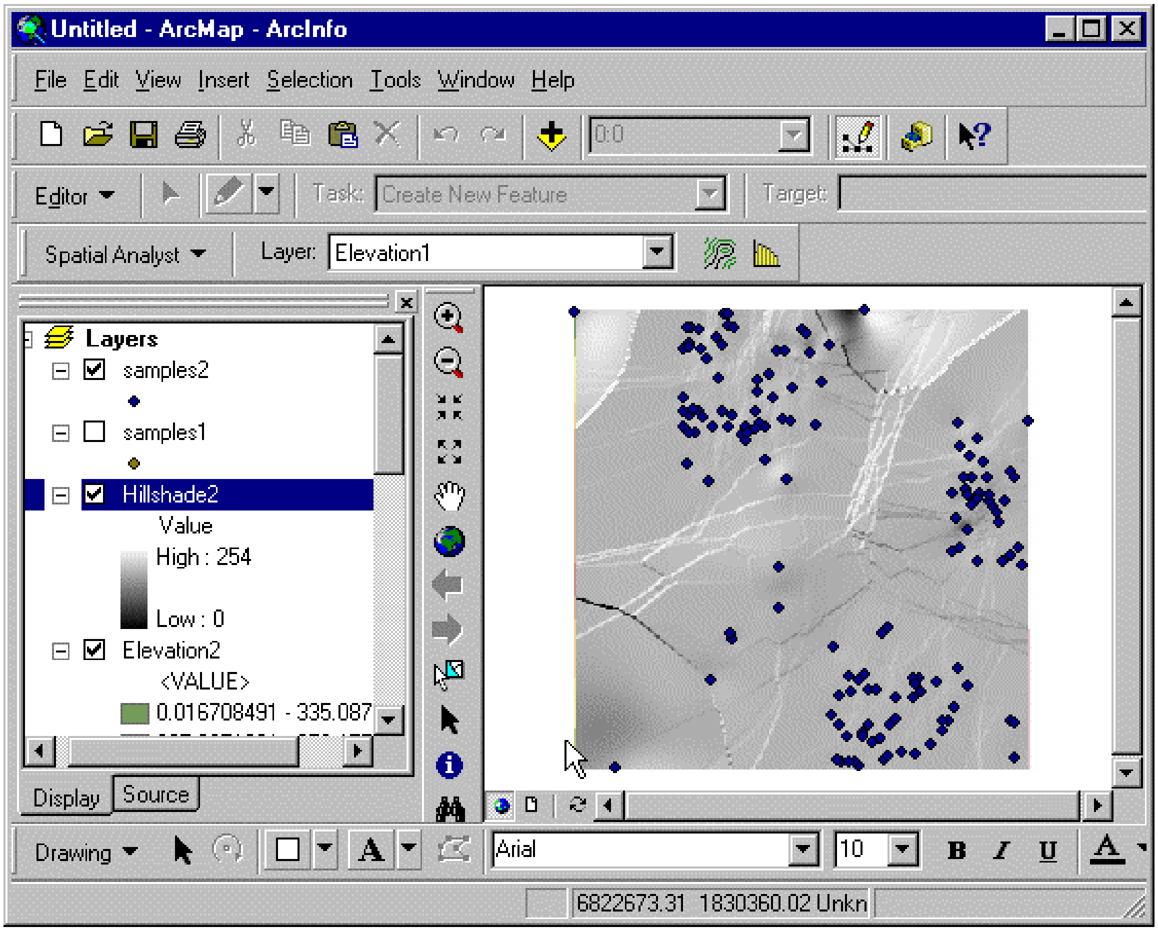
1.8



1.9

Examine the Elevation2 layer. Areas shown in white are high elevation areas, and areas shown in green are low elevation areas. Notice that the lighter colored areas (higher elevation) are toward the right side of the layer.

You'll be able to see the differences between the results from the two samples more easily by comparing hillshade layers.



1.10

Notice that there is not a distinct hill near the center of Hillshade2.

Optional EXAMPLE:

Comparing the discrete object and field views for lakes in Minnesota

This simple example illustrates the difference between the discrete object and field conceptualizations.

Suppose you were hired for the summer to count the number of lakes in Minnesota and told that your answer would appear on every license plate issued by the state. The task sounds simple and you were happy to get the job. But on the first day you started to run into difficulty. What about small ponds, do they count as lakes? What about wide stretches of rivers? What about swamps that dry up in the summer? What about a lake with a narrow section connecting two wider parts, is it one lake or two?

Your biggest dilemma concerns the scale of mapping, since the number of lakes shown on a map clearly depends on the map's level of detail—a more detailed map almost certainly will show more lakes.

Your task reflects a discrete object view of the phenomenon. The action of counting implies that lakes are discrete, two-dimensional objects littering an otherwise empty geographic landscape. In a field view, on the other hand, all points are either lake or non-lake. Moreover, we could refine the scale a little to take account of marginal cases; for example, we might define the scale shown in the table below, which has five degrees of lakeness.

Lakeness	Definition
1	Location is always dry under all circumstances
2	Location is sometimes flooded in Spring
3	Location supports marshy vegetation
4	Water is always present to a depth of less than 1m
5	Water is always present to a depth of more than 1m

Fig 1.10

the complexity of the view would depend on how closely we looked, of course, and so the scale of mapping would still be important.

But all of the problems of defining a lake as an object would disappear (though there would still be problems in defining the levels of the scale). Instead of counting, our strategy would be to lay a grid over the map and assign each grid cell a score on the lakeness scale.

The size of the grid cell would determine how accurately the result approximated the value we could theoretically obtain by visiting every one of the infinite number of points in the state.

At the end, we would tabulate the resulting scores, counting the number of cells having each value of lakeness or averaging the lakeness score. We could even design a new and scientifically more reasonable license plate—"Minnesota, 12% Lake" or "Minnesota, Average Lakeness 2.02."

Summary

Geographic data link place, time, and attributes and geographic representation in digital form is becoming increasingly important in our society. Digital representations have enormous advantages over paper maps, written reports, or spoken accounts, but there are problems that must be resolved when representing geographic data digitally. Special attention must be given to the questions of what to represent and how to represent that which is being represented.

How to represent geographic data

There are two fundamental ways of conceptualizing geographic data: discrete objects and fields. Discrete objects represent the world as objects with well-defined boundaries in empty space. Fields represent the world as a continuous surface, in which a variable is defined at every possible position.

There are also two fundamental ways of digitally representing geographic data: raster and vector. In principle, both can be used for fields and discrete objects, but in practice there is a strong association between raster and fields and between vector and discrete objects.

In a raster representation, geographic space is divided into cells that have properties associated with them. In a vector representation, features are captured as points, which are connected to form lines or boundaries around areas.

What geographic data to represent

The world is infinitely complex, but based on Tobler's Law, you can use representational assumptions (such as inverse distance weighting) to infer information about the gaps between sample observations. Tobler's Law states that everything is related to everything else, but near things are more related than distant things. Spatial autocorrelation is the formal property that measures the degree to which near and distant things are related. Spatial autocorrelation can be positive (features similar in location are also similar in attributes), negative (features that are similar in location are dissimilar in attributes), or neutral (attributes are independent of location).

When selecting sample points, it is important to select representative samples. Classical statistics emphasizes the importance of random samples, but random samples don't always work well with geographic data. Many times you may need to use a systematic or application-specific sampling scheme.

Spatial interpolation is the process of filling in the gaps between sample points. By a strict interpretation of Tobler's Law, distance has a smooth and continuous effect on the values of attributes between sample observations. Although this is appropriate for many applications, there are circumstances in which variation is not smooth and continuous. In these circumstances, other interpolation methods may be necessary.