Spatial Analysis with GIS  
Lecture 1: Introduction

GIS期刊Important GIS-related journals

* International Journal of Geographical Information Science
* Transactions in GIS
* Landscape and Urban Planning
* Computers, Environments and Urban Systems
* Annals of the American Association of Geographers
* Applied Geography
* International Journal of Applied Earth Observation and Geoinformation
* International Journal of Digital Earth

GIS and spatial analysis  
- Brief introduction of concepts

Roger Tomlinson (1933-2014): "father of GIS”

GIS概念

GIS: In the strictest sense, a GIS is **a computer system** capable of **assembling, storing, manipulating, and displaying geographically referenced information**, i.e. data identified according to their locations. Practitioners also regard the total GIS as including operating personnel and the data that go into the system （USGS）

GIScience: the generic, fundamental issues that surround the use of GIS technology, impede its successful implementation, or emerge from an understanding of its potential capabilities. **It involves both research on GIS and research with GIS** (Goodchild, 1992).

* is the science behind the technology
* considers fundamental questions raised by the use of systems and technologies
* is the science needed to keep technology at the cutting edge
* is a multidisciplinary field

空间分析概念

Spatial analysis or spatial statistics includes any of the formal techniques which study entities using their topological, geometric, or geographic properties.

The results depend on locations of objects being analyzed.

Requires attributes and locations, which are stored by GIS.

Michael F. Goodchild (National Center for Geographic Information and Analysis)

* + A set of techniques for analyzing spatial data
  + used to gain insight as well as to test models
  + ranging from inductive to deductive
    - finding new theories as well as testing old ones
  + can be highly technical or mathematical
    - can also be very simple and intuitive

Other definitions:

"A set of techniques requiring access both to the locations of objects and also to their attributes"

* + requires methods for describing locations (i.e. a GIS)
  + some techniques do not look at attributes
  + mapping is a form of spatial analysis?
  + move the objects, and the results change
    - e.g. move the people, and the US Center of Population moves
    - e.g. move the people, and average income does (not) change
  + most statistical techniques are invariant under changes of location
    - compare the techniques in SAS, SPSS, Systat etc.
* Traditionally, geographic information systems are considered to perform four basic functions on spatial data: input, storage, analysis and output (Goodchild, 1987).
* Typically, **a variety of map description and manipulation functions are defined by commercial vendors** as being "spatial analysis," but this has little to do with the usual interpretation of the concept in the academic world.
* There, **the emphasis is on spatial modeling and statistical analysis**, e.g., as illustrated in the review in Fischer and Nijkamp (1992).
* The analysis function of a GIS was further divided into four components in Anselin and Getis (1992):
  + selection (sampling of data from the data base)
  + manipulation (partitioning, aggregation, overlay, buffering, and interpolation)
  + exploration
  + confirmation
* The first two of these are typical features of existing GIS
* Methods for working with spatial data “A set of methods whose results change when the locations of the objects being analyzed change”.
  + to detect patterns, anomalies
  + to find answers to questions
  + to test or confirm theories
    - deductive reasoning
  + to generate new theories and generalizations
    - inductive reasoning
* Methods for adding value to data
  + in doing scientific research
  + in trying to convince others
* Turning raw data into useful information
* A collaboration between human and machine
  + Human directs, makes interpretations and inferences
  + Machine does tedious, complex stuff

空间分析的目的Goals of spatial analysis

* Identifying spatial patterns
* Identifying spatial objects that are potential generators of patterns
* Identifying information relevant for explaining the spatial pattern (and hiding irrelevant information)
* Presenting the information in a way that is intuitive and supports further analysis

Types of spatial analysis  
- Terms, methods and tools

Types of spatial analysis include:

* Query analysis
* Spatial distribution
* Geometric analysis
* Buffer analysis
* Overlay analysis
* Network analysis
* Surface analysis
* Spatial interpolation
* Spatial statistics analysis
* DEM analysis…

**Basic categories (Goodchild):**

1. Queries & Reasoning

Most basic – simple questions posed by user

No changes occur to database, no new data created

2. Measurements

Numerical values that describe aspect of data

Length, area

3. Transformations

Spatial analyses that change datasets, combine, compare

Convert raster ←→ vector

4. Descriptive summaries

Centers dispersion

Spatial dependence

Fragmentation

5. Optimization

Best routes

Best locations

6. Hypothesis testing

Inference from sample to population

**1. Queries & Reasoning**

Queries

* Request for information by:
  + typed question
  + menu command
  + mouse click
* Based on database views:
  + catalog view

for Physical structure of a database, such as

1. storage devices 2. files (shapefiles, tables)  
Query for metadata & information about file properties.   
 last modified? coordinate system? text or numeric field?

* + map view

Simple queries for:

X, Y coordinates of a location

Distance between two locations

Attributes of an object

* + table view

Query tables with a special syntax that uses logical and arithmetic operators. eg. SQL (Structured Query Language) .

The query result would select records in the table that met both conditions. If the table view were linked to a map view, the parcels would also be selected on the map.

* + Histogram

Compares attribute values to their relative frequencies, count of how many records in the table fall within equal intervals of the attribute value range.

Eg. Shows whether a dataset has a normal or skewed distribution

Pie Chart: Each slice of pie represents the population of one country in proportion to the population of all

* + scatterplot .

Compares one attribute value (x-axis) to another (y-axis).

Distribution of the points shows whether there is a fixed relationship between the attributes.

* + linked views

Exploratory Data Analysis - Gain insight by working with dynamically linked views. Link is a dynamic relationship between two views of data in which a query made in one view returns information from both views.

* Reasoning with GIS

We spend our lives in the vague world of human discourse

"is Santa Barbara north of LA?"

a GIS needs to know exactly what is meant by "north of“

is Reno east or west of San Diego?

we tend to think of the US as a square, with two N-S coasts

how to design a GIS to provide driving directions?

to direct people through airports?

a GIS would be easier to use if could "think" and "talk" more like humans

or if there could be smooth transitions between our vague world and its precise world

in our vague world, terms like "north of" are context-specific

geographically relevant terms like "across" or "in" have many meanings

**2. Measurement**

* Simple geometric measurements associated with objects **area**, **distance**, **length**, **perimeter**, **shape**
* The **ability make measurements between map features** is the main reason that early GIS development was funded by the Canadian government.
* Such measurements are tedious and inaccurate if made by hand, instead, using GIS and digital databases is fast, reliable, and accurate

Measurement of distance and length

* Calculation from metric coordinates
* Straight-line distance on a flat plane : Pythagorean Distance
* Distance on a spherical globe

Given their latitudes and longitudes

The Great Circle gives the shortest distance between points on spherical earth

Measuring distance

1.Between points

Euclidian distance: D

Manhattan distance: a+b

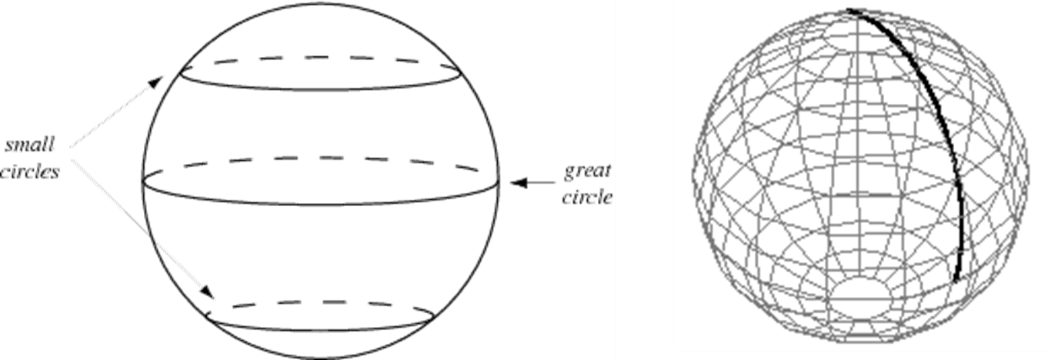
Network distance: a+b+c+d

2.Between Objects:

Shortest distance

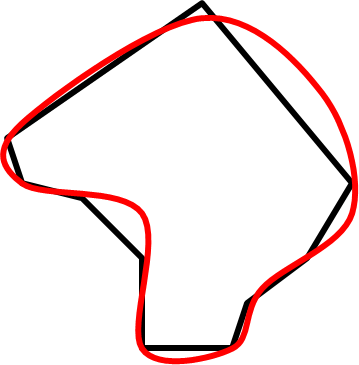
3.Distance on a spherical globe

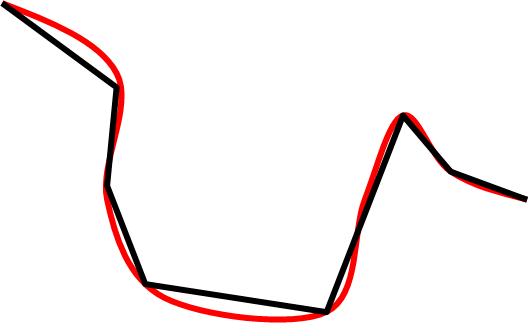
* From (*lat*1,*long*1) to (*lat*2,*long*2) *R*  is the radius of the Earth, roughly 6378 km
* *d* = *R* arccos [sin *lat*1 sin *lat*2 + cos *lat*1 cos *lat*2 cos (*long*1 - *long*2)]



Measurement of perimeter

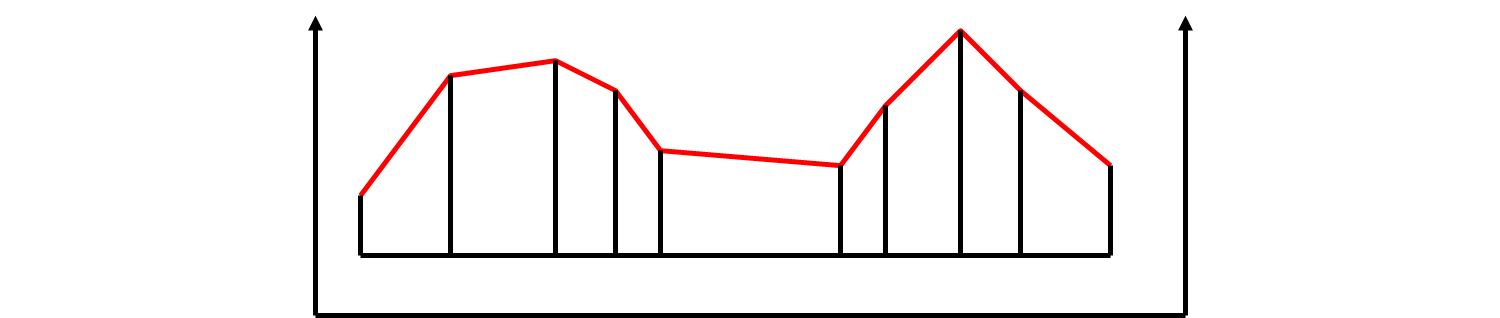
* Algorithm used for the Perimeter function is based on Dana Tomlin's algorithm in "Geographic Information Systems and Cartographic Modeling"
* Add the lengths of polyline or polygon segments



**distortions **

* If segments are straight, length will be underestimated in general:

The length of a true curve is almost always longer than the length of its polyline or polygon representation

* Measurements in GIS are often made on horizontal plane
  + length and area may be substantially lower than on a true three-dimensional surface
  + underestimated in hilly areas
  + applies to surveyed land also

Measurement of area

* Proceed in clockwise direction around the polygon

For each segment

drop perpendiculars to the *x* axis

compute the area of the trapezium :difference in *x* times average of *y*

keep a cumulative sum of areas

At the end, the sum will be the area of the polygon

* The algorithm will fail when:
  + Islands must all be scanned clockwise
  + Holes must be scanned anticlockwise
    - Holes have negative area
  + First should make Y positive
  + Because of limited computer precision
    - Results could be wrong if the area is very small and the coordinate values are very large , e.g. in UTM or SPC
    - Need double precision for calculations but not for results

Measurement of shape

* A compact shape has a small perimeter for a given area
* Compare perimeter to the perimeter of a circle of the same area
* Shape = perimeter / [3.54 sqrt (area)]

Usage of shape measures

* Gerrymandering
  + Creating oddly shaped districts to manipulate the vote
  + Named for Elbridge Gerry
  + Today GIS is used to design districts
* Other types of districts designed
  + Administrative regions
  + Sales districts

Modifiable Area Unit Problem (MAUP)

* The MAUP is a source of statistical bias that can radically affect the results of statistical hypothesis tests. It affects results when point-based measures of spatial phenomena (e.g., population density) are aggregated into districts. The resulting summary values (e.g., totals, rates, proportions) are influenced by the choice of district boundaries.
* For example, census data may be aggregated into census enumeration districts, or postcode areas, or police precincts, or any other spatial partition (thus, the "areal units" are "modifiable").

**3.Transformations**

* Transform GIS data into useful products.
* Reveal aspects of data that are not immediately obvious.
  1. Buffering
  2. Point in polygon
  3. Polygon overlay
  4. Spatial Interpolation
  5. Density estimation

Discrete objects and fields：

Two ways of conceptualizing geographic variation

The most fundamental distinction in geographic representation

Discrete objects：

* Points, lines, and areas
* Objects with well-defined boundaries
* Countable
  + Biological organisms
  + Trees
  + Human-made objects
  + houses, fire hydrants

Field：

* Properties that vary continuously over space
* Archetype--Elevation: A single value at every point on the Earth’s surface

Atmospheric temperature, pressure

* Any field can have slope, gradient, peaks, pits

Buffers

* Common GIS operation
* Encloses point, line, or polygon at specified distance.
* New spatial objects created.

Point in Polygon

* Determine whether a point lies inside or outside a polygon

Generalization: assign many points to containing polygons

Used to assign crimes to police precincts, voters to voting districts, accidents to reporting counties

Points are assigned the attributes of the polygons within which they fall

Polygon overlay

Stamps outlines of polygons in one layer onto polygons in another layer.

Can be applied with discrete objects or fields.

* Two cases: for discrete objects and for fields
* Discrete object case:

Find polygons formed by the intersection of two polygons

When discrete polygons are overlaid, new spatial objects are created by the lines of intersection. The new objects may have the attributes of either or both of the original polygons.

When polygons that make up fields are overlaid, new spatial objects are again created from areas of intersection. The new objects have the attributes of both layers.

Spurious or Sliver Polygons

In any two such layers there will almost certainly be boundaries that are common to both layers e.g. following rivers

The two versions of such boundaries will not be coincident

As a result large numbers of small sliver polygons will be created

* + these must somehow be removed
  + this is normally done using a user-defined tolerance

Spatial interpolation

* The estimation of unknown values on a surface from a sample of known values.
* Can be used to estimate such as Elevation, Rainfall, Temperature, etc...
* Underlying principle - Tobler's Law: all places are related but nearby places are more closely related than distant places
* Interpolation is commonly a **raster operation**, but it can also be done in a **vector** environment using a TIN surface model.
* **The better the input data is, the better the results will be.**

IDW (Inverse distance weighting)

* Most common method of spatial interpolation. Not the most accurate method, but easiest to understand
* The unknown value of a field at a point is estimated by taking an average over the known values
  + Weighting each known value by its distance from the point, giving greatest weight to the nearest points
  + An implementation of Tobler’s Law

Kriging

* A technique of spatial interpolation firmly grounded in geostatistical theory
* Like IDW, Kriging estimates unknown values from a weighted average of nearby sample points.
* In Kriging, however, the sample point weights are determined statistically.
* Kriging can be used to interpolate surfaces that have trends (hillsides, for example) or directional influence in the data.

**4. Descriptive Summaries**

* Attempt to summarize properties of data sets in one or two statistics
* Descriptive Summaries
  1. Centers: Measures of central tendency and dispersion
  2. Dispersion: Histograms, pie charts, and scatterplots
  3. Spatial dependence
  4. Fragmentation
  5. Density estimation

The Centroid

* Common measure of central tendency:
* Centroid - location that minimizes the total squared distances of a set of points.
* The balance point

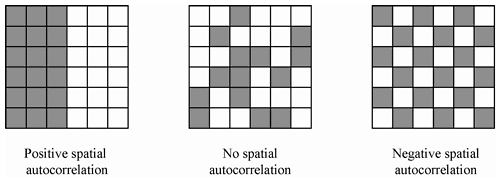
Dispersion

* A measure of the spread of points around a center
* For spatial data, a simple measure of dispersion is mean distance from the centroid.
* Spread of numbers in the dataset (variance).
* **Common measures of dispersion;**
  + Range - difference between highest and lowest values.
  + Variance - mean squared deviation from the mean.
  + Standard deviation - square root of the variance.

Spatial autocorrelation

Tobler's Law: things that are close to one another tend to be more alike than things that are far apart.

* Positive spatial autocorrelation - locations near a given location tend to be similar to it.
  1. For ratio, interval, and ordinal data, "similar" means close in value.
  2. For nominal data, it means of the same kind or a closely related kind.
     1. land use cover, animal habitat
* Negative spatial autocorrelation - locations near a given location tend to be dissimilar to it.
  1. Rare and usually human-made
     1. Checkerboard pattern of land use
     2. Natural example might be a lake dotted with many small islands.



Fragmentation

* Measure the patchiness of data sets
  1. Extent to which an area is broken up into pieces.
  2. Unfragmented - large expanses of uniform type
* Habitat fragmentation - important in determining the success of animal and bird populations
* Populations are less likely to survive in highly fragmented landscapes

Density estimation using kernels

* each point being replaced by a pile of sand of constant shape
* add the piles to create a surface
* Spatial interpolation fill the gaps in field
* Density estimation creates field from discrete objects
  1. field value =estimate of discrete object density
  2. e.g., e.g., estimating population density (a field) from individual people (discrete objects)
* **Density estimation and spatial interpolatio**n applied to the same data
* Density estimation of ozone measuring stations vs. Interpolating surface based on *locations* of ozone measuring stations
* Kernel function
* Each discrete object is replaced by a mathematical function known as a kernel
* Kernels are summed to obtain a composite surface of density
* The smoothness of the resulting field depends on the width of the kernel
  1. narrow kernels produce bumpy surfaces
  2. wide kernels produce smooth surfaces

**5. Optimization**

* Spatial analysis can be used to solve many problems of design
* A spatial decision support system (SDSS) is an adaptation of GIS aimed at solving a particular design problem

Optimizing point locations

**Location-allocation problems**

* Design locations for services, and allocate demand to them, to achieve specified goals
* Goals might include:
  1. minimizing total distance traveled
  2. minimizing the largest distance traveled by any customer
  3. maximizing profit
  4. minimizing a combination of travel distance and facility operating cost

**Routing problems**

* Search for optimum routes among several destinations
* The traveling salesman problem
  1. find the shortest tour from an origin, through a set of destinations, and back to the origin

**Optimum paths**

* Find the best path across a continuous cost surface
  1. between defined origin and destination
  2. to minimize total cost
  3. cost may combine construction, environmental impact, land acquisition, and operating cost
  4. used to locate highways, power lines, pipelines
  5. requires a raster representation

**Summary: Usefulness of GIS and spatial analysis**

* Accurate
* Easy in operation
* Great analytical capabilities
* Applied to precision agriculture, land Use planning, environmental protection, forest planning etc…

New progress  
- geosimulation, big data…

Geosimulation

* “The best way to understand a phenomenon is to mimic, replicate, reproduce, model it”.
* Models and simulation remain the best way to explore, query, and observe (albeit in a counterfeit manner) things that are simply inaccessible in the real world, because of the time-scales involved, the spatial scales and ranges, because of their hazardous nature, or because of the social and economic repercussions of interfering with them.
* Geosimulation has emerged as a trend in spatial simulation, focused on tackling the inherent complexity of geographical phenomena with focus on entity representation, entity behavior, entity interaction in space and time, and entity dynamics and evolution.
* 1. Cellular automata (CA) and its applications

Gird + Neighborhood interaction Rules= State Change

* 2. Agent-based modeling (ABM) and its applications
* **Big data** is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications.

Raster data analysis

Raster Data in ESRI GRID

* An Esri grid is **a raster GIS file format** developed by Esri, which has two formats:  
  - ARC/INFO GRID, ARC GRID and many other variations  
  - ASCII format, also known as an ARC/INFO ASCII GRID
* The grid defines geographic space as an array of equally sized square grid points arranged in rows and columns. Each grid point stores a numeric value that represents a geographic attribute (such as elevation or surface slope) for that unit of space. Each grid cell is referenced by its x,y coordinate location.

**Cell-based modelling**

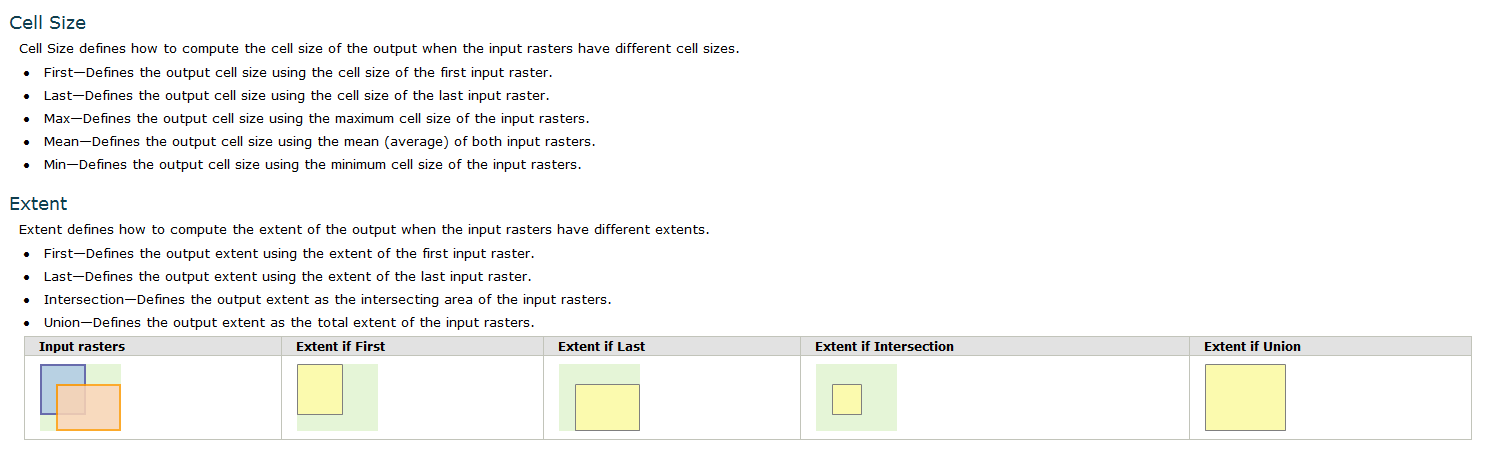
* Grid-cell processing provides efficient spatial-storage facilities and a powerful modeling environment for analyzing geographic features.
* It is especially effective for storing and representing continuous spatial surfaces and points, lines and polygons in the same format.
* Modeling locational problems is an inherent strength of the cell-based storage and analysis environment.
* Grid systems treat points, lines, polygons and surfaces, and their locational structures the same way: as cells in a grid.
* When all the data types are in the same structure, one semantical language can be used.
* More importantly, the different data types can be mixed with no prior preparation.
* An environment that integrates data types provides the user greater flexibility when modeling.
* Because the grid-based system’s foundation uses uniform grids, the mathematics are very simple and very fast when completing analysis between grids.
* Once registered, computing or deriving a value for an output cell from two or more input cells is a matter of direct value computation.
* No geometric detection, topology building and error checking is necessary.
* Because the grid-based system’s foundation uses uniform grids, the mathematics are very simple and very fast when completing analysis between grids.
* Once registered, computing or deriving a value for an output cell from two or more input cells is a matter of direct value computation.
* No geometric detection, topology building and error checking is necessary.

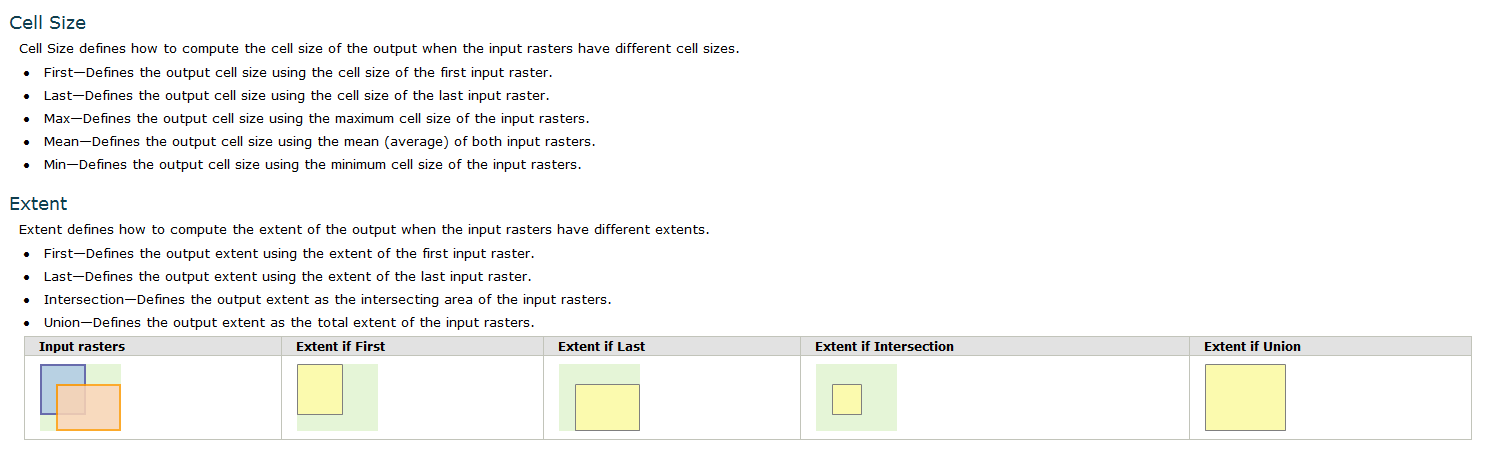
Functions (Map algebra)

* Arithmetic operations
* Logical operations
* Relational operations
* Reclassification
* Focal and zonal analysis
* Euclidean distance
* Cost distance
* Transformations

Before doing any analysis…

* Make sure these properties are consistent between data :  
  - Spatial reference  
  - Extent  
  - Cell size





Logical operations (boolean operations)

* Logical operations (AND, OR, XOR, NOT) are useful for answering questions like the following:  
  - Which areas are both grasslands AND sandy soils?  
  - Which areas are either grasslands OR sandy soils?

Arithmetic operations: Plus Minus Times Divide Power Exp Square\_root Ln

Logical operations (boolean operations): AND OR XOR NOT

Relational operations

* Equal To (==)、Greater Than (>)、Greater Than And Equal To (>=)、Less Than (<)
* Less Than And Equal To (<=)、Not Equal To (!=)

其中一个为NULL，值为NULL

Reclassifies：Reclassifies (or changes) the values in a raster.

Focal statistics: calculates for each input cell location a statistic of the values within a specified neighborhood around it.---Rectangular neighborhood、Circular neighborhood、Annulus neighborhood（环形）、Wedge neighborhood（楔）

example: calculate the density of built-up areas

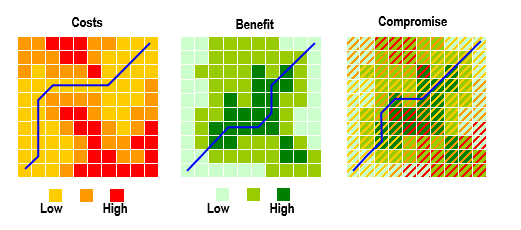
Zonal statistics: calculates statistics on values of a raster within the zones of another dataset.

* Example: summarize built-up areas by each county

Euclidean distance

Cost distance: calculates **the least accumulative cost distance** for each cell **to the nearest source** over a cost surface.

* + Space is not isotropic for most phenomena.
  + Absolute barriers.
    - * Stop movements / interactions completely.
      * Mountain ranges.
      * Rivers / oceans.
  + Relative barriers.
    - * Friction that varies according to direction and to features of space.
      * Slope.
      * Type of roads.
      * Border.



Transformations: Raster to features(Data types)

Multi-criteria evaluation techniques for GIS(多准则评价)

Introduction

* MCE techniques began to emerge to solve decision-making and planning problems in the early 1970s
* Numerous factors or criteria should be involved to solve real-world problems
* MCE can be employed to handle a number of criteria in decision-making
* The planning process is becoming more complicated in technical, physical, social and economic aspects
  + how about in geography?
* The analysis of alternative planning options is an essential part of plan-making (Landis, 1995)
* MCE can be used for analyzing the complex trade-offs between different alternatives (van Delft and Nijkamp, 1977)
* It is useful to generate and compare a set of alternatives before planners can choose a suitable plan
  + Scenario analysis
* The overlay of various layers of maps is a classic example of a site search exercise
* GIS can be used to facilitate map overlay analysis(叠置分析)
* However, traditional GIS overlay analyses have **limitations when multiple and conflicting criteria and objectives are involved** (Carver, 1991)

Limitations of overlay analyses (Carver, 1991):

* Usually, a few factors are considered because there are difficulties in comprehending more than four to five factors (Janssen and Rietveld, 1990)
* Most overlay procedures in GIS do not allow that variables may not equally important
* The outcome of area screening depends strongly on the choice of threshold values
* The use of threshold values to map continuous variables will lead to a substantial loss of information

The integration MCE with grid-based GIS

* It is quite straightforward to carry out MCE based on grid-based GIS although it is not essential
* The division of feasible areas on a grid basis is the most convenient and acceptable solution
* The raster format has advantages in handling multiple layers of GIS data
* It can much simplify the calculation

***Selecting Criteria***

* Selecting evaluation criteria should be based on the problem situation.
* There are two tendencies in defining the set of criteria:
  + *A formidable number* of criteria may be included when the decision model describes the problem situation as closely as possible.
  + *A small number* of criteria are used due to the oversimplification of the decision problem.
* What is needed is a balanced approach that responsibly surveys all the possible evaluation criteria and provides a reasonable mechanism for selecting the set of criteria.
* The procedure is to select a set of attributes based on the desirable properties of attributes.
* Each attribute must be:
  + comprehensive (it must be unambiguous and understandable);
  + measurable (e.g. the level of convenience and comfort is very difficult to measure).
* A set of attributes should be:
  + Complete (they cover all aspects of a decision problem);
  + Operational (they are meaningful in analysis);
  + Decomposable (they can be broken into parts to simplify the process);
  + Nonredundant (they avoid the problems of double counting); and
  + Minimal (the number of attributes are used as small as possible).
* There are no universal techniques for determining a set of criteria.
* For example, the set of criteria for evaluating sites for nuclear plants will be different from those for evaluating sites for residential development.
* Usually, the procedure for identifying a set of evaluation criteria should be a multistep iteration process.
* The iterative procedure may include:
  + The elimination of redundant criteria;
  + The combination of two or more criteria into one criterion;
  + The decomposition of a criterion into a number of attributes to facilitate the measurement process;

**Creating Criteria**

* The way of measurement is essential for producing meaningful and accurate analysis results
* Measurement is the process of assigning symbols to the attributes of objects, events, or states
* There are four levels of measurement:   
  *nominal, ordinal, interval, and ratio*



Land use types: forest, paddy, and bare land. Population size:small, medium, and large.

Temperature:300C(its zero is set arbitrarily);the distance of 20-300C is the same as that of 50-600C

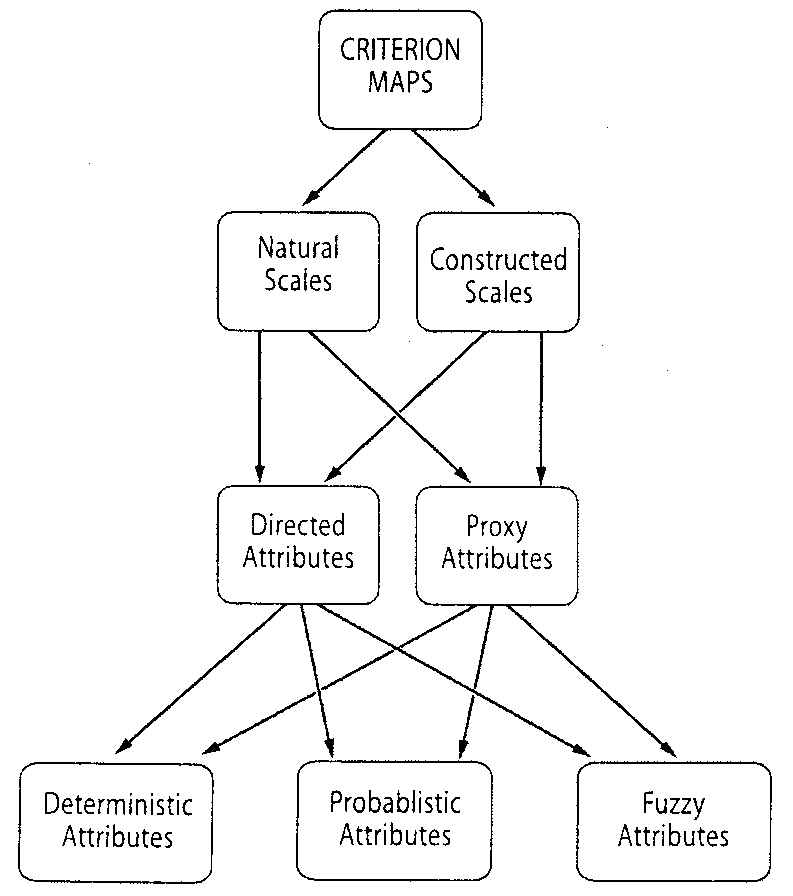
Time and mass: hours, seconds, tons or kilograms. 100kg body has twice the mass of a 50kg body.

* Natural-scale criteria: They are objective scales. No value judgment is needed in constructing natural scales. (e.g. distance in kilometers; number of animal species per acre of forest)
* Constructed-scale criteria: They are called subjective scales. (e.g. There is no natural scale for measuring the aesthetic impact of establishing a residential building in forest areas. Another example is suitability scores.)

Quantitative scale

(natural scale) 40min (constructed scale)7point (Nominal measurement) good

* Both natural and constructed scales can be further subdivided into tow categories:
  + direct; and
  + proxy scales
* A direct scale implies that the criterion can be measured directly. (e.g. for the objective ‘minimize the cost of establishing a faculty’, a direct-scale criterion map is the cost of establishing the faculty)
* A proxy scale can be used in the situations that obvious attributes are unavailable
* For example, GDP may be used as a proxy for land demand, or…



GIS Operations on criteria maps

* *Arithmetic operation*
  + Addition (+)
  + Subtraction (-)
  + Multiplication (×)
  + Division (/)
* *Logical operation*
  + Intersection (*AND*)
  + Union (*OR*)
  + Complement (*NOT*)
* *Probabilistic operation*
* *Fuzzy operation*



**Standardization of Criteria**

* Because of the different scales upon which criteria are measured, it is necessary that factors be standardized.
* Various criterion map layers must be transformed to comparable units.
* There are a variety of procedures for standardization.
* The two most often used standardization procedures are based on:
  + the maximum score; and
  + the minimum and maximum values.



* 1) The maximum score
* 
* 2) the minimum and maximum values
* 3) Fuzzy sets
* The process of standardizing evaluation criteria can also be seen as one of recasting values into of set membership (Eastman, 1997).
* Indeed, researchers have suggested that blindly using a linear scaling (or any other scaling) between the minimum and maximum values of the image is ill advised.
* For example, if the map were to cover a range of perhaps 100 km from the reserve, then the farthest point away from the reserve would be given a value of 1.0. Using a linear function, then, a location 5 km from the reserve would have a standardized value of only 0.05.
* When the primary issue was noise and minor disturbance from local citizens, a distance of only 5 kilometers would have been equally as good as being 100 km away.Thus the standardized score should really have been 1.0.
* If an MCE were undertaken using the blind linear scaling, locations in the range of a few 10s of km would have been severely devalued.
* Four fuzzy set membership functions:
  + Sigmoidal: The Sigmoidal ("s-shaped") Membership function is perhaps the most commonly used function in Fuzzy Set theory. It is produced using a cosine function.
  + J-Shaped: The J-Shaped function is also quite common, although in most cases it would seem that a sigmoidal function would be better.

it should be pointed out that with the J-shaped function, the function approaches 0 but only reaches it at infinity. Thus the inflection points a and d indicate the points at which the function reaches 0.5 rather than 0

* + Linear : This function is used extensively in electronic devices advertising fuzzy set logic, in part because of its simplicity, but also in part because of the need to monitor output from essentially linear sensors
  + User-defined : When the relationship between the value and fuzzy membership does not follow any of the above three functions, the user-defined function is most applicable

The control points used in this function can be as many as necessary to define the fuzzy membership curve

**Calculating Weights**

* MCE problems usually involve criteria of varying importance to decision makers
* This requires the information about the relative importance of the criteria
* The derivation of weights is a central step for solving the decision-making problem
* The larger the weight, the more important is the criterion in the overall utility

Ranking method

* The simplest method for estimating weights is to arrange them in rank order according to the decision maker’s preference.
* Once the ranking is established, several procedures can be used to generate numerical weights from rank-order information.
  + **Rank sum,**
  + **rank reciprocal**
  + ***Rank exponent***



Pairwise comparison method

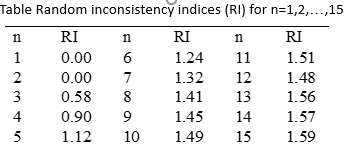
* The pairwise comparison method was proposed by Saaty (1980) in the context of the analytic hierarchy process (AHP).
* The procedure consists of three major steps:
  + generation of the pairwise comparison matrix,
  + the criterion weight computation, and
  + the consistency ratio estimation.
* Example: Consider a site-suitability problem. The involves evaluating a set of sites for development based on three criteria: price (p), slope (s), and view (v). It requires assessing the relative importance of the three criteria.
* Step 1: Development of the pairwise comparison matrix
* The first step is to create a ratio matrix by pairwise comparisons. The comparison employs an underlying scale with values from 1 to 9 to rate the relative preferences for two criteria (Table).
* Suppose that price is moderately to strongly preferred over the slope attributes; That is, the comparison results in a value of 4.
* If criterion A receives a score of 4 relative to criterion B, criterion B should receive a score of ¼ when compared to criterion A.
* 
* Step 2: Computation of the criterion weights
* (a) sum the values in each column of the comparison matrix; (b) divide each element in the matrix by its column total (the resulting matrix is refereed to as the normalized pairwise comparison matrix; and (c) compute the average of elements in each row of the normalized matrix. These averages provide an estimate of the relative weights.



* Step 3: Estimation of the consistency ratio
* determine the weighted sum vector by multiplying the weight for the first criterion (price) times the first column of the original pairwise comparison matrix, then multiply the second weight (slope) times the second column, the third criterion times the third column of the original matrix, finally, sum these values over the rows; and (b) determine the consistency vector by dividing the weighted sum vector by the criterion weights.



* Before calculating consistency ratio (CR), we need to compute values for two more terms, lambda (λ) and the consistency index (CI). The value for lambda is simply the average of the consistency vector:
* λ= (3.247+3.10+3.022)/3 = 3.126
* The calculation of CI is based on the following formula:
* Finally, the consistency ratio (CR), which is defined as follows:
* Where *RI* is the random index, the consistency index of a randomly generated pairwise comparison matrix. RI depends on the number of elements being compared (Table).
* The consistency ratio (CR) is designed in such a way that if CR < 0.10, the ratio indicates a reasonable level of consistency in the pairwise comparisons. If, however, CR ≥ 0.10, the values of the ratio indicate inconsistent judgments. In such cases one should reconsider and revise the original values in the pairwise comparison matrix.



Deriving a Scale of Priorities from Pairwise Comparisons