GEOG 204

LECTURE 10

Course Review

Final Exam: Thur Nov 28. Cumulative, 45 mins

Last tutorial: Today/Monday 25th

Project Submission: By Fri Dec 3

Housekeeping

Definition

- GIS: system used to store, manipulate, analyze, manage, and present geographic data

- Importance of GIS
  - Makes it possible to efficiently integrate, manage, and analyze geographical information from maps, remote sensors, data collectors and text.

Common GIS Operations

- Data Acquisition
- Data Attribute Management
- Data Display
- Data Exploration (exploratory spatial analysis)
- Data analysis and modeling

4

QN: List four operations that make GIS an important too for spatial analysis

Geographic phenomena

- Spatial Variation
  - Tobler's law
  - Homogeneity/Uniform distribution
  - Random distribution
  - Clustering
- Temporal(time) Variation

5

QN: What is Tobler's Law of Geography?
QN: What does Tobler's Law tell us about geographical phenomena?

Representing geographic space

- The two basic forms
  - Objects and Fields
    - Vector and Raster
- The three geometry types
  - Points, lines, polygons
- The different data formats for Raster and Vector
  - e.g. shp, geotiff, GeoJSON, rst,

6

QN: What are the two spatial data models?

Location and Attributes (Fields)

Attribute data: allows us to ask the question ... "what is it ?"

> Every layer has an associated table

> These are linked to spatial location by a code number

➢Attributes are stored in columns as *items*

➢Rows display the attributes for each feature = *records*

➢Entries may be text strings, integers, float (decimal) or dates

Item (Description)

Record (place) ->

POLYGON	ESA_1	SPC1	PCT1	SPC2	PCT2	AGE_CL	HT_CL_IN	SITE_IDX	CRNCL_CL	SitePrep	Dist	YearDist	Regen	STTEND
67		HW	40	S	40	2	1	16.6	8	B	R	1985	1999	F
133			0		0	0	0	0	0			0	0	
199		HM	40	HW	30	9	3	7.2	5		L	1980	0	
353		HW	90	BA	10	9	4	11.6	1	B	L	1980	1999	F
229		HW	70	HM	20	9	3	9.5	5	B	L	1980	1999	F
264		HM	50	HW	30	9	3	7.5	5	H	L	1980	1999	F
162			0		0	0	0	0	0			0	0	
393		HW	60	HM	20	9	3	8.5	5	H	L	1980	1999	R
165		HM	80	BL	20	9	3	7	4	H	L	1980	1999	R

QN: How are attributes represented in a Raster Data model Vs in a Vector Data model?

Scale

- Scale: The ratio or relationship between a distance on a map and the corresponding distance the ground.
- Map details depend on the scale of the data
  - The larger the scale, the more the detail
    - Large scale is synonymous with fine scale (fine granularity)
  - The smaller the scale, the less the detail
    - Small scale is synonymous with course scale (course granularity)

8

QN: What is the scale of Geographic data?
QN: Calculate cartographic scale
qN: What is a scale bar?

Scale

- Spatial Scale **could be making reference to**
  - Resolution: the size of your pixel (raster)
  - Extent: the size of your study area
    - "small-scale" operation covers a small area
    - "large-scale" operation covers a large area
- Cartographic Scale
  - 1:1000000 (1 cm = 10 km)
- Scale Bar

9

Census Data

- Geographic Units
  - Dissemination Block
  - An area bounded on all sides by roads and/or boundaries of standard geographic areas. The dissemination block is the smallest geographic area for which **population and dwelling counts** are disseminated.
  - Dissemination Areas
    - composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons. It is the smallest standard geographic area for which **all census data** are disseminated.
    - Enumeration Areas: term used for the same geographic unit prior to 2001
  - Census Tracts
    - Larger areas that have a population between 2,500 and 8,000 persons.
    - They are located in centres of 50,000 or more in the previous census.
- What is a unique aspect about the 2011 Census?
- What are boundary files?
- What is the difference between a census block, dissemination area, and census block?

	Nontopological	Topological	
		Featured-based	Layer-based
Logical operations	• Attribute database query		• Reclassification and aggregation
Arithmetic operations	• Change mapping • Summary statistics		
Overlay operations	• Address geocoding	• Overlay analysis	• Overlay analysis
Geometric property operations	• Calculation of areas, perimeters, and distances		• Network analysis
Geometric transformation operations	• Coordinate and geometric transformation • Surface interpolation		
Geometric derivation operations		• Buffering	

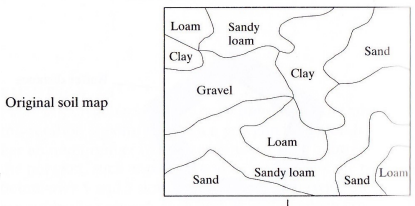
Source: C.P. Lo, A.K.W. Yeung (2007) Concepts and Techniques of Geographic Information Systems

Topological Functions

- Reclassification
  - Database simplification process that reduces (modifies) the categories in attribute data
  - Two-step process
    - Nontopological: select attribute e.g. a range of values and assign them a new class
    - Topological: dissolve according to new classification scheme

12

QN: What is data reclassification?
QN: Explain an example of when data reclassification would be helpful

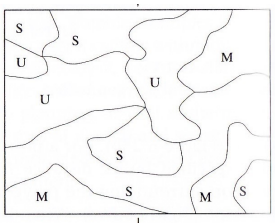


Original soil map

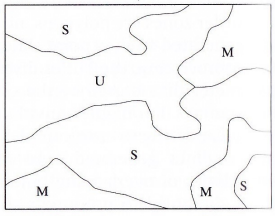
Clay	Unsuitable (U)
Gravel	
Loam	Suitable (S)
Sandy loam	
Sand	Moderately suitable (M)

Reclassification criteria

Land suitability map



Land suitability map with boundaries dissolved if adjacent polygons have the same characteristics



Source: C.P. Lo, A.K.W. Yeung (2007) Concepts and Techniques of Geographic Information Systems

13

Topological Functions

- Topological Overlay Operators
  - Types:
    - **Point** layer in **polygon** layer
    - **Line** layer in **polygon** layer
    - **Polygon** layer in **Polygon** layer
  - Operators:
    - UNION, INTERSECT, IDENTITY, CLIP, ERASE, SPLIT, BUFFER

Raster Operations

- Raster operations: grouped according to the way raster cells are used in the analysis
  - Local Operations:
    - value of the cell in the output layer is a function of the cell at the same location in the input layer
  - Neighborhood Operations:
    - value of the cell in the output layer is a function of the cells neighboring the cell at the same location in the input layer
  - Extended Neighborhood Operations:
    - value of the cell in the output layer is a function of the cells neighboring and beyond the immediate neighborhood of the cell at the same location in the input layer
  - Regional Operations:
    - the output layer us generated by identifying cells that intersect with or fall within each region on the input layer

15

QN: Distinguish between raster-based Local and Neighborhood operations?

Local Operations

- Reclassification
  - Create a new raster layer by applying changes to the attribute values of the cells in the input layer
  - Logical or arithmetic operations

Binary masking; Classification reduction; Classification Ranking; Changing Measurement Scales

- Overlay Analysis
  - Logical or arithmetic operations
    - AND, OR, XOR; addition, subtraction, multiplication, division, assignment
  - Two or more input layers

16

Is it possible to do a buffer operation on raster dataset?
Is it possible to do arithmetic operations on Raster dataset?
Is it possible to do arithmetic operations on the vector dataset?

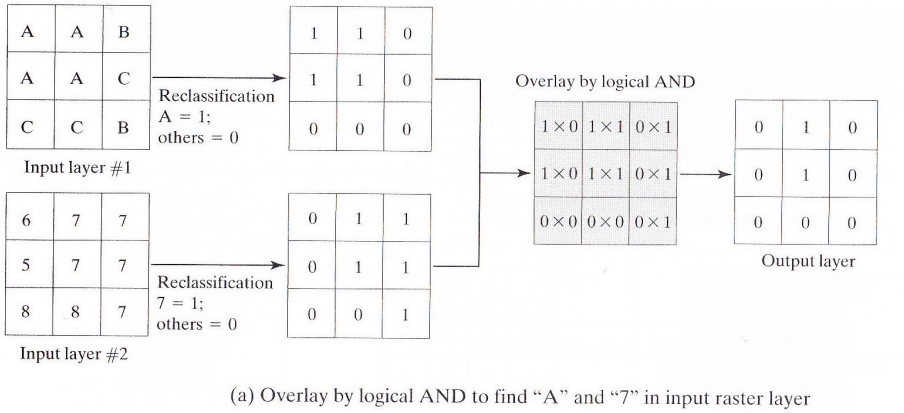
Raster Operations

	Local Operations	Neighborhood Operations	Extended Neighborhood Operations	Regional Operations
Logical Operations	• Reclassification			
Arithmetic Operations	• Reclassification	• Aggregation • Filtering	• Statistical analysis	
Overlay Operations	• Logical •Arithmetic			• Category-wide overlay
Geometric Property Operations		• Slope and aspects	• Distance, proximity, and connectivity	• Area • Perimeter • Shape
Geometric Transformation Operations			• Rotation • Translation • Scaling	
Geometric Derivation Operations			• Buffering • Viewshed analysis	• Identification and reclassification

17

Why is slope/aspect a neighborhood operation?

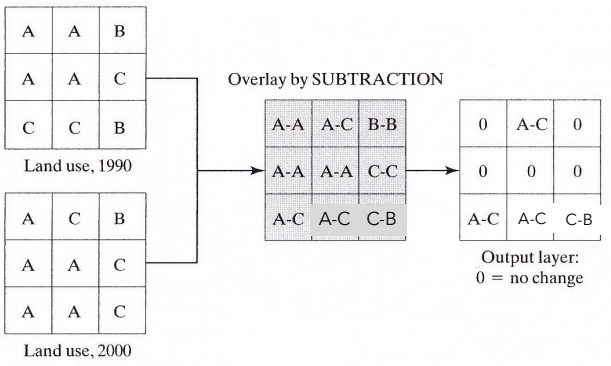
Overlay Analysis



(a) Overlay by logical AND to find "A" and "7" in input raster layer

18

Overlay Analysis



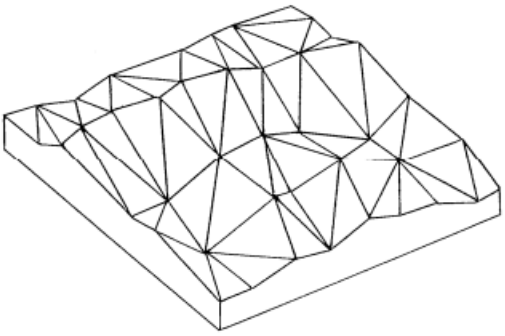
(a) Overlay by arithmetic SUBTRACTION to detect land-use change

19

Spatial Interpolation

- Triangulated Irregular Network (TIN)
  - Consist of z-value nodes that are connected by edges to form contiguous and non-overlapping triangles
  - The edges in TINs can be used to capture the position of linear features that play an important role in the definition of the surface (e.g. ridgelines or stream courses)

Triangulated Irregular Networks (TIN): vectors



TIN: a series of triangles capturing the topography .. x, y, z at nodes

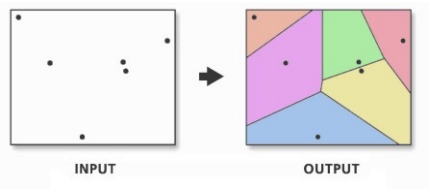
Each triangle has a uniform slope and direction (aspect)

Advantage: **significant points or lines** can be encoded e.g. peaks, ridges, valleys

Disadvantage: more complex, needs more processing to generate, when a new point is added, the TIN needs to be rebuilt

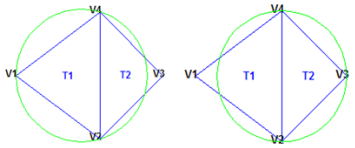
Spatial Interpolation

- Thiessen Polygons



Any location within a Thiessen polygon is closer to its associated point than to any other point input feature

All points are triangulated into a triangulated irregular network (TIN) that meets the Delaunay criterion: the circumcircle of every triangle is empty, that is, there is no other point in its interior.



The perpendicular bisectors for each triangle edge are generated, forming the edges of the Thiessen polygons. The location at which the bisectors intersect determine the locations of the Thiessen polygon vertices.

22

Spatial Interpolation

- Contouring
  - Contours are lines that connect locations of equal values for a given continuous phenomenon.
    - as elevation, temperature, precipitation, pollution, or atmospheric pressure.
    - Contour lines are often generally referred to as isolines but can also have specific terms depending on what is being measured (e.g. isobars for pressure, isotherms for temperature, and isohyets for precipitation)

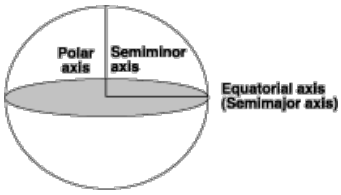
Spatial Interpolation

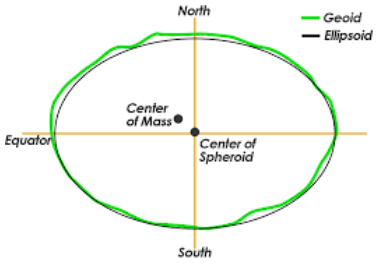
- Contouring
  - The distribution of the contour lines shows how values change across a surface.
    - Little change in a value, the lines are spaced farther apart.
    - Great change, the lines are closer together.

What is a TIN?
What are Thiessen/voronoid Polygons?
What are contours lines?
Give examples of phenomena that can be represented with the aid of contours lines?
What does the distribution(spread far apart and aligned close together) tell us about the change in the phenomenon's values?

Modeling the Earth's shape

- the Ellipsoid/Spheroid
- the Geoid (approximates earth's gravity)





- the Mean Sea Level
  - Over oceans MSL = geoid; differ on landmass
- the Terrain

25

Distinguish between the MSL, Ellipsoid, Geoid.

Types of Coordinate Systems

- Geographic coordinate systems
  - Global or spherical coordinate systems such as latitude-longitude.
    - Prime meridian
    - Latitude, Longitude
    - Parallels, meridians
    - Units of measure
- Projected coordinate systems
  - Examples?
- Datum
  - Examples?

Map Projections

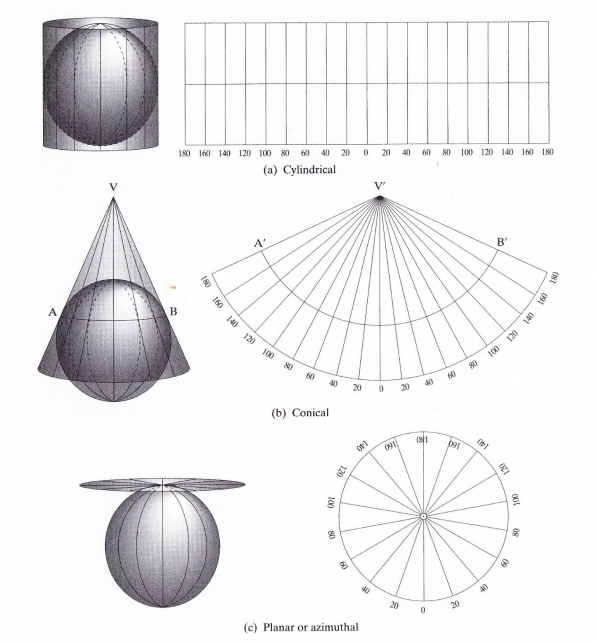
- Basic Concepts
  - "Project" features on the Earth's curved surface onto a flat surface (e.g. sheet of paper)
  - Therefore "map projection" is about preserving the properties of real-world features when they are depicted on a 2D map.
    - The properties are:
      - Area
      - Shape
      - Distance
      - Direction

27

List and describe the map properties that are affected by map projections.

Map Projections Classes

- There are different ways to classify map projections
  - One is based on developable surface on which a network of meridians and parallels is projected
    - Cylindrical
    - Conical
    - Planar/Azimuthal



Three basic types of map projections

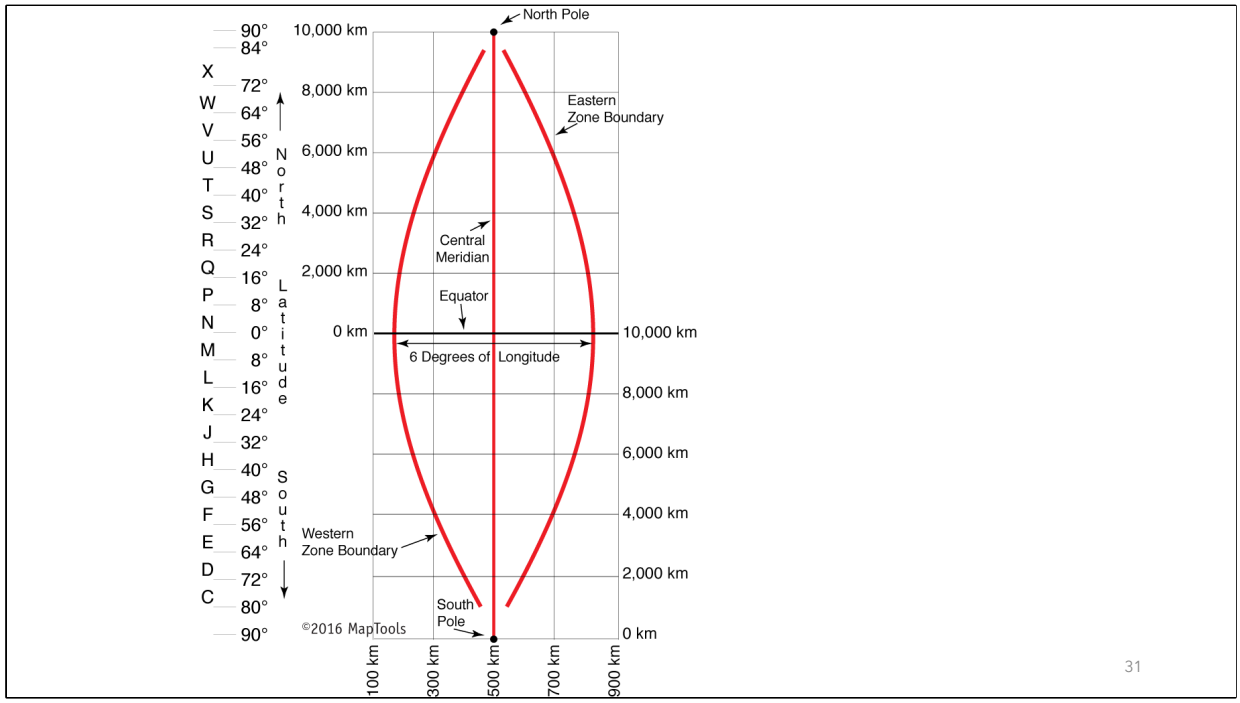
Source: C.P. Lo and A.K.W Yeung

29

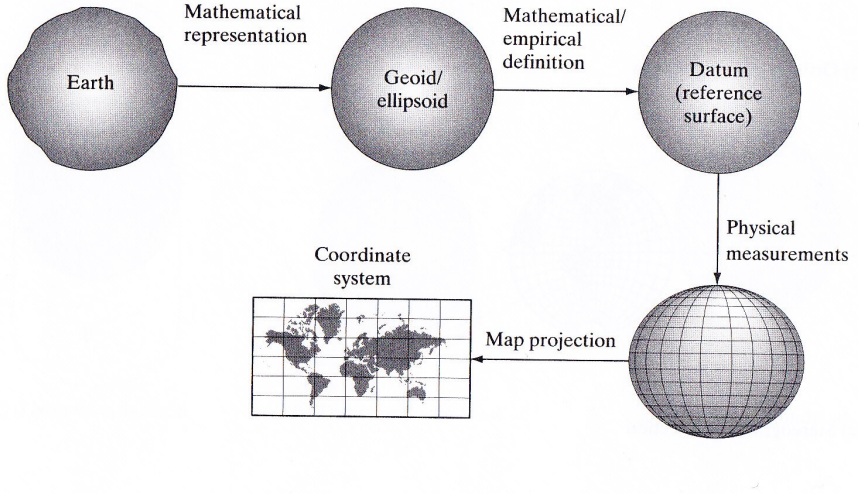
Major Projection Classes

- The other is based on the projection properties
  - Conformal
    - Local shapes/angles are correct, areas allowed to vary
  - Equal-area
    - Areas are correct, shapes allowed to vary
  - Equidistant
    - Distance to certain point correct
    - Areas and shapes allowed to vary
  - Azimuthal
    - Directions are accurate

List and describe the major Projection classes



What is the Universal Transverse Mercator Coordinate System?



Georeferencing

Source: C.P. Lo and A.K.W Yeung

32

Geocoding

- Geocoding
  - The process of assigning spatial locations to descriptive data for those locations
    - from location description to spatial locations
- The most common type of geocoding is *address matching*
- Address matching
  - requires two sets of data
    - The first data set contains individual street addresses in a table, one record per address
    - The second is a reference database that consists of a street map and attributes for each street segment such as the street name, address ranges, and postal code
  - Three phases: preprocessing, matching, plotting
    - To plot the location, the system performs linear interpolation

33

What is Geocoding and reverse geocoding?

Given a street address how would find the coordinates of that address?

Given a street segment whose starting and end coordinates are known, how would you find the coordinates of an address 2/3 of the way along the segment?

Data Quality: Sources of Errors in Data

- Sources of Errors:
  - **Human errors** include mistakes, such as reading an instrument incorrectly, and faulty judgments (e.g. ambiguous boundaries such as high water mark).
  - **Environmental characteristics**, such as variations in temperature can result in measurement errors
  - **Instrument errors** Measurements are as precise as the instrument's capabilities.
    - The smallest measurement that can be made is the instrument's resolution.

34

Elements of Data Quality

- Accuracy
  - Positional accuracy
    - closeness of locational information (usually coordinates) to the true position
  - Thematic/attribute accuracy
    - the closeness of attribute values to their true value
- Lineage
  - a record of the data sources and of the operations which created the database
    - how were they digitized, from what documents?
    - when were the data collected? By who?
  - is often a useful indicator of accuracy
- Logical consistency
  - refers to the consistency of the data model (particularly the topological consistency)
    - is the database consistent with its definitions?
    - is there exactly one label for each polygon?
    - are there nodes wherever arcs cross, or do arcs sometimes cross without forming nodes?

35

Distinguish between Accuracy, Precision and Uncertainty

Elements of Data Quality

- Completeness
  - degree to which the data exhausts all the possible items
    - are all possible objects included within the database?
  - affected by rules of selection, generalization and scale
- Temporal quality
  - The quality of temporal attributes and temporal relationship of features.
- Data usability
  - Suitability to an application and its related functional requirement

Data Quality - Key Issues

- Key Concepts
  - Accuracy, Precision and Uncertainty
- Accuracy:
  - closeness of the measurements, computations to the true values (or values accepted to be true)
    - spatial data are a generalization of the real world, the "true value" is thus an estimate of the real world
  - ~ absence of errors

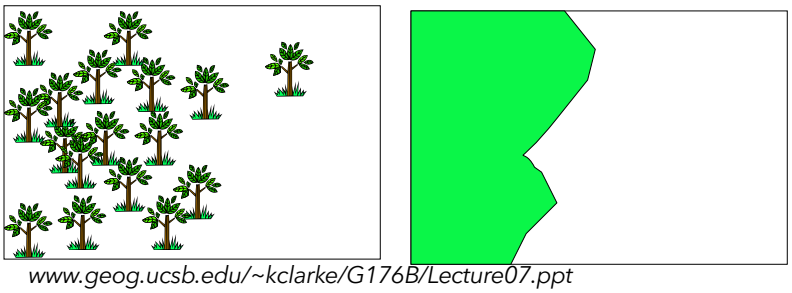
37

Data Quality - Key Issues

- Precision:
  - the number of decimal places or significant digits in a measurement
    - precision is not the same as accuracy - a large number of significant digits doesn't necessarily indicate that the measurement is accurate
    - Precision is related to how detailed a description one has
  - a GIS works at high precision, mostly much higher than the accuracy of the data itself

Data Quality - Key Issues

- Uncertainty: our imperfect and inexact knowledge of the world
  - Positional uncertainty
  - Attribute uncertainty
  - Definitional uncertainty
  - Measurement uncertainty



www.geog.ucsb.edu/~kclarke/G176B/Lecture07.ppt

39

Terms

- Choropleth Map
- Topology
- Topography
- Temporal
- GIS Functions/Operations vs Analysis Functions/Operations
- Geographic Coordinate Systems
- Projected Coordinate Systems