Massachusetts Institute of Technology
Department of Urban Studies and Planning

11.520: A Workshop on Geographic Information Systems

11.188: Urban Planning and Social Science Laboratory

Spatial Data Models: Spatial Analysis II (Raster Models)

26 October 2005 - Joseph Ferreira

Raster vs. Vector Data Models for GIS

Vector (model *boundaries* of spatial features) - our focus so far

- **Vector Feature Types:**
  - **Points**
    - The fundamental building block
  - **Lines**
    - Built from at least two points at the ends of the line: the nodes
    - Extra points between the nodes--vertices--may add shape to the line
  - **Polygons**
    - A closed object with an interior and exterior
    - Build from one or more lines
    - May have islands
- **Vector Data Formats (for ArcGIS):**
  - **ArcInfo Coverages**
    - One directory per layer containing its geometry files (**`.adf`**files)
    - Geometry includes topological relationships among features
    - A workspace typically contains several coverage directories
    - Database tables for all coverages in the workspace are stored in "Info" tables in a shared **`info`** directory
    - This shared **`info`** directory impedes data management
    - Coverages must be moved using ArcGIS, not operating system file management commands!

- **ArcGIS Shapefiles**
    - **`.shp`**, **`.shx`**, **`.dbf`** files (and possibly others)
    - Topological relationships are not stored in the layer, but computed on the fly when needed
    - Shapefiles are easily moved or copied within the OS; just copy or move **`layer.*`**
- **Spatial Database Engine (SDE)**
    - Retrieved dynamically from a database server
    - Relies on a heavy-duty RDBMS such as Oracle
- **Other GIS packages (MapInfo, Intergraph, TransCAD, Maptitude**) use their own proprietary data formats, making for a Babel of GIS data
- **Standards**: SDTS (spatial data transfer standard for archival file format); Open Geospatial Consortium protocols for web services and Application Programming Interface(Web Mapping Service and Web Feature Service); Geographic Markup Language (GML) for xml-based data interchange; etc.

Raster (model *properties* of uniformly spaced grid cells)

- Typically represented as a two-dimensional X-Y array
- [Goodchild's illustration of raster geometry:]()

- More dimensions (Z for height, T for time) also possible, but harder to visualize
- Most rasters assign a single **scalar** value to each grid cell

- Value of the cell may represent
    1. an average value over the entire cell area
    2. the value at the center of the cell (ArcView does it this way)
    3. the value at the grid node (a corner) Goodchild's Illustration:
- Possible to have multiple values--a **vector** of values--assigned to each cell
- Goodchild's discussion of rasters
- Georeferenced images (e.g., orthophotos) are another type of raster data
    - Orthophotos
        - Cell value is pixel brightness in orthophoto
    - Scanned maps
- ArcGIS (and earlier ArcView and ArcInfo) use a common data format called a **grid**
- ArcGIS's toolkit for raster analysis is the optional (and expensive) **Spatial Analyst** extension

Comparing Field and Object Models

Object (Vector) Model

- Each feature is a discrete object with vectors representing object boundaries
- "treats the information space as populated by discrete, identifiable entities, each with a georeference" (Worboys, p. 149)
- Michael F. Goodchild's definition (from the NCGIA Core Curriculum in GIScience)

Field (Raster) Model

- Labels discrete chunks of space and records the properties of each chunk
- Good where values vary continuously over space; the raster approximates these variations with discrete "samples"
- "[geographic] information as collections of spatial distributions" (Worboys, p. 149)
- Examples:
- Temperature
- Rainfall
- Elevation
- Depth
- Concentration of a chemical in the air, water, or soil
- Fields are actually **functions** that map spatial locations to values
- Representing continuously varying 'fields'
- Representing fields (Goodchild's discussion)
- Different field representations (Goodchild's illustration):

a) rectangular cells	d) digitized contours
b) rectangular grid of points	e) polygons
c) irregularly spaced points	f) triangulated irregular network (TINs)

- Examples where the field model works well (from Goodchild)
- Weather modeling example at the National Center for Atmospheric Research (NCAR):
  MM5 (mesoscale model, fifth-generation)
- Issues with storing discrete objects in rasters (from Goodchild)

Examples of the contents of a single layer

- Output from one band of a remote sensing satellite (or a panchromatic aerial photo)
  - gives the level of radiation received by the satellite in that band, recorded as a number between 0 and 255 (8-bit)
- A classified scene in which satellite output has been assigned to one of a number of classes denoting various land uses
  - e.g. 1=urban, 2=cultivated land, 3=water.
  - many image processing and pattern recognition algorithms are used to classify/categorize imagery

- A digital elevation model
    - values denote elevation of each cell's center point (above mean sea level in meters)
- A representation of the presence of roads
    - e.g. 1=road present, 0=no road
- A soil or flood plain map
    - value = predominant type of soil in grid cell
    - value = 50 if greatest flood risk in cell is 1-in-50 year flood; 100 if 1-in-100 year flood; etc.

Raster Analyses: Neighbors and 'Map Algebra'

- Edge-neighbors are four neighboring cells that share an edge with the cell.
- Adding diagonals yields nine-nearest neighbors
- Map Algebra (phrase coined by Dana Tomlin)
    - Often useful to compute algebraic function of neighbors
        - Smooth distributions (recompute cell value to be average of neighbors)
        - Model water flow (accumulate water from neighbors that are higher up)
        - Plume dispersion model
    - Useful to construct new raster layer where each cell's value is an algebraic function of neighbors
    - The regular structure of the grid cells can simplify spatial modeling and analysis

Raster Difficulties: Edge Effects

- Some cells on the border that have only two-three edge-neighbors.
- Map algebra models will behave differently at boundary where there are fewer neighbors - edge effects
- Common fixes for edge effects
    - Run the model with an expanded coverage area for theraster, but then throw away the borders.
    - Weight cells to compensate for missing neighbors (but difficult to determine the weight)
    - Declare that a cell on the bottom border of the raster actually neighbors a cell on the top border.

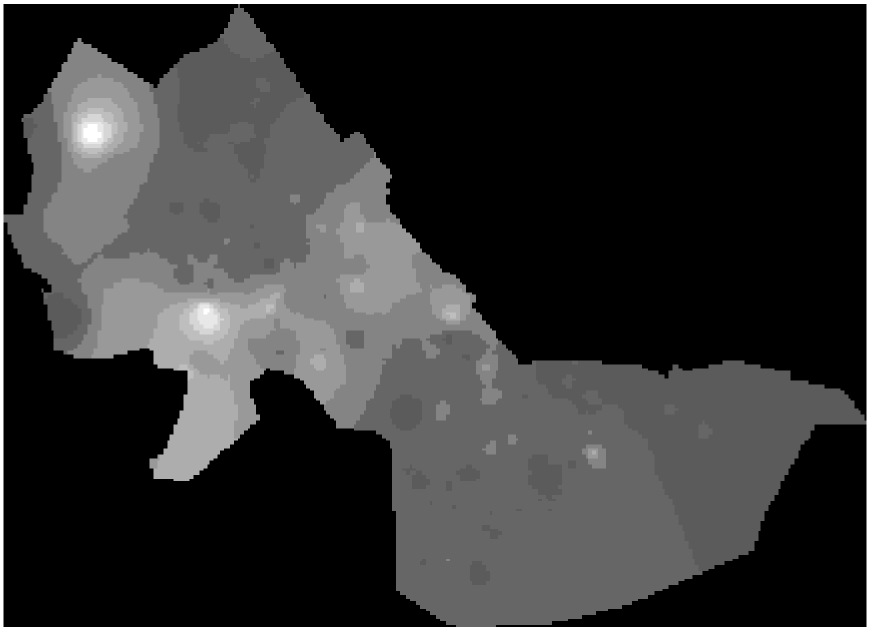
Rasters in Practice

- There are many practical applications of rasters within and outside GIS.
    - a computer display is a raster
    - digital cameras use rasters
    - images on the Web are rasters

- Certain kinds of data always come in raster form
    - digital elevation models
    - remote sensing images
- Raster data standard
    - geoTIFF is an adaptation of the general-purpose TIFF image standard that includes the necessary hooks for registering the raster to the Earth, plus other geographic features.
    - Orthophotos are georeferenced raster images
    - Scanned maps may look like raster images but could easily be distorted
- When NOT to Use Raster Representations
    - Rasters are less useful for representing networks where topology/connectivity is important and can't be captured at grid cell scale
    - Example 1: modeling sewer lines as a raster layer
        - code 1 in cells where a sewer is present, 0 elsewhere
        - if two adjacent cells both have 1, that's no guarantee the sewers they contain are connected
    - Example 2: Representing land ownership parcels as a raster layer
        - by definition, the boundary between two survey points is a mathematically straight line
        - the jagged appearance of a raster representation would be unacceptable
    - Rasters cell size is a direct indicator of level of geographic detail
        - Sometimes a plus - better indication of relevant data resolution
        - To double spatial resolution, there may be four times as many cells

Example: using grids with Cambridge data from forthcoming Lab 7:

- Computing a housing value 'surface' for Cambridge
    - Using the Census block group data:
        - Rasterize Cambridge into 100 meter grid cells
        - Vector-to-raster conversion of (median) housing value for Cambridge block groups
        - Smooth the 'surface' using neighborhood averages
    - Using the sales89 data:
        - Compute cell values from sales within the cell
        - Adjust the 'surface' based on neighboring sales
    - Combine the two estimates:
        - Average the cell estimates (does this make sense? use weights?)
        - Understand use and limitations of map algebra models



Interpolated Grid of Inflation-Adjusted 1989 Cambridge, MA Housing Sales Prices

Other Spatial Data Models: Network Models of Linear Features, Triangulated Irregular Networks (TIN), etc.

- Representing Roads:
  - Centerlines
  - Polygons
  - Voids
  - Directed Graphs
  - Raster
- Raster vs. vector models of roads
- Lanes as spatial features vs. attribute values
- Segmentation issues (and dynamic segmentation strategies)
- TIGER files and Tiger file Technical Documentation (see page 26 for a good diagram)
- F. Benjamin Zhan's discussion of Representing Networks
  - Zhan's sample network
  - Zhan's list of Common Network Operations and Applications
- Network modeling tools:
  - Caliper's TransCAD
  - ArcView's Network Analyst

Review of Current and Forthcoming Class Work

- Homework Sets 2 & 3: Site selection analysis for Cambridge senior citizen center
    - Homework 2 (due with short project proposal)
        - Vector-based analysis of 'suitable' sites (based on several criteria)
    - Short project proposal
    - Homework 3
        - Additional consideration of rasterized senior-density surface
- Lab 7: Raster Modeling with ArcView's Spatial Analyst
- Lab 8: Address matching
- Lab 9: ArcIMS (web mapping)
- In-class test
- Project presentation

Suggested Additional Readings on Raster Models

The NCGIA Core Curriculum in GIScience			
Unit	**TOC Section**	**Unit**	**Author**
Table of Contents (TOC)			
Representing Fields	2.4	054	Michael F.Goodchild
Rasters	2.4.1	055	Michael F. Goodchild
Representing Networks	2.6	064	F. Benjamin Zhan

Worboys, Michael F. *GIS: A Computing Perspective.* London: Taylor & Francis, 1995.
Chapter 4: Models of Spatial Information

More abstract, general, and mathematical than the NCGIA core curriculum notes

(Minimal discussion of raster models in the Ormsby 'Getting to Know ArcGIS' book)