Geography 38/42:286
GIS 1

Topic 5:
Raster and Vector Data Models

Chapters 3 & 4: Chang
(Chapter 4: DeMers)

1

The Nature of Geographic Data

- Most features or phenomena occur as either:
  - discrete entities (objects/PLPs)
  - continuously varying phenomena (fields/CS)
  - some could go either way

2

GIS Data Representation

- How PLPs and CS are stored
- Two primary GIS data models
  - Raster
  - Vector

3

Data Structures and Data Models

- "Raster" and "Vector" refer to a particular data model
- A data structure refers to a particular implementation of either the raster or vector model

4

Topology

- The spatial relationships between things
- An important distinction between data models and data structures
- Can be stored or calculated "on the fly"

5

Raster – Vector Data Models

- Numerous differences in terms of:
  1.
  2.
  3.
  4.
  5.
  6.

6

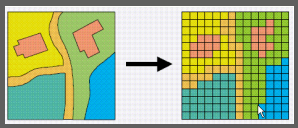
Database Models

- Also differences in how attribute data are stored

7

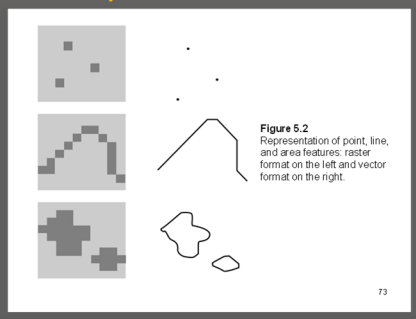
Raster Data Model

- Study area divided up into a regular array of discrete grid cells of uniform shape and size



8

Raster Representation of PLPs



9

Sources of Raster Data

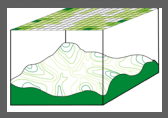
- Common raster data sources include:
  - Imagery (satellite, aerial photography)
  - Classified imagery
  - DEMs, DSMs, DTMs, etc.
  - Scanned products (e.g. maps)
  - Interpolated or generated "surfaces"

10

Thematic vs. Discrete

- Thematic rasters - discrete representations

- Continuous rasters - continuous representations



11

Accuracy of Raster Data

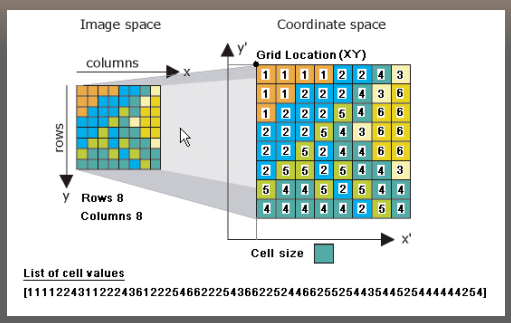
- Determined by grid cell size
- Referred to as resolution
  - User defined
  - Sensor design

12

Projecting Raster Data

- Grid cells represent same location on the ground
- Georeferencing aligns raster layer to real world coordinate system

13



14

Advantages of Raster Model

- Advantages:
  - Easy to understand
  - Mathematical operations
  - Representing continuous phenomena

15

Disadvantages of Raster Model

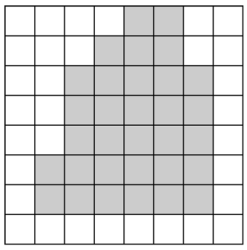
- Disadvantages:
  - Representing discrete phenomena
  - Data redundancy
  - No topology (for discrete data)
  - Spatial accuracy

16

Raster Data Structures

- Method by which raster model is implemented in a particular GIS software application
- Typically, user is unaware of how data structure works

17

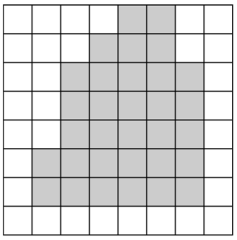


Cell by Cell

Row 1: 0 0 0 0 1 1 0 0
Row 2: 0 0 0 1 1 1 0 0
Row 3: 0 0 1 1 1 1 1 0
Row 4: 0 0 1 1 1 1 1 0
Row 5: 0 0 1 1 1 1 1 0
Row 6: 0 1 1 1 1 1 1 0
Row 7: 0 1 1 1 1 1 1 0
Row 8: 0 0 0 0 0 0 0 0

Figure 5.8
The cell-by-cell data structure records each cell value by row and column.

79



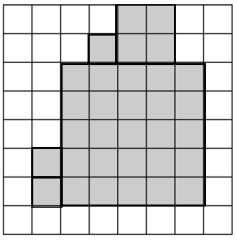
Row 1: 0 0 0 0 1 1 0 0
Row 2: 0 0 0 1 1 1 0 0
Row 3: 0 0 1 1 1 1 1 0
Row 4: 0 0 1 1 1 1 1 0
Row 5: 0 0 1 1 1 1 1 0
Row 6: 0 1 1 1 1 1 1 0
Row 7: 0 1 1 1 1 1 1 0
Row 8: 0 0 0 0 0 0 0 0

Run Length Encoding

8,8,1
0,4,1,2,0,2
0,3,1,3,0,2
0,2,1,5,0,1
0,2,1,5,0,1
0,2,1,5,0,1
0,1,1,6,0,1
0,1,1,6,0,1
0,8

Figure 5.9
The run length encoding method records the cell values in runs. Row 1, for example, has two adjacent cells in columns 5 and 6 that are gray or have the value of 1. Row 1 is therefore encoded with one run, beginning in column 5 and ending in column 6. The same method is used to record other rows.

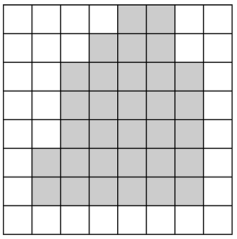
80



Row 1: 0 0 0 0 1 1 0 0
Row 2: 0 0 0 1 1 1 0 0
Row 3: 0 0 1 1 1 1 1 0
Row 4: 0 0 1 1 1 1 1 0
Row 5: 0 0 1 1 1 1 1 0
Row 6: 0 1 1 1 1 1 1 0
Row 7: 0 1 1 1 1 1 1 0
Row 8: 0 0 0 0 0 0 0 0

Block Codes

1,1,2,4,6,2,7,2
4,1,1,5
25,1,3,3



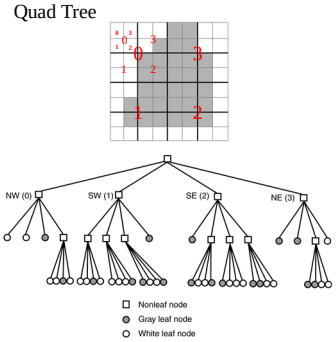
Row 1: 0 0 0 0 1 1 0 0
Row 2: 0 0 0 1 1 1 0 0
Row 3: 0 0 1 1 1 1 1 0
Row 4: 0 0 1 1 1 1 1 0
Row 5: 0 0 1 1 1 1 1 0
Row 6: 0 1 1 1 1 1 1 0
Row 7: 0 1 1 1 1 1 1 0
Row 8: 0 0 0 0 0 0 0 0

Chain Codes

0=N, 1=E, 2=S, 3=W

1,5,1
1,1,2,2,1,1,2,4,3,5,0,1,1,1,0,3
1,1,0,1,1,1,0,1

Quad Tree



(02, 032), (102, 113, 120, 123, 13), (20, 210, 213, 220, 230, 231), (30, 31, 320, 321)

Figure 5.10
The regional quad tree method divides a raster into a hierarchy of quadrants. The division stops when a quadrant is made of cells of the same value (gray or white). A quadrant that cannot be subdivided is called a leaf node. In the diagram, the quadrants are indexed spatially: 0 for NW, 1 for SW, 2 for SE, and 3 for NE. Using the spatial indexing method and the hierarchical quad tree structure, the gray cells can be coded as 02, 032, and so on. See text for more explanation.

81

Mr SID

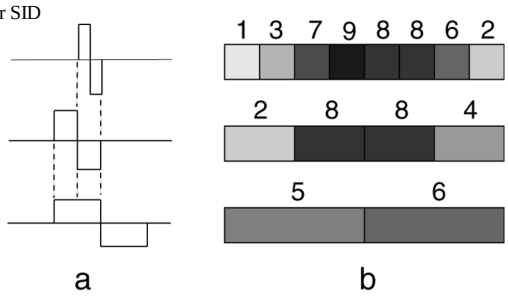
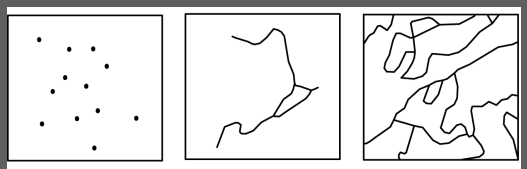


Figure 5.11
The Haar wavelet and the wavelet transform. (a) Three Haar wavelets at three scales (resolutions). (b) A simple example of the wavelet transform.

82

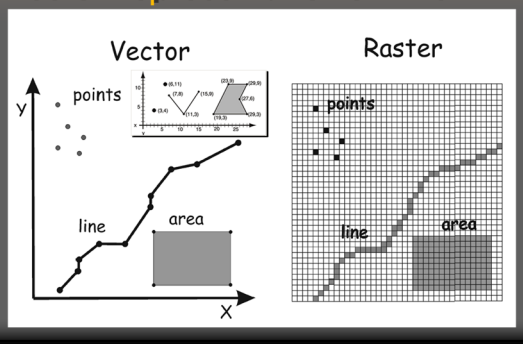
Vector Data Model

- Study area is referenced to a coordinate system that is used to identify the location of points, lines, and polygons using x, y, (z) coordinates



24

Vector Representation of PLPs



25

Vector Data Sources

- Common vector data sources include:
  - NTDB data (National Topographic Database)
  - MLI data (1:20k roads, hydrology, etc.)
  - Tiger Data (U.S. Cesus)
  - GPS or Total Station survey data
  - Digitized data

26

Vector Data Model

- Advantages:
  - Representation of discrete objects
  - No data redundancy
  - Spatial accuracy
  - Topology

27

Vector Data Model

- Disadvantages:
  - Mathematical analysis
  - Not suitable for continuously distributed data

28

Vector Data Models

- Non-topological
- Topological

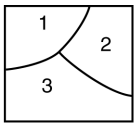
29

Vector Data Models

- Today there are two basic vector data models
  - Georelational Vector Data Model
    - Arc/Info coverages & ArcView shapefiles
  - Object-based Vector Data Model
    - ArcGIS geodatabases

30

Georelational Data Model



Graphic Files

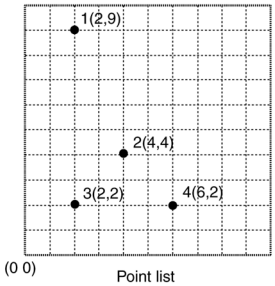
Polygon/arc list
Arc-coordinate list
Left/right list
.
.
.

INFO File

Polygon-ID	Field 1	· · ·
1		
2		
3		

Figure 3.2
Based on the georelational data model, an ArcInfo coverage has two components: graphic files for spatial data and INFO files for attribute data. The label connects the two components.

40

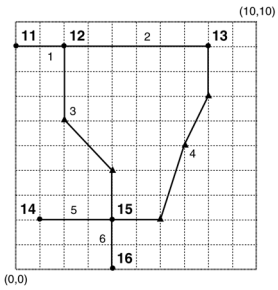


Point list

ID	*x,y*
1	2,9
2	4,4
3	2,2
4	6,2

Figure 3.7
The data structure of a point coverage.

45



Arc-node list

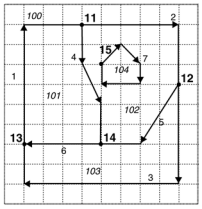
Arc#	F-node	T-node
1	11	12
2	12	13
3	12	15
4	13	15
5	15	14
6	15	16

Arc-coordinate list

Arc#	*x,y* Coordinates
1	(0,9) (2,9)
2	(2,9) (8,9)
3	(2,9) (2,6) (4,4) (4,2)
4	(8,9) (8,7) (7,5) (6,2) (4,2)
5	(4,2) (1,2)
6	(4,2) (4,0)

Figure 3.8
The data structure of a line coverage.

46



Left/right list

Arc#	L-poly	R-poly
1	100	101
2	100	102
3	100	103
4	102	101
5	103	102
6	103	101
7	102	104

Polygon/arc list

Polygon #	Arc#
101	1,4,6
102	4,2,5,0,7
103	6,5,3
104	7

Arc-coordinate list

Arc#	*x,y* Coordinates
1	(1,3) (1,9) (4,9)
2	(4,9) (9,9) (9,6)
3	(9,6) (9,1) (1,1) (1,3)
4	(4,9) (4,7) (5,5) (5,3)
5	(9,6) (7,3) (5,3)
6	(5,3) (1,3)
7	(5,7) (6,8) (7,7) (7,6) (5,6) (5,7)

Figure 3.9
The data structure of a polygon coverage.

47

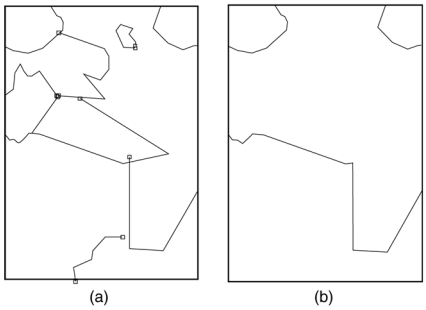


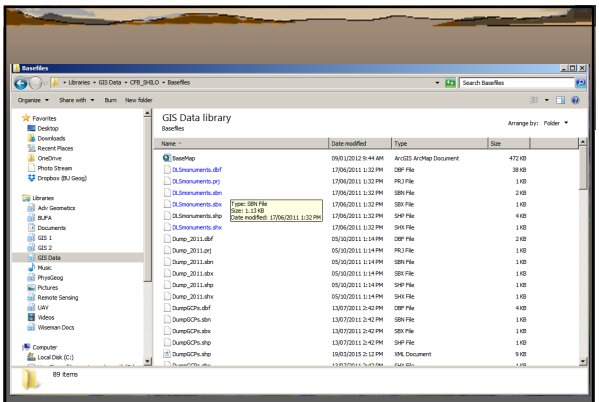
Figure 3.10
A polygon coverage, shown in *a*, has topological errors. Each small square symbol represents an error caused by lines that do not meet correctly. The shapefile, shown in *b*, is converted from the polygon coverage.

48

Georelational Data Model

- ArcView shapefiles
  - Non-topological vector data model
  - Topology on the fly
  - Advantages

36



37

Object-based Model

- Based on object oriented programming theory
- Objects:
  - represent features
  - have defined properties and methods
  - can be organized into classes

38

Geodatabase Model

- Geospatial features represented using objects
- Features have properties and methods
- Features and attributes stored in single file/database

39

Binary field storing feature geometry

Objectid	Shape	Landuse_ID	Category	Shape_Length	Shape_Area
1	Polygon	1	5	14607.7	5959800
2	Polygon	2	8	16979.3	5421216
3	Polygon	3	5	42654.2	21021728

Figure 4.1
The object-based data model stores each land use polygon in a record. The Shape field stores the spatial data of land use polygons. Other fields store attribute data such as Landuse_ID and Category.

78

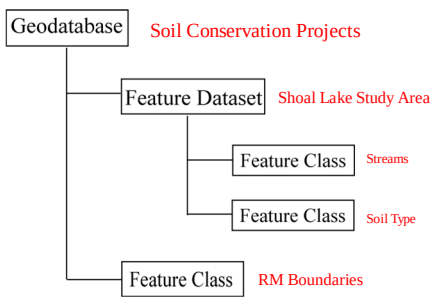


Figure 4.8
In a geodatabase, feature classes can be standalone feature classes or members of a feature dataset.

89

Feature Type	Rule
Polygon	must not overlap, must not have gaps, must not overlap with, must be covered by feature class of, must cover each other, must be covered by, boundary must be covered by, area boundary must be covered by boundary of, and contains point
Line	must not overlap, must not intersect, must not have dangles, must not have pseudo-nodes, must not intersect or touch interior, must not overlap with, must be covered by feature class of, must be covered by boundary of, endpoint must be covered by, must not self overlap, must not self intersect, and must be single part
Point	must be covered by boundary of, must be properly inside polygons, must be covered by endpoint of, and must be covered by line

Table 4.1 Topology rules in the geodatabase data model

94

ArcGIS Geodatabase Structure

- Advantages
  - Validation rules
    - Topology Rules
    - Attribute Domains
    - Relationship Rules
    - Connectivity Rules
    - Custom Rules

43

Raster – Vector Conversion

- Rasterization
- Vectorization

44

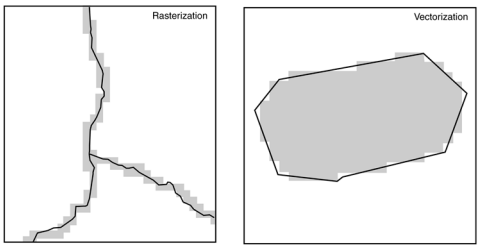


Figure 5.12
On the left is an example of conversion from vector to raster data, or rasterization. On the right is an example of conversion from raster to vector data, or vectorization.

115

Why Convert?

- Rasterize
  - statistical or other mathematical operations
- Vectorise
  - extract thematic information from classified RS imagery (roads, rivers, land cover classes)
  - scanned maps (ArcScan)

46

Why Integrate?

- As an image background
- Image hotlinks
- To subset and classify imagery
- Still seldom used in analysis together

47

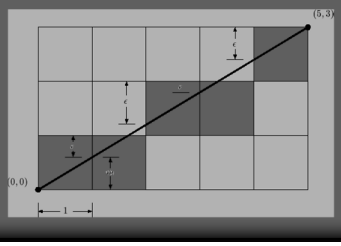
Rasterization

- specify output grid cell size
- attribute to be used as cell value
- grid cell size v. important
  - ½ min. dimension of smallest feature
- inevitable distortion

48

Line Rasterization

- Straight line rasterization
- Bresenham's Line Algorithm

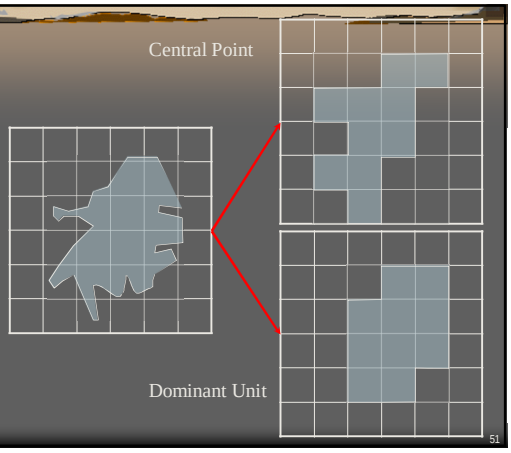


49

Polygon Rasterisation

- How are boundary pixels handled?
  - Central Point
  - Dominant Unit
  - Ranked List

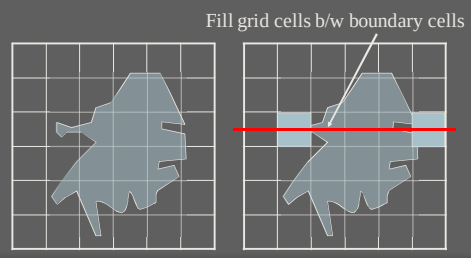
50



51

Scan Line Coherence

- Common polygon rasterisation algorithm



52

Vectorisation

- V. imp for feature extraction
  - Scanned map products
  - RS imagery second largest source of geospatial data

53

Figure 6.11
Large format drum scanners. (Courtesy of GTCO Calcomp, Inc.)

136

Scanning Maps

- Considerations:
  - Colour depth
  - Resolution

55

Vectorization

- For classified RS images
  - Reclassification
- For scanned maps
  - Image preprocessing
  - Thresholding
  - Bilevel images

56

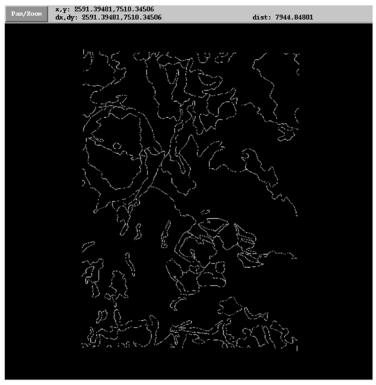


Figure 6.12
A binary scanned file: the lines are soil lines, and the black areas are the background.

138

Vectorization

- Editing/pre-processing boolean image
  - Remove gaps
  - Noise
  - Line thinning

58

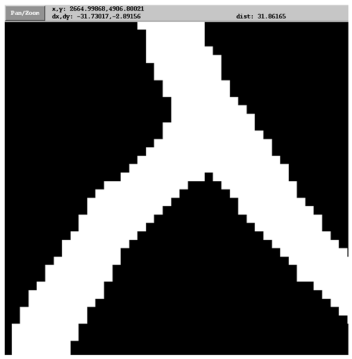


Figure 6.13
A raster line in a scanned file has a width of several pixels.

139

Vectorisation

- Vectorization of boolean image
  - Node detection
- Cleaning
- Build topology

60