



The Raster GIS

presented by:

Tim Haithcoat

**University of Missouri
Columbia**



12

From materials of:

Dan Tomlin, Ohio State University
Micha Pazner, University of Manitoba

The Data Model

- Geographical variation in the real world is infinitely complex
 - The closer you look, the more detail you see, almost without limit
- It would take an infinitely large database to capture the real world precisely
 - Data must somehow be reduced to a finite and manageable quantity by a process of generalization or or abstraction
 - Geographical variation must be presented in terms of discrete elements or objects

The Data Model Defined

- Data Model: the rules used to convert real geographical variation into discrete objects
- Tscichritzis & Lochovsky (1977) define a data model as “as set of guidelines for the representation of the logical organization of the data in a database...(consisting) of named logical units of data and the relationships between them.”

The Data Model

- Current GISs differ according to the way they organize reality through the data model
- Each model tends to fit certain types of data and applications better than others
- The data model chosen for a particular project or application is also influenced by:
 - The software available
 - The training of the key individuals
 - Historical precedent
- 2 major choices of data models: Raster & Vector



Raster Model

- Divides the entire study area into a regular grid of cells in specific sequence
 - The conventional sequence is row by row from the top left corner
 - Each cell contains a single value
 - Is space-filling since every location in the study area corresponds to a cell in the raster
 - One set of cells and associated values is a layer
 - There may be many layers in a database
 - Examples: soil type, elevation, land use, land cover
- Tells what occurs everywhere - at each place in the area

Vector Model

- Uses discrete line segments or points to identify locations
 - Discrete objects (boundaries, streams, cities) are formed by connecting line segments
 - Vector objects do not necessarily fill space, not all locations in space need to be referenced in the model
- Tells where everything occurs - gives a location to every object

Rasters are the simplest
of the available data models.

thus, we begin our examination
of GIS data and operations
with the raster model and will
consider vector models after
the fundamental concepts have
been introduced....

Let's Begin!

Creating a Raster

- Consider laying a grid over a geologic map
 - Create a raster by coding each cell with a value that represents the rock type which appears in the majority of that cells areas
 - When finished, every cell will have a coded value
- In most cases, the values that are to be assigned to each cell in the raster are written into a file, often coded in ASCII
 - This file can be created manually by using a word processor, database or spreadsheet program or it can be created automatically
 - Then it is normally imported into the GIS so that the program can reformat the data for its specific processing needs

Cell by Cell Entry

- Direct entry of each layer cell by cell is simplest
 - Entry may be done within the GIS or into an ASCII file for importing
 - Each program will have specific requirements
- The process is normally tedious and time consuming
 - Layer can contain millions of cells
 - Average landsat image is around 7.4×10^6 pixels, average TM scene is about 34.9×10^6 pixels

Cell by Cell Entry (continued)

- Run length encoding can be more efficient
 - Values often occur in runs across several cells
 - This is a form of spatial autocorrelation - tendency or nearby things to be more similar than distant things
 - Data entered as pairs, first run length, then value

- Example: this array would be

entered as 3 0 2 1 2 0 3 1 2 0 3 1 1 0 4 1

- This is 16 items to enter, instead of 20
- In this case the saving is 20%, but much higher savings occur in practice

0	0	0	1	1
0	0	1	1	1
0	0	1	1	1
0	0	1	1	1
0	1	1	1	1

Cell by Cell Entry (continued)

- Imagine a database of 10,000,000 cells and a layer which records the county containing each pixel
 - Suppose there are only two counties in the area covered by the database
 - Each cell can have only one of only two values so the runs will be very long
- Only some GISs have the capability to use run length encoded files

Digital Match

- Much raster data is already in digital form, as images, etc.
 - however, resampling will likely be needed in order that pixels coincide in each layer
- Because remote sensing generates images, it is easier to interface with a raster GIS than any other type
- Elevation data is commonly available in digital raster form from agencies such as the US Geological Survey

Cell Values: Types of Values

- The type of values contained in cells in a raster depend upon both the reality being coded and the GIS
- Different systems allow different classes of values, including:
 - Whole number (integers)
 - Real (decimal) values
 - Alphabetic values
 - Many systems only allow integers, others which allow different types restrict each separate raster layer to a single kind of value

Cell Values: Types of Values (continued)

- If systems allow several types of values, (e.g. some layers numeric, some non-numeric), they should warn the user against doing unreasonable operations
 - Example: it is unreasonable to try to multiply the values in a numeric layer with the values in a non-numeric layer
- Integer values often act as code numbers, which “point” to names in an associated table or legend
 - Example: the first example might have the following legend identifying the name of each soil class:

0 = “no class”	1 = “fine sandy loam”
2 = “coarse sand”	3 = “gravel”

Cell Values: One Value per Cell

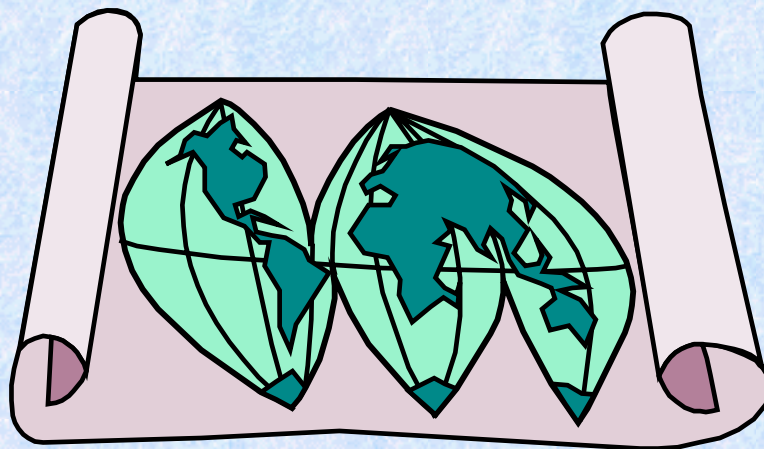
- Each pixel or cell is assumed to have only one value
 - This is often inaccurate -the boundary of two soil types may run across the middle of a pixel
 - In such cases the pixel is given the value of the largest fraction of the cell, or the value of the middle point in the cell
- Note, however, a few systems allow a pixel to have multiple values
 - The NARIS system developed at the University of Illinois in the 1970s allowed each pixel to have any number of values and associated percentages
 - Example: 30% a, 30% b, 40% c

Map Layers

- The data for an area can be visualized as a set of maps of layers
 - A map layer is a set of data describing a single characteristic for each location within a bounded geographic area
- Only one item of information is available for each location within a single layer - multiple items of information require multiple layers
 - On the other hand, a topographic map can show multiple items of information for each location
 - Example: elevation (contours), counties (boundaries), roads, railroads, urbanized areas (gray tint)
 - These would be 5 layers in raster GIS

Map Layers (continued)

- Typical raster databases contain up to a hundred layers
 - Each layer (matrix, lattice, raster, array) typically contains hundred or thousands of cells
- Important characteristics of a layer are its resolution, orientation and zone(s)

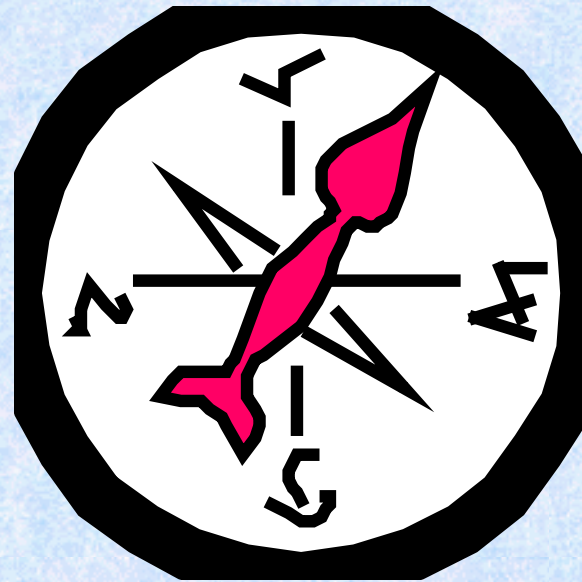


Resolution

- In general, resolution can be defined as the minimum linear dimension of the smallest unit of geographic space for which data are recorded
- In the raster model the smallest units are generally rectangular (occasionally systems have used hexagons or triangles)
- The smaller units are known as cells, pixels
- Note: high resolution refers to rasters with small cell dimensions
 - High resolution means lots of detail, lots of cells, large rasters, small cells

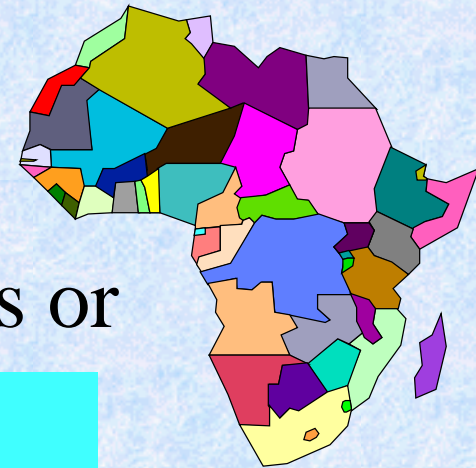
Orientation

- The angle between true north and the direction defined by the columns of the raster



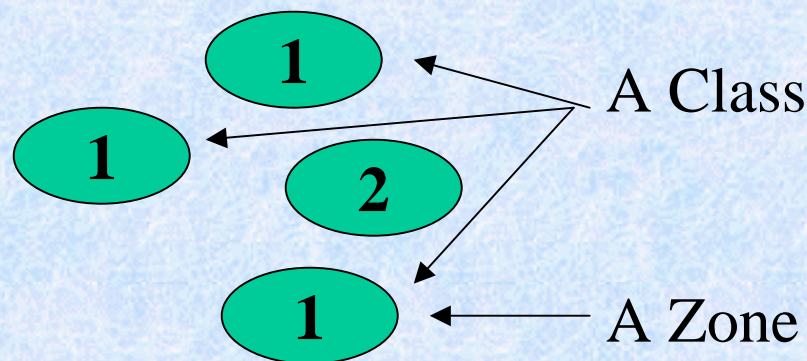
Zones

- Each zone of a map layer is a set of contiguous locations that exhibit the same value
- These might be:
 - Ownership parcels
 - Political units such as counties or nations
 - lakes or islands
 - Individual patches of the same soil or vegetation type



Zones (continued)

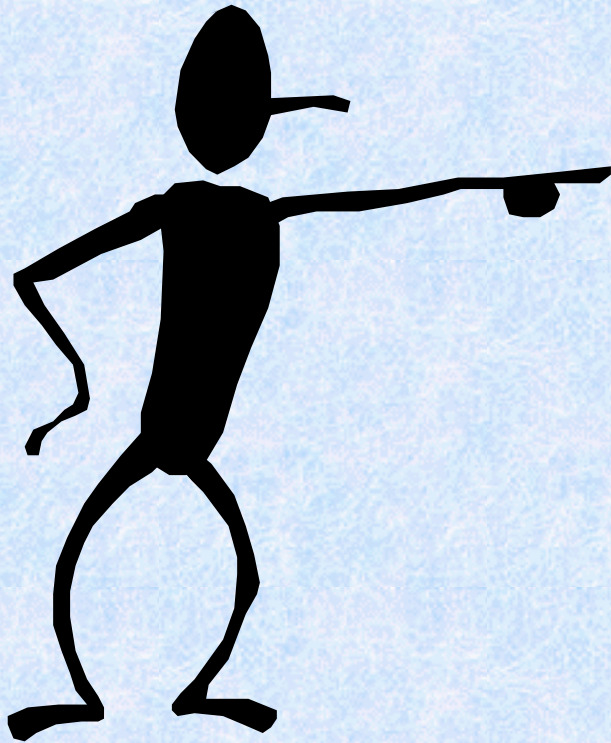
- There is considerable confusion over terms here
 - Other terms commonly used are patch, region, polygon
 - Each of these terms, however, have different meanings to individual users & different definitions in specific GIS packages
 - In addition, there is a need for a second term which refers to all individual zones that have the same characteristics
 - Class is often used for this concept:



Zones (continued)

- Note that all map layers will have zones, cell contents may vary continuously over the region making every cell's value unique
 - Example: satellite sensors record a separate value for reflection from each cell
- Major components of a zone are its value and locations

Value

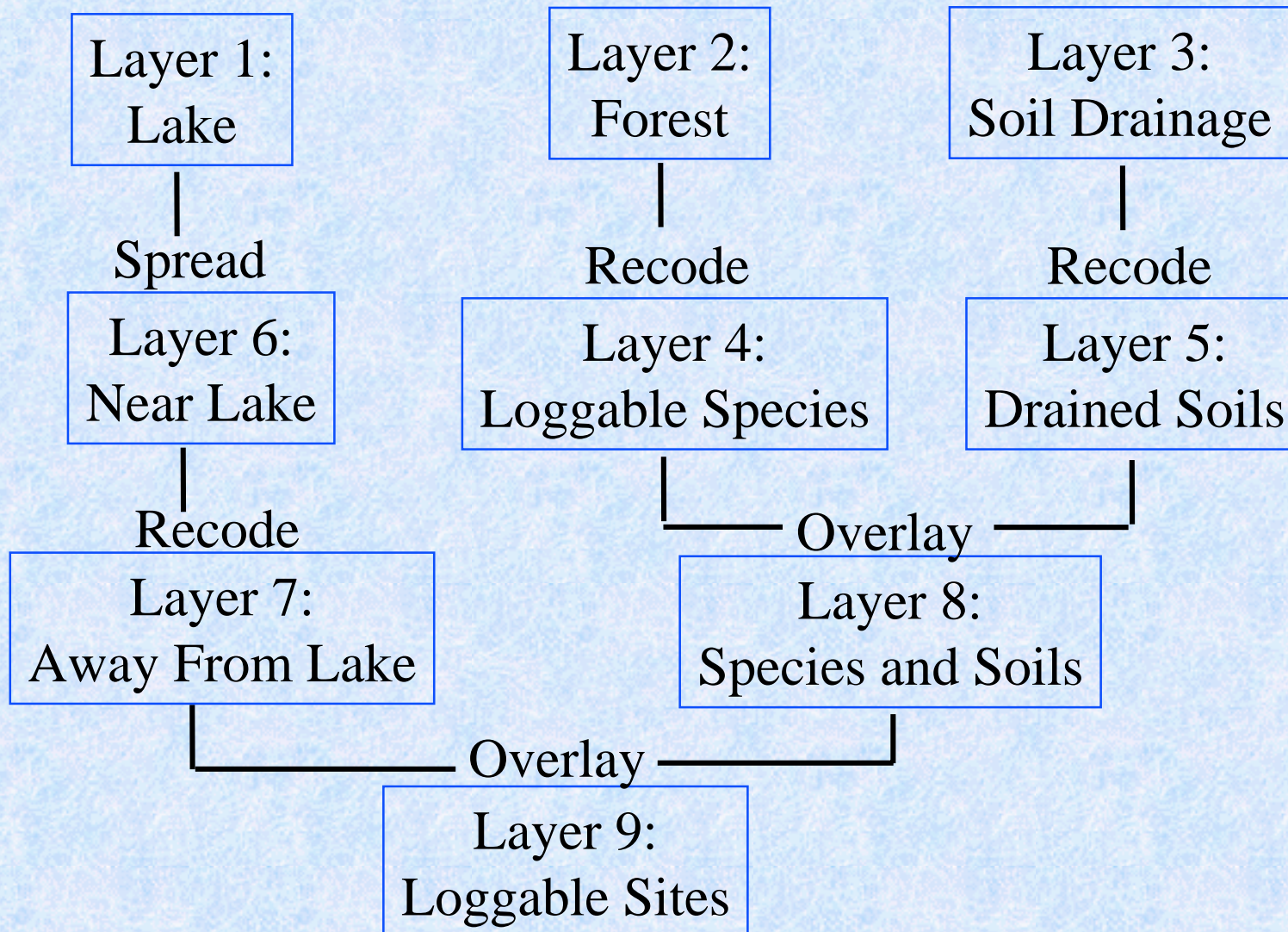


- Is the item of information stored in a layer for each pixel or cell
- Cells in the same zone have the same value

Location

- Generally location is identified by an ordered pair of coordinates (row and column numbers) that unambiguously identify the location of each unit of geographic space in the raster (cell, pixel, grid cell)
- Usually the true geographical location of one or more of the corners of the raster is also known

Example Project Steps



Example Analysis: Using A Raster GIS

- **OBJECTIVE:** Identify areas suitable for logging
- An area is suitable if it satisfies the following criteria:
 - is Jackpine (Black Spruce are not valuable)
 - is well drained (poorly drained and waterlogged terrain cannot support equipment, logging causes unacceptable environmental damage)
 - is not within 500 m of a lake or watercourse (erosion hazard)

PROCEDURE

Example Project Details (1 of 4)

RECODE LAYER 2 - Y IF VALUE 2 (JACKPINE), N IF OTHER VALUE

LAYER 2: FOREST SPECIES

0	0	2	2	2
0	0	2	2	2
0	1	1	1	2
1	1	1	1	1
1	1	1	1	1

LAYER 4: LOGGABLE SPECIES

n	n	y	y	y
n	n	y	y	y
n	n	n	y	y
n	n	n	n	n
n	n	n	n	n

RECODE LAYER 3 - Y VALUE 2 (GOOD), N IF OTHER VALUE

LAYER 3: SOIL DRAINAGE

0	0	2	2	2
0	0	1	1	2
0	0	1	1	2
0	1	1	1	2
1	1	1	2	2

LAYER 5: DRAINED SOIL

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	y
n	n	n	y	y

PROCEDURE

Example Project Details (2 of 4)

SPREAD THE LAKE ON LAYER 1 BY ONE CELL (500 M)

LAYER 1: LAKE

1	1	0	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0
0	0	0	0	0

LAYER6: NEAR LAKE

1	1	1	0	0
1	1	1	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0

RECODE LAYER 3 - Y VALUE 2 (GOOD), N IF OTHER VALUE

LAYER 6: NEAR LAKE

1	1	1	0	0
1	1	1	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0

LAYER 7: AWAY FROM LAKE

n	n	n	y	y
n	n	n	y	y
n	n	y	y	y
n	y	y	y	y
y	y	y	y	y

PROCEDURE

Example Project Details (3 of 4)

OVERLAY LAYERS 4 & 5 - IF BOTH 4 & 5 ARE Y, N OTHERWISE

LAYER 4: LOGGABLE SPECIES

n	n	y	y	y
n	n	y	y	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

LAYER 5: DRAINED SOIL

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	y
n	n	n	y	y

LAYER 8: SPECIES & SOILS

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

PROCEDURE

Example Project Details (4 of 4)

OVERLAY LAYERS 7 & 8 - IF BOTH 7 & 8 ARE Y, N OTHERWISE

LAYER 7: AWAY FROM LAKE

n	n	n	y	y
n	n	n	y	y
n	n	y	y	y
n	y	y	y	y
y	y	y	y	y

LAYER 8: SPECIES & SOILS

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

LAYER 9: LOGGABLE SITES

n	n	n	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

RESULT:

The loggable cells are
y on layer 9.

Operations Used

RECODE

OVERLAY

SPREAD

- We could have achieved the same result using the operations in other sequences, or by combining recode and overlay operations
 - Example: overlay layers 2 and 3, coding as follows
 - Y if layer 2 is 2 and layer 3 is 2, n otherwise
 - This would replace two recodes and an overlay
 - Some systems allow layers to be overlaid 3 or more at a time
- The names given to operations vary from system to system, but most of the operations themselves are common across systems

Raster GIS Capabilities

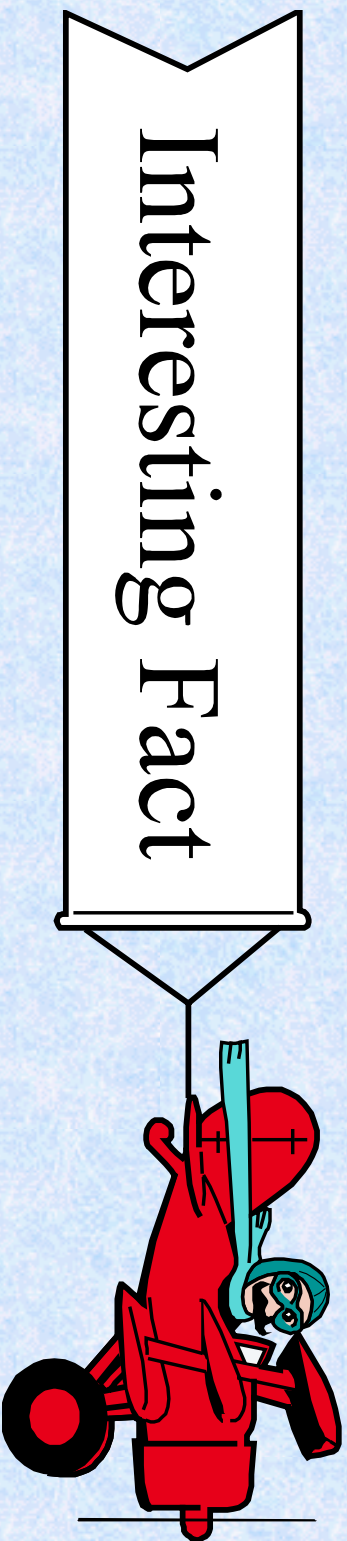
- A raster GIS **MUST** have capabilities for:
 - Input of data
 - Various housekeeping functions
 - Operations on layers (recode, overlay, spread)
 - Output of data and results
- The range of possible functions is enormous, current raster GISs only scratch the surface
 - Because the range is so large, some have tried to organize functions into a consistent scheme, but no scheme has been widely accepted yet
- Each raster GIS uses different names for the functions

Displaying Layers: Basic Display

- The simplest type of values to display are integers
 - On a color display each integer value can be assigned a unique color
 - There must be as many colors as integers
- If the values have a natural order we will want the sequence of colors to make sense
 - example: elevation is often shown on a map using the sequence blue-green-yellow-brown-white for increasing elevation
- On a dot matrix printer, shades of gray can be generated by varying the density of dots
- If there are too many values for the number of colors, may have to recode the layer before display

Displaying Layers: Other Types

- It may be appropriate to display the data as a surface
- Contours can be “threaded” through the pixels along lines of constant value
 - The searching operation for finding contours is computer-intensive so may be slow
- The surface can be shown in an oblique, perspective view
 - This can be done by drawing profiles across the raster with each profile offset & hidden lines removed
 - The surface might be colored using the values in a second layer (a 2nd layer can be “draped” over the surface defined by the first layer), the result can be effective
 - These operations are also computer-intensive because of the calculations necessary to simulate perspective and remove hidden lines



“LA the Movie” was created by Jet Propulsion Lab, Pasadena, CA, by draping a Landsat scene over a DEM of LA, then simulating the view from a moving aircraft.

Logical Operations

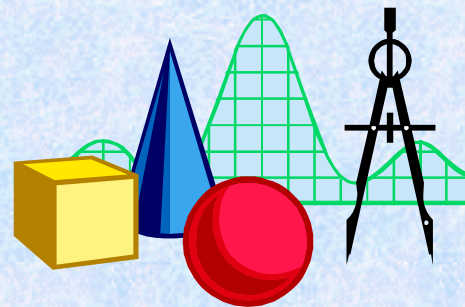
- Produce a new layer from one or more input layers
- The value of each new pixel is defined by the values of the name pixel on the input layer(s)
 - Neighboring or distant pixels have no effect
- Note: arithmetic operations make no sense unless the values have appropriate scales of measurement
 - Example: You cannot find the “average” of soils types 3 and 5, nor is soil 5 “greater than” soil 3

Recoding

- Using only one input layer
- Examples:
 - Assign a new value to each unique value on their input layer
 - Useful when the number of unique input values is small
 - Assign new values by assigning pixels to classes or ranges based on their old values
 - Example: 0-4999 becomes 1, 500-999 becomes 2, >1000 becomes 3
 - Useful when the old layer has different values in each cell (e.g. elevation or satellite images)
 - Sort the unique values found on the input layer and replace by the rank of the value
 - example: 0,1,4,5 on input layer become 1,2,3,4 respectively
 - Applications: assigning ranks to computed scores of capability, suitability, etc.
 - Some systems allow a full range of mathematical operations
 - Example: $\text{newvalue} = (2 * \text{oldvalue} + 3)^2$

Overlaying Layers

- An overlay occurs when the output value depends on two or more input layers
 - Many systems restrict overlay to two input layers only
- Examples:
 - Output value equals arithmetic average of input values
 - Output value equals the greatest (or least) of the input values
 - Layers can be combined using arithmetic operations
 - x and y are the input layers, z is the outputs
 - Some examples:
 - $z = x + y$
 - $z = x^y$
 - $z = s/y$



Overlaying Layers (continued)

- Examples (continued)
 - Combination using logical conditions
 - Example: if $y > 0$, then $z = y$, otherwise $z = x$
 - Note: in many raster packages logical conditions cannot be done directly from input layers
 - Must first create reclassified input images so that cells have 0 if they do not meet the condition and 1 if they do
 - Assign a new value to every unique combination of input values

- Example:

<u>Layer 1</u>	<u>Layer 2</u>	<u>Output Layer</u>
1	A	1
1	B	2
2	A	3
2	B	4

Logical Operations on Raster: LOGICAL AND

WHAT CELLS ARE BOTH A AND 7?

A	A	B
A	A	B
C	C	B

RECLASSIFY

1	1	0
1	1	0
0	0	0

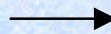
6	7	7
6	7	7
8	8	7

RECLASSIFY

0	1	1
0	1	1
0	0	1

MULTIPLY

1x0	1x1	0x1
1x0	1x1	0x1
0x0	0x0	0x1



0	1	0
0	1	0
0	0	0

Logical Operations on Raster: LOGICAL OR

WHAT CELLS ARE BOTH A OR 7?

A	A	B
A	A	B
C	C	B

6	7	7
6	7	7
8	8	7

RECLASSIFY

1	1	0
1	1	0
0	0	0

RECLASSIFY

0	1	1
0	1	1
0	0	1

ADD

1+0	1+1	0+1
1+0	1+1	0+1
0+0	0+0	0+1

1	2	1
1	2	1
0	0	1

RECLASSIFY

1	1	1
1	1	1
0	0	1

Operations on Local Neighborhoods

- The value of a pixel on the new layer is determined by the local neighborhood of the pixel on the old layer



Filtering

- A filter operates by moving a “window” across the entire raster
 - e.g. many windows are 3 x 3 cells
- The new value for the cell at the middle of the window is a weighted average of the values in the window
- By changing the weights we can produce two major effects:
 - Smoothing (a “low pass” filter, removes or reduces local detail)
 - Edge enhancement (a “high pass” filter, exaggerates local detail)
- Weights should add to 1

Filtering Examples

.11 .11 .11

.11 .11 .11

.11 .11 .11

- Replaces each value by the simple average of it & its eight neighboring values
- Severely smooths the spatial variation on the layer

.05 .05 .05

.05 .60 .05

.05 .05 .05

- Gives the pixel's old value 12 times the weight of its neighboring values
- Slightly smooths the layer

-.1 -.1 -.1

-.1 1.8 -.1

-.1 -.1 -.1

- Slightly enhances local detail by giving neighbors negative weights

Filters can be useful in enhancing detail on images for input to GIS, or smoothing layers to expose general trends.

Slopes and Aspects

- If the values in a layer are elevations, we can compute the steepness of slopes by looking at the difference between a pixel's value and those of its adjacent neighbors
- The direction of steepest slope, or the direction in which the surface is locally “facing”, is called its aspect
 - Aspect can be measured in degrees from North or by compass points - N, NE, E, etc.

Slopes & Aspects (continued)

- Slope & Aspect are useful in analyzing vegetation patterns, computing energy balances and modeling erosion or runoff
 - Aspect determines the direction of runoff
 - This can be used to sketch drainage paths for runoff



Operations on Extended Neighborhoods

- Distance
 - Calculate the distance of each cell from a cell or the nearest of several cells
 - Each pixel's value in the new layer is its distance from the given cell(s)
- Buffer Zones
 - Buffers around objects and features are very useful GIS capabilities
 - Example: build a logging buffer 500 m wide around all lakes and watercourses
 - Buffer operations can be visualized as spreading the object spatially by a given distances

Operations on Extended Neighborhoods

(continued)

- Buffer Zones (continued)
 - the result could be a layer with values:
 - 1 if in original selected object
 - 2 if in buffer
 - 0 if outside object and buffer
 - Applications include noise buffers around roads, safety buffers around hazardous facilities
 - The rate of spreading may be modified by another layer representing “friction”
 - Example: the friction layer could represent varying cost of travel
 - This will affect the width of the buffer - narrow in areas of high friction, etc.

Operations on Extended Neighborhoods

(continued)

- Buffer Zones (continued)
 - In many programs, the buffer operation requires the user to first do a distance operation, then a reclassification of the distance layer
- Visible area or “viewshed”
 - Given a layer of elevations and one or more viewpoints, compute the area visible from at least one viewpoint
 - example: value = 1 if visible, 0 if not
 - Useful for planning location of unsightly facilities such as smokestacks, or surveillance facilities such as fire towers or transmission facilities

Operations on Zones ~ Groups of Pixels

- Identifying zones
 - By comparing adjacent pixels, identify all patches or zones having the same value
 - Give each such patch or zone a unique number
 - Set each pixel's value to the number of its patch or zone
- Areas of zones
 - Measure of area of each zone and assign this value to each pixel instead of the zone's number
 - Alternatively output may be in the form of a summary table sent to the printer or a file

Operations on Zones ~ Groups of Pixels

(continued)

- Perimeter of zones
 - Measure the perimeter of each zone and assign this value to each pixel instead of the zone's number
 - Alternatively output may be in the form of a summary table sent to the printer or a file
 - length of perimeter is determined by summing the number of exterior cell edges in each zone
 - Note: the values calculated in both area & perimeter are highly dependent upon the orientation of objects (zones) with respect to the orientation of the grid
 - However, if boundaries in the study area do not have a dominant orientation such errors may cancel out

Operations on Zones ~ Groups of Pixels

(continued)

- Distance from zone boundary
 - Measure the distance from each pixel to the nearest part of its zone boundary and assign this value to the pixel
 - Boundary is defined as the pixels which are adjacent to pixels of different values
- Shape of zone
 - measure the shape of the zone and assign this to each pixel in the zone
 - One of the most common ways to measure shape is by comparing the perimeter length of a zone to the square root of its area
 - By dividing this number by 3.54 we get a measure which ranges from 1 for a circle (the most compact shape possible) to 1.13 for a square to large numbers for long, thin, wiggly

Operations on Zones ~ Groups of Pixels

(continued)

- Shape of zone (continued)
 - Commands like this are important in landscape ecology
 - Helpful in studying the effects of geometry and spatial arrangement of habitat
 - Example: size and shape of woodlots on the animal species they can sustain
 - Example: value of linear park corridors across urban areas in allowing migration of animal species



Commands

to describe contents of layers

- It is important to have ways of describing a layer's contents
 - Particularly new layers created by GIS operations
 - Particularly in generating results of analysis

ONE LAYER

*Generate statistics
on a layer*

Examples:
mean, median, most
common value, other
statistics

MORE THAN ONE LAYER

*Compare two maps
statistically*

Examples:
Is pattern on one map
related to pattern on the
other?, chi-square test,
regression, analysis of
variance

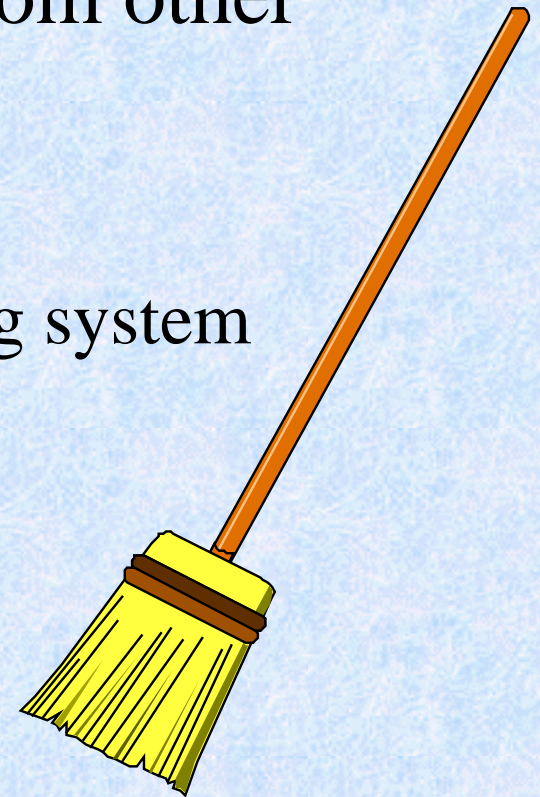
ZONES ON ONE LAYER

*Generate Statistics
for the zones on a
layer*

Examples:
largest, smallest,
number, mean area

Essential Housekeeping

- List available layers
- Input, copy, rename layers
- Import and export layers to and from other systems
 - Other raster GIS
 - Input of images from remote sensing system
 - Other types of GIS
- Identity resolution, orientation



Essential Housekeeping (continued)

- “Resample”
 - change cell size, orientation, portion of raster to analyze
- Change colors
- Provide help to the user
 - Exit from the GIS (the most important command of all!)

