GE2215 Lecture 10 Spatial Analysis - Raster Data Analysis

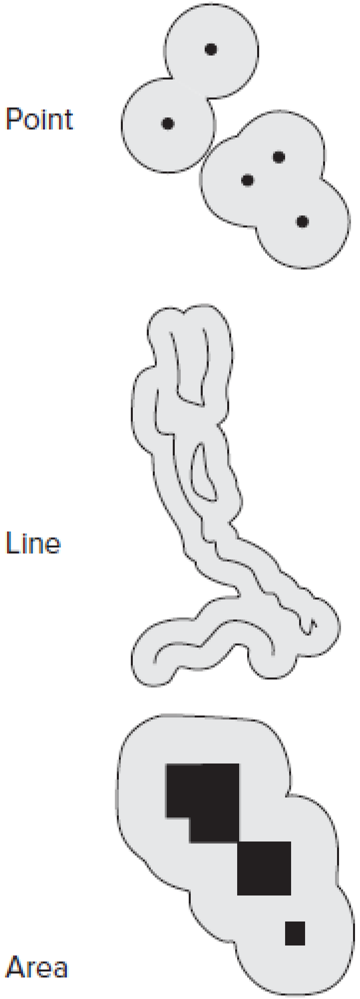
Dr. Yan Yingwei

Department of Geography

National University of Singapore

Recap: Buffering analysis

- Features for buffering
  - Points
  - Lines
  - Polygons
- The buffer zones are saved to a new layer
- Parameters of buffering
  - Buffer size



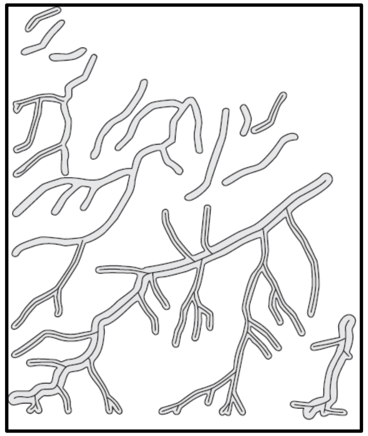
Circular buffer zones

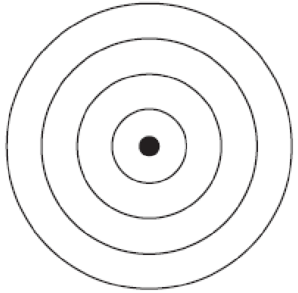
Elongated buffer zones

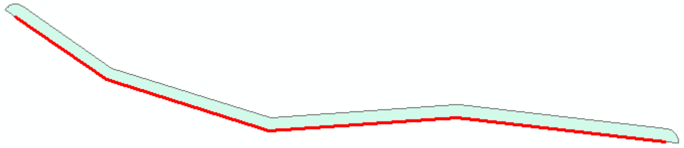
Buffer zones that extend outward from the polygon boundaries

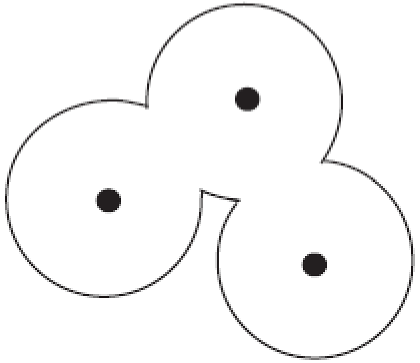
Recap: Variations in buffering

- **Variation 1**: Varying buffer sizes
- **Variation 2**: Multiple buffer zones
- **Variation 3**: Buffer on one side or both sides
- **Variation 4**: Dissolves buffer zones or not



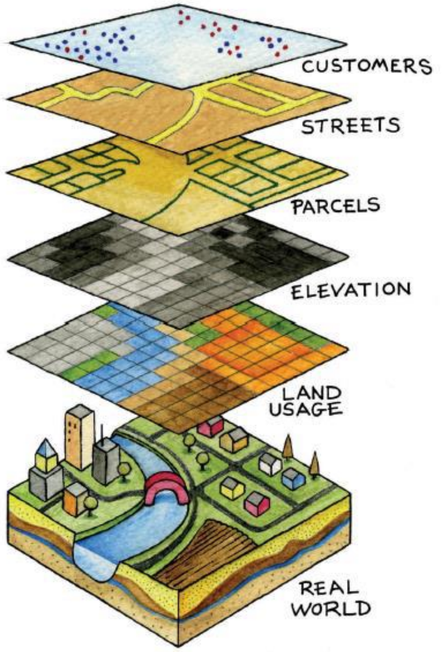






Recap: Overlay analysis

- In GIS, an overlay is the process of taking two or more different maps of the same area and placing them on top of one another to form a new map
- Overlay as editing tool *vs* analytical tool
  - Overlay as an editing tool works only with feature geometries
  - Overlay as an analytical tool works with both geometries and attributes
- Overlay vs buffering
  - Overlay works on multiple feature layers
  - Buffering works on a single feature layer



Recap: Overlay types

- **Point-in-polygon** overlay
- **Line-in-polygon** overlay
- **Polygon-on-polygon** overlay

Recap: Overlay operations

- Union
- Intersection
- Symmetrical difference
- Clip
- Erase
- Identity
- Split

Features to be overlaid must be spatially registered and based on the same coordinate system

Recap: Distance measurement

- **Distance Measurement** refers to measuring straight line distances between features
  - Points in a layer to points in another layer
  - Points in a layer to its nearest point in another layer
- **Distance measures** can be used directly for data analysis
- **Distance measures** can also be used as inputs to data analysis
  - Pattern analysis: analyzing the patterns (random or dispersed or clustered) of distributions of points

Recap: Other feature manipulations

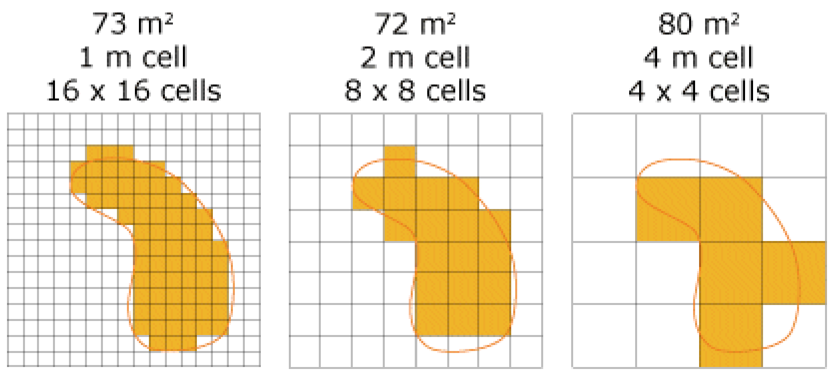
- Single layer analysis
  - Dissolve
  - Select
  - Eliminate

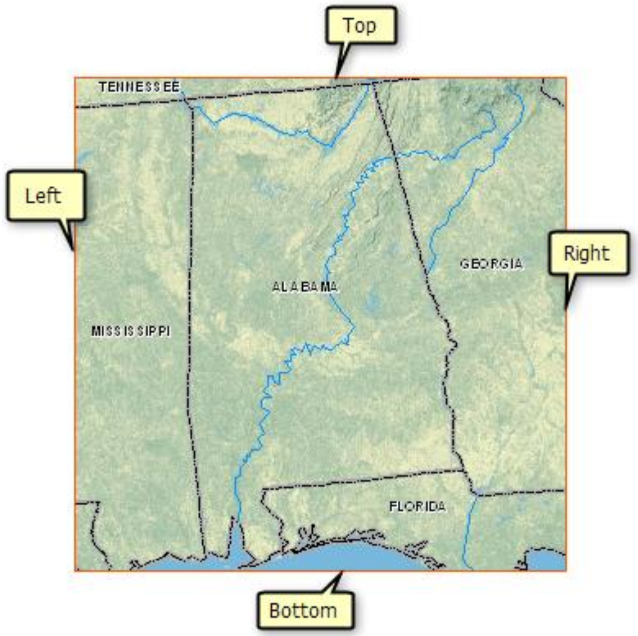
Outline of this lecture

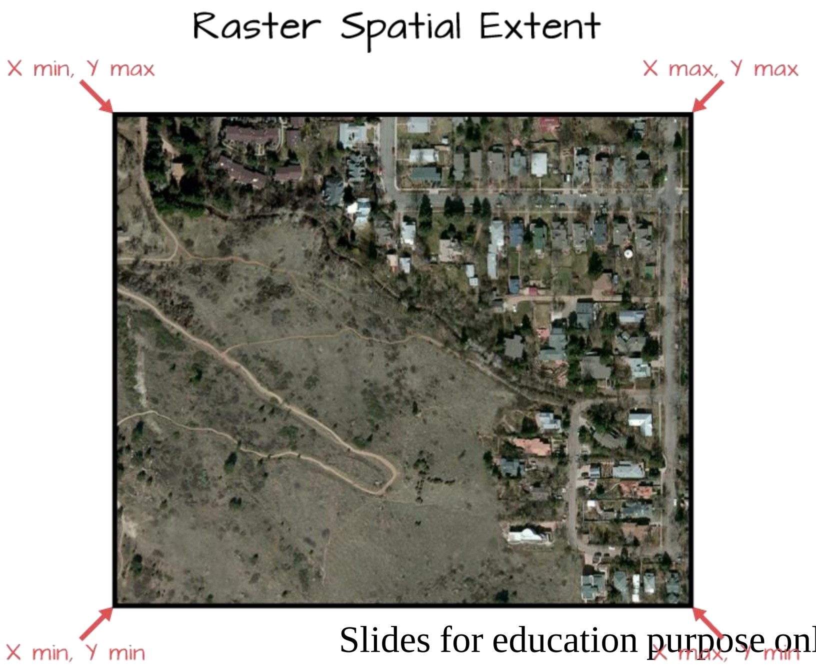
- Data analysis environment
- Raster analysis functions
  - Local operation
  - Focal operation
  - Zonal operation
  - Global operation
- Other raster data operations
- Map algebra

Data analysis environment

- Data analysis environment is defined by two parameters: output cell size and area extent
- An area extent can be a rectangle defined by minimum and maximum x, y coordinates

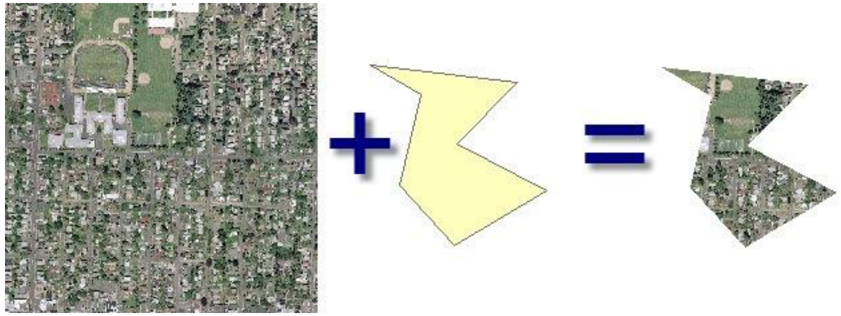


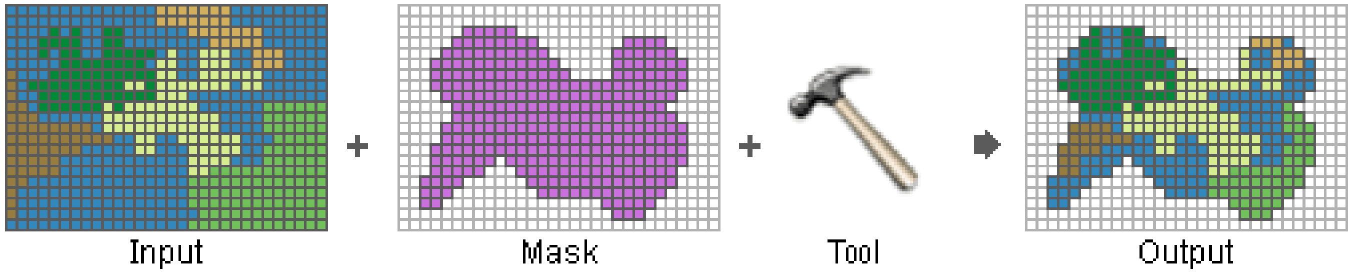




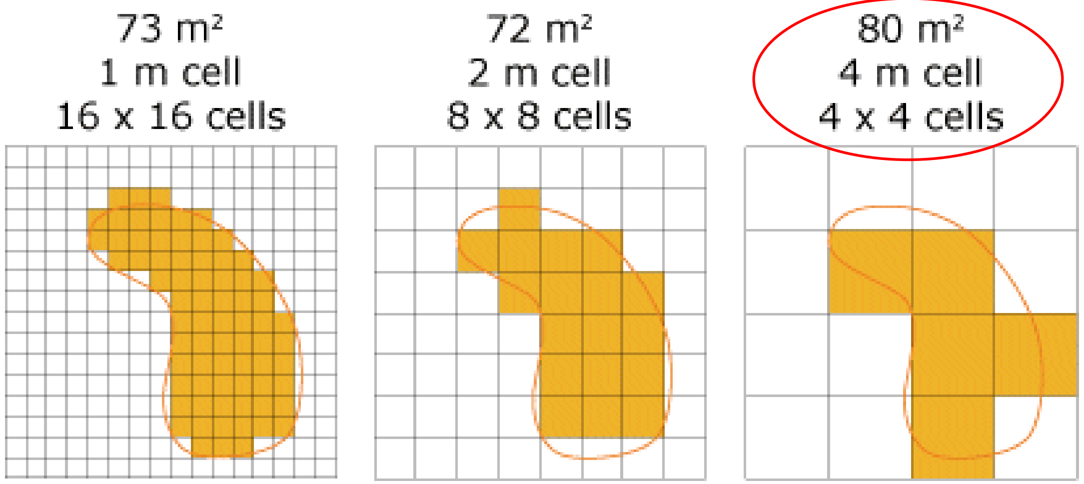
Data analysis environment

- Data analysis environment is defined by two parameters: output cell size and area extent
- An area extent can also have irregular shape defined by a mask, which can be set by:
    - A feature layer
    - A raster





Data analysis environment



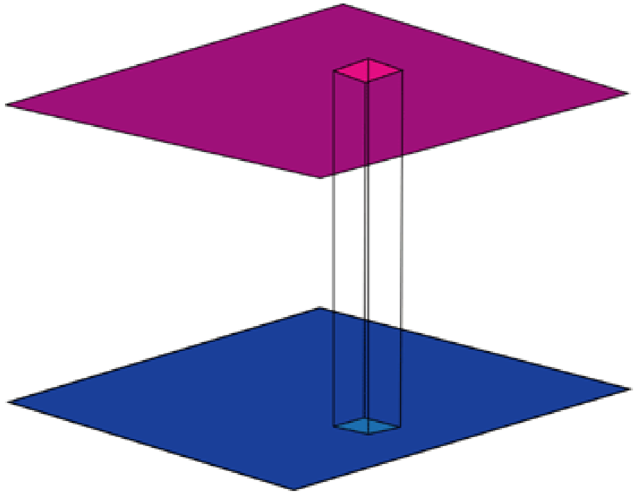
- Output cell size
  - Theoretically, the output cell size can be set at any scale
  - Typically, the output cell size is set to **be equal to**, or **larger than**, the **largest** cell size among the input raster layers
  - The resolution of the output should correspond to that of the lowest-resolution input raster
  - For example, if there are two input raster layers, one with 1 m cell size, and the other one with 4 m cell size, the output cell size should be 4 m or larger

Outline of this lecture

- Data analysis environment
- Raster analysis functions
  - Local operation
  - Focal operation
  - Zonal operation
  - Global operation
- Other raster data operations
- Map algebra

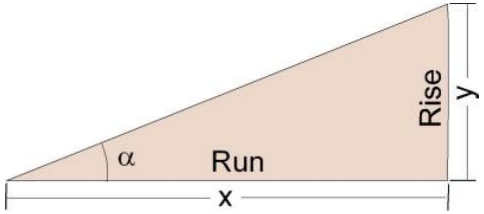
Local operation

- Local operations are cell-by-cell operations
- Local operations can be operated on single or multiple rasters
  - When used on a single raster, a local operation usually takes the form of applying some mathematical transformation to each individual cell
  - When used on multiple rasters, mathematical operations are performed on cells at the same locations of different rasters



Local operations: Single Layer

- Data type conversion



(A) Slope Ratio $= \frac{\text{Rise}}{\text{Run}} = \frac{y}{x}$

(B) Slope Percentage $= \frac{y}{x} \times 100$

(C) Slope Angle (α) $= \tan^{-1}\left(\frac{y}{x}\right)$

1.3	1.2	0.1	0.8
	1.8	2.5	2.7
4.4	-1.9	-0.5	2.9
4.6	0	1.7	1.5

→

1	1	0	0
	1	2	2
4	-1	0	2
4	0	1	1

VALUE=NODATA

Convert a float raster to an integer raster

15.2	16.0	18.5
17.8	18.3	19.6
18.0	19.1	20.2

→

8.64	9.09	10.48
10.09	10.37	11.09
10.20	10.81	11.42

Slope in percentage　　Slope in degree

A local raster can convert a slope raster from percent to degrees

Local operations: Single Layer

- Reclassification
  - Reclassification creates a new raster by classification
    - One-to-one change

Input Land Use Map		Output Raster
Building	→	1
Water		2
Road		3
Vegetation land		4
Bare land		5

Input Land Use Map **Output Raster**

Local operations: Single Layer

- Reclassification
  - Reclassification creates a new raster by classification
    - Assigns a new value to a range of cell values

Input Map with population density		Output Raster
[0, 25)	Persons/ square mile →	1
[25, 50)		2
[50, 75)		3
[75, 100)		4
>=100		5

Input Map with population density **Output Raster**

Local operations: Multiple layers

- Local operations with multiple rasters are the equivalent to vector-based overlay operations
- Local operations with **multiple rasters** include:
  1. Mathematical operations: $+$, $-$, $\times$, $\div$
  2. Summary statistics: maximum, minimum, range, sum, mean, median, and standard deviation, which apply to numeric data only
  3. Tabulation of measures: majority, minority, combine and number of unique values, which apply to both numeric and categorical data

Local operations: Multiple layers

- Mathematical operations

Roads

2	0	0
0	2	0
0	0	2

Roads = 2
Other areas = 0

+

Forestry

0	5	5
0	0	0
5	5	5

Forest = 5
Other area = 0

=

Result

2	5	5
0	2	0
5	5	7

Neither road nor forest = 0
Road, not in forest = 2
Forest, no road = 5
Road in forest area = 7

0	1	1
0	0	0
1	1	1

Forest = 1
Other areas = 0

×

1	1	1
1	1	0
0	0	0

Resort area = 1
Other area = 0

=

0	1	1
0	0	0
0	0	0

Forest *and* resort = 1
Other areas = 0

Local operations: Multiple layers

- Summary statistics

(a)

5	2	3
	2	2
3	1	1

(b)

1	3	2
4	7	5
1	1	

(c)

3	4	1
4	3	2
2	1	1

(d)

3	3	2
	4	3
2	1	

The cell value in (*d*) is the mean calculated from three input rasters (*a*, *b*, and *c*) in a local operation. The shaded cells have no data.

Input

(a)

5	2	3
	2	2
3	1	1

(b)

1	3	2
4	7	5
1	1	

(c)

3	4	1
4	3	2
2	1	1

Output

(d)

5	4	3
	7	5
3	1	

The cell value in (*d*) is the max value calculated from three input rasters (*a*, *b*, and *c*) in a local operation. The shaded cells have no data.

Local operations: Multiple layers

- Tabulation of measures

(a)

3	2	3
	2	2
3	1	1

(b)

1	3	2
1	2	1
1	1	

(c)

3	3	2
1	3	2
1	1	1

(d)

3	3	2
	2	2
1	1	

The cell value in (*d*) is the majority statistic derived from three input rasters (*a*, *b*, and *c*) in a local operation. The shaded cells have no data.

Combination assigns a unique output value to each unique combinations of input values

(a)

3	2	1
2	1	2
1	2	3

(b)

3	2	4
3	2	4
2	4	1

(c)

1	3	6
2	4	5
4	5	7

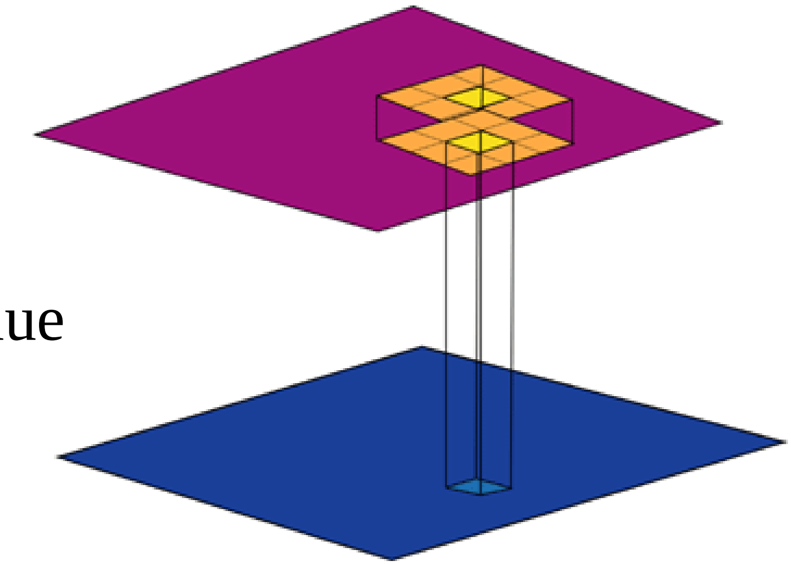
(d)

Combine code	1	2	3	4	5	6	7
(slope, aspect)	(3,3)	(2,3)	(2,2)	(1,2)	(2,4)	(1,4)	(3,1)

Each cell value in (*c*) represents a unique combination of cell values in (*a*) and (*b*). The combination codes and their representations are shown in (*d*).

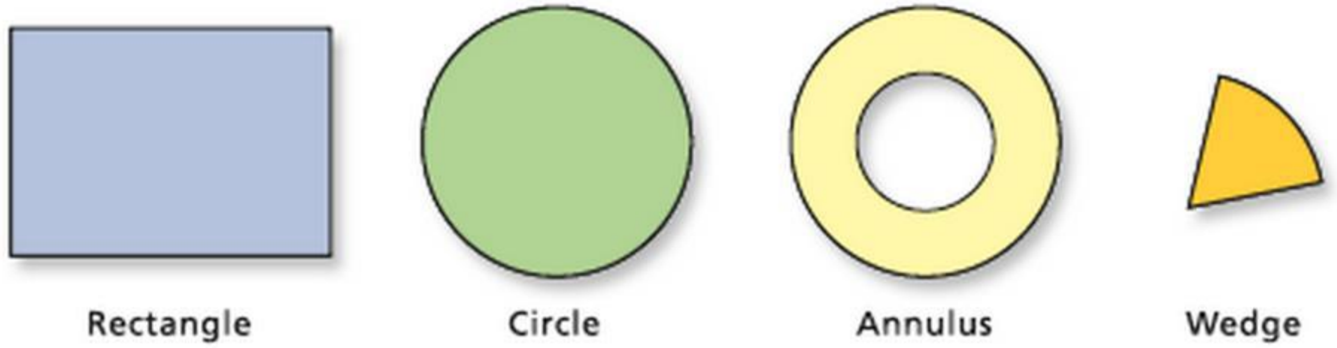
Focal operations

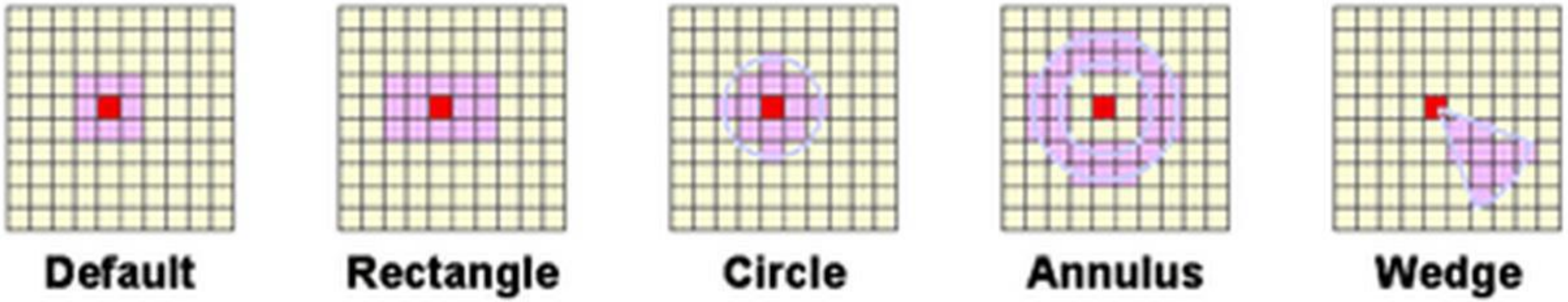
- A focal operation, also called a neighbourhood operation, involves a focal cell and a set of its surrounding cells.
- Focal functions produce an output raster dataset in which output value at each location is a function of the input value at a cell location and the values of the cell in a specified neighbourhood around the location
- How is a neighbourhood defined?



Neighborhood shapes

- Neighborhoods can be defined by rectangle, circle, annulus, and wedge shape





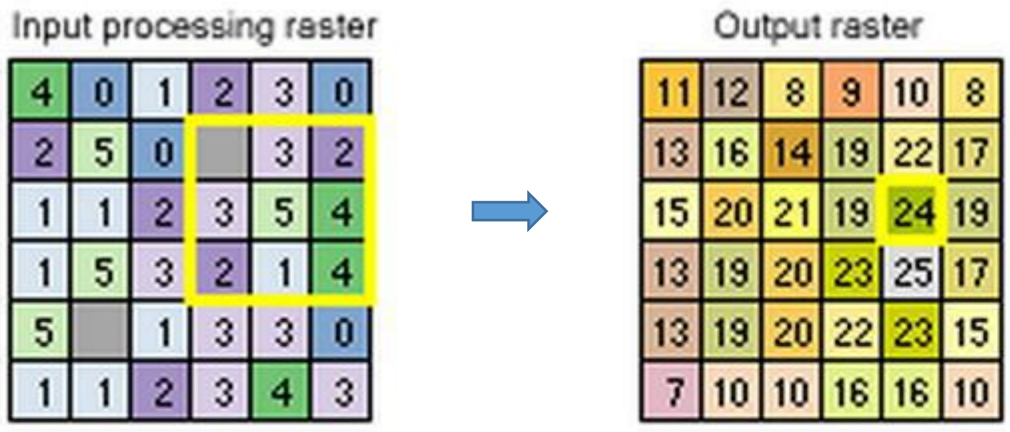
3-by-3 5-by-3

Focal operations

- Focal operations are usually performed on a single layer
- The output from a focal operation can show:
    - Summary statistics: maximum, minimum, range, sum, mean, median, and standard deviation
    - Tabulation of measures: majority and minority
- The above statistics and measures are similar with those from local operations with multiple layers
- Instead of using cell values from multiple input rasters, a focal operation uses the cell values from a defined neighborhood

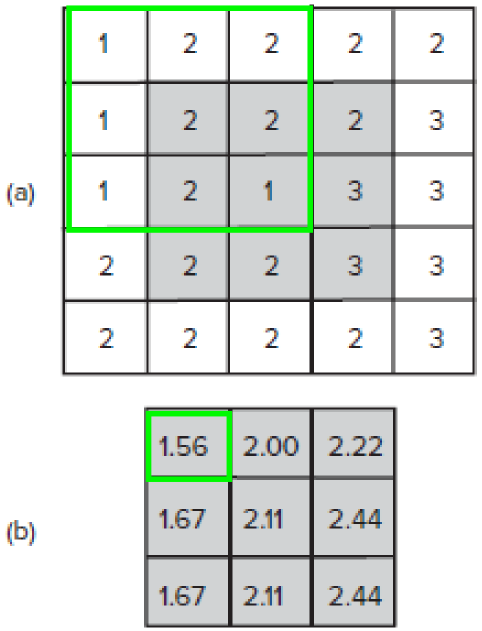
Focal operations

- Summary statistics



Sum operation

The cell values in (*b*) are the neighborhood sums of the shaded cells in (*a*) using a 3-by-3 neighborhood.



Mean operation

The cell values in (*b*) are the neighborhood means of the shaded cells in (*a*) using a 3-by-3 neighborhood.

Focal operations

- **Range statistics**
  - The range measures the difference between the maximum and minimum cell values within the defined neighborhood

Focal operations

- **Range statistics**

(a)

200	200	110	210	210
200	200	110	210	210
150	150	100	170	170
140	140	130	160	160
140	140	130	160	160

(b)

100	110	110
100	110	110
50	70	70

The cell values in (*b*) are the neighborhood range statistics of the shaded cells in (*a*) using a 3-by-3 neighborhood.

- A high value indicates the existence of an edge (e.g., boundary between grass land and water body) within the neighborhood
- Therefore, range statistic is often used for edge enhancement, which is a typical image processing method

Focal operations

- **Range statistics**

(a)

200	200	110	210	210
200	200	110	210	210
150	150	100	170	170
140	140	130	160	160
140	140	130	160	160

(b)

100	110	110
100	110	110
50	70	70

- How to deal with cells on the margin?
  - Abandon cells on the margin
  - Only use cell values available in the neighborhood
  - …

The cell values in (*b*) are the neighborhood range statistics of the shaded cells in (*a*) using a 3-by-3 neighborhood.

Focal operations

- Tabulation of measures
  - The opposite of edge enhancement is a smoothing operation based on the majority measure
  - Smoothing operation is often used to reduce noise within an image

(a)

1	2	2	2	2
1	2	2	2	3
1	2	1	3	3
2	2	2	3	3
2	2	2	2	3

(b)

2	2	2
2	2	3
2	2	3

The cell values in (*b*) are the neighborhood majority statistics of the shaded cells in (*a*) using a 3-by-3 neighborhood.

Applications of focal operations

- Data simplification
  - For example, the moving average method reduces the level of cell value fluctuation in the input layer
- Image processing
  - Edge enhancement, smoothing operation
- Terrain analysis
  - The slope, aspect and surface curvature of a cell are derived using elevations from its adjacent neighbors
- Site selection
  - Focal operation has the ability to summarize statistics within an area

Zonal operations

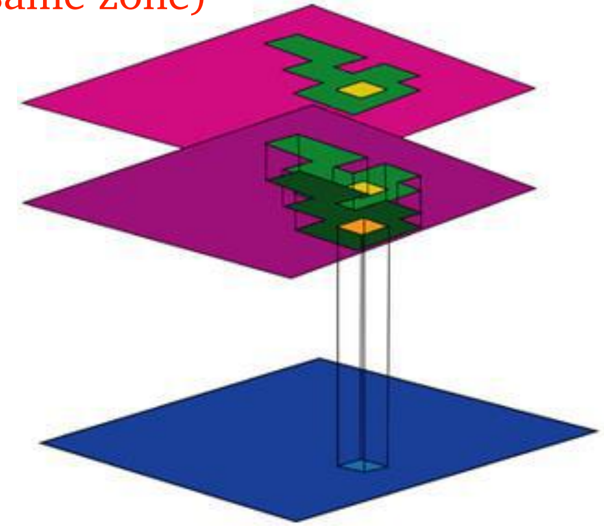
- A zonal operation works with groups of cells of same values or like features.

  These groups are called zones

- Zones may be contiguous or non-contiguous

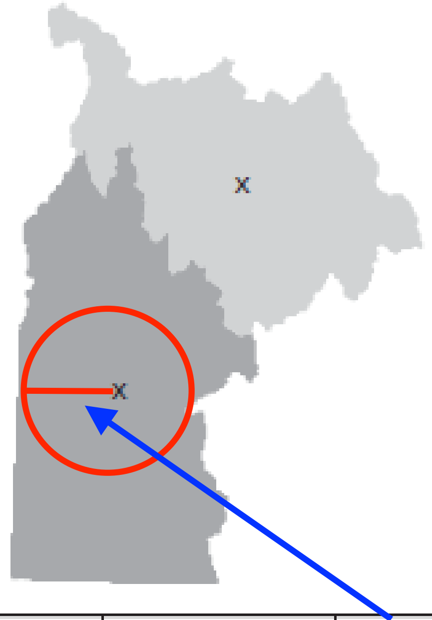
Pixels do not need to connect to each other (can be in different locations, but considered the same zone)

- Zonal functions are similar to focal functions except that the definition of the neighborhood in a zonal function is the configuration of zones



Zonal operations

- A zonal operation may work with a single raster or two rasters
    - Given a single input raster, zonal operations measure the geometry of each zone in the raster, such as area, perimeter, thickness, and centroid

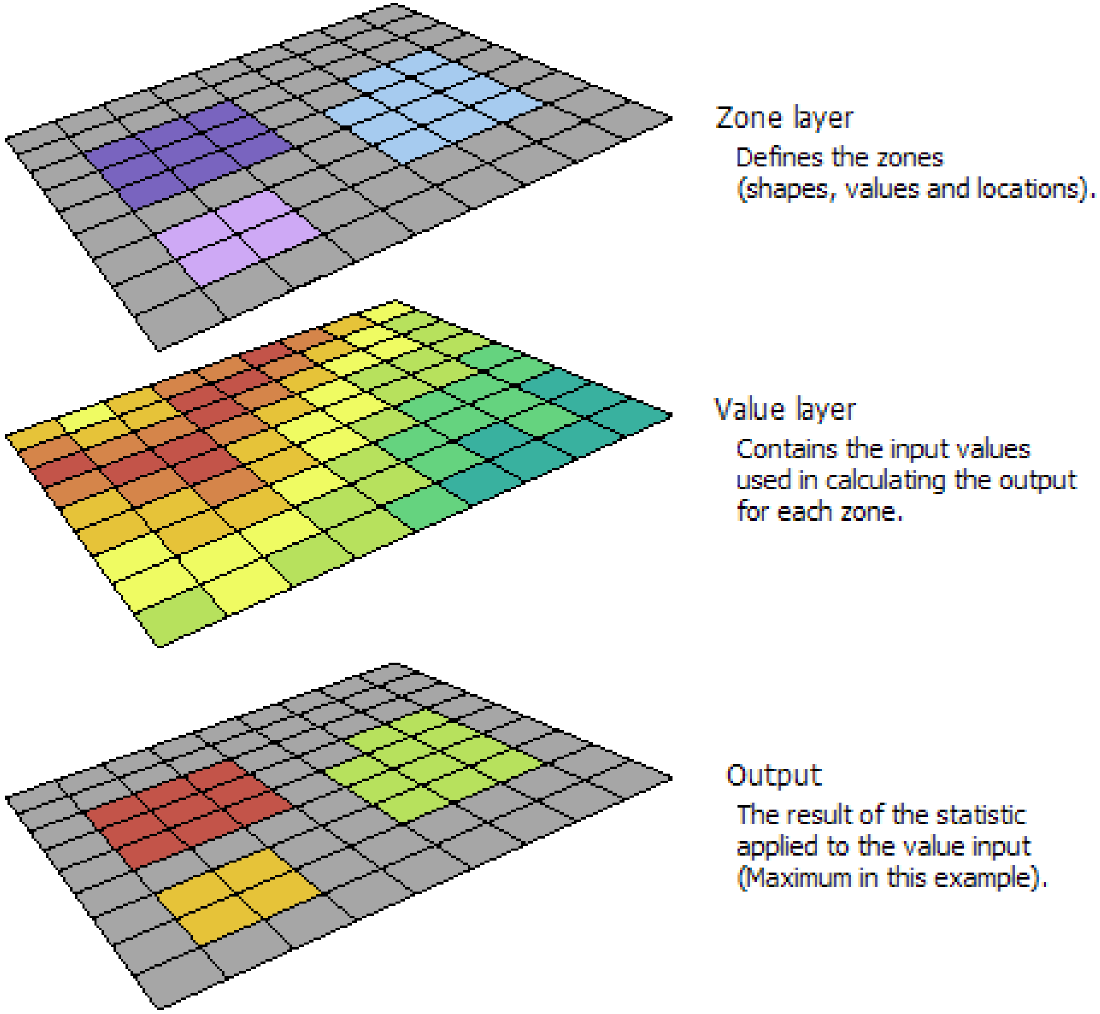


Zone	Area	Perimeter	Thickness
1	36,224	1708	77.6
2	48,268	1464	77.4

- The area is the sum of the cells that fall within the zone times the cell size
- The perimeter of a contiguous zone is the length of its boundary, and the perimeter of a non-contiguous zone is the sum of the length of each part
- The thickness calculates the radius of the largest circle that can be drawn within each zone
- The centroid is the geometric center of a zone

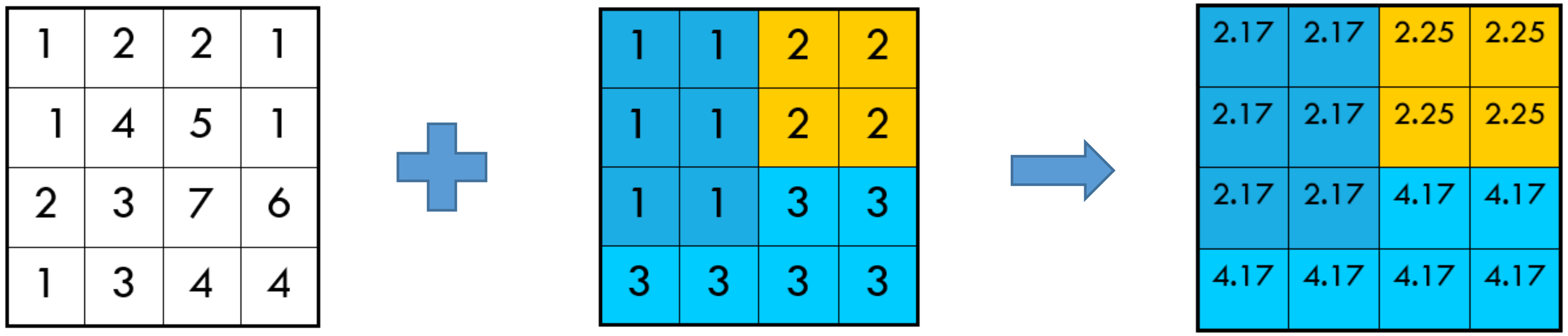
Zonal operations

- A zonal operation may work with a single raster or two rasters
    - Given two rasters in a zonal operation, one input raster and one zonal raster, a zonal operation produces an output raster, which summarizes the cell values in the input raster for each zone in the zonal raster



Zonal operations

- A zonal operation may work with a single raster or two rasters
  - The summary statistics include area, minimum, maximum, sum, range, mean, standard deviation, median, majority, minority, and variety

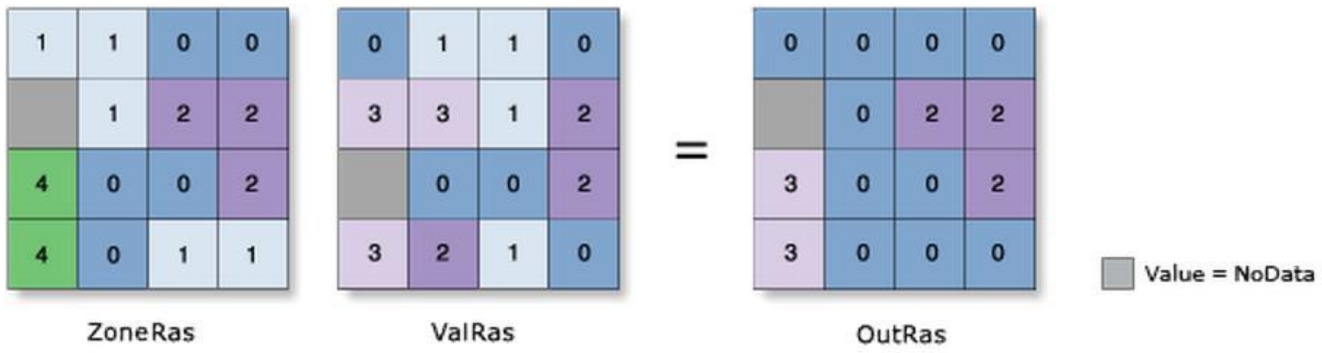


(a) Input raster (b) Zonal raster (c) Output raster

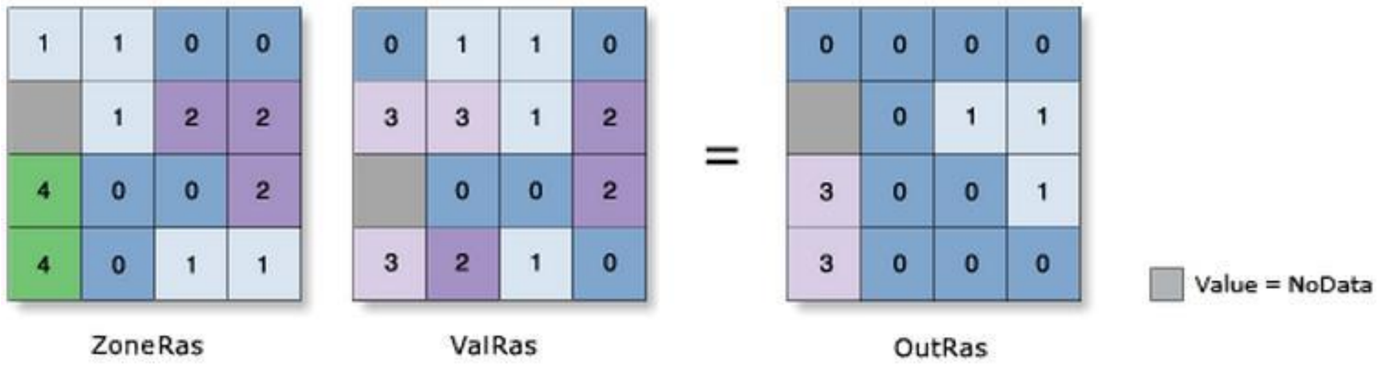
The cell values in (c) are the zonal means derived from an input raster (a) and a zonal raster (b)

Zonal operations

- Zonal majority



- Zonal minimum

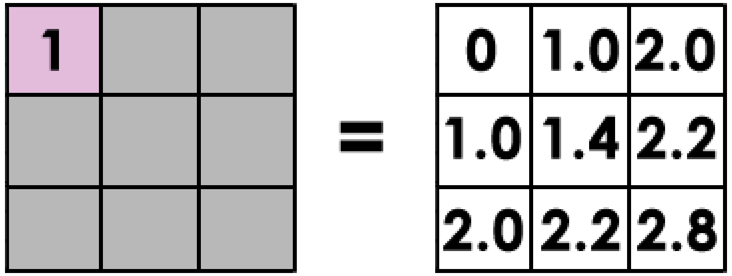


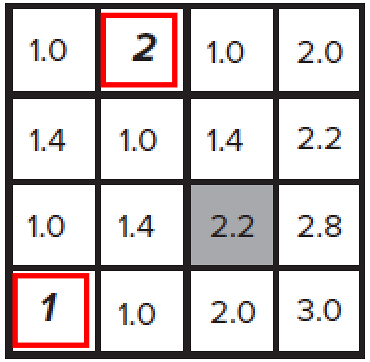
Global operations

- Global operations are similar to zonal operations whereby the entire raster dataset's extent represent a single zone
- Global operations apply a bulk change to all cells in a raster
- Global, or per-raster, operations compute an output raster dataset in which the output value at each cell location is potentially a function of all the cells combined from the various input raster datasets.

Global operations: Euclidean distance

- Euclidean distance is an example of a global operation. By calculating the closest distance away from a source, it applies the function globally in a raster





It calculates the distance from each cell to the closest source cell

Global operations: Allocation and Direction

- Allocation: The cell value in an allocation raster corresponds to the closest source cell for the cell
- Direction: The cell value in a direction raster corresponds to the direction in degrees that the cell is from the closest source cell

2	2	2	2
2	2	2	2
1	1	1	2
1	1	1	1

It shows the allocation of each cell to the closest source cell

90	2	270	270
45	360	315	287
180	225	243	315
1	270	270	270

It shows the direction in degrees from each cell to the closest source cell

Outline of this lecture

- Data analysis environment
- Raster analysis operations
  - Local operation
  - Focal operation
  - Zonal operation
  - Global operation
- Other raster data operations
- Map algebra

Other raster data operations

- Raster data clip
- Raster data combine
- Raster data extraction
- Raster data generalization

Raster data clip

- There are two methods to clip a raster
  - Specify an analysis mask (clip by mask)
  - Specify the minimum and maximum x-, y-coordinates (clip by rectangle)

(a) Input raster

(b) An analysis mask

(c) Clip-by-mask result

(d) A rectangle

(e) Clip-by-rectangle result

Raster data combine

- The process of combining multiple input rasters into a single raster is also called raster mosaic
- Raster mosaics require that the input rasters have overlaps (seamless combination)
- Different options are provided for filling in the cell values in the overlapping areas
  - Use the first or the second input raster's data
  - Use the blending of data from the input rasters

Raster data extraction

- **Raster data extraction** creates a new raster by extracting data from an existing raster
  - Extract by location
  - Extract by attribute
  - Extract by color
  - Extract by other characteristics
    - Shape, texture, …

Extraction by location

210	190	170	155	140	135
204	183	165	145	125	120
200	175	160	122	110	100
208	187	165	150	126	120

DEM, cell with height > 150

How do we automatically extract building roof in the image?

Raster data generalization

- **Raster data generalization** is to simplify raster data
  - Resampling: filling each pixel of the new image with a value or derived value from the original image (the new image has a larger cell size)
    1. Nearest neighbor resampling
    2. Bilinear interpolation (distance-weighted) **Refer to Lecture 6**
    3. Cubic convolution (distance-weighted)

Raster data generalization

- **Raster data generalization** is to simplify raster data
  - Aggregate: calculating each output cell value as the mean, median, sum, minimum or maximum of the input cells

1	3	2	5
1	3	2	7
1	1	2	5
2	4	3	2

(a)

2	4
2	3

(b)

An Aggregate operation creates a lower-resolution raster (*b*) from the input (*a*). The operation uses the mean statistic and a factor of 2

The output cell size is the integral multiples of the input cell size

Outline of this lecture

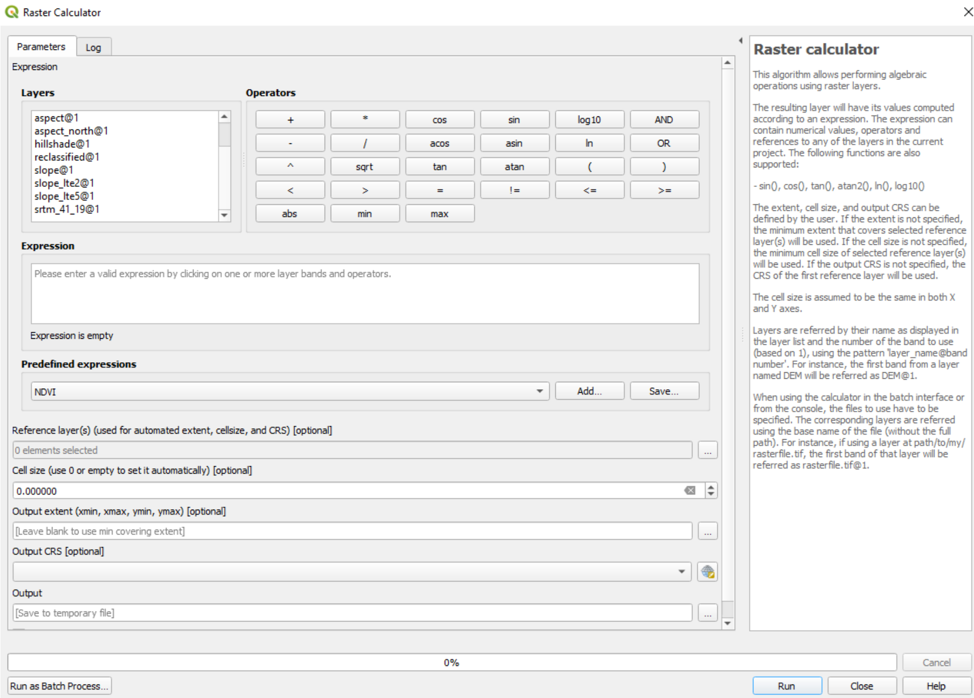
- Data analysis environment
- Raster analysis functions
  - Local operation
  - Focal operation
  - Zonal operation
  - Global operation
- Other raster data operations
- Map algebra

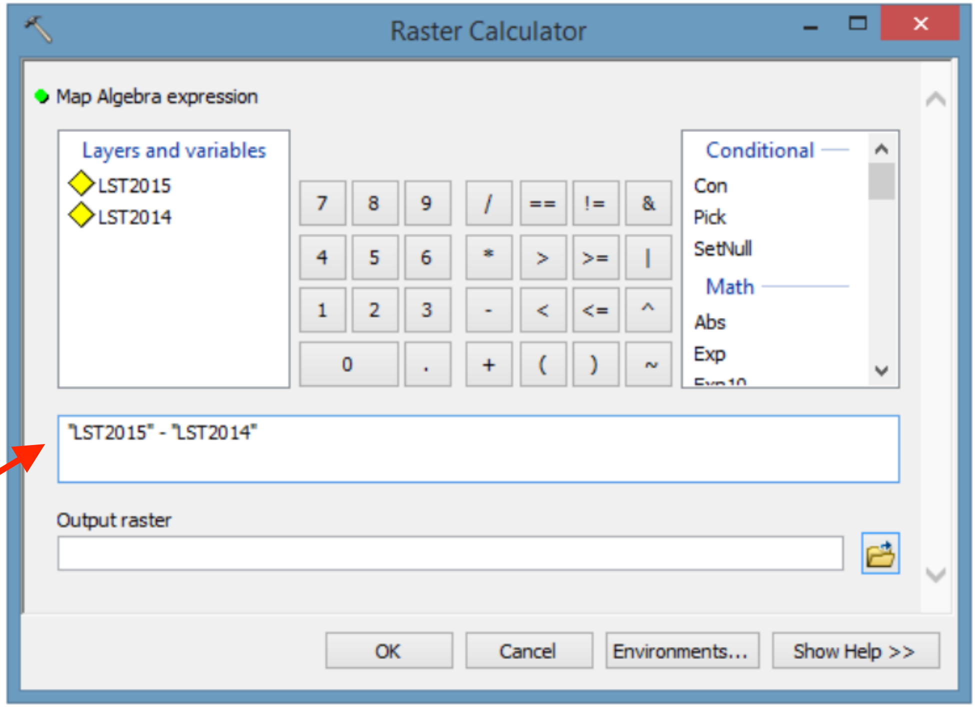
Map algebra

- Map algebra basically involves doing math with maps
- Map algebra only applies to rasters
- Map algebra can be local, focal, zonal and global

Map algebra

- QGIS provides the tool of map algebra under **Raster Analysis**
- ArcGIS provides the tool of map algebra under **Spatial Analyst Tools**
    - Select your input raster data sets and operators
    - Set your function
    - Save as new raster layer





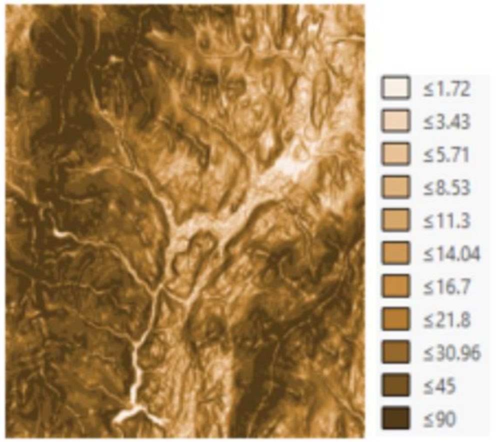
LST: Land surface temperature

(Local Operation - consider pixel by pixel)

Commonly used mathematical functions

- Arithmetic operations (addition, subtraction, multiplication, division)
- Statistical operations (minimum, maximum, average, median)
- Relational operations (greater than, smaller than, equal to)
- Trigonometric operations (sine, cosine, tangent, arcsine)
- Exponential and logarithmic operations (exponent, logarithm)

Example of map algebra



- Example 1:

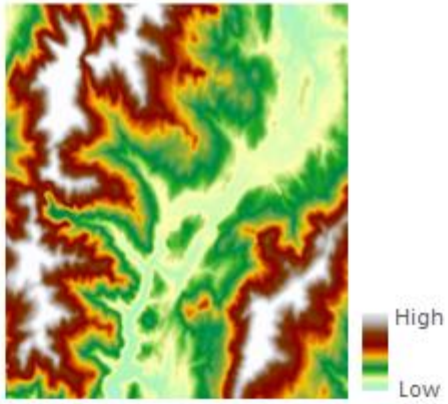
ExtractByAttributes(“slope_d”, “Value < 20”)

- “slope_d” is a slope raster measured in degree
- It means to extract from “slope_d” a raster that has slope values lower than 20°

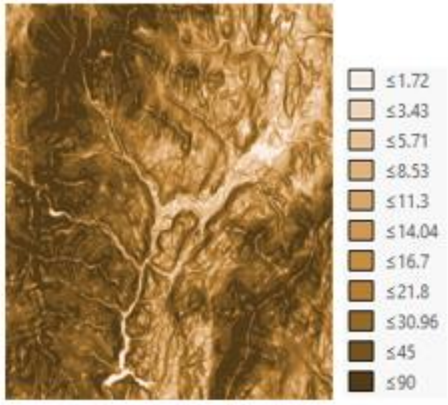
Can be used for site selection, and reclassify the values (1 or 0) depending if slope value is lower or equal and more than 20, then use the data for analysis

Example of map algebra

- Example 2:



Input elevation raster



Output slope raster
(in degrees)

ExtractByAttributes(Slope("emidalat", "DEGREE"), "Value < 20")

- "emidalat" is an elevation raster
- The above expression has two steps
  - It derives a slope raster measured in degree, from "emidalat"
  - It extracts from the derived raster a new raster that has slope values lower than 20°

Summary

- Data analysis environment
  - Area extent
  - Output cell size
- Raster analysis functions
  - Local operation: cell-by-cell operation
  - Focal operation: neighborhood operation
  - Zonal operation
    - Zone: a group of cells of same values of like features
  - Global operation: apply a bulk change to all cells in a raster

Summary

- Other raster data operations
  - Raster data clip
  - Raster data combine
  - Raster data extraction
  - Raster data generalization
- Map algebra: doing math with rasters
  - Select your input raster data sets and operators
  - Set your function
  - Save as new raster layer

THANK YOU