Ch-3-Spatial Data analysis and interpretation

Biruk G.
April 20, 2020

Introduction

***Analysis* of spatial data**

- can be defined as computing new information that provides new insight from the existing, stored data.

Classification of the Analytical Techniques

1. Measurement functions

- Allow the calculation of distances, lengths and areas.

2. Retrieval functions

- Selective searches of data

3. Classification

- Allows the assignment of features to a class based on attribute values.

Classification of the Analytical Techniques…contd

4. Generalization

Is a function that joins different classes of objects with common characteristics

5. Overlay functions

- Combination of two or more spatial data layers.

6. Neighborhood functions

- Evaluate the characteristics of an area *surrounding* the features location

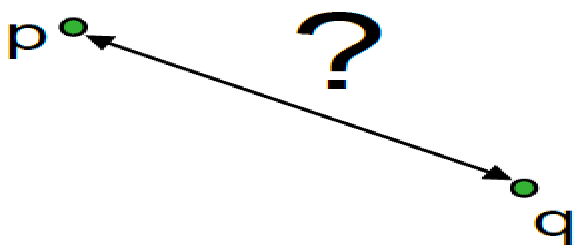
GIS Functionality Vector Analysis

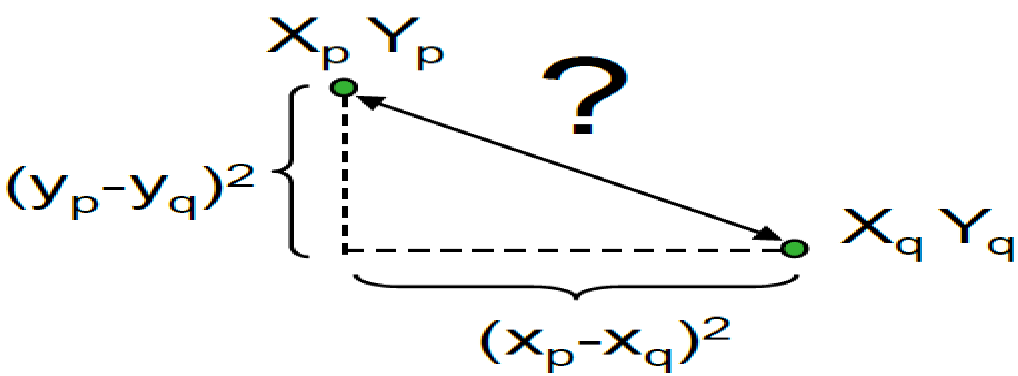
Vector Analysis: Measurement

- Geometric measurement on spatial features includes counting, distance and area-size computations.
- The primitives of vector datasets are **point, (poly)line and polygon**.
- Related geometric measurements are **location, length, distance & area size**.
- Some of these are geometric properties of a feature in isolation (location, length, area size); others (distance) require **two features** to be identified.

Vector Analysis: Measurement…contd

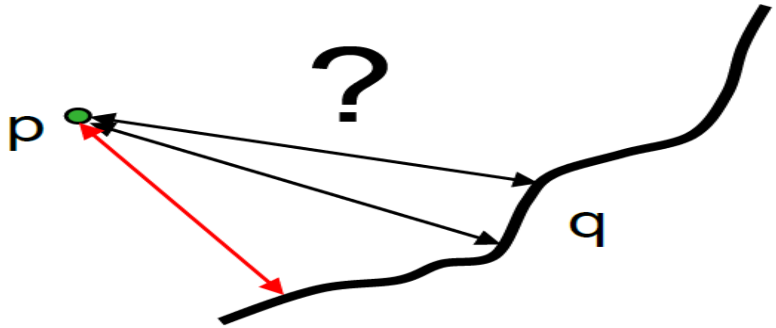
- Distance between two points →Pythagorean distance fun.

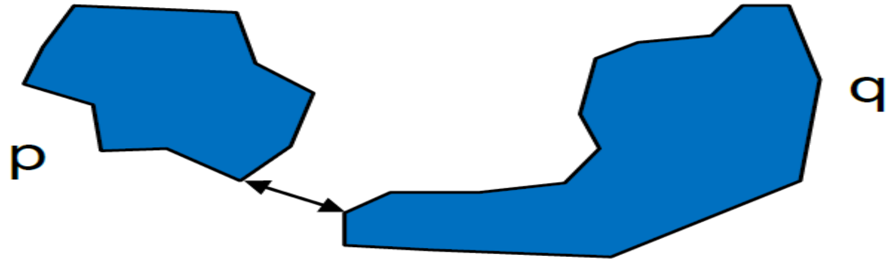




Vector Analysis: Measurement

- If **one or both features** are not a point we will measure the ***minimal distance*** between the two features





Vector Analysis: Retrieval Function

- Spatial Selection by Attribute conditions
  - Relational operators
  - Logical operators (and, or, not)
  - Combining attribute conditions
- Spatial selection using topological relationships
  - Selecting features that are inside selection objects
  - Selecting features that intersect
  - Selecting features adjacent to selection objects
  - Selecting features based on their distance

A) Spatial Selection by Attribute conditions

- It is selecting features by using selection conditions on **feature attributes**.
- These conditions are formulated in **SQL (if the attribute data reside in a relational database**) or in a software-specific language (if the data reside in the GIS itself).

A) Spatial Selection by Attribute conditions..contd

- This type of selection answers questions like
- **“Where are the features with . . . ?”**
- Display the result both on the map and in the attribute table.

Operators used to set condition	
< (less than)	> (greater than)
= (equals)	>= (greater than or equal)
<= (less or equal than)	<> (does not equal)

A. Spatial Selection by Attribute conditions…contd

Composite conditions use logical connectives:

AND (returns true if both expressions a and b are true)
OR (returns true if one or both of the expressions a and b is true)
NOT (returns true if expression is false)

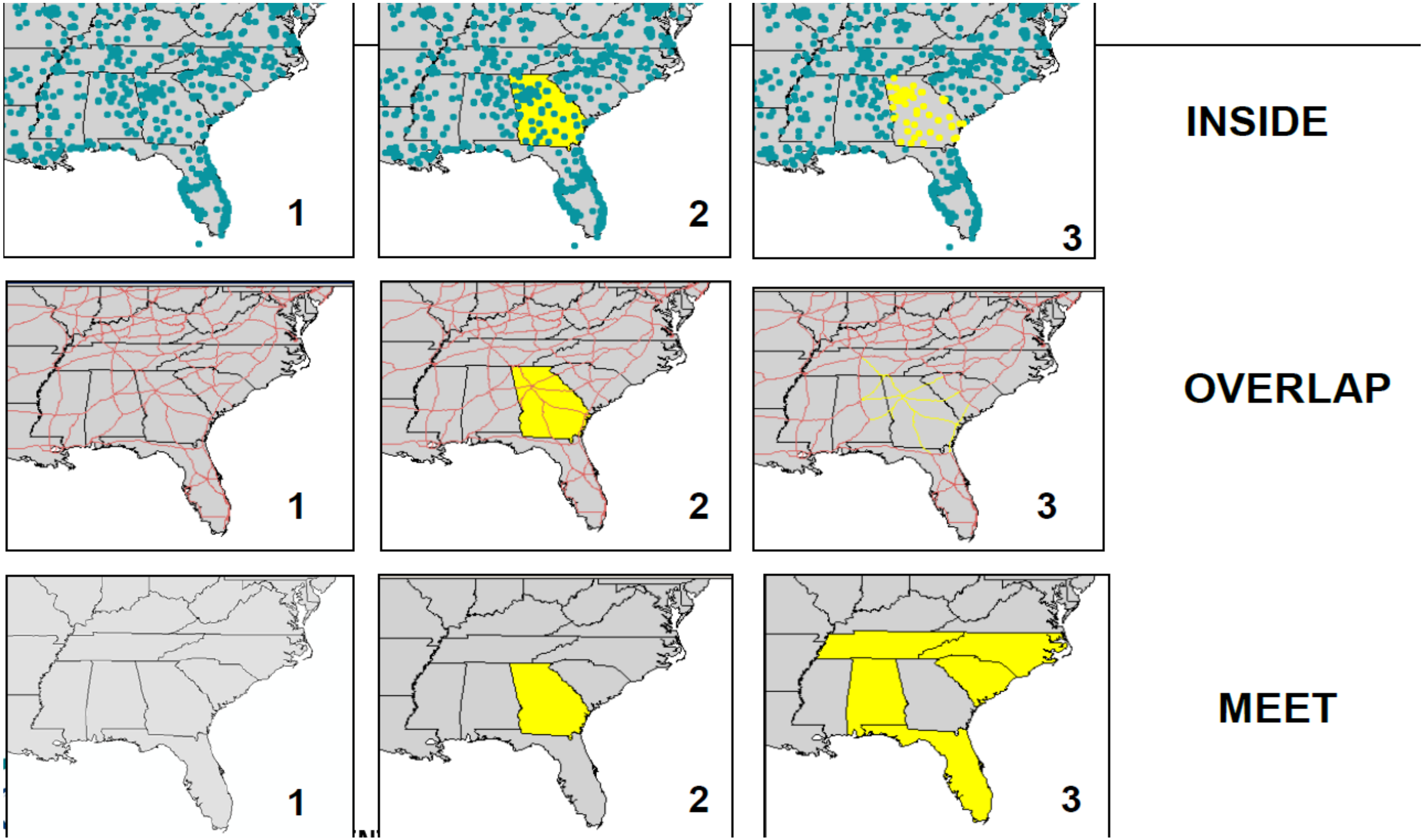
Area < 400,000	(atomic condition)
(Area < 400,000) AND (landuse =80)	(composite cond.)
(Area < 400,000) OR (landuse = 80)	(composite cond.)
NOT (landuse = 80)	(negate condition)

20:48

B. Spatial Selection by Using Topological Relationships

- Various forms of topological relationships can be used in spatial selections:
- Selecting features that are inside selection objects (containment relationship)
- Selecting features that intersect
- Selecting features adjacent to selection objects
- Selecting features base on their distance

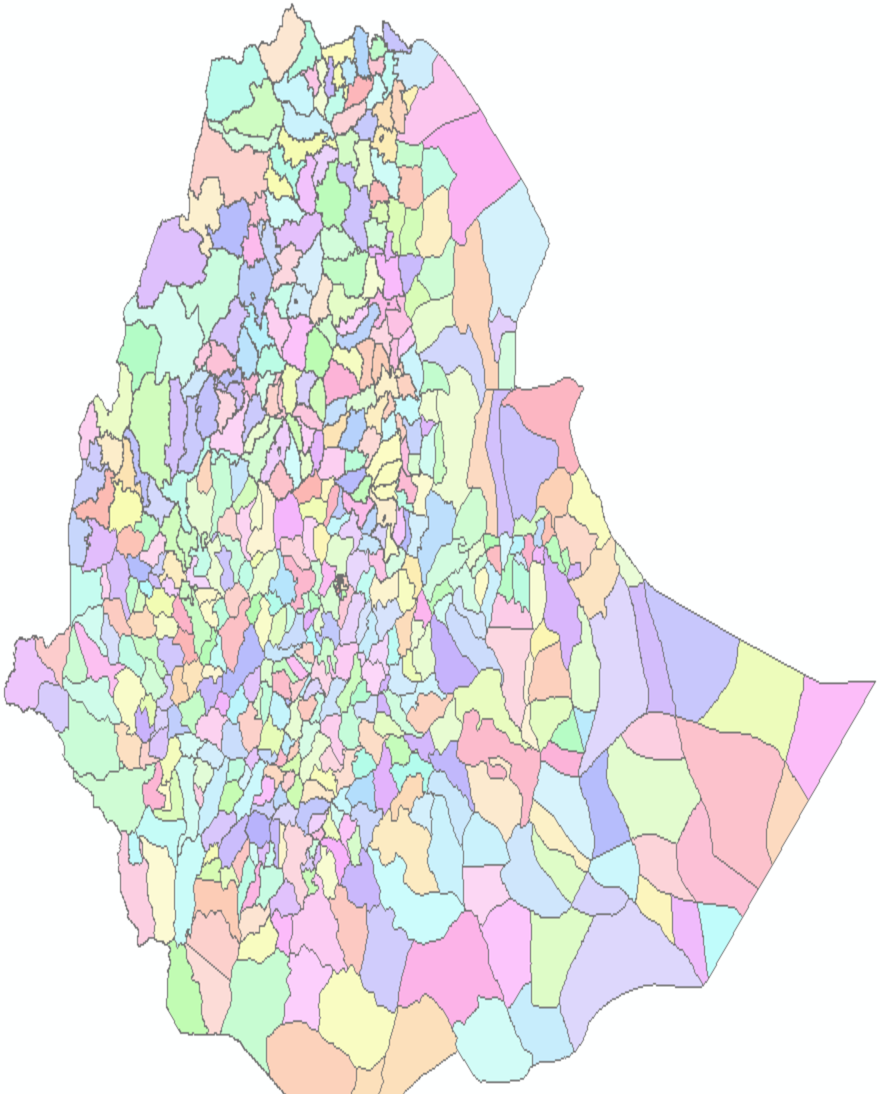
B. Spatial Selection by Using Topological Relationships…contd

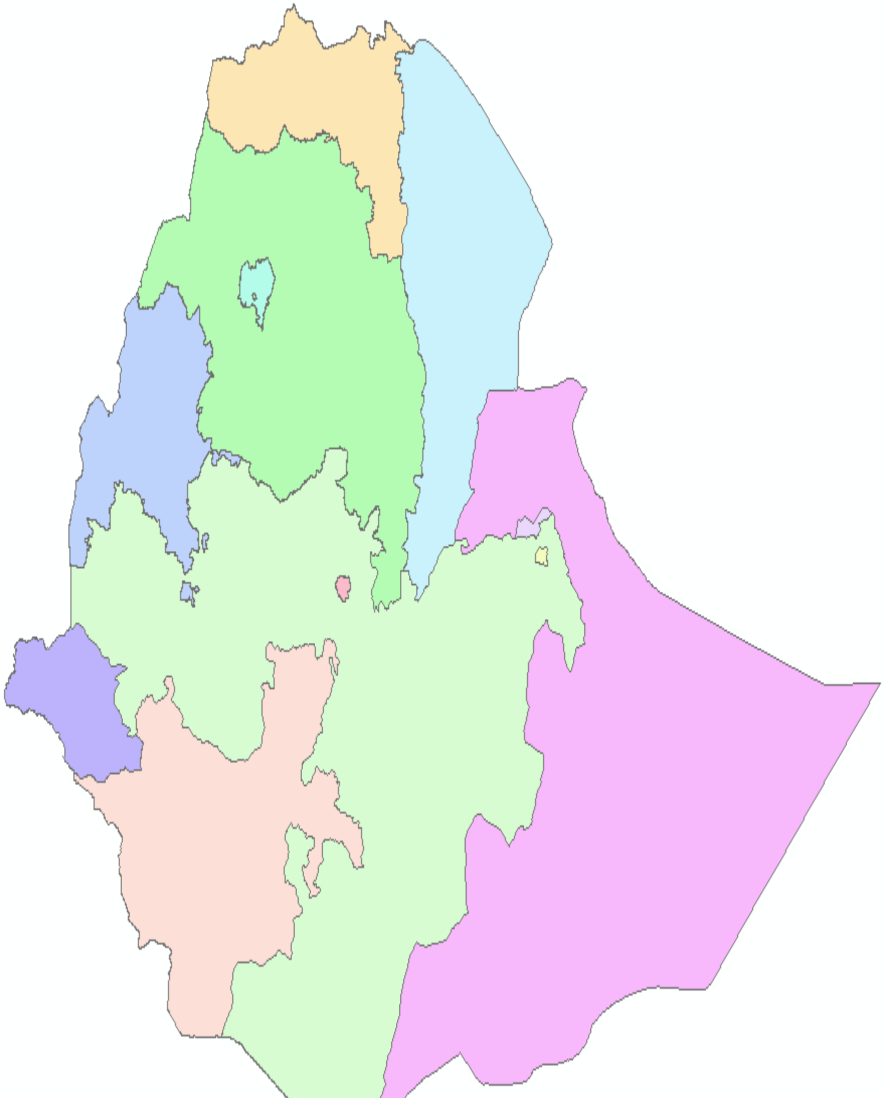


Vector Analysis: Classification

- Classification is a technique of purposely removing detail form an input dataset, in the hope of revealing patterns (of spatial distribution)
- We do so by assigning a characteristic value to each element in the input set, raster, points, lines or polygons

Vector Analysis: Classification…contd





Vector Analysis: Classification->post-processing

Left, without post-processing, right with post-processing.

- Vector classification can be performed with (or without) post-processing.
- With post-processing we mean that adjacent features that after classification have the same class are merged together.
- This is called ***spatial merging***, ***aggregation*** or ***dissolving***.

Vector Analysis: Overlay

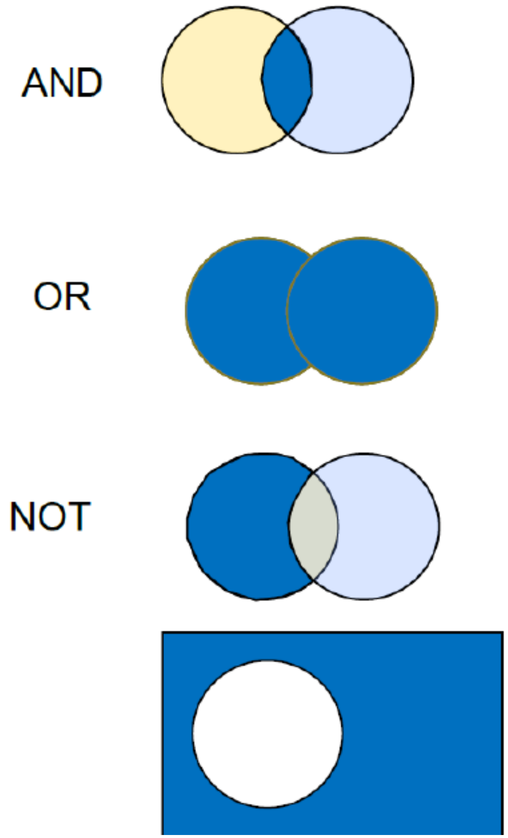
- Standard overlay operators take two input data layers and assume that:
  - they are **geo-referenced in the same system** and
  - **they overlap** in study area.
- If either of these requirements is not met, the use of an overlay operator is pointless.
- The principle of spatial overlay is to compare the characteristics of the same location in both data layers and to produce a result for each location in the output data layer.

Vector Analysis: Overlay...contd

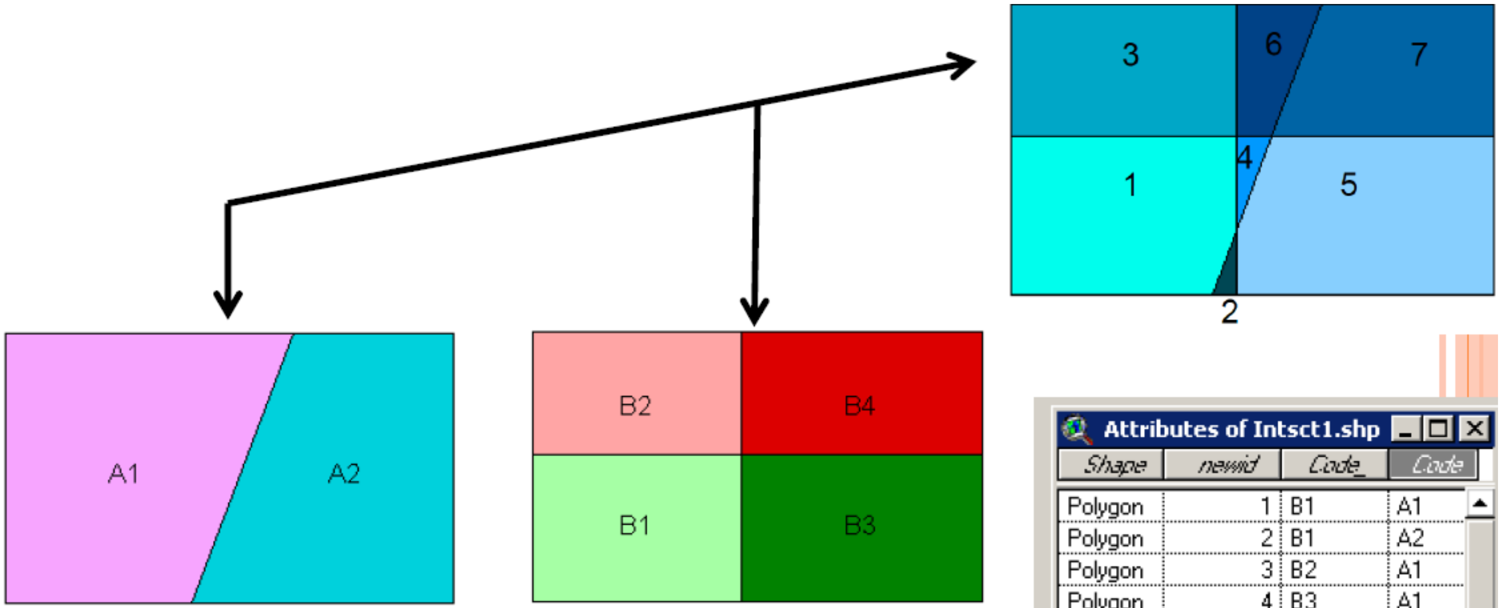
- The specific result to produce is determined by the user.
- It might involve a **calculation or some other logical function** to be applied to every area or location.
- Use logical connectors
  - Intersection
  - Union
  - Difference
  - Complement

Vector Analysis: Overlay...contd

- The potato fields on clay soils (*<u>intersection</u>*)
- The fields where potato or maize is the crop (*<u>union</u>*)
- The potato fields not on clay soils (*<u>difference</u>*)
- The fields that do not have potato as crop (*<u>complement</u>*)



Vector Analysis: Overlay...contd



Attributes of Overlay1.shp

Shape	ID	Code
Polygon	1	A1
Polygon	2	A2

Attributes of Overlay2.shp

Shape	ID	Code
Polygon	1	B1
Polygon	2	B2
Polygon	3	B3
Polygon	4	B4

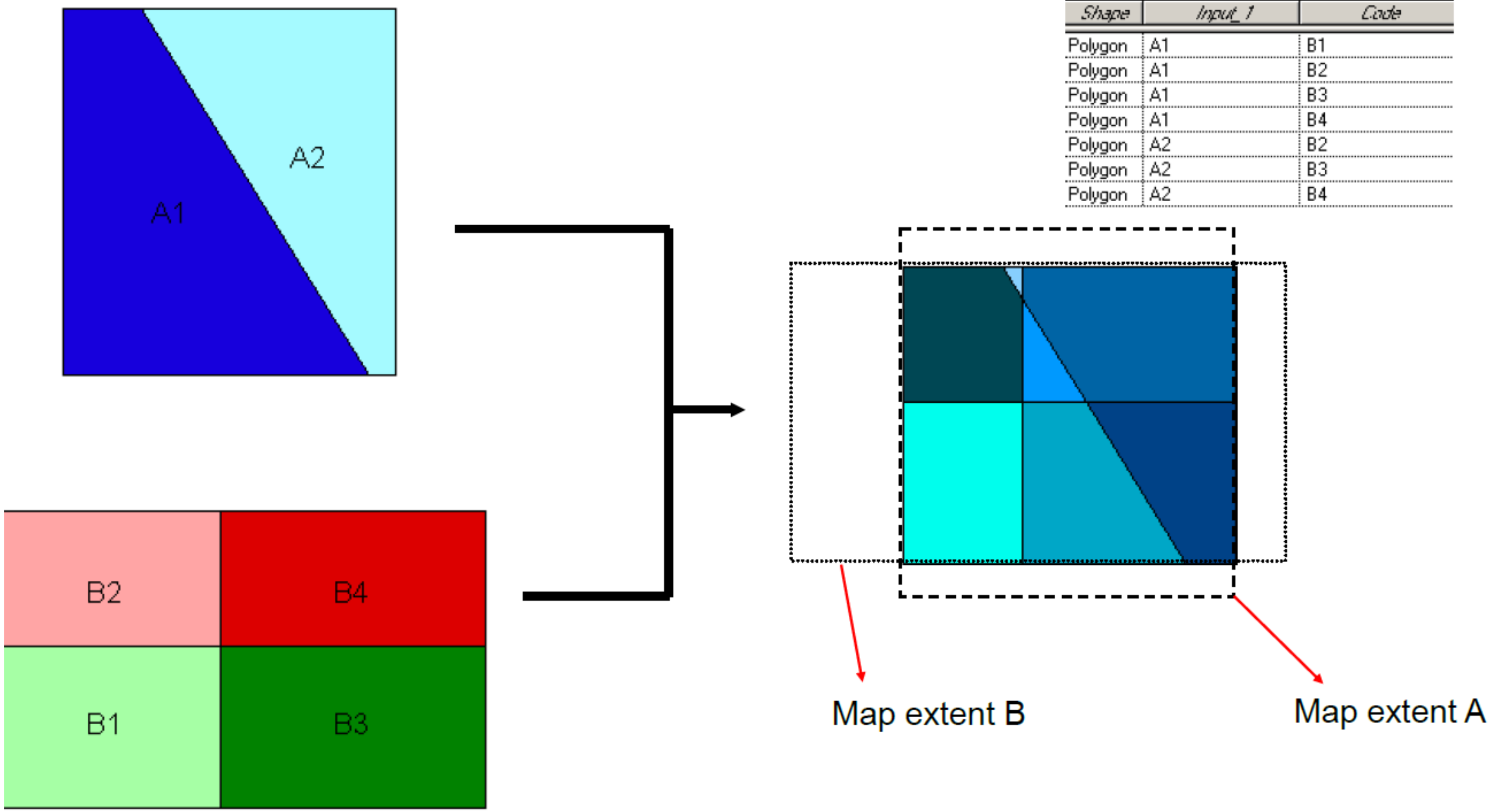
Attributes of Intsct1.shp

Shape	newid	Code_	Code
Polygon	1	B1	A1
Polygon	2	B1	A2
Polygon	3	B2	A1
Polygon	4	B3	A1
Polygon	5	B3	A2
Polygon	6	B4	A1
Polygon	7	B4	A2

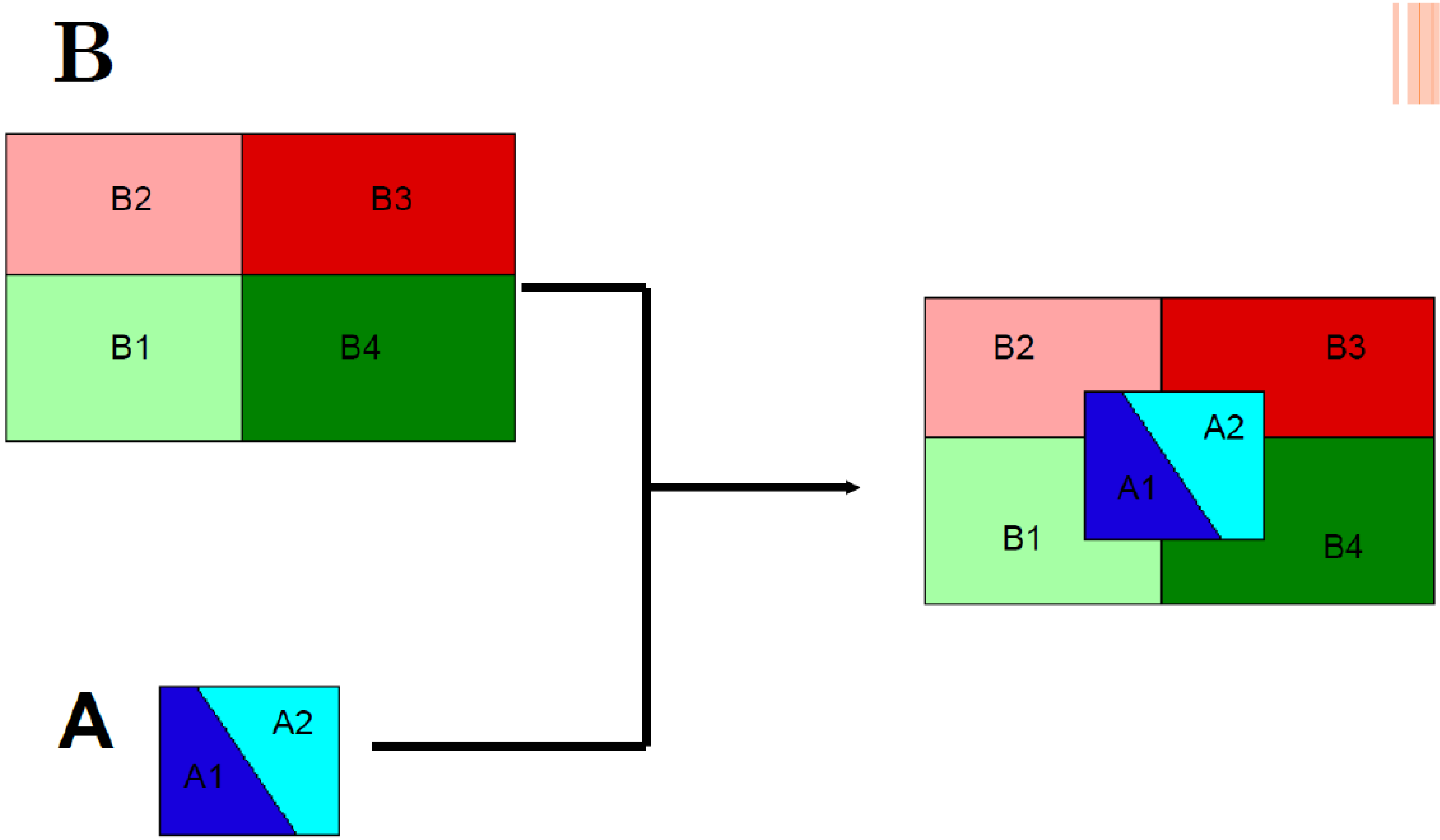
Overlay Intersect

Attributes of Intsct3.shp

Shape	Input_1	Code
Polygon	A1	B1
Polygon	A1	B2
Polygon	A1	B3
Polygon	A1	B4
Polygon	A2	B2
Polygon	A2	B3
Polygon	A2	B4



Overlay Overwrite



Vector Analysis: Neighborhood Operations

- Will find out characteristics of the vicinity (neighborhood) of a location
- To perform neighborhood analysis we must:
- State which target locations are of interest to us, and what is their spatial extent
- Define how to determine the neighborhood for each target

Vector Analysis: Neighborhood Operations…contd

- Define which characteristic(s) must be computed for each neighborhood
  - Proximity computations
  - Buffer zone generation
  - Thiessen polygons

Vector Analysis: Neighborhood Operations…contd

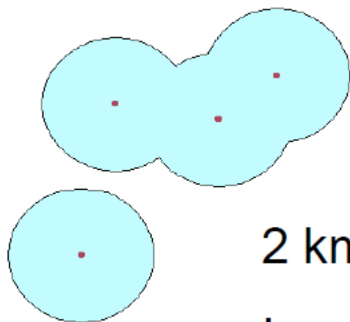
- **Proximity computations**

  - For instance, our target might be a Water Point. Its neighborhood could be defined as
    - an area within 2 km travelling distance; or
    - all roads within 500 m travelling distance; or
    - all other water points (WPs) within 10 minutes travelling time;
    - all residential areas for which the WP is the closest WP.

Vector Analysis: Neighborhood Operations…contd

1. Target: medical clinics

2. Neighborhood



2 km distance

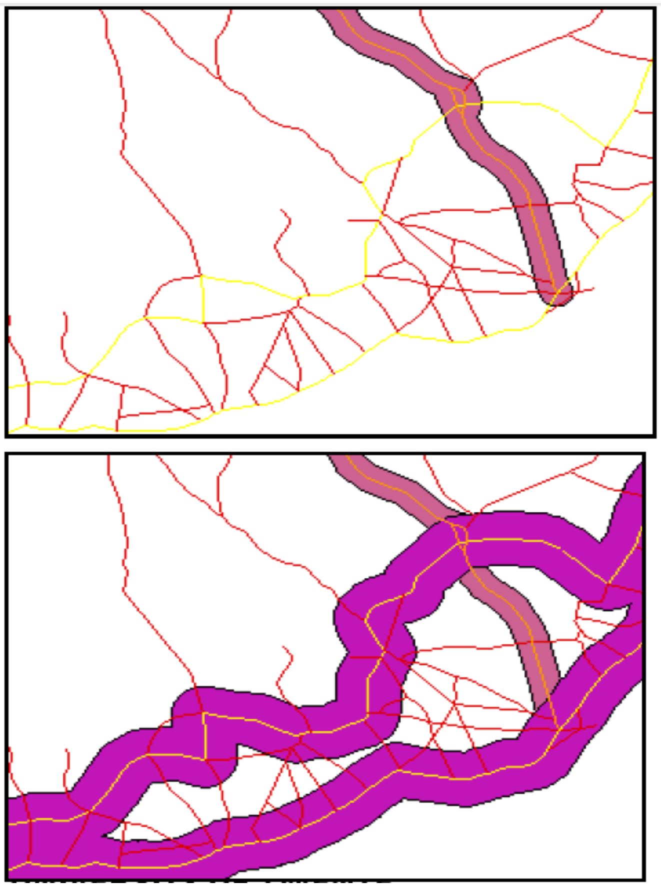
In a straight line

2 km travel distance

3. Characteristics

- How many people live in the area
- What is their average household income
- Are there any high-risk industries located in the neighborhood

Vector Analysis: Neighborhood Operations - Buffer …contd

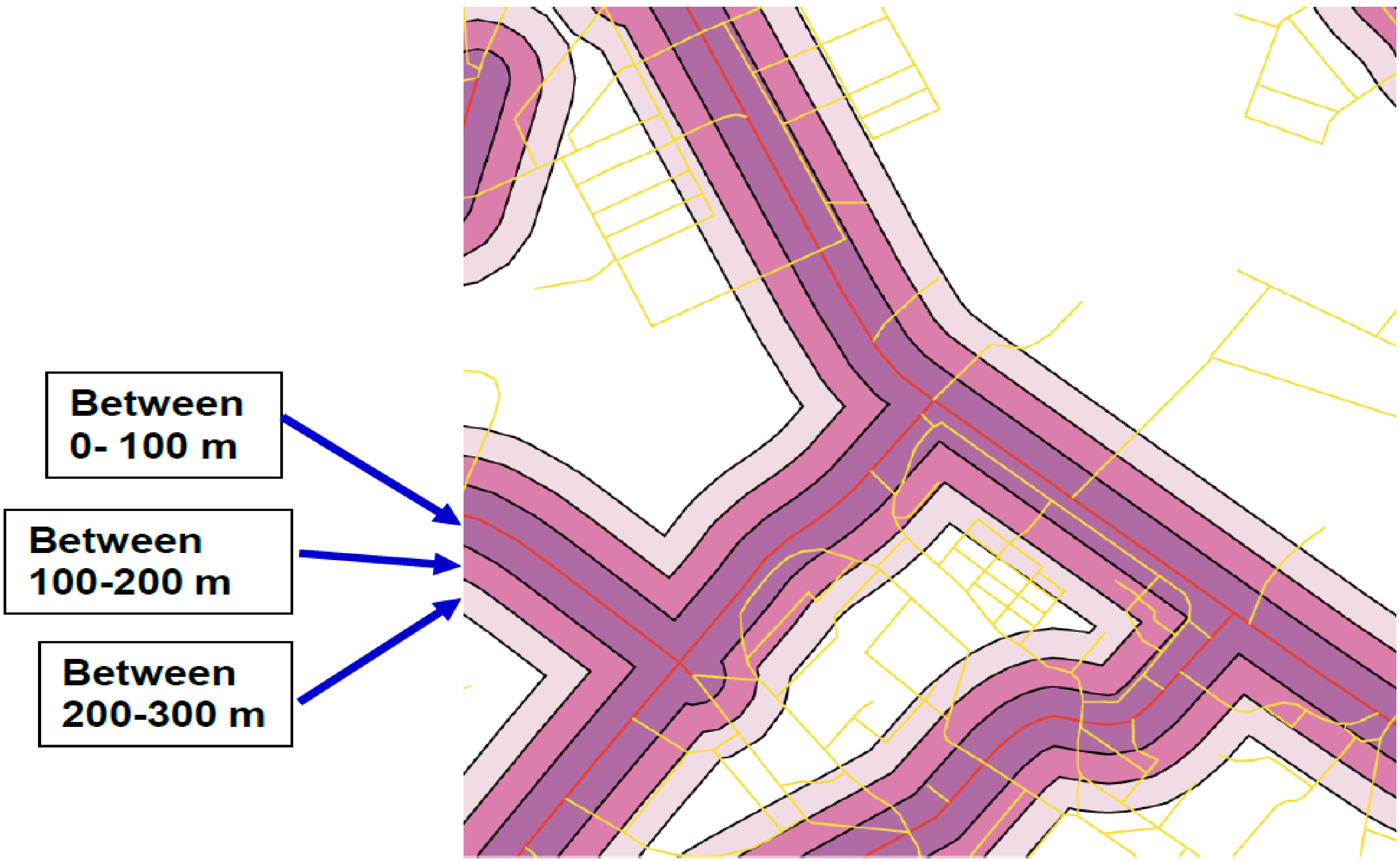


Principle:

1. Select one or more target locations
2. Determine the area within a certain distance

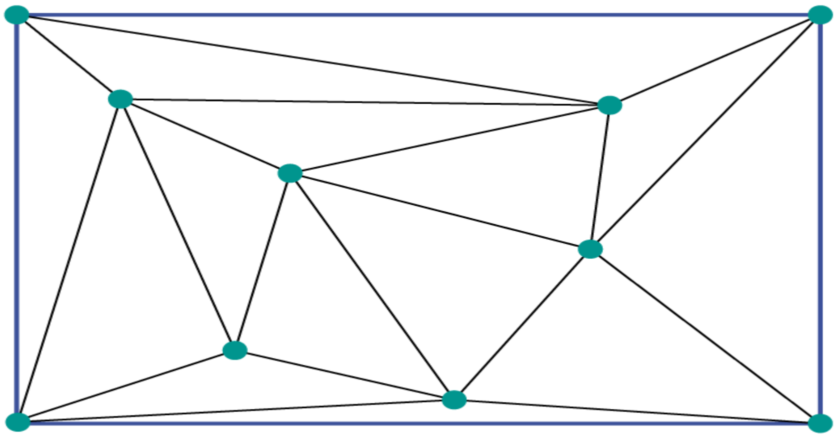
- Buffers become polygons
- Zonated buffers are possible

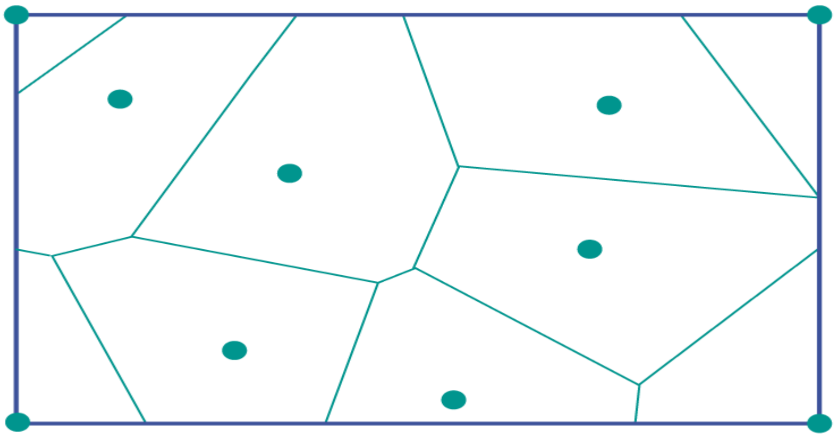
Vector Analysis: Neighborhood Operations - Buffer …contd



Vector Analysis: Neighborhood Operations - Thiessen polygons…contd

- Thiessen polygon partitions make use of geometric distance to determine neighborhoods.
- This is useful if we have a spatially distributed set of points as target locations and we want to know the closest target for each location in the study.
- This technique will generate a polygon around each target location that identifies all those locations that ‘belong to’ that target.





GIS Functionality Raster Analysis

Measurements

- Location
- Distance
- Area size

Classification

- User controlled
- Automatic

Overlay

- Arithmetic operators
- Comparison operators
- Logical operators
- Conditional expressions
- Decision tables

Neighborhood functions

- Buffer
- Flow

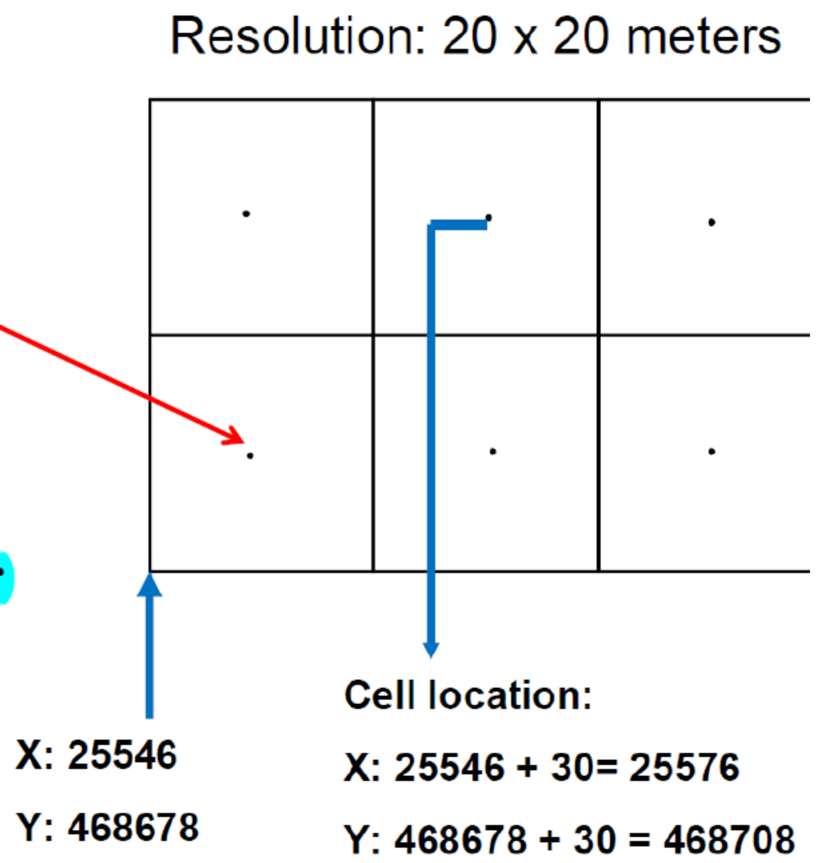
Raster based surface analysis

- Slope
- Aspect

Raster Analysis: Measurement - Location

Location

- Location of an individual cell is derived **from *anchor* point & resolution**
- A cell's location can be defined by the coordinates of its ***lower left corner* or its *midpoint***

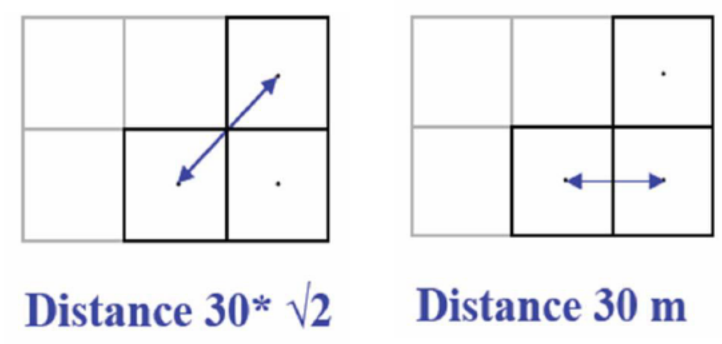


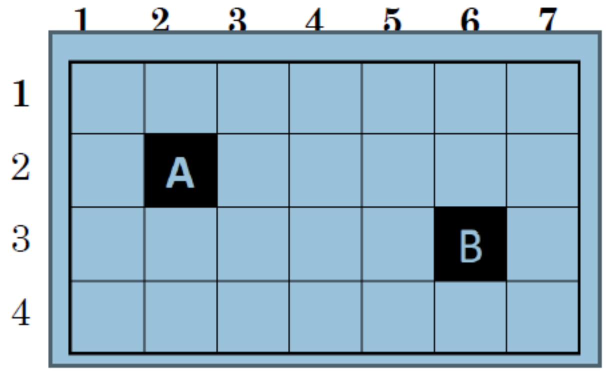
Raster Analysis: Measurement - Distance

Distance

- The Euclidean distance in a raster layer is the distance between the anchor points (for example mid-points) of two cells
- When a raster is used to represent line features as strings of cells, the length of a line is computed as the sum of the distances between the cells

Cell size = 30 m x 30 m

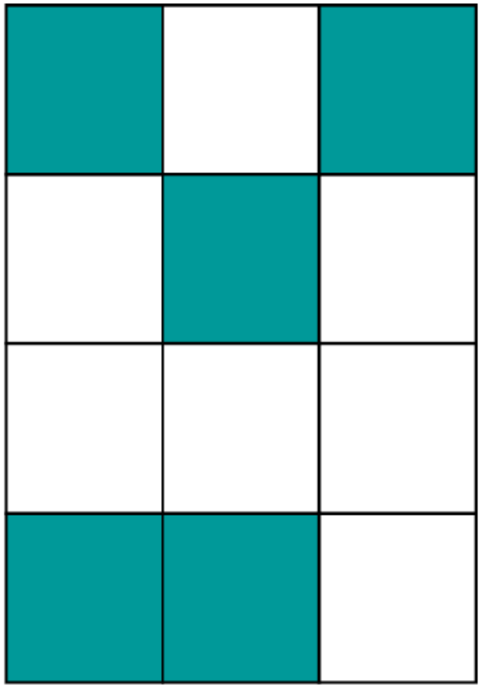




$$D_{A \rightarrow B} = \sqrt{((6pixel - 2pixel)^2 + (3pixel - 2pixel)^2)} \sim 4.1pixel$$

Raster Analysis: Measurement – Area Size

- Area Size
  - number of cells * cell size
  - The number of cells is also called the frequency or count



Cell size: 30 m x 30 m

900 * 5 = 4500 m^2

Raster Analysis: (Re) Classification

- Classification is a technique of removing detail from an input dataset in order to reveal patterns
- Classification will produce a new output dataset
  - a new attribute in vector
  - a new raster layer in raster
- If the input dataset itself is a classification we speak of reclassification

Raster Analysis: (Re) Classification

A) User Controlled

Two Examples of classification tables:

Old value	New value
391 - 2474	1
2475 - 6030	2
6031 - 8164	3

Code	Old value	New value
10	Planned Residential	Residential
20	Industrial	Commercial
30	Commercial	Commercial

The top table, the original values are ranges, in the lower table the old values already were a classification.

- In user-controlled classification we indicate the classification ***attribute*** and the classification ***method***.
- This is normally done via a ***classification_table***.

Raster Analysis: (Re) Classification

B) Automatic

- User specifies the number of output classes.
- Computer decides the class break points.
- Two techniques of determining the class breaks are discussed:
- *Equal frequency*
- *Equal interval*

Raster Analysis: (Re) Classification

B) Automatic Equal interval

- is calculated as (vmax –vmin) /n
- Vmax is the maximum attribute value, vmin is the minimum attribute value and n is the number of classes.
- In our example: (10-1) /5 ≈ 2
- Each class will have two values.

1	1	1	2	8
4	4	5	4	9
4	3	3	2	10
4	5	6	8	8
4	2	1	1	1

(a) original raster

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

(b) equal interval classification

original value	new value	# cells
1,2	1	9
3,4	2	8
5,6	3	3
7,8	4	3
9,10	5	2

Raster Analysis: (Re) Classification

B) Automatic Total number of cells 25 / 5 = 5 cells per class Equal frequency

- is also called quantile
- Total number of features / number of classes (n)
- The objectives is to create categories with roughly equal number of features (or cells).

1	1	1	2	8
4	4	5	4	9
4	3	3	2	10
4	5	6	8	8
4	2	1	1	1

(a) original raster

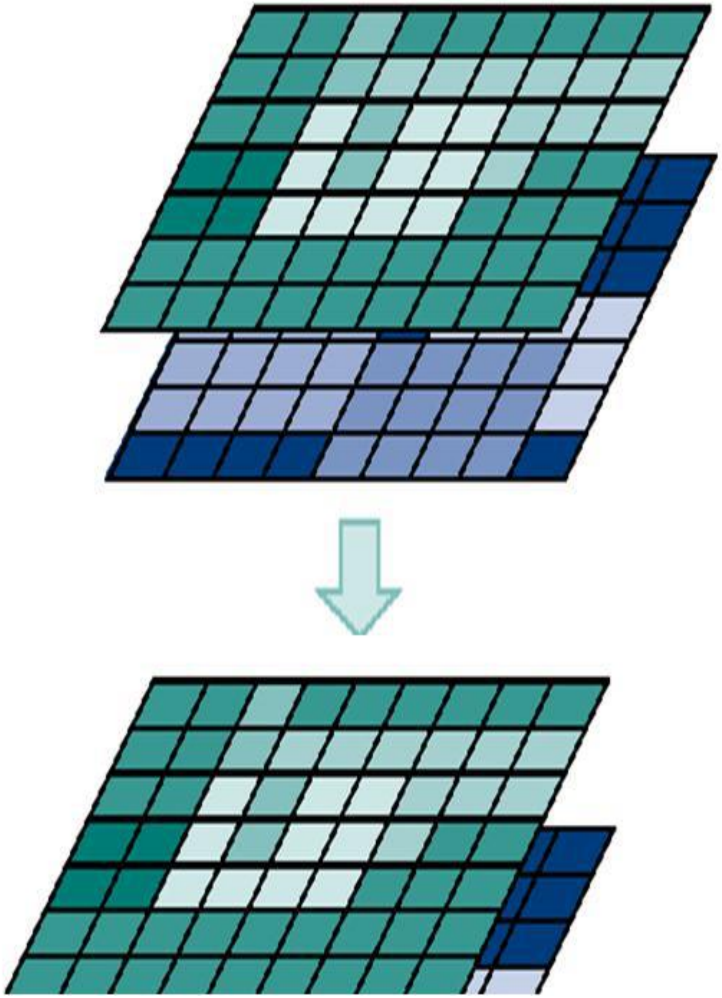
1	1	1	2	5
3	3	4	3	5
3	2	2	2	5
3	4	4	5	5
3	2	1	1	1

(c) equal frequency classification

original value	new value	# cells
1	1	6
2,3	2	5
4	3	6
5,6	4	3
8,9,10	5	5

Raster Analysis: Raster Overlay

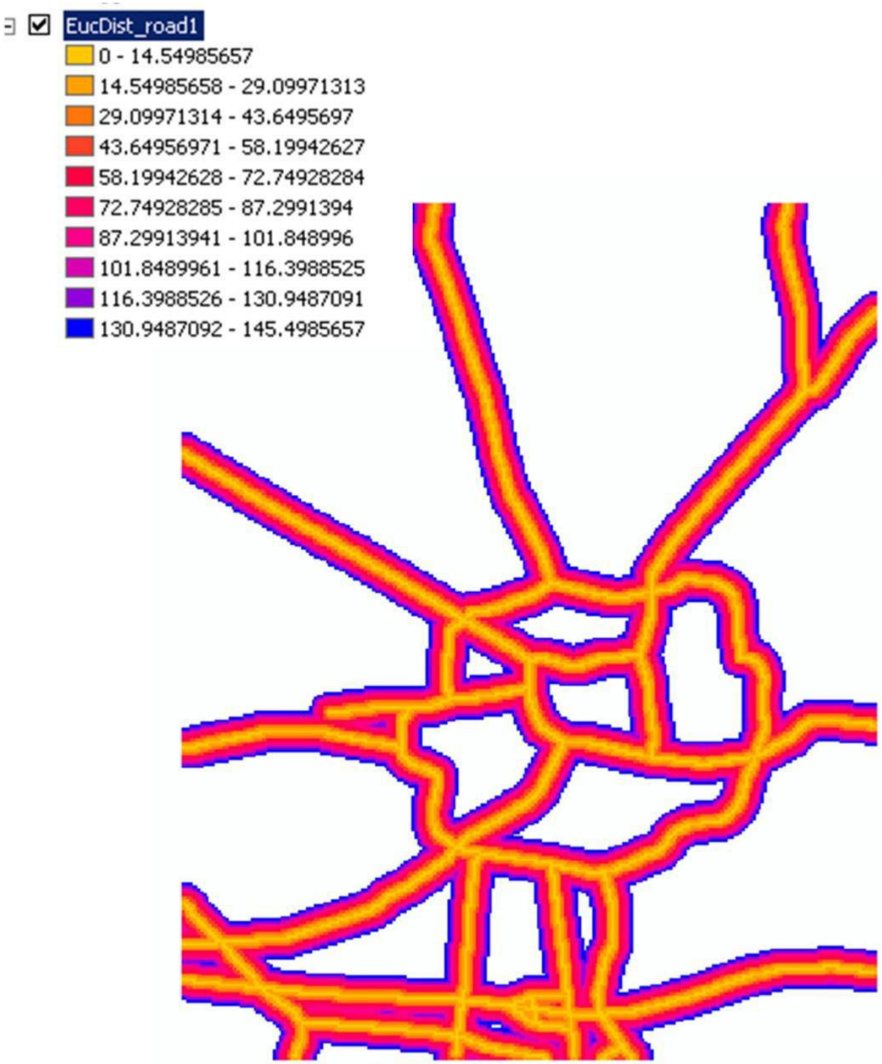
- New cell values are calculated using *map* ***algebra*** **/** ***raster calculus*****.**
- Performed on *cell-by-cell* basis.
- Operators:
  - Arithmetic overlay operators
  - Comparison and logical operators
  - Conditional expressions
  - Decision table



Raster Analysis: Neighborhood Functions

Proximity computation: Buffer

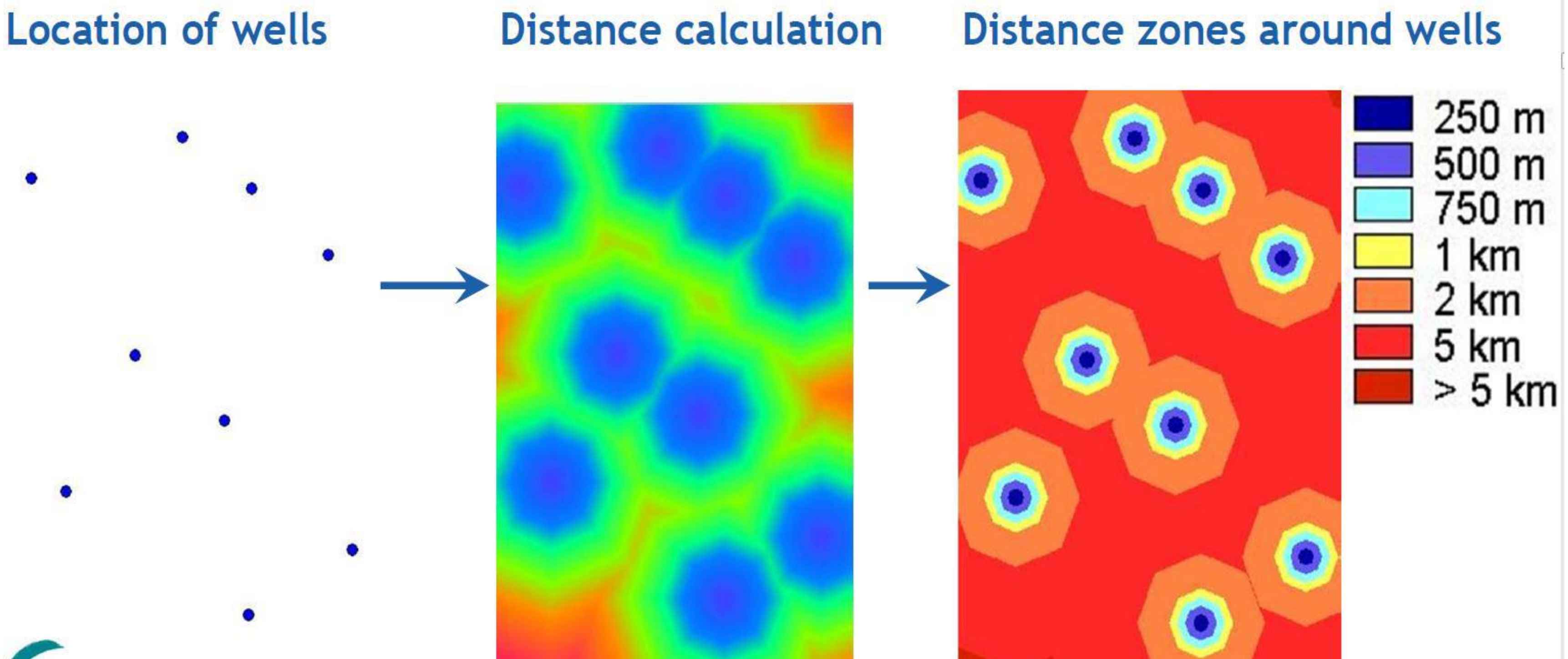
- Need target cell(s)
- The distance function applies the Pythagorean distance between the cell centers.
- Using cell resolution as the unit distance
- The distance from a non-target cell to the target is the minimal distance one can find between that non-target cell and any target cell.



Raster Analysis: Neighborhood Functions

Proximity computation: Buffer

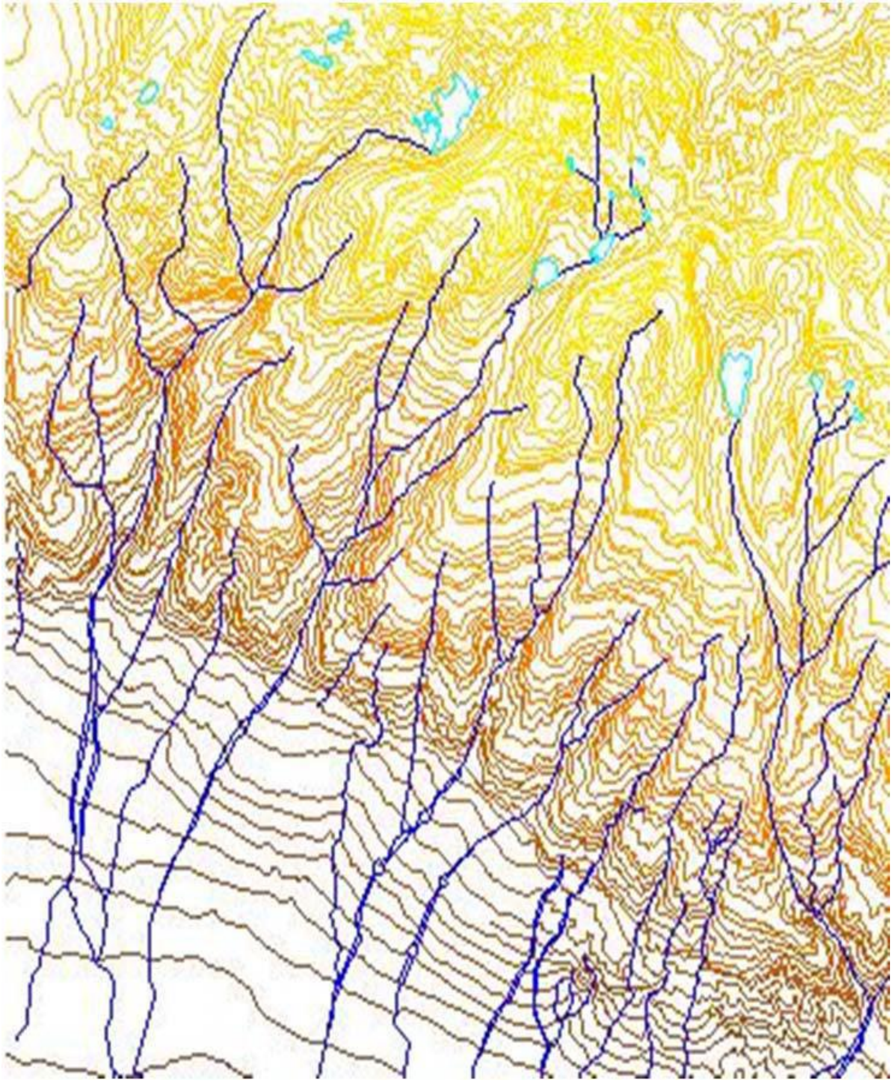
Determine the distance at 250m, 500m, 750m, 1 km, 2 km, 5 km and > 5km from a well



Raster Analysis: Neighborhood Functions

Flow Computation

- Also called **seek** computation
- applies when a phenomenon does not spread in all directions, but chooses a ***least-cost path***.
- Typical example:
  - Determination of drainage patterns in a catchment



Raster Analysis: Neighborhood Functions

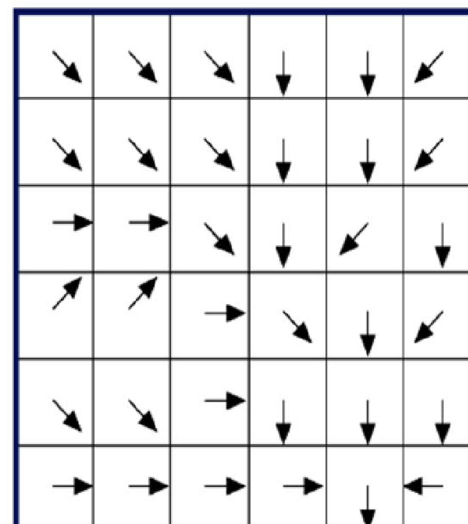
Flow Computation

- Input for a flow computation is an **elevation raster**
- For each cell, the **steepest downward slope** to a neighbor cell is determined
- The direction of this downward slope is stored in the **flow direction raster**

156	144	138	142	116	98
148	134	112	98	92	100
138	106	88	74	76	96
128	116	110	44	62	48
136	122	94	42	32	38
148	106	68	24	22	24

Input Elevation Raster

Output Directional Raster or Flow Direction



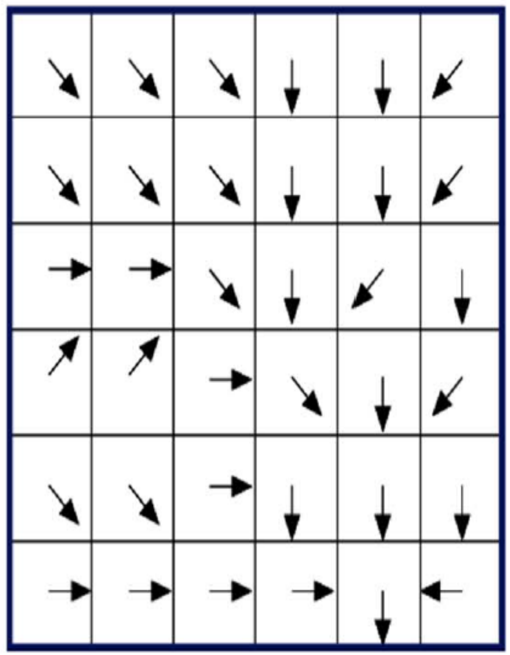
Raster Analysis: Neighborhood Functions

Flow Computation

Identify for each cell the steepest downhill slope

156	144	138	142	116	98
148	134	112	98	92	100
138	106	88	74	76	96
128	116	110	44	62	48
136	122	94	42	32	38
148	106	68	24	22	24

(a Elevation



(b Flow direction

142 – 138 = 4   4

142 – 112 = 30 $30 / \sqrt{2} = 21.2$

142 – 98 = 44   44

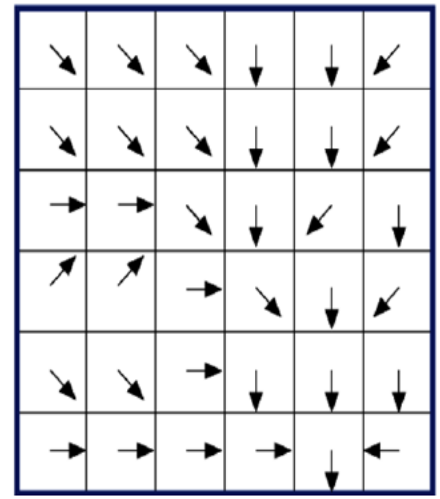
142 – 92 = 50 $50 / \sqrt{2} = 35.4$

142 – 116 = 26   26

Raster Analysis: Neighborhood Functions

Flow Computation – Accumulated Flow

- From a flow direction map, a **flow accumulation map** can be derived.
- The accumulated flow for each cell is **the number of cells that flow into this particular cell**.
- Cells with a **value 0** have no other cells flowing into them, and represent **higher areas**.
- Cells with a **high accumulated flow** count represent **streams**.



0	0	0	0	0	0
0	1	1	2	2	0
0	3	7	5	4	0
0	0	0	20	0	1
0	0	0	1	24	0
0	2	4	7	35	1

Raster Analysis: Neighborhood Functions

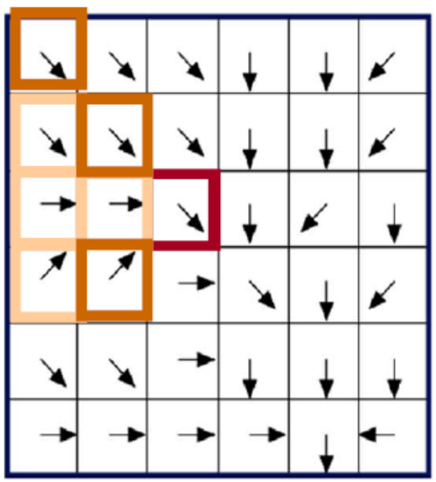
Flow Computation – Accumulated Flow

- the number of cells that flow into this particular cell

156	144	138	142	116	98
148	134	112	98	92	100
138	106	88	74	76	96
128	116	110	44	62	48
136	122	94	42	32	38
148	106	68	24	22	24

(a

Elevation



(b

Flow direction

0	0	0	0	0	0
0	1	1	2	2	0
0	3	7	5	4	0
0	0	0	20	0	1
0	0	0	1	24	0
0	2	4	7	35	1

(c

Accumulated flow count

Raster Analysis: Surface Analysis

Slope and Aspect

- Calculated from a grid of elevations (a digital elevation model-DEM)
- Slope and aspect are calculated at each point in the grid, by comparing the point's elevation to that of its neighbors
- usually its eight neighbors
- but the exact method varies
- in a scientific study, it is important to know exactly what method is used when calculating slope, and exactly how slope is defined

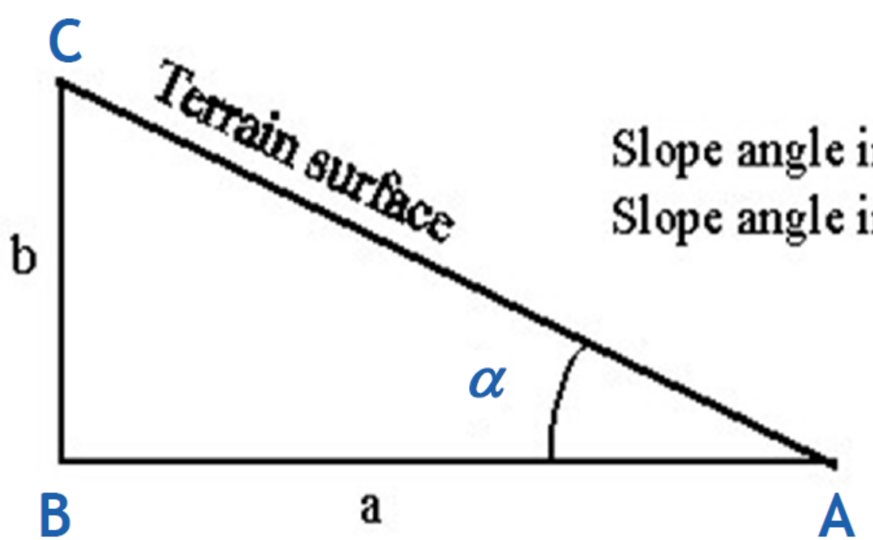
Raster Analysis: Surface Analysis

- **Slope and Aspect**
- **Slope angle calculation**
- the calculation of the slope steepness, expressed as an angle in degrees or percentages, for any or all locations.
- **Slope aspect calculation**
- The calculation of the aspect (or orientation) of the slope in degrees (between 0 and 360), for any or all locations.

Raster Analysis: Surface Analysis

- **Slope angle**

- Change along vertical direction, *CB* is b
- Change along horizontal direction, *AB* is a
- Equations to estimate slope angle α are given

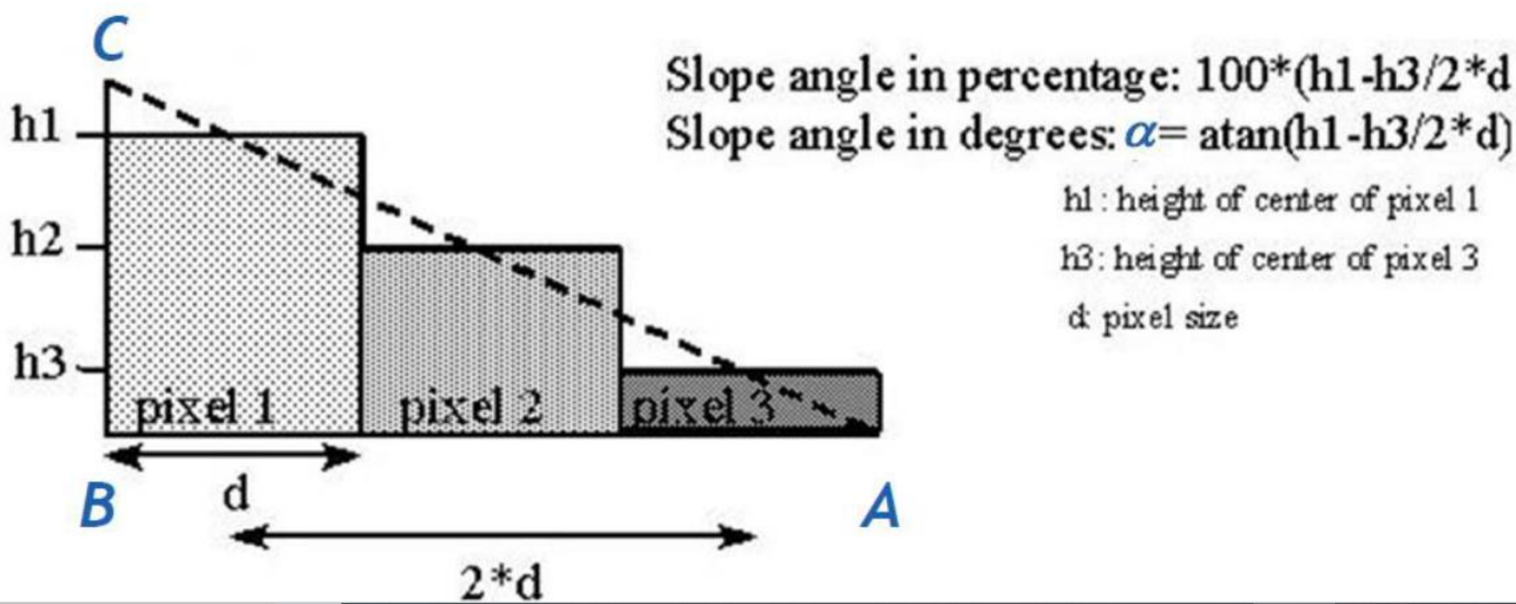


Slope angle in percentage: 100*(b/a)
Slope angle in degrees: α= atan(b/a)

Note: 45° = 100%. The steepness of a slope can exceed the 100%

Raster Analysis: Surface Analysis

Slope Angle Schematic example for raster data:
Gradient in vertical direction, ***CB***, is equal to h1-h3
Gradient in horizontal direction, ***AB***, is equal to 2*d
Equations to estimate slope angle alpha ($\boldsymbol{\alpha}$)are given



The end of the slide