UNIVERSITY OF TWENTE.

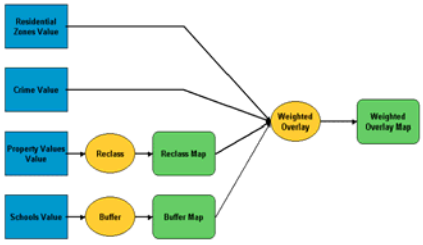
SPATIAL DATA ANALYSIS (1)

ELLEN-WIEN AUGUSTIJN. ADAPTED BY GABRIEL PARODI

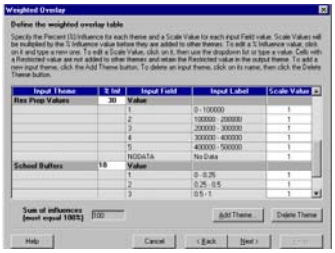
STUDY MATERIAL: PRINCIPLES OF GEOGRAPHIC INFORMATION SYSTEMS
AN INTRODUCTORY TEXTBOOK
CHAPTER 6

ITC FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

0. Introduction

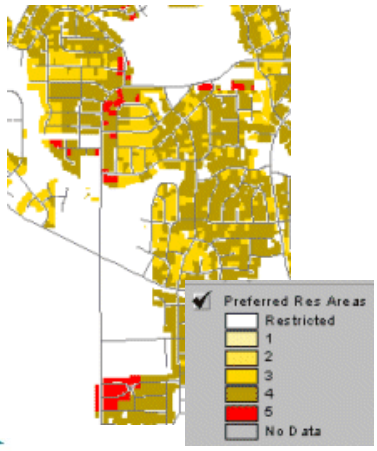


more precisely described in an application model.



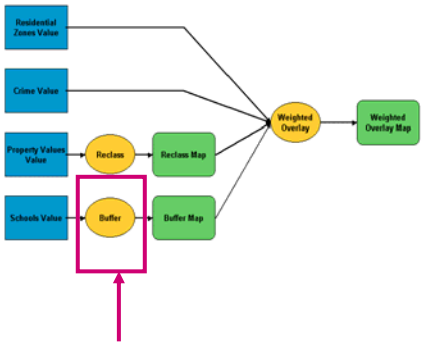
0. Introduction

Site selection – prescriptive



- Application model: description of the behavior of relevant geographic phenomena, it tries to simulate the geographic phenomena that are present.
- Models for planning and site selection are prescriptive they quantify environmental, economic and social factors

0. Introduction



How does this operation work?

- In predictive models, a forecast is made of the likelihood of future events, for example, of a future landslide.
- This lesson does not discuss analytical modeling, it discusses analytical functions that form the building blocks for analytical models.

0. Introduction

- There are many ways to classify analytical functions.
- The one suggested by Aronoff is used here:
  - Measurements, retrieval and classification
  - Overlay functions
  - Neighborhood functions
  - Connectivity functions
- Functions are split over 2 lessons (9 and 10)
- Every function is discussed for vector and for raster data

UNIVERSITY OF TWENTE.

0. Introduction

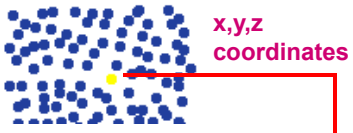
- Overview of this lesson:
- Measurement operations
  - 1.1. Vector measurements
  - 1.2. Raster measurements
- Spatial selection queries
  - 2.1 Spatial Selection by Attribute conditions
  - 2.2 Spatial Selection using topological relationships
- Reclassification
  - 3.1 User controlled classification
  - 3.2 Automatic classification
- Overlay techniques

UNIVERSITY OF TWENTE.

1. Measurements

- Only Geometric Measurements are discussed, no measurements on attribute values
- All measurement functions are discussed both in raster and vector (on points, lines and polygons)
- Vector measurements include: location, length, distance and area size

1.1 Measurements - Vector

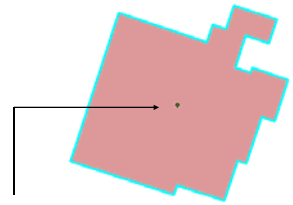


Attributes of Points.shp

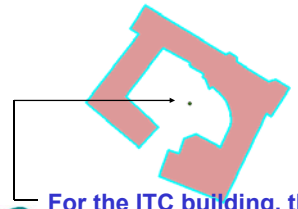
Shape	x-coord	y-coord	z-coord
PointZ	6294767.7100	1979391.8500	968.430
PointZ	6294751.4100	1979316.3100	967.640
PointZ	6294770.5500	1979263.7300	966.520
PointZ	6294795.7500	1979211.2400	969.800
PointZ	6294743.2100	1979187.5000	984.560
PointZ	6294766.8200	1979147.8300	968.360
PointZ	6294723.1100	1979125.0200	969.020
PointZ	6294730.6300	1979073.7500	946.590
PointZ	6294853.6900	1979120.7900	943.900
PointZ	6294879.7500	1979171.0900	943.570

- Vector measurements include: location, length, distance and area size
- Location → always stored by GIS
  - One coordinate pair for points
  - List of pairs for lines and polygons

1.1 Measurements - Vector



The centroid of a square building will be in the middle



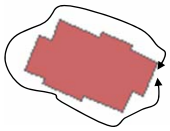
For the ITC building, the centroid is outside the polygon

- Polygons have a centroid.
- The centroid is the center of mass of the polygon, the point on which it would balance under a homogeneously applied force like gravity.
- Note that the centroid can be outside the polygon
- The location of the centroid of a polygon can be either stored or computed on the fly

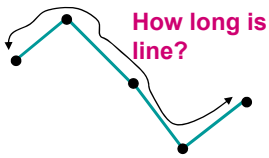
ITC UNIVERSITY OF TWENTE.

1.1 Measurements - vector

How long is the boundary?



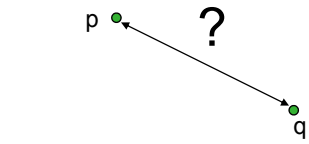
How long is the line?

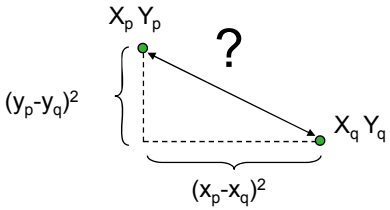


- Length is associated with polylines, and with polygon boundaries.
- It can be stored by the GIS or computed on the fly

ITC UNIVERSITY OF TWENTE.

1.1 Measurement - Vector



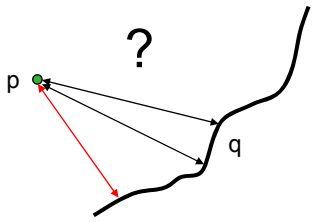


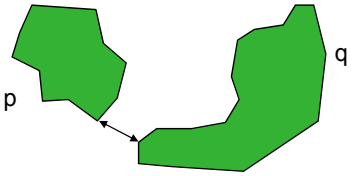
- Distance between two points → Pythagorean distance function:

$$dist(p,q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$$

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

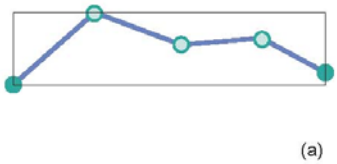
1.1 Measurement - Vector



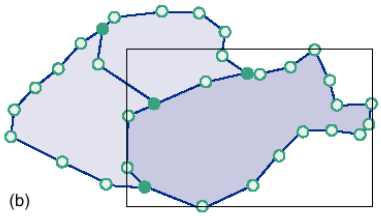


- If one or both features are not a point we will measure the minimal distance between the two features

1.1 Measurement - Vector



(a)

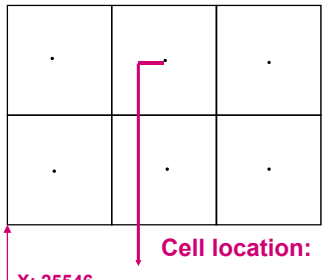


(b)

- A bounding box is the minimal rectangle that covers the feature.
- Bounding box computation is a Support function
- Example: when the bounding boxes do not intersect, the features do not intersect

1.2 Measurement - Raster

Resolution: 20 x 20 meters



X: 25546

Y: 468678

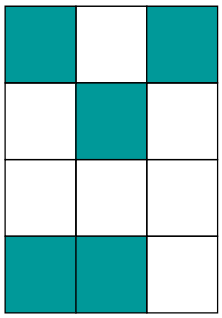
Cell location:

X: 25546 + 30= 25576

Y: 468678 + 30 = 468708

- Raster measurements include: location, distance and area size
- Location of an individual cell → derived from anchor point and resolution
- The cell's location can be its lower left corner or midpoint

1.2 Measurement - Raster



Cell size: 30 m x 30 m

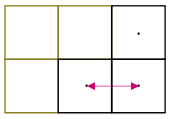
900 * 5 = 4500 m^2

- Area size → number of cells * cell size
- When you know the resolution you can calculate the area of a single cell. In this example 30 x 30 meters = 900 m2.
- The number of cells is also called the frequency or count.

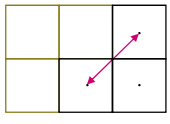
UNIVERSITY OF TWENTE.

1.2 Measurement - Raster

Cell size: 30 m x 30 m



Distance 30 m



Distance 30* √2

- Distance → standard distance function applied to the locations of their mid-points.
- 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.

UNIVERSITY OF TWENTE.

1. Measurement - Summary

- Vector:
- Vector measurements include: location (coordinates and centroids), length, distance and area size
- Another geometric measurement is the minimal bounding box of lines and polygons.
- Pythagorean distance function is used
- Raster:
- Raster measurements include: location, distance and area size
- standard Pythagorean distance function applied to the locations of their mid-points.

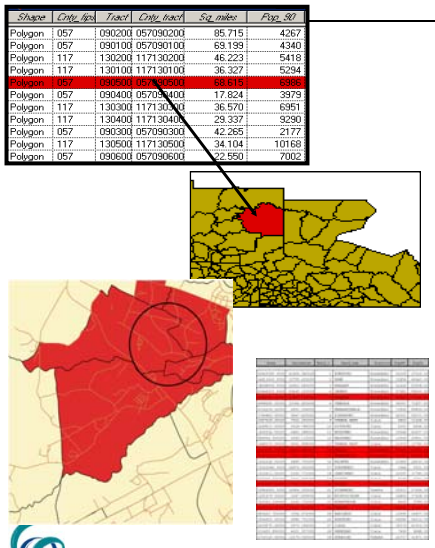
UNIVERSITY OF TWENTE.

2.1 Spatial selection queries

- Interactive
- Spatial Selection by Attribute conditions
  - Relational operators
  - Logical operators
  - 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

UNIVERSITY OF TWENTE.

2.1 Spatial selection queries



- Interactive spatial selection is a selection in which you select features by clicking on the screen (on the feature to select) or drawing a graphic, to select all objects within this graphic

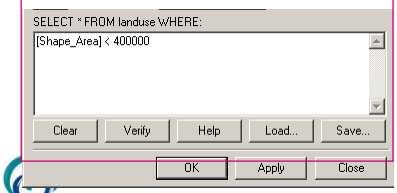
UNIVERSITY OF TWENTE.

2.1 Spatial Selection by Attribute conditions

Example SQL:

```
SELECT  *
FROM                Landuse
WHERE   Area < 4000000
```

Software specific example:



- Attribute queries are selections in which we use a selection condition on the features attributes
- The condition is specified in a query language
- This query language can be SQL when the data is stored in a relational database (lesson 3) or in a software specific language.
- Answers questions : Where are the features with …… ?

UNIVERSITY OF TWENTE.

2.1 Spatial Selection by Attribute conditions

Predictive symbols:

- < (less than)
- = (equals)
- <= (less or equal than)
- > (greater than)
- >= (greater than or equal)
- <> (does not equal)

- A condition that tests a single criterion is called an atomic condition.
- Atomic conditions use a predictive symbol such as < (less than).
- Any of these symbols is combined with an expression on the left and on the right

Expression on the right **<** **Expression on the right**

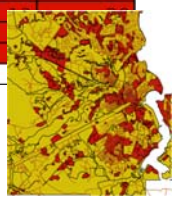
Area < 400,000

UNIVERSITY OF TWENTE.

2.1 Spatial Selection by Attribute conditions

Area < 400,000

Area	IDs	LandUse
174308.7000	2	30
2066475.000	3	70
214582.5000	4	80
29313.8600	5	80
73328.0800	6	80
53303.3000		80
614530.1000	8	20
1637161.000	9	80
156357.4000	10	70
59202.2000	11	20
83289.5900	12	80
225642.2000		
28377.3300		
228930.3000		
986242.3000		



- The expression can be an attribute name (Area), a constant (4000000), or an arithmetic expression:

Ranking > MinRanking + 5

Area / Length < 1

UNIVERSITY OF TWENTE.

2.1 SPATIAL SELECTION BY ATTRIBUTE CONDITIONS

- AND returns true if both expressions a and b are true

Should be true

- (Area < 400,000) AND (landuse = 80)

Should also be true

- Logical connectives are used to combine two atomic conditions to one composite condition
- Examples of logical connectives are: AND, OR and NOT

UNIVERSITY OF TWENTE.

2.1 Spatial Selection by Attribute conditions

(Area < 400,000) OR (landuse = 80)

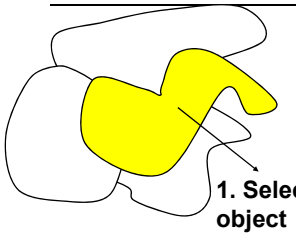
true	false	true
false	true	true
true	true	true
false	false	false

The total expression is…..

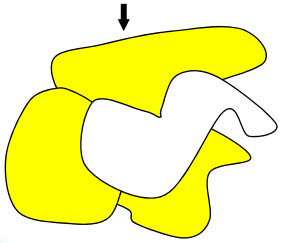
- OR returns true if one or both of the expressions a and b is true
- NOT returns true if the expression is false

UNIVERSITY OF TWENTE.

2.2 Spatial Selection using topological relationships



2. Meets relationship

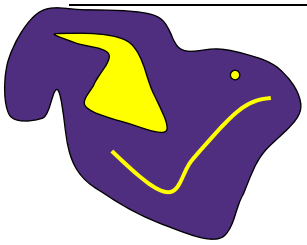


- Spatial selections use topological relationships
- The steps carried out are:

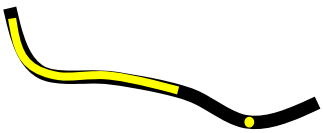
1. Select one or more selection objects.
2. Apply a chosen spatial relationship to determine the features that have that relationship with the selection objects.

ITC UNIVERSITY OF TWENTE.

2.2 Spatial Selection using topological relationships



A polygon can contain another polygon, line or point

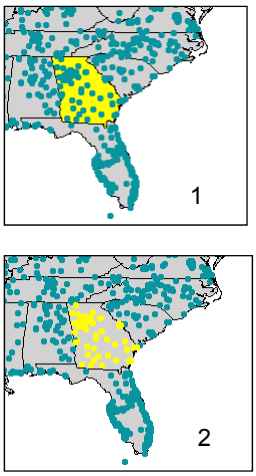


A line can contain a line or a point

- Selecting features that are inside selection objects
- Polygons can contain polygons, lines or points, and lines can contain lines or points

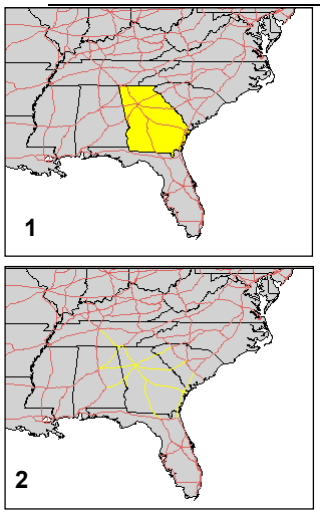
ITC UNIVERSITY OF TWENTE.

2.2 Spatial Selection using topological relationships



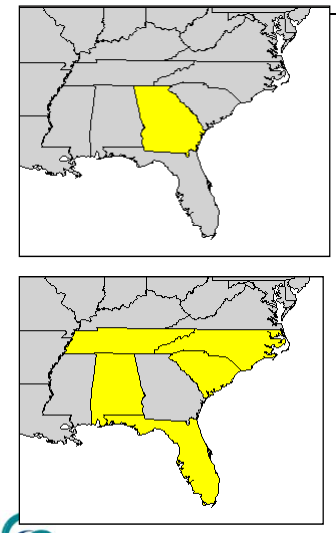
- An example of a selection using the **inside** relationship is shown on the left.
- In picture 1 the state of Georgia is selected, this is the selection object.
- In picture 2 all the cities that are inside this state are selected.

2.2 Spatial Selection using topological relationships



- Selecting features that intersect.
- In picture 1 the state of Georgia is selected
- In picture 2, the interstates that run through Georgia (intersect).

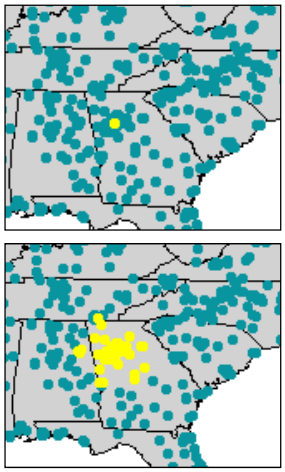
2.2 Spatial Selection using topological relationships



- Selecting features adjacent to selection objects.
- Adjacency is the same as the meets relationship.

UNIVERSITY OF TWENTE.

2.2 Spatial Selection using topological relationships



- Selecting features based on their distance.
- Will search within a given distance from the selection object, at a given distance, or beyond a given distance.
- Example (left) select all the cities within 100 kilometers from Atlanta.

UNIVERSITY OF TWENTE.

2. Spatial Selection - Summary

- Two types of selections, attribute selections and selections using topological (spatial) relationships
- In attribute queries, predictive symbols and Logical connectives are used.
- We distinguish atomic conditions and composite condition.

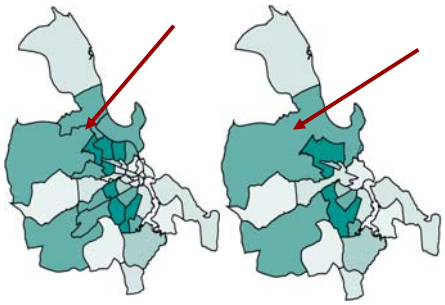
UNIVERSITY OF TWENTE.

3. (Re)Classification

- Remove detail from an input dataset to reveal important spatial patterns.
- Reduce the number of classes and eliminate details.
- If the input dataset itself is the result of a classification we call it a reclassification
- Spatial selection is a two step process:
  - Select one or more selection objects.
  - Apply a chosen spatial relationship to determine the features that have that relationship with the selection objects.

UNIVERSITY OF TWENTE.

3. (Re) Classification – post processing



- Vector reclassification 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.

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

UNIVERSITY OF TWENTE.

3. (Re) Classification – post processing

34.6	45.8	51.3	56.9
43.7	49.9	54.0	59.3
47.6	54.2	56.8	62.5
49.8	58.8	65.2	66.1

3	4	5	5
4	4	5	5
4	5	5	6
4	5	6	6

- For raster datasets, post-processing is not possible, only the value of the cell will change.
- Post-processing of vector data is only sensible for lines and polygons.

UNIVERSITY OF TWENTE.

3. (Re) Classification

- 3.1 User controlled classification
  - Classification table
- 3.2 Automatic classification
  - Equal interval technique
  - Equal frequency technique

UNIVERSITY OF TWENTE.

3.1 User controlled classification

Two Examples of classification tables:

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

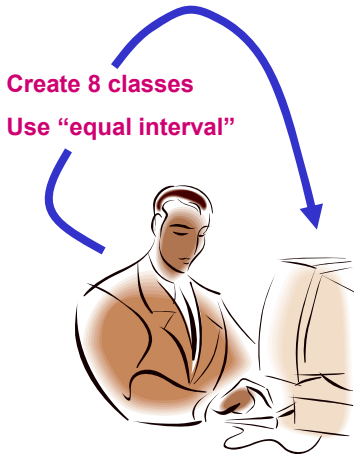
Code	Old value	New value
10	Planned Residential	Residential
20	Industrial	Commercial
30	Commercial	Commercial

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.

UNIVERSITY OF TWENTE.

3.2 Automatic classification



- User specifies the no. of output classes.
- Computer decides the class break points.
- Two techniques of determining the class breaks are discussed:
  - Equal interval:
  - Equal frequency

3.2 Automatic classification

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

Two original values per class

- 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.

3.2 Automatic classification

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

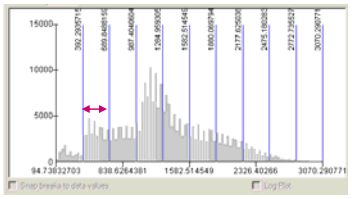
original value	new value	# cells
1	1	6
2,3	2	5
4	3	6
5,6	4	3
8,9,10	5	5

Total number of cells 25 / 5 = 5 cells per class

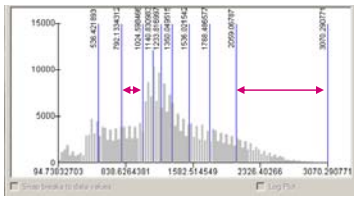
- Equal frequency, is also called quantile.
- Total number of features / number of classes (n)
- The objectives is to create categories with roughly equal number of features (or cells).

UNIVERSITY OF TWENTE.

3.2 Automatic classification



Equal interval technique



Equal frequency technique

- The two techniques are illustrated here again. On the horizontal axis the original values. The classes are the blue vertical lines. In the equal interval example, the blue lines are equally spaced. On the vertical axis the frequency. In the equal frequency technique, when the frequency goes up, the classes become smaller.

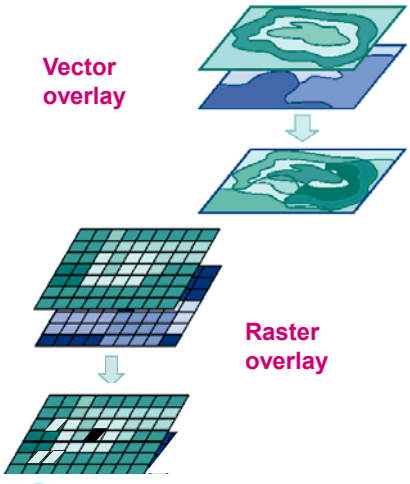
UNIVERSITY OF TWENTE.

3. (Re) Classification - Summary

- Reclassification is used to reduce the complexity of a layer in order to show patterns
- Two types of classifications, user controlled classification and automatic classification
- In user controlled classification the user will create a classification table.
- In automatic classification the user will determine the number of classes and the classification technique.
- Two classification techniques were discussed:
  - Equal interval
  - Equal frequency

ITC UNIVERSITY OF TWENTE.

4. Overlay



- Standard overlay operators take two input data layers, and assume they are geo-referenced in the same system, and overlap in study area. If either condition is not met, the use of an overlay operator is senseless.
- The principle of spatial overlay is to compare the characteristics of the same location in both data layers, and to produce a new output value for each location.

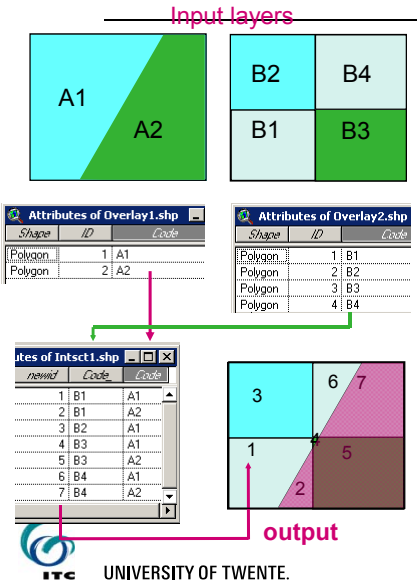
ITC UNIVERSITY OF TWENTE.

4. Overlay

- Overview of the topics discussed:
- 4.1 Vector overlay techniques
  - Intersection
  - Clip by
  - Overwrite by
- 4.2 Raster overlay techniques
  - Arithmetic operators
  - Comparison and logical operators
  - Conditional expressions
  - Decision table

UNIVERSITY OF TWENTE.

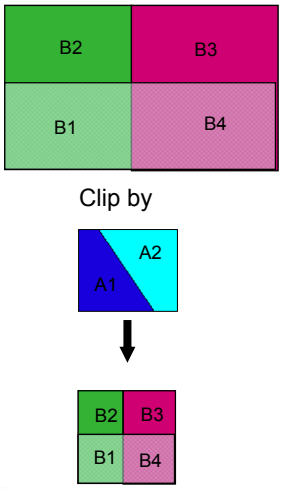
4.1 Vector overlay - intersection



- The standard operator for two layers of polygons is the polygon intersection operator.
- The result of this operator is the collection of all possible polygon intersections
- The attribute table combines the information of the two input tables (spatial join).

UNIVERSITY OF TWENTE.

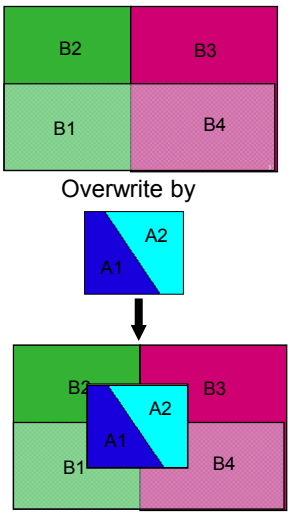
4.1 Vector Overlay Clip



- Clip takes a polygon data layer and restricts its spatial extent (the area that it covers) to the outer boundary of a second input layer (clip layer).
- No other polygons from the clip layer play a role in the result.
- This technique can be used to reduce the area of a thematic layer to that of the study area.

UNIVERSITY OF TWENTE.

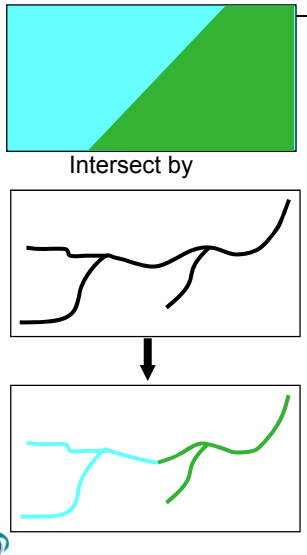
4.1 Vector Overlay Overwrite



- The polygon overwrite creates a layer with the polygons of the first layer except where polygons exist in the second layer (as they take priority)
- This operator can be used to overwrite a layer with “updates“ stored in a second layer.

UNIVERSITY OF TWENTE.

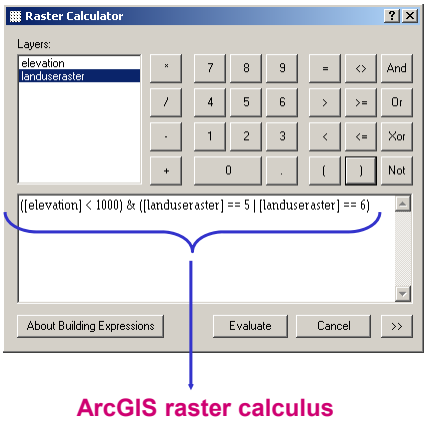
4.1 Vector Overlay



- Vector overlays are usually also defined for point and line data layers.
- When a polygon layer is intersected by a line layer the result will be a line layer (the layer of the lowest order).

UNIVERSITY OF TWENTE.

4.2 Raster overlay operators



ArcGIS raster calculus

- Vector overlay operators are geometrically complicated, raster overlays are mostly cell by cell computations.
- GISs that support raster processing have a full language to express operations. This is called a raster calculus
- The syntax used here is based on the ILWIS raster calculus.

UNIVERSITY OF TWENTE.

4.2 Raster overlay operators

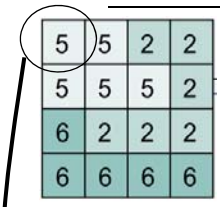
- New cell values are calculated using calculus - map algebra.
- Performed on cell-by-cell basis.
- No geometric calculation

Output_raster_name := Raster_calculus_expression

4.2 Raster overlay operators

- Overview:
- Arithmetic overlay operators
- Comparison and logical operators
- Conditional expressions
- Decision table

4.2 Raster overlay - arithmetic

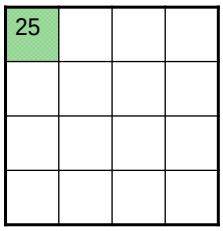


5	5	2	2
5	5	5	2
6	2	2	2
6	6	6	6

Input raster layer (Raster1)

Raster 2:= Raster1 * 5

? := 5 * 5



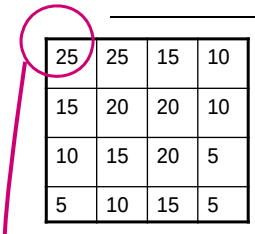
25			

- Arithmetic operators
  - +, -, *, /
  - MOD (modulo division)
  - DIV (integer division)
  - goniometric operatos: sin, cos, tan, asin, acos, atan.
- For example:

Raster2 := Raster1 * 5

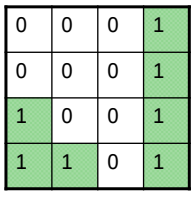
UNIVERSITY OF TWENTE.

4.2 Raster overlay – comparison/logical



25	25	15	10
15	20	20	10
10	15	20	5
5	10	15	5

Output_raster := Input_raster < 15



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

True – 1

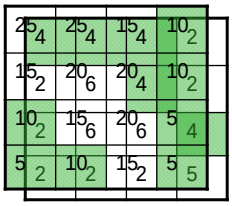
False - 0

- Comparison operators
  - <, <=, =, >=, >, <>
  - C:=A<>B is true when the cell's value in A differs from the cell's value in B. It is false if they are the same.
- Logical operators
  - AND,OR,NOT, XOR (exclusive)
    - a XOR b is true if either a or b is true, but not both.

UNIVERSITY OF TWENTE.

4.2 Raster overlay – comparison/logical

A B



Output_raster := A < 15 AND B = 4

- the cell values of the output raster is either true or false. Other values are not possible.

True – 1

False - 0

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

ITC UNIVERSITY OF TWENTE.

4.2 Raster overlay – conditional expr.

A

25	25	15	10
15	20	20	10
10	15	20	5
5	10	15	5

B

4	4	4	2
2	6	4	2
2	6	6	4
2	2	2	5

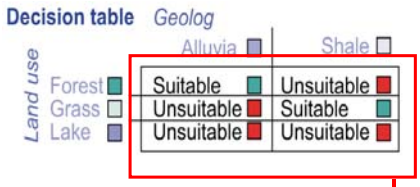
Output:= IFF (A < 15 OR B = 2, A, 40)

40	40	40	10
15	40	40	10
10	40	40	5
5	10	15	5

- Conditional expressions test a condition, if the expression in true, the then_expression will be evaluated, if it is false the else_expression.
- A conditional expression can lead to many different types of outputs, numeric value but also true/false
- Format: output_raster:= IFF (condition, then_expression, else_expression)

ITC UNIVERSITY OF TWENTE.

4.2 Raster overlay – decision table



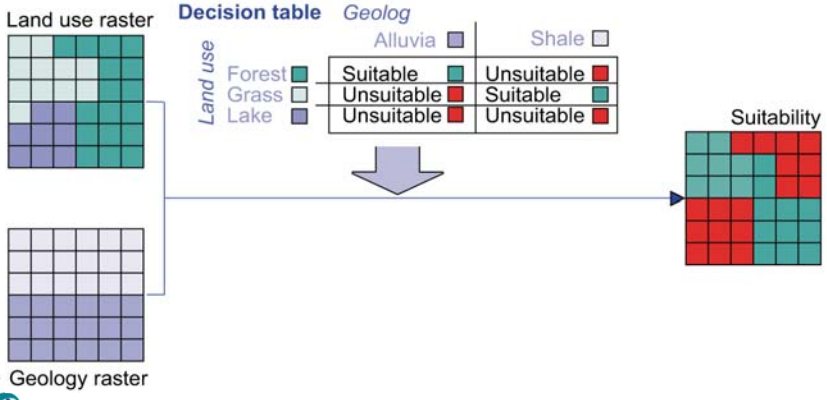
The table above lists two input layers, Landuse and Geolog. Landuse has three values (Forest, Grass, Lake) and Geolog has two values (Alluvia and Shale).

For each combination it lists the output values.

- Decision tables are the same statement as in a Conditional statement, but presented in a different way.
- The decision table will guide the overlay process
- It lists all possible combinations of input values, and the output values.

4.2 Overlay operations - RASTER

- Overlay using a decision table
  - for complicated conditional expressions.



4. Overlay - Summary

- Vector overlay techniques, intersection, clip by and overwrite by.
- Intersection is the fundamental operator, the attribute table is a spatial join (fields from both input tables.
- Clip by is like a cookie cutter, cutting out the map extend of the second layer.
- For types of raster overlay techniques, using arithmetic operators, comparison and logical operators, conditional expressions and decision table.
- Comparison and logical operators only evaluate to true and false.
- Conditional expressions and decision table lead to the same result

UNIVERSITY OF TWENTE.

Lecture End

UNIVERSITY OF TWENTE.