ISA

Geographic Information Systems 2026/2027

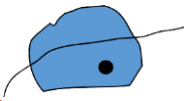
Lesson 13 - Spatial Data Structures - Raster

U LISBOA | UNIVERSIDADE DE LISBOA INSTITUTO SUPERIOR DE AGRONOMIA

Topics

- The vector data structure
- The raster data structure
- Vector to raster and raster to vector data conversions
- Raster datasets (overview)
  - The reclassification operation
  - Boolean datasets
  - Map Algebra
  - Histograms
  - Legends

Vector data structure



So far we have been representing geographic objects as points, lines or polygons (singlepart or multipart)

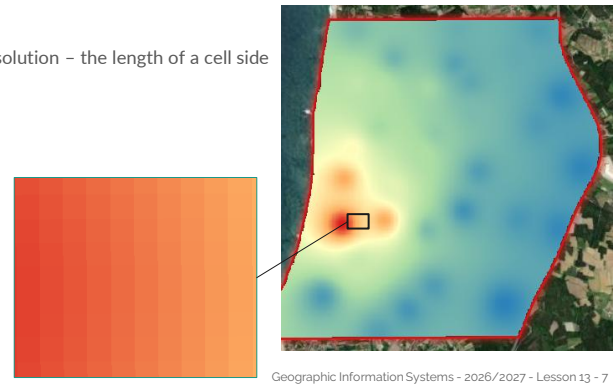
- Every point is defined by an ordered pair of coordinates referenced to a Coordinate Reference System
  - The pair of coordinates is composed with numbers (in the CRS map units) called longitude/meridian/X & latitude/parallel/Y
  - When elevation is also considered, points are defined by ordered triplets such as (x, y, z) where z is the height value (in the CRS height units)
- Every line is defined by an ordered set of points (in this context also called vertexes or vertices)
  - Lines can also be 3D when their vertices are 3D
- Every polygon is defined by a line which starting and ending point coincide (also called closed line)
  - Polygons can also be 3D when their vertices are 3D
- Singlepart features are elementary features (point, line or polygon); multipart features are composed by elementary features with a common geometry (point, line or polygon)

Raster data structure



In raster data structure, the area is divided into rectangular (usually squares) cells – or pixels (short for picture elements):

- cells are of fixed size
- cells size is also called spatial resolution – the length of a cell side as measured on the ground



Raster data structure



In raster data structure, the area is divided into rectangular (usually squares) cells – or pixels (short for picture elements):

- cells are of fixed size
- cells size is also called spatial resolution – the length of a cell side as measured on the ground

All geographic variation is expressed by assigning **one number to every cell**

- each cell value represents the value of a given attribute
- cells values may also be the unique identifier assigned to objects fully described elsewhere
  - usually a non-spatial table

The **geographic location** is defined by:

- the geographic coordinates of one of the 4 grid vertices
- the grid direction (for instance, South-North)
- and the raster resolution

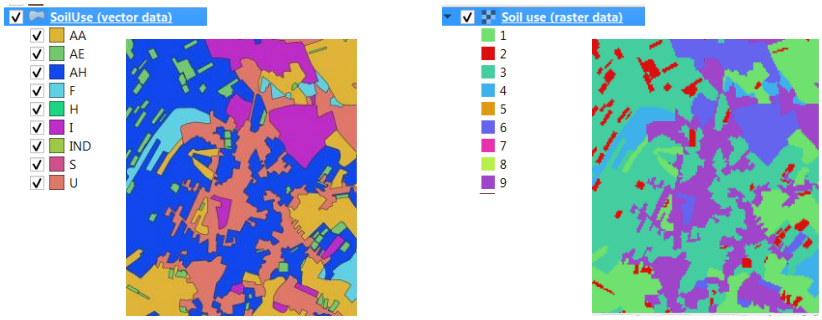
Data structures vs. file formats

Despite vector & raster data are usually stored in different file formats, data structure and file format have different meanings:

- **Vector data** can be stored using several file formats: **geoPackage**, **file geodatabase**, **shapefile**, **Drawing Exchange Format (dxf)**, **DGN**, **DGW**, ...
- Raster data can be stored using several file formats: **geoTiff**, **ILWIS**, **IMG**, **geoPackage**, **file geodatabase**

Raster & Vector visualization

Soil use in a given region represented by vector and raster data (spatial resolution 10 m)

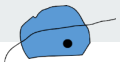




Vector data – advantages and disadvantages

ADVANTAGES

- The shape of geographic objects can be **accurately represented** – even very small objects, objects with **irrelevant area** and non-physical objects
  - examples in many applications: communication lines, water pipes, electrical wires, streams, …
  - examples of non-physical objects: country & other administrative boundaries, mountain & hill ridges, contour lines, or atmospheric pressure …



Vector data – advantages and disadvantages

ADVANTAGES

- The shape of geographic objects can be **accurately represented** – even very small objects, objects with **irrelevant area** and non-physical objects
- Better suited to **represent discrete geographic objects** (objects with distinct boundaries)
- Each dataset can represent **several different variables** – the number of attributes is (theoretically) unlimited
- Spatial data editing is allowed
- Lower volume of data storage

DISADVANTAGES

- Continuous geographic variables **representation is difficult** & usually requires significant data simplification
- Most spatial analysis functions require **heavy data processing**
  - e.g., algorithms to compute the intersection of polygons and lines are complex

Raster data – Advantages & disadvantages



ADVANTAGES

- Spatial analysis involving logical & arithmetic operators – **Map Algebra** – & functions to model space phenomena are easily processed
- Better suited to represent **continuous geographic variables**
  (variables representing real numbers that smoothly change over space)
  - E.g., elevation, slope, temperature, precipitation…
  - Some spatial data sources only use the raster structure
    - Satellites were conceived to provide huge volumes of raster spatial data
    - Estimation of variable values by spatial interpolation of a sample of punctual observations

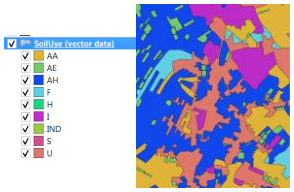
DISADVANTAGES

- **Larger volume of data**
- The shape of spatial objects can be artificial
  - can be overcome by increasing the raster spatial resolution (decreasing the cell size)
    - For instance, a spatial resolution of 0.5 m can represent real world objects almost like they actually are

Object view vs. Field view

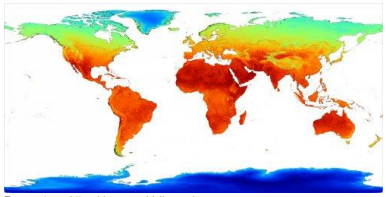
VECTOR

- Vector data structure is best suited to represent the geographic reality that is composed by **objects with distinct boundaries** (sometimes spread over space)
  - Aim: represent & describe only some relevant locations – the object view: **where is what**?
  - Geographic reality considered as discontinuous



RASTER

- Raster data structure is best suited to represent a **geographic reality existing everywhere** on a given region
  - Aim: represent all transition values over space of a given variable – the field view: what is where?
  - Geographic reality considered as continuous

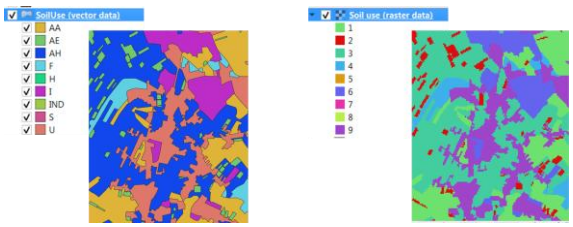


Temperature - https://www.worldclim.org/

Vector data conversion to raster data

Raster data structure **only represents one variable.** It is possible to convert vector dataset to raster dataset representing **one** of the vector gds attributes:

- This attribute must be **numeric** (integer or real)
- If the attribute is not numeric, it is necessary to do a **conversion**:
  - assigning a different number to each unique value of that attribute
    - example: if the attribute values are A, B and C then A → 1, B → 2 & C → 3
  - in the output raster, pixel values must be interpreted according to the rule established above



Vector data conversion to raster data

Raster data structure **only represents one variable.** It is possible to convert vector dataset to raster dataset representing **one** of the vector gds attributes:

- This attribute must be **numeric** (integer or real)
- If the attribute is not numeric, it is necessary to do a **conversion**:
  - assigning a different number to each unique value of that attribute
    - example: if the attribute values are A, B and C then A → 1, B → 2 & C → 3
  - in the output raster, pixel values must be interpreted according to the rule established above
- **Spatial resolution** of the output raster dataset is also required in order to execute this kind of data conversion
- **CRS** of the output raster dataset is also required
  - usually, by default, the CRS of the input vector dataset is the CRS of the output raster dataset but this option can be overridden

Raster data conversion to vector data

It is possible to convert a raster dataset to a vector dataset … with precautions:

- Output vector gds is always geometric type **polygon** (usually singlepart)
- Polygons are built with adjacent cells (pixels) **sharing a same value** – thus, the output polygons shape follow the pixels edges
- If the raster dataset represents a truly continuous variable, **there is a low probability** that adjacent pixels share the same value over large regions;
  - in this case, output vector gds will **contain many small polygons** (worst case scenario, the number of output polygons corresponds to the number of input pixels!)
    - This gds will be **too difficult to use vector spatial operations** to process the output vector data
    - in this kind of situation, before executing the conversion, a reclassification of the raster dataset is always mandatory
- The **CRS** of the output vector dataset is also required
  - usually, by default, the CRS of the input raster dataset is the CRS of the output vector dataset but this option can be overridden

Reclassification operation

- Only applies to raster datasets
- Allows to change the pixel values according to a given set of rules
  - Usually these rules assign a number (representing a class) to every interval of input pixel values
    - these intervals must be disjoint and the union of all intervals must contain the domain of the input pixel values
    - Example:

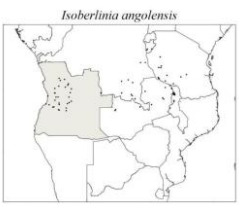
      [1 ; 99.999]→ 50 [100 ; 199.999]→ 150 [200 ; 299.999]→ 250 …

      or

      [1 ; 199.999]→ 1 [200 ; 299.999]→ 0
- The reclassification operation is often used either to reduce the number of pixel unique values or to create a Boolean dataset

Boolean datasets

- Boolean datasets are raster datasets where the domain of the pixels is {0 , 1}
- Usually Boolean datasets are used to represent location suitability concerning a given criteria:
  - value 1 (1 = true) means pixel location verifies a given condition & value 0 (0 = false) means pixel location does not verify this condition
    - example: given a raster dataset representing the average yearly temperature (units degrees Celsius), locations suitable to forest are those with temperature lower than 18°C;
      the Boolean dataset corresponding to this rule would require the following reclassification operation of the raster dataset representing the average yearly temperature [1 ; 18]→ 1





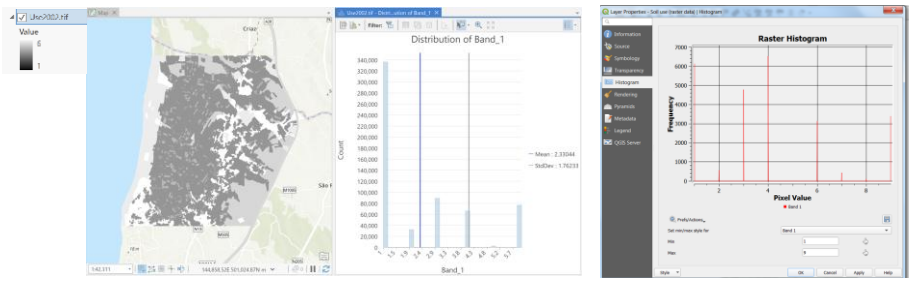
Map Algebra

- Map Algebra refers to a set of arithmetic & logical operations involving pixel values of one or more raster datasets
- Operations: logical conjunction, disjunction, negation, arithmetic addition, subtraction, division, multiplication, exponentiation, logarithm, trigonometric sine, cosine, tangent, co-tangent, …
  - Usually these kind of operations apply to input raster dataset with the same extent and spatial resolution (otherwise additional precautions must be taken)
  - Map Algebra operations are computed for every input values of pixels representing the same location. Example:

0	-5	3	+	1	2	1	=	1	-3	4
4	2	1		3	-1	0		7	1	1
-2	1	3		-2	0	5		--4	1	8
-1	0	1		3	1	-2		2	1	-1

Raster datasets: histograms

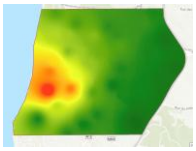
- Histograms are a usual technique to explore the contents of a raster dataset
- Histograms represent the frequency distribution (y-axis) of pixel values (x-values)
  - Example of a histogram:

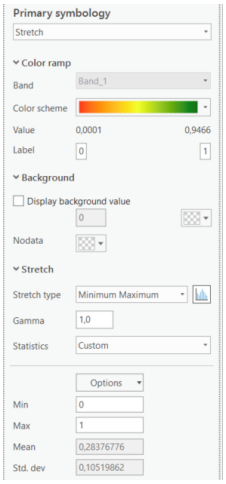


Raster datasets: legends (ArcGIS)

Raster datasets legends can be:

- **Stretch** – suited to continuous variables
  - uses the statistics of the raster layer (needs to be calculated in advance)
  - stretches the layer values along a colour ramp (grey or colour)
  - by default, the lowest colour value is assigned to the pixel's minimum value & the highest colour to the maximum value of the pixels
  - stretch methods: **StdDev**, **MinMax**, **PercentMinMax**, **Sigmoid**, **None**

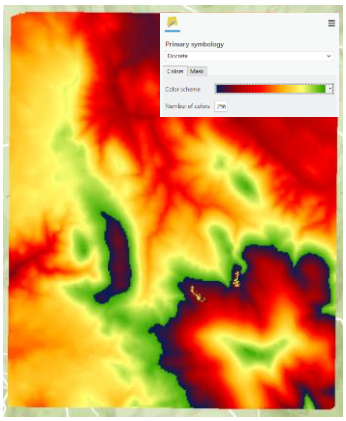




Raster datasets: legends (ArcGIS)

Raster datasets legends can be:

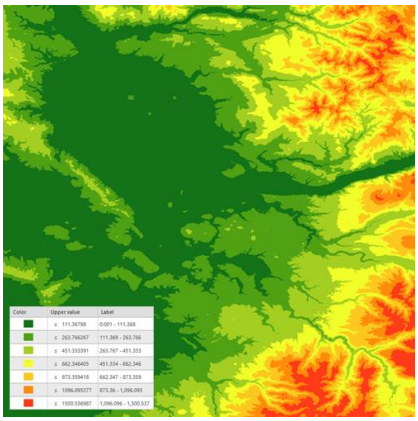
- **Stretch** – suited to continuous variables
- **Discrete** - its a continuous colour schema used on grouped data based in the number of colours provided



Raster datasets: legends (ArcGIS)

Raster datasets legends can be:

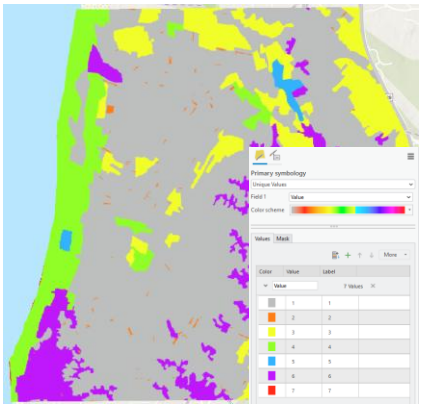
- **Stretch** – suited to continuous variables
- **Discrete** - its a continuous colour schema used on grouped data based in the number of colours provided
- **Classify** - Similar to Discrete, in allowing to define the number of classes, but allows to define limits for each class



Raster datasets: legends (ArcGIS)

Raster datasets legends can be:

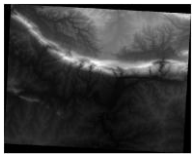
- **Stretch** – suited to continuous variables
- **Discrete** - its a continuous colour schema used on grouped data based in the number of colours provided
- **Classify** - Similar to Discrete, in allowing to define the number of classes, but allows to define limits for each class
- **Unique values** – a different colour is assigned to each unique pixel value; suited to raster datasets with a small number of unique pixel values (representing a discrete variable)

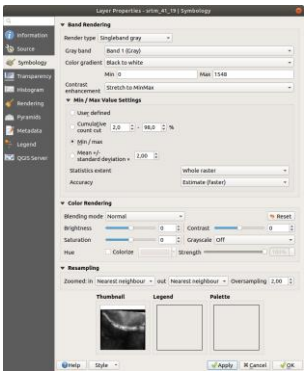


Raster datasets: legends (QGIS)

Raster datasets legends can be:

- **Singleband gray** – suited to continuous variables
  - by default, black colour is assigned to the pixels' minimum value & the white colour is assigned to the maximum value of the pixels
  - a palette between black & white (gray) is assigned to pixel values between the minimum & maximum values using a linear function
    (option "contrast enhancement" -> "Stretch to MinMax")

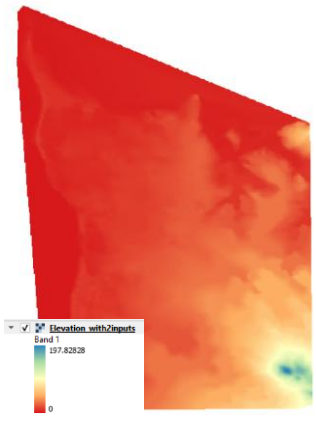




Raster datasets: legends (QGIS)

Raster datasets legends can be:

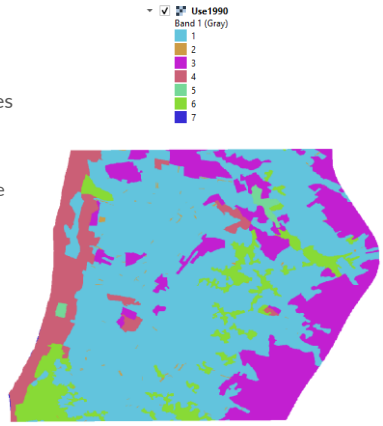
- **Singleband gray** – suited to continuous variables
- **Singleband pseudocolour** – suited to continuous variables
  - a coloured palette can now be used to assigned pixel values between the minimum & maximum values using a linear function
    (option "colour ramp" to select colour palette)



Raster datasets: legends (QGIS)

Raster datasets legends can be:

- **Singleband gray** – suited to continuous variables
- **Singleband pseudocolour** – suited to continuous variables
- **Palleted/ Unique values** – different colours are assigned to each unique pixel value
  - suited to raster datasets with a small number of unique pixel values (representing a discrete variable)
  - similar to ArcGIS Pro "Unique values"



Raster datasets: legends (QGIS)

Raster datasets legends can be:

- **Singleband gray** – suited to continuous variables
- **Singleband pseudocolour** – suited to continuous variables
- **Palleted/ Unique values** – different colours are assigned to each unique pixel value
- **Hillshade** – uses hillshading (details later on)
- **Multiband colour** – suited to multiband raster datasets
  - the analysis of the raster data histogram usually helps to elect the legend type best suited for the raster dataset

Credits

The structure and content of the course, including most of the slides and exercises was developed by Professor Graça Abrantes (SIG 2021_2022) and Professor Rui Figueira (SIG 2023_2024)