Food and Agriculture Organization of the United Nations
Rural Development Administration

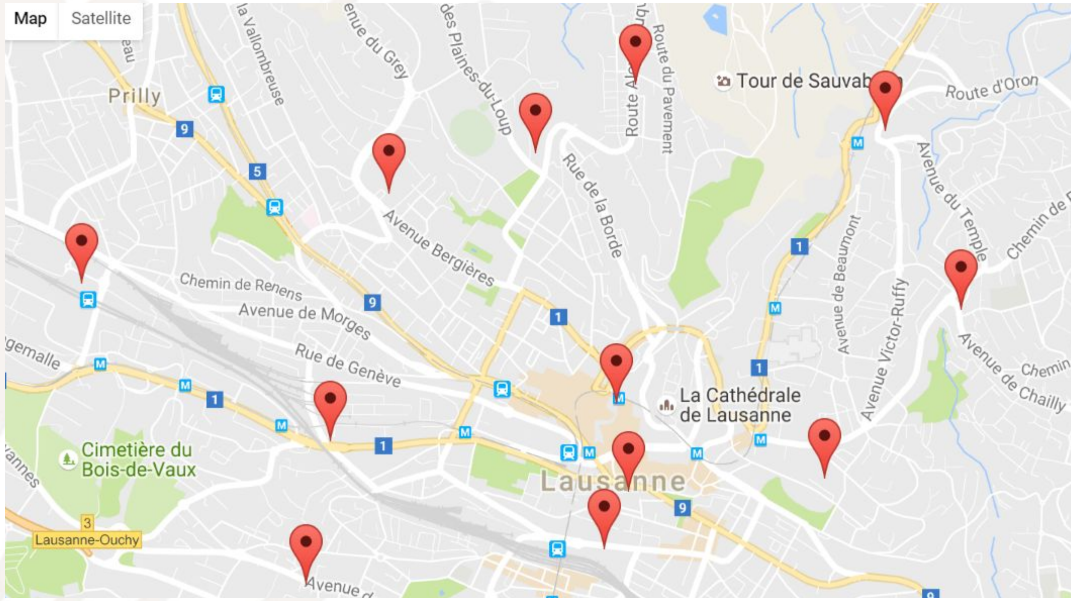
Introduction to Spatial Data and Digital Soil Mapping

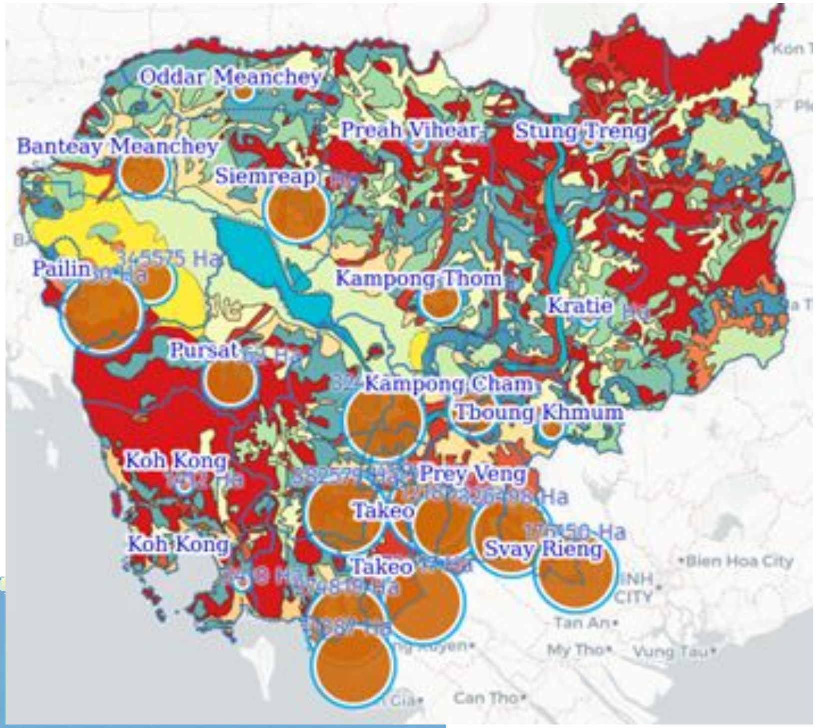
Isabel Luotto

Table of contents:

What is spatial data?

- **Spatial data** (geospatial data, geographic information) it is the data or information that identifies the **geographic location** of features and boundaries on Earth
- Spatial data is stored with **coordinates** to describe location;
- **Maps** are a representation of the spatial data;
- Spatial data is often accessed, manipulated or analyzed through **Geographic Information Systems** (GIS).





educational platform

Spatial phenomena

Spatial phenomena can generally be thought of as either :

- **discrete objects** with clear boundaries (e.g. river, road, town)
- or as a **continuous** phenomenon that can be observed everywhere, but does not have natural boundaries (e.g. elevation, temperature, and air quality).

Elevation is a continuous phenomena: it exists in every point

A lake is a discrete object: it has clear boundaries

Spatial phenomena

Questions:

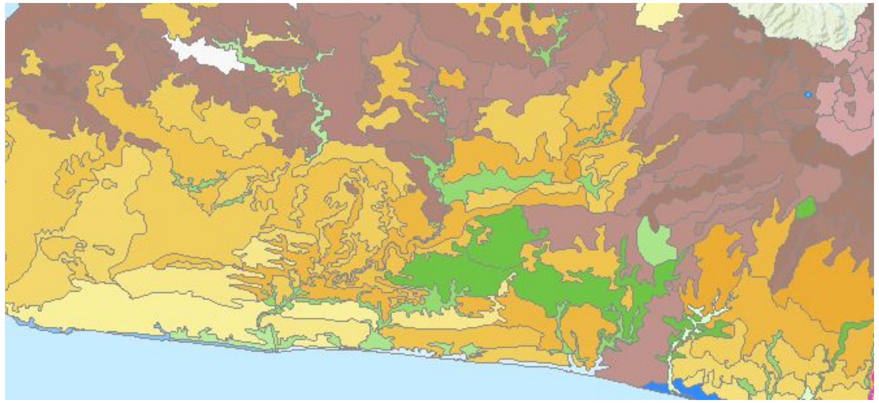
- Is **soil** a discrete or continuous phenomenon?
- Are **soil properties** (carbon, texture, pH, etc...) discrete or continuous?
- Are **soil measurements** discrete or continuous?

Spatial phenomena

Answers:

- **Soil** is a **continuous** phenomenon. It covers almost all land surface.

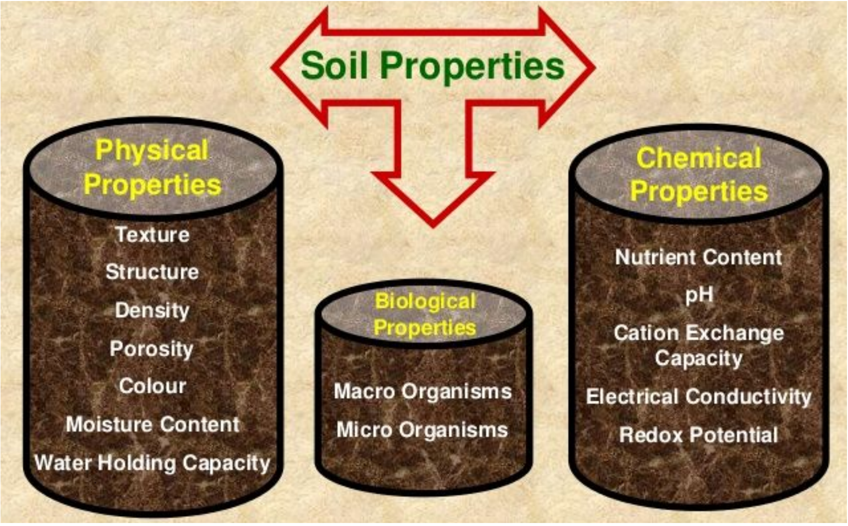
- *However **we classify** soils into **discrete** soil types that have boundaries*



Spatial phenomena

Answers:

- Most of the **soil properties** are continuous. They exist **in every point** of the soil.
- **Soil measurements** are discrete. We **sample** soil and measure soil properties in **discrete locations**.



Spatial Data Models

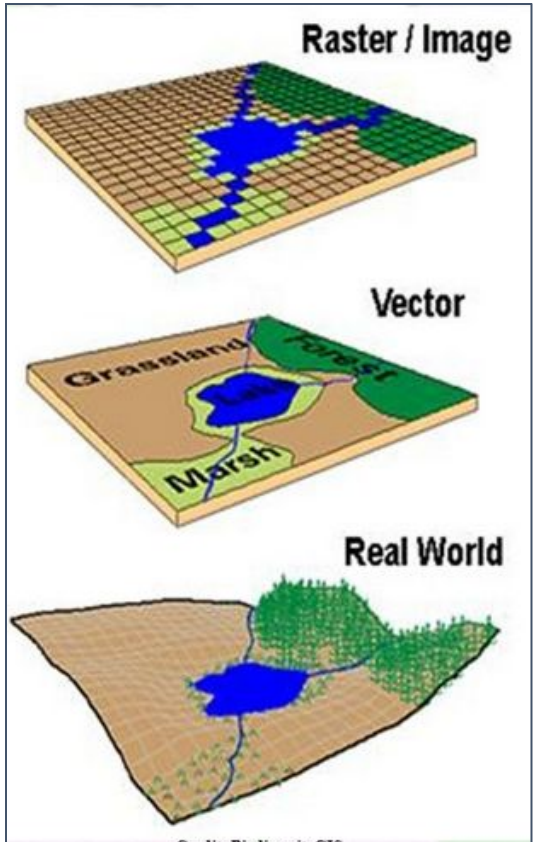
There are 2 ways to represent spatial data: **Vector** and **Raster**

Vector model:

- **Objects**, represented as points, lines and polygons
- Good for storing **discrete data**

Raster model:

- **Surface** divided into a regular grid of cells (pixels)
- Can be used both for storing **discrete** and **continuous data**

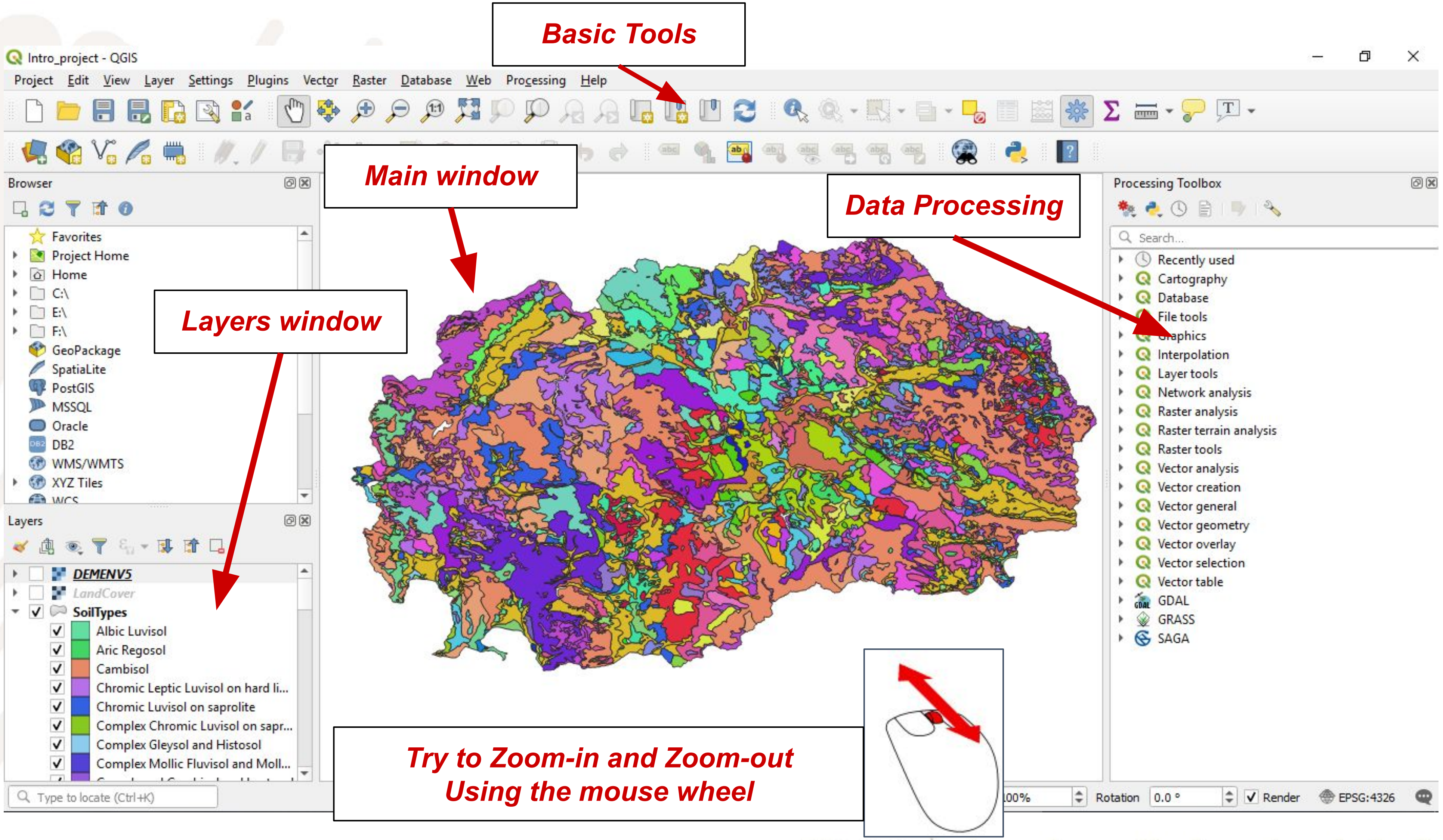


Credit: BiodiversityGIS

Introduction QGIS

Now, let’s see how spatial soil data works **GIS**!

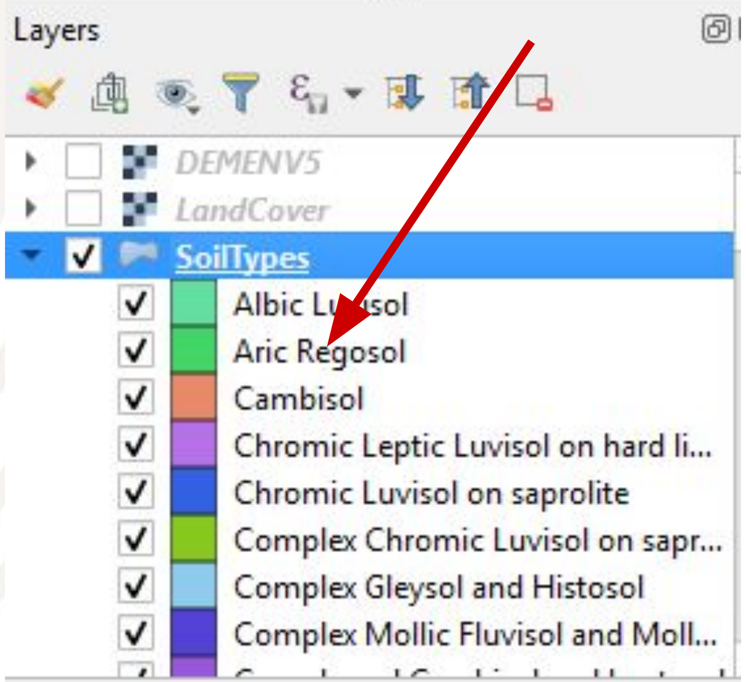
- Find the folder **QGIS Introduction**
- Open file **Intro_project.qgz**
- Wait while QGIS is loading...



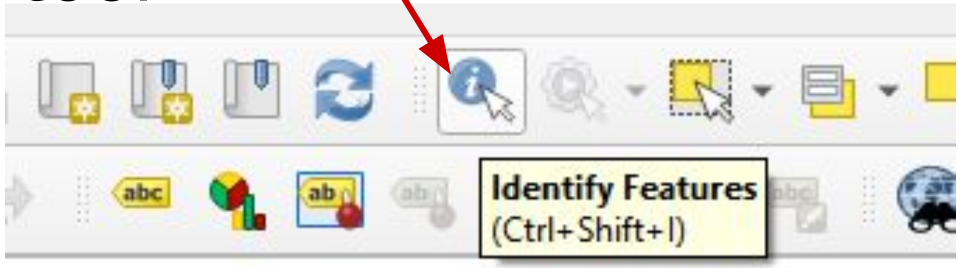
Basic Tools
Main window
Data Processing
Layers window
Try to Zoom-in and Zoom-out
Using the mouse wheel
Intro_project - QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Browser
Favorites
Project Home
Home
C:\
E:\
F:\
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
XYZ Tiles
Layers
DEMENV5
LandCover
SoilTypes
Albic Luvisol
Aric Regosol
Cambisol
Chromic Leptic Luvisol on hard li...
Chromic Luvisol on saprolite
Complex Chromic Luvisol on sapr...
Complex Gleysol and Histosol
Complex Mollic Fluvisol and Moll...
Type to locate (Ctrl+K)
Processing Toolbox
Search...
Recently used
Cartography
Database
File tools
Graphics
Interpolation
Layer tools
Network analysis
Raster analysis
Raster terrain analysis
Raster tools
Vector analysis
Vector creation
Vector general
Vector geometry
Vector overlay
Vector selection
Vector table
GDAL
GRASS
SAGA
Rotation 0.0 °
Render
EPSG:4326

Vector data

- Click on the **SoilTypes** layer to select it.



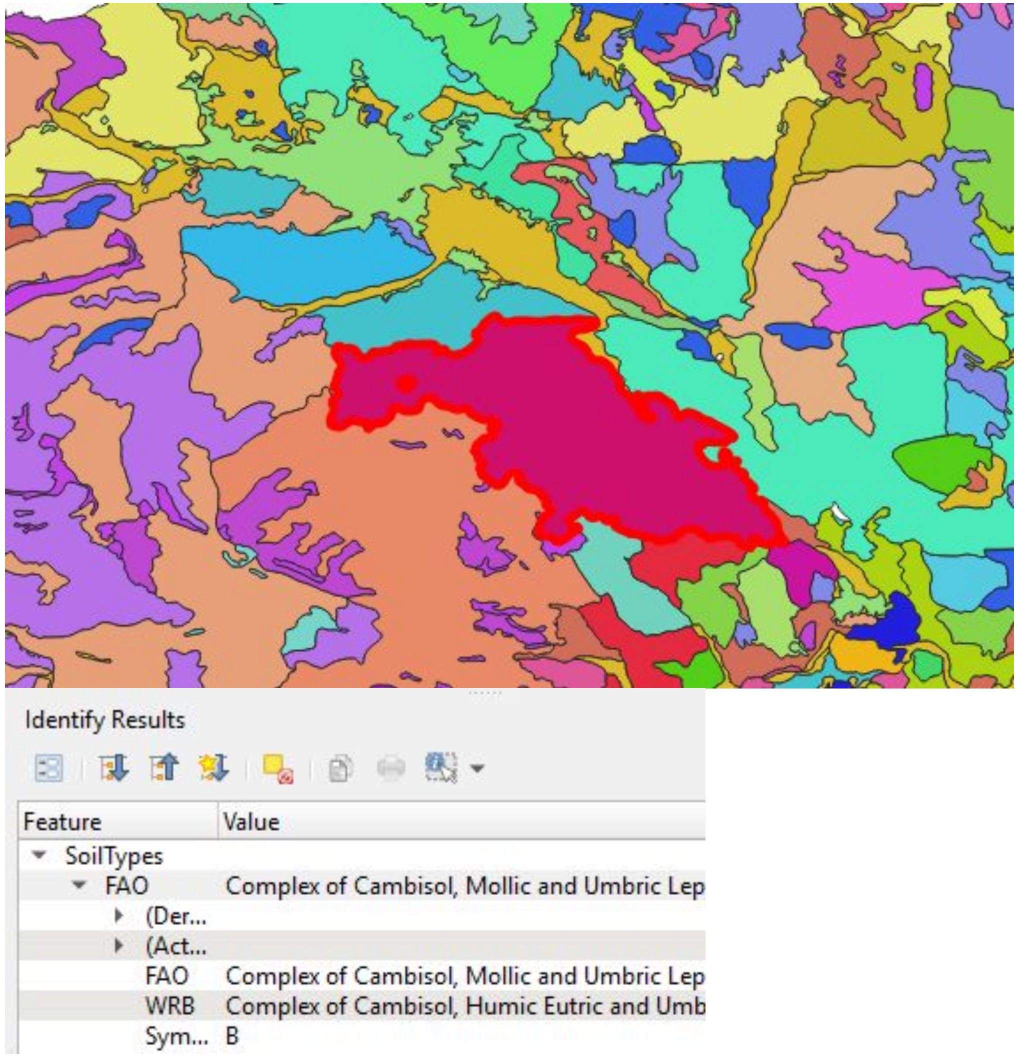
- Then select **Identify Features** tool



l platform GLOBAL SOIL PARTNERSHIP

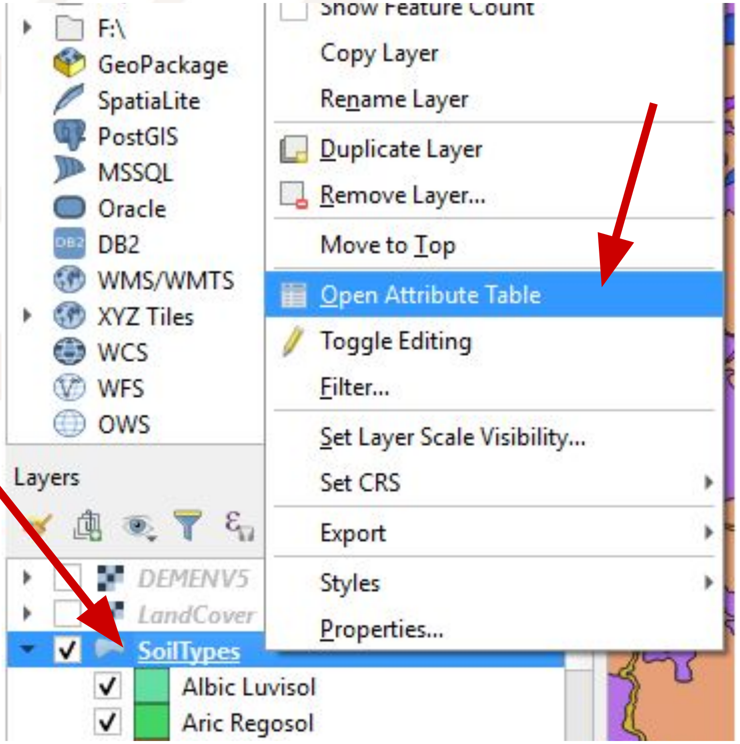
Vector data

- Now, click on the map to select any **soil polygon**;
- On the right, a new window will appear, representing **attributes** of the selected polygon.
- This is **Vector** data type.
- In vector data, every feature is a **discrete object** that has **attributes** (e.g. names soil types).



Attribute table

- Right click on the SoilTypes layer
- Open Attribute table



SoilTypes :: Features Total: 2516, Filtered: 2516, Selected: 0

	FAO	WRB	Symbol
1	Populated Site	Populated Site	NaM
2	Populated Site	Populated Site	NaM
3	Populated Site	Populated Site	NaM
4	Populated Site	Populated Site	NaM
5	Populated Site	Populated Site	NaM
6	Populated Site	Populated Site	NaM
7	Populated Site	Populated Site	NaM
8	Populated Site	Populated Site	NaM
9	Populated Site	Populated Site	NaM
10	Populated Site	Populated Site	NaM
11	Populated Site	Populated Site	NaM
12	Albic Luvisol	Albic Luvisol	La
13		Humic Calcaric...	Rz
14	Chromic Luviso...	Chromic Luviso...	Lc
15	Vertisol	Vertisol	V

Vector data

Key characteristics of **Vector data**:

- Discrete spatial objects: points, lines or polygons
- Attribute table attached to spatial objects

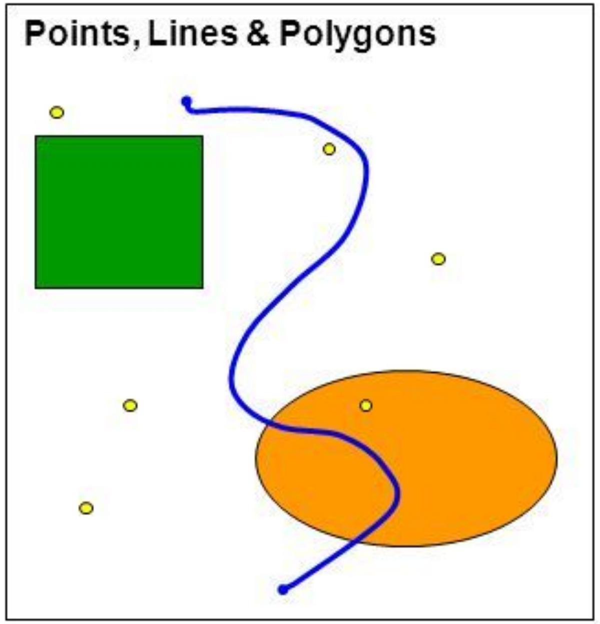
Common Vector data formats: shapefile (.shp)

Advantages of Vector data:

- Good for storing discrete data (e.g. points)
- High geographic accuracy of points and boundaries

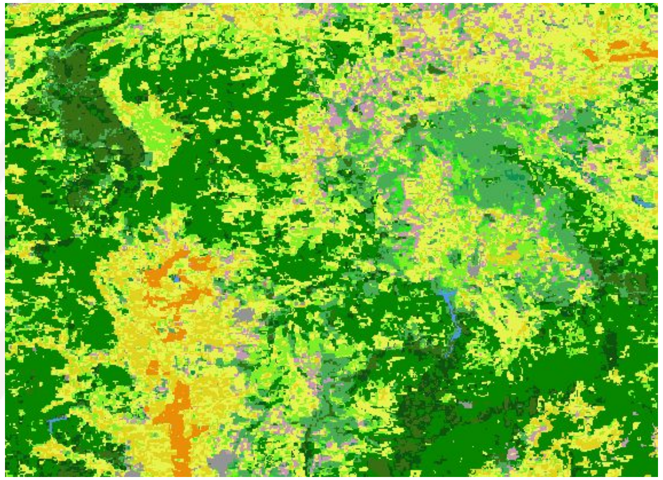
Disadvantages of Vector data:

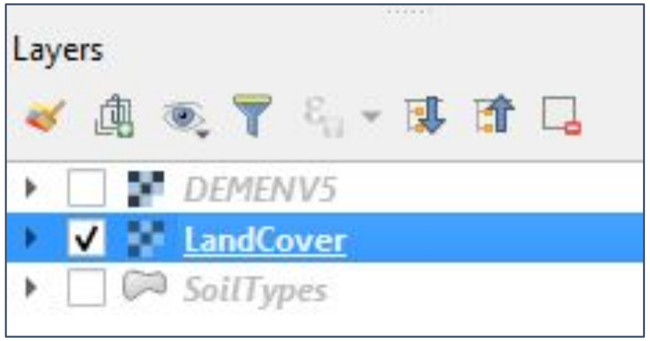
- Continuous data is poorly stored (needs classification)
- Hard to make calculations with multiple layers

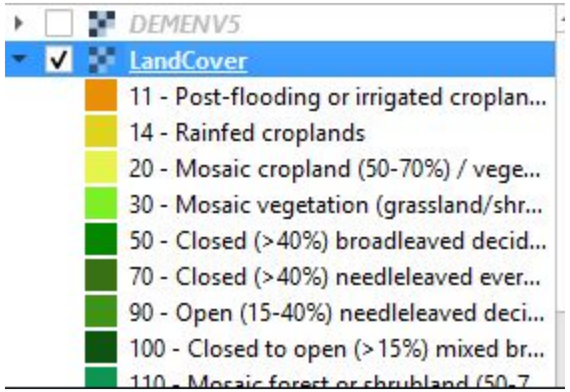


Discrete raster data (categorical)

- Now, deselect polygon by clicking on white space
- Switch off **SoilTypes** in the **Layers** window, and switch on **LandCover**
- Expand the legend by clicking on a black triangle
- Explore the Land cover map

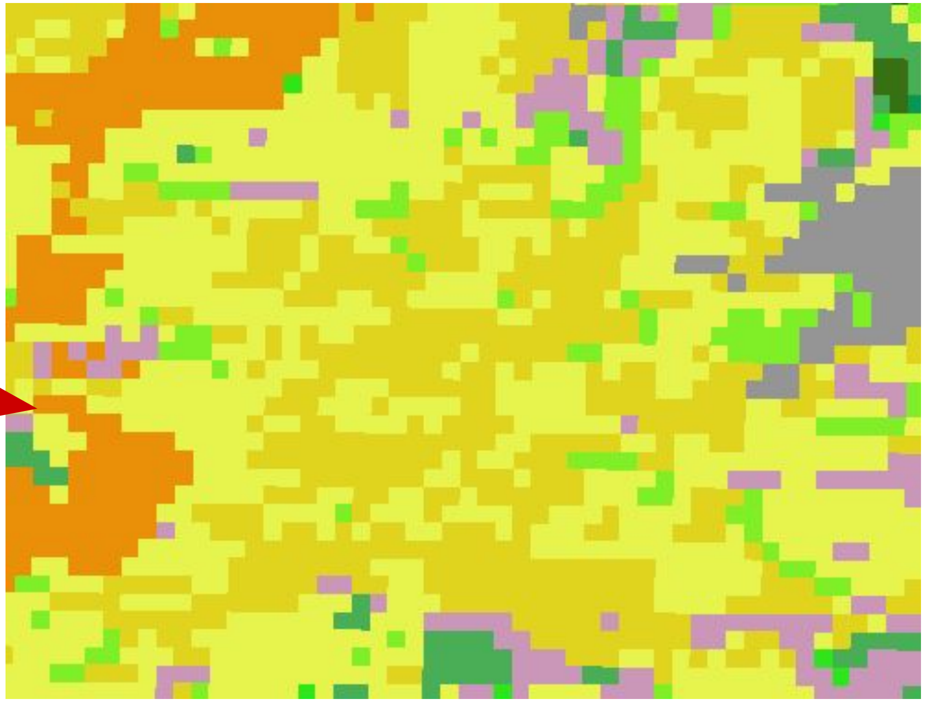


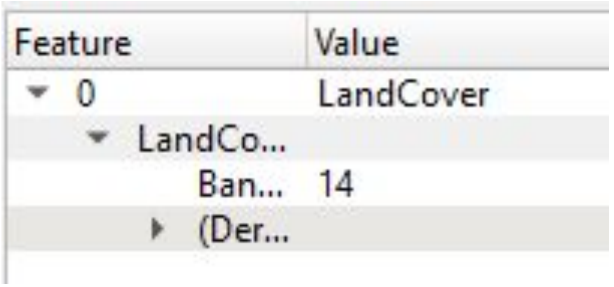




Discrete raster data (categorical)

- Zoom-in to see that the image consists of rectangular pixels (**raster cells**);
- Click on the image to see values;
- There is no attribute table only single **value in each raster cell**
- These values are **not real numbers**, but coded **categories**
- You can see the meaning of each coded category in the legend, e.g: *14 - Rainfed croplands*



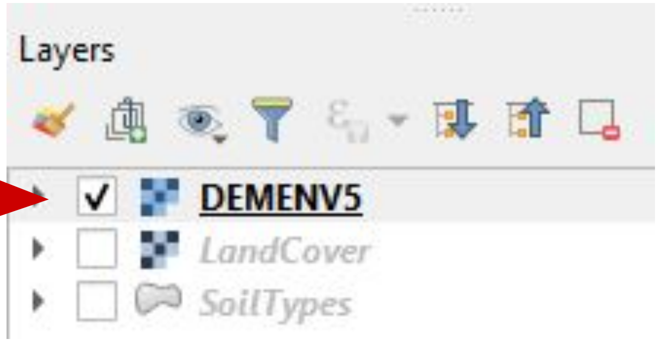


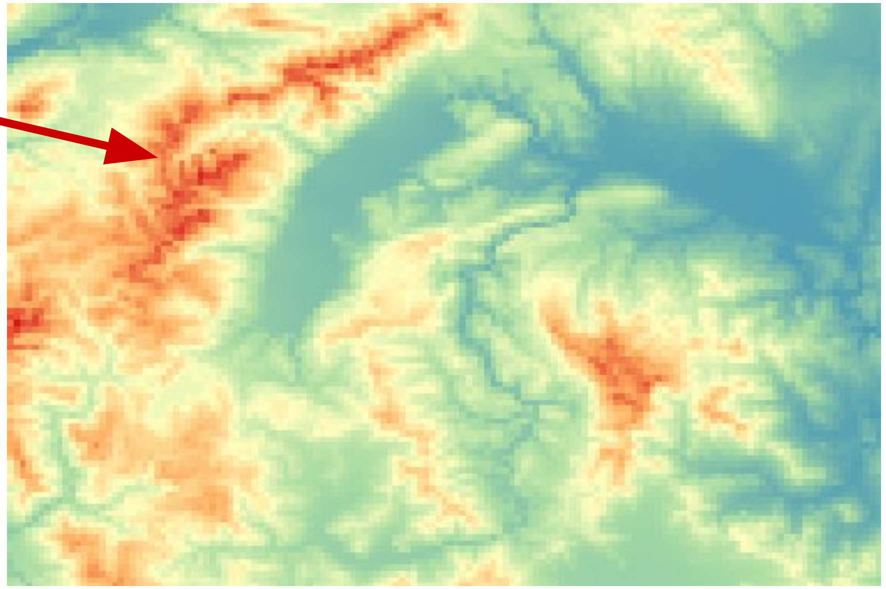
EduSoils | e-learning soil educational platform

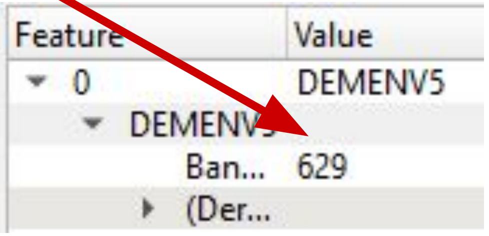
GLOBAL SOIL PARTNERSHIP

Continuous raster data

- Zoom-out, switch off **LandCover,** and switch on **DEMENV5** layer
- This is a **digital elevation model**
- Click on different parts of the map to see the **values**
- Value of each cell represents **real elevation (altitude)** in meters above sea level.





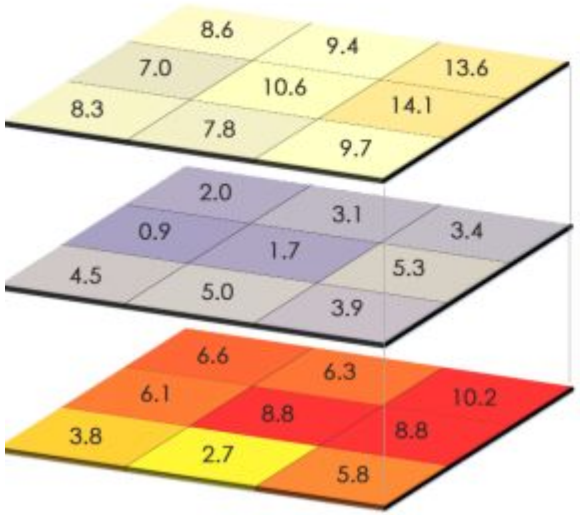


Raster data

Key characteristics of **Raster data**:

- A regular grid of cells (pixels) with a value in each cell
- Raster resolution defines cell size and accuracy
- Higher resolution = smaller cell size = better quality

Common Raster data formats: GeoTiff (.tif, .tiff);



Advantages of Raster data:

- Can store discrete data (**categories**) and continuous data (**values**)
- **Calculations** with multiple rasters is quick and easy to perform

Disadvantages of Raster data:

- Low resolution rasters are not accurate

Raster vs Vector

Questions:

- How can you represent **soil types**: with raster or with vector?
- Which way is **better**?
- How can you represent **soil properties** (e.g. Carbon, pH, texture): with raster or with vector?
- Which way is **better**?

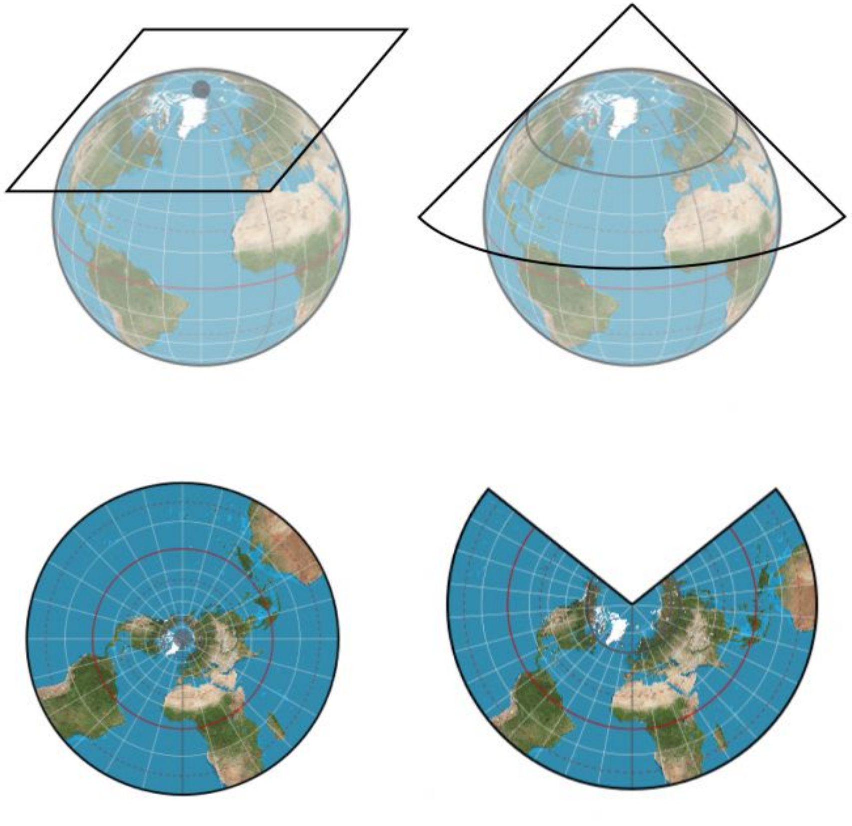
Raster vs Vector

Answers:

- **Soil types** can be represented both as a **Vector** (soil polygons) or as a **Discrete Raster**;
- **Vector** format is better for accurate representation of polygons;
- **Soil properties** can be represented both as a **Vector** (attributes of soil polygons) or a **Continuous Raster**;
- **Raster** format is better for representing variability of soil properties, calculations and modelling.

Map projections

- All spatial data are displayed using a certain **projection**



Coordinate Reference Systems (CRS)

Coordinate Reference System (**CRS**) is an **essential property** of spatial data:
it defines, how coordinates of our data are stored.

2 types of CRS:

Geographic:

Coordinates are stored as latitude and longitude - **in degrees**

- Good for **storing data** (independently of projection)
- Cannot be used for **spatial calculations** (e.g. distance, area)

e.g. **WGS 84** (EPSG:4326)

Projected:

Coordinates are stored according to a projection - **in meters**

- Can be used for **spatial calculations** (e.g. distance, area)
- Knowledge of the **exact CRS** is required to interpret the data

e.g. **UTM zone 48S** (EPSG:32748)

Information on **EPSG** codes: **http://www.spatialreference.org**

Soil observations

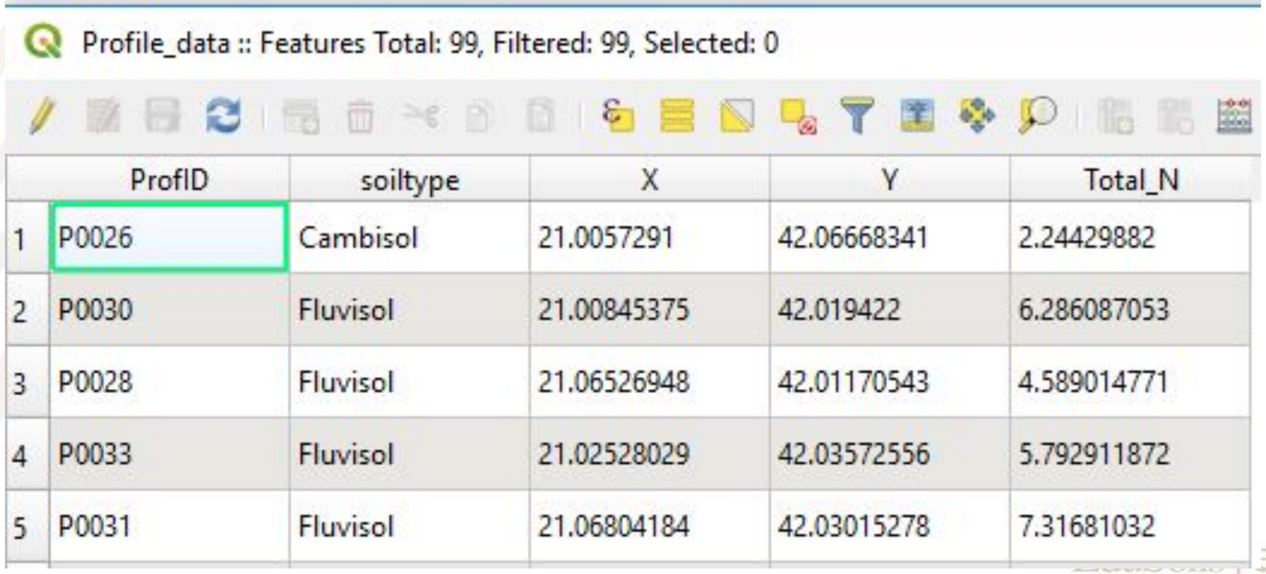
- What about data of **soil observations** (soil profiles, surface sampling)?
- Usually, data of soil observations are stored as a **table** (xlsx, csv)
- When a table has geographic **coordinates** (X, Y), it can be converted into **Vector** data type - **Points**.

Profile	Depth	Total	$Ca(NO_3)_2$	CH_3COONa
	cm			
1	0-10	426.3	1.8	154.7
	10-20	87.4	nd	19.4
	20-40	33.3	nd	12.0
2	0-10	3,988.8	63.5	2,415.5
	10-20	2,995.4	9.7	1,957.9
	20-40	3,123.2	7.7	1,968.8
3	0-10	15,293.1	717.0	13,000.8
	10-20	17,913.2	834.5	21,041.7
	20-40	19,434.0	856.6	19,359.9
4	0-10	3,892.7	3.6	1,976.3
	10-20	7,227.9	3.1	2,208.3
	20-40	5,021.2	3.4	2,275.0
5	0-10	14,530.1	3.0	8,514.8
	10-20	5,806.8	3.1	4,102.7
	20-40	10,140.1	3.7	8,937.4
6	0-10	695.9	175.0	283.5
	10-20	240.9	20.0	226.8
	20-40	105.4	2.8	128.7
7	0-10	2,570.6	14.7	1,919.8
	10-20	1,266.7	3.7	685.2
	20-40	1,270.6	1.5	544.8

Adding a table to the map

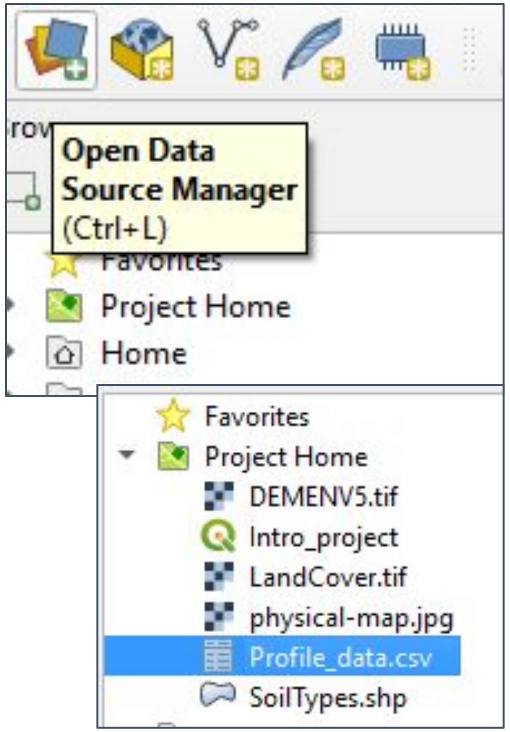
- Click on **Open Data** tool in the to left corner
- Double-click on **Profile_data**.csv
- Table added to **Layers**. You may close the window
- Right click on the **Profile_data**

Open Attribute table



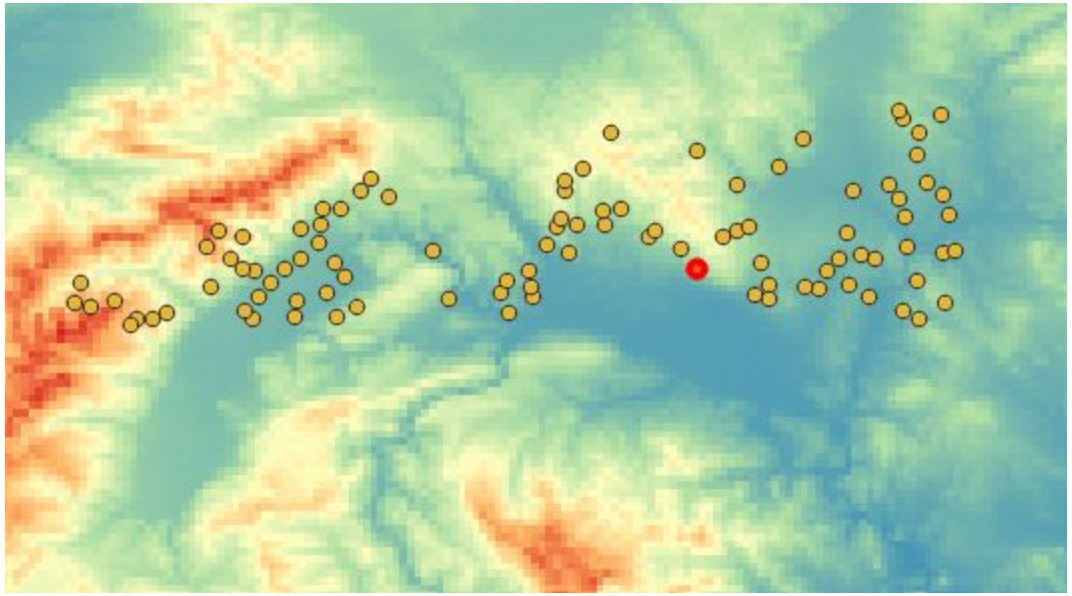
Profile_data :: Features Total: 99, Filtered: 99, Selected: 0

	ProfID	soiltype	X	Y	Total_N
1	P0026	Cambisol	21.0057291	42.06668341	2.24429882
2	P0030	Fluvisol	21.00845375	42.019422	6.286087053
3	P0028	Fluvisol	21.06526948	42.01170543	4.589014771
4	P0033	Fluvisol	21.02528029	42.03572556	5.792911872
5	P0031	Fluvisol	21.06804184	42.03015278	7.31681032



Point data (soil observations)

- You can explore a new **Vector** data layer with point observations;
- You can select points and see the **attributes**



Feature	Value
Points from...	
Title	P0087
(Der...	
(Act...	
ProfID	P0087
soilt...	Complex of Cambisol, Mollic and Umbric Lep
X	21.52601467
Y	42.06665761
Tota...	7.166966797

Soil properties data

We have:

- **Discrete** soil observations as point data in **sampling locations**

We need:

- **Continuous** estimation of soil properties in **every point of the land surface**

Task of the Digital Mapping of Soil Properties:

- To **predict continuous soil properties** in every point of the land surface based on **discrete measurements** of soil sampling. ***But how to do it?***

Drivers of soil formation

V. Dokuchaev (1883):

Genesis and evolution *of soils is the result of the interaction of a number of* ***environmental parameters****:*

- Climate
- Organisms
- Parent Material
- Relief
- Time

Drivers of soil formation

H. JENNY (1941):

- Conceptualization of soil as an <u>state equation of soil formation</u>.
- Soil and soil properties are a function of a number of environmental parameters named soil forming factors:

$$S = f(cl, o, r, p, t)$$

- climate
- organisms
- relief
- parent material
- time

CLORPT MODEL

Digital Soil Mapping (DSM)

Definition of Digital Soil Mapping (DSM)

The creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observations and knowledge and from related environmental variables (Lagacherie and McBratney, 2007).

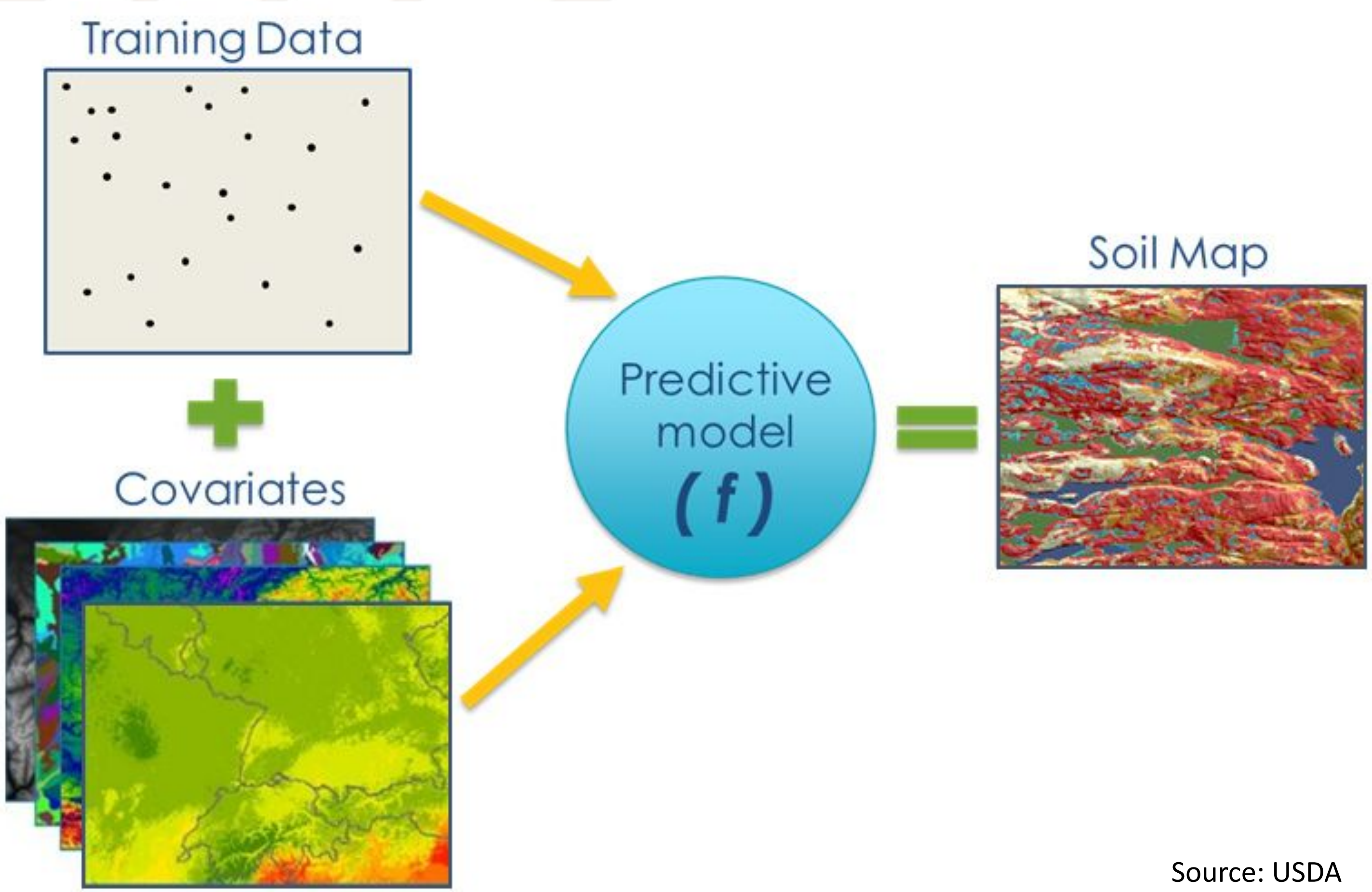
McBRATNEY, 2012: Conceptualization of forming factors. Soil and soil properties are a function of a number of environmental parameters named soil forming factors:

$$S = f(s,\ c,\ o,\ r,\ p,\ a,\ n) + \varepsilon$$

SCORPAN MODEL

- S: soil attribute to predict
- s: soil properties
- c: climate
- o: organisms
- r: relief
- p: parent material
- a: age
- n: location
- ε: function residuals (errors)

Digital Soil Mapping Workflow



Source: USDA

Digital Soil Mapping Workflow

Steps to Map Soil Properties with DSM approach:

1) Prepare **soil observations** (soil profile data);
2) Prepare environmental **covariates** (raster layers);
3) Fit the **prediction model** using a target soil property (e.g. SOC) as a dependent variable and the covariates as predictors;
   - Linear modelling;
   - Geostatistical modelling;
   - Machine learning;
4) **Map** the soil property using the prediction model;
5) **Assess** the quality and uncertainty of the map.

Rural Development Administration

AFACI
Asian Food & Agriculture Cooperation Initiative

Food and Agriculture Organization of the United Nations

EduSoils

GLOBAL SOIL PARTNERSHIP