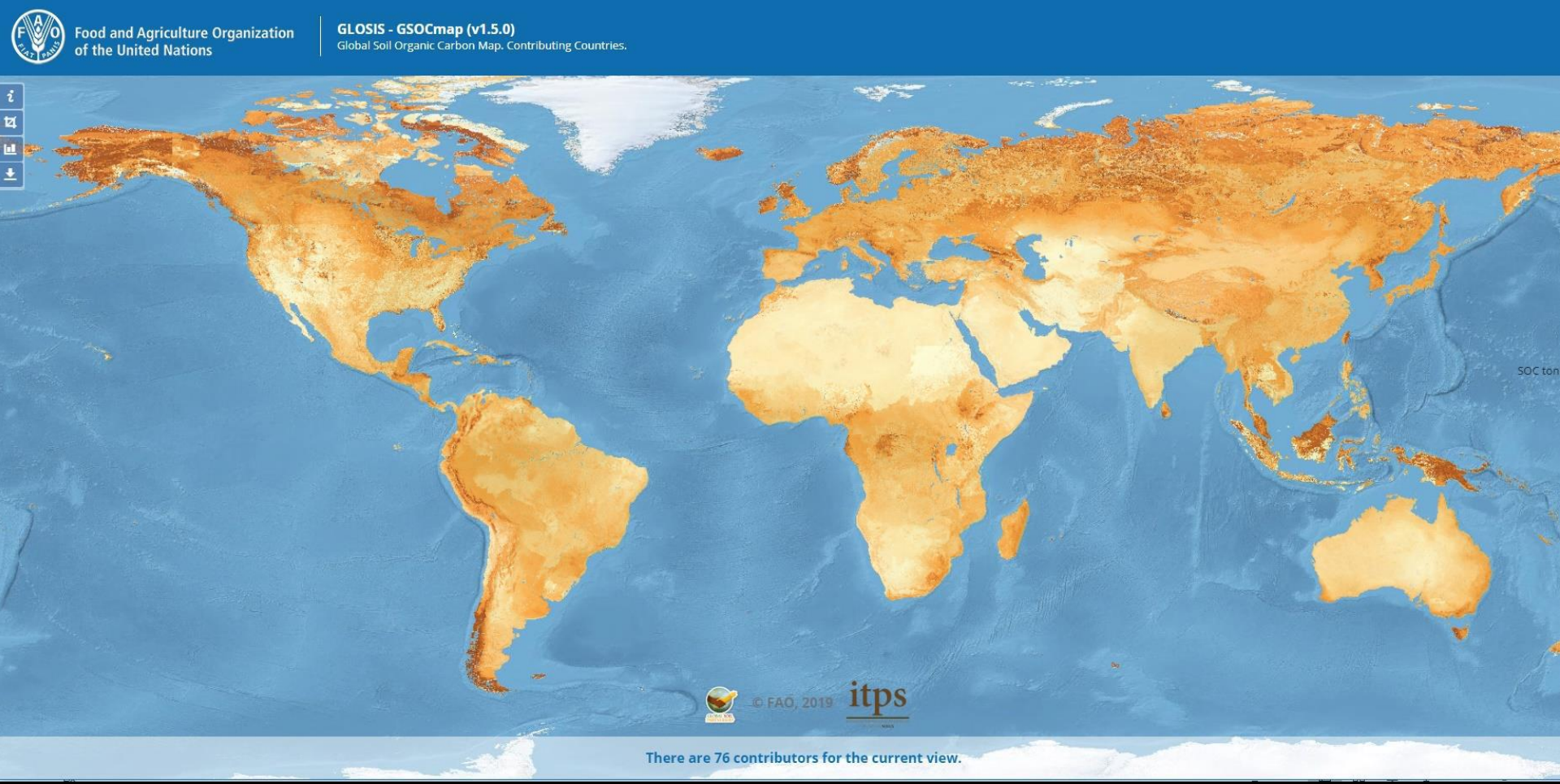


Preparing Environmental Covariates for Digital Soil Mapping

Ruhollah Taghizadeh
Soil Science and Geomorphology

E-mail: *ruhollah.taghizadeh-mehrjardi@mnf.uni-tuebingen.de*

Content

- **Part 1**
  - **Digital Soil Mapping**
  - **Environmental Covariates**
- **Part 2**
  - **Lets practice!**

SCORPAN model

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

S : Soil, at a specific point in space and time: soil classes, Sc or soil attributes, Sa

From Jenny's Equation

c : climate, climate properties of the environment;
o : organisms, vegetation;
r : topography, landscape attributes;
p : parent material, lithology;
a : age or time factor;

Additions:

s : soil, prior knowledge of the soil at a point;
n : space, relative spatial position;
ε : auto-correlated random spatial variation.

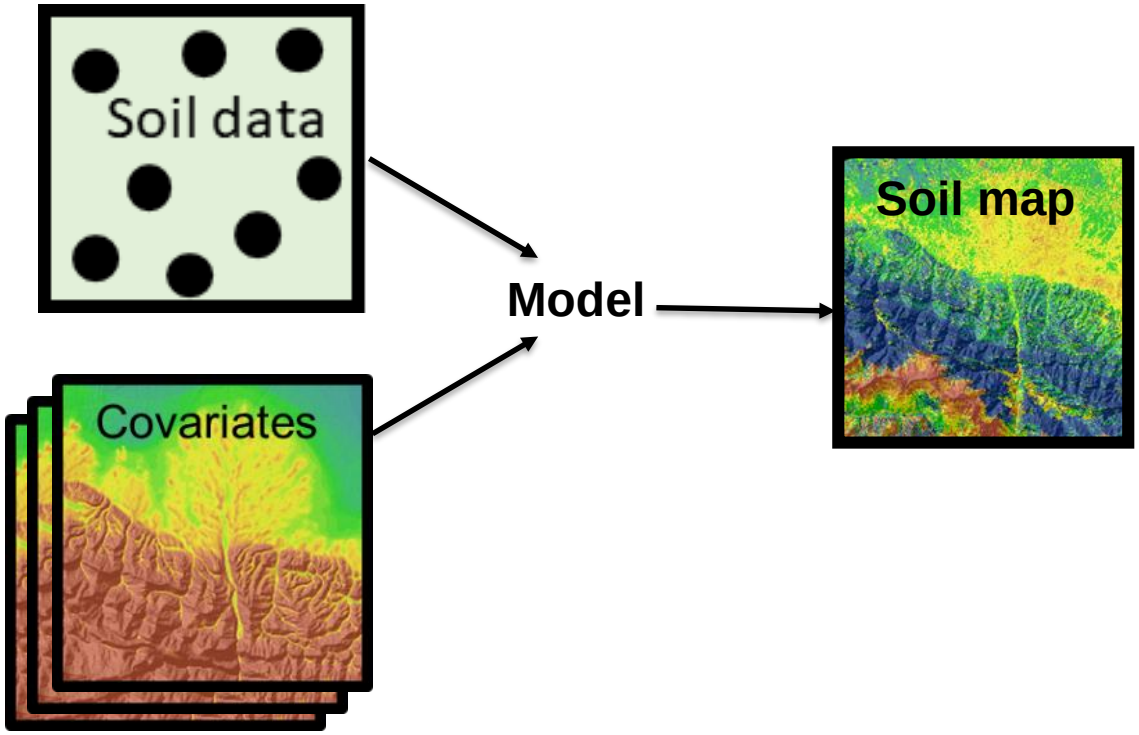
f() : Quantitative function *f* linking S to ***scorpan*** factors

McBratney et al., 2003

SCORPAN model

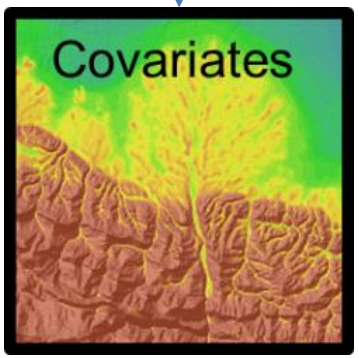
$$S = f(s, c, o, r, p, a, n) + \varepsilon$$

- S: Soil data
- f: Model
- (s, c, o, r, p, a, n): Covariates
- ε: spatially dependent residuals



SCORPAN Model

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



Environmental Covariates

s
c
o
r
p
a
n

Possible sources of information to represent seven scorpan factors

scorpan factor	Possible representatives
s	Legacy soil maps, point observations, expert knowledge
c	Temperature and precipitation records
o	Vegetation maps, species abundance maps, yield maps, land use maps
r	Digital elevation model, terrain attributes
p	Legacy geology maps, gamma radiometric information
a	Weathering indices, geology maps
n	Latitude and longitude or easting and northing, distance from landscape features, distance from roads, distance from point sources of pollution

- Proximal sensing, Remote sensing, and digital elevation models are three prominent examples of high-resolution environmental covariates can be used to represent various scorpan factors

McBratney et al., 2003

Proximal Soil Sensing (PSS)

s c o r p a n

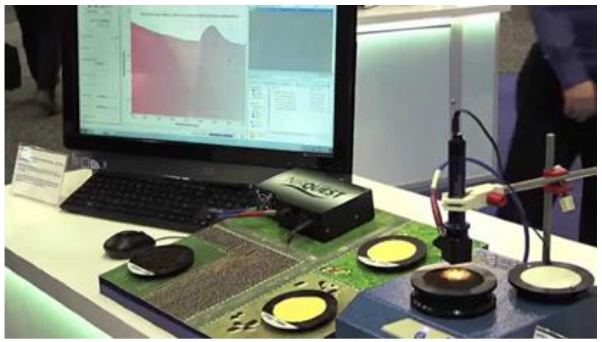
- Proximal soil sensing is a set of techniques developed to measure the physical, chemical and biological properties of soil when placing the sensor in contact, or at a proximal distance (within 2 m) to, the soil being characterized (Viscarra Rossel et al., 2011)

- The proximal soil sensors are more time- and cost effective than conventional laboratory analyses

- The proximal soil sensors facilitate the collection of larger amounts of (spatial) data

EM38

PXRF

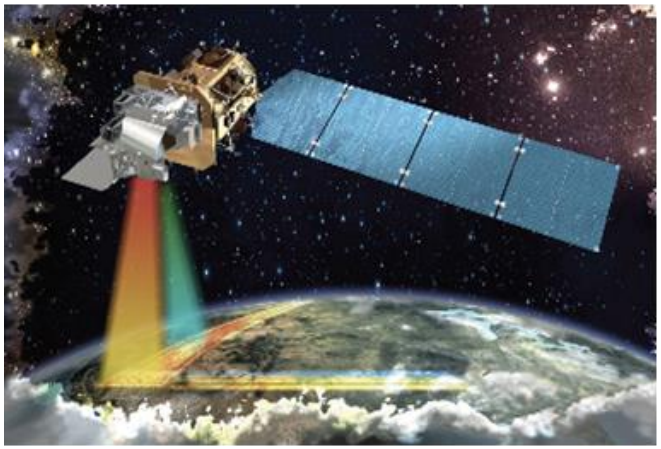


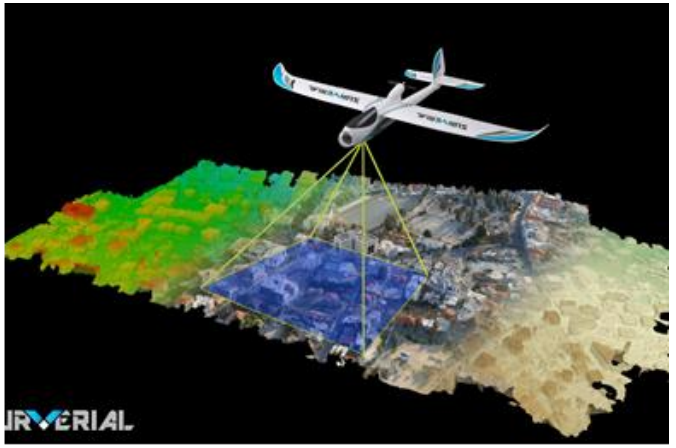
NIR

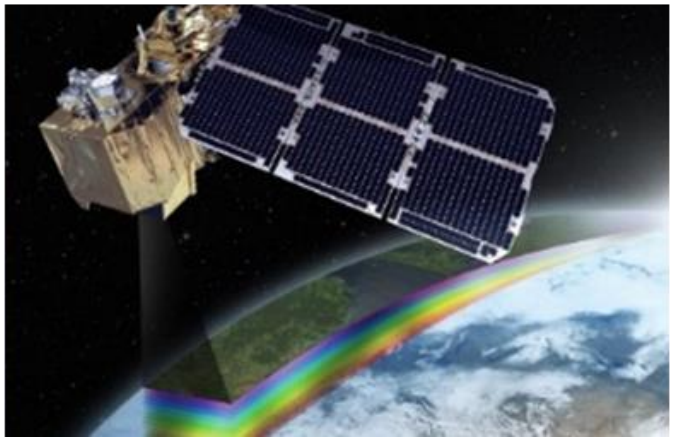
Remote Sensing

s c o r p a n

- The remotely sensed data has been shown to be an efficient means of assessing the condition of natural resources

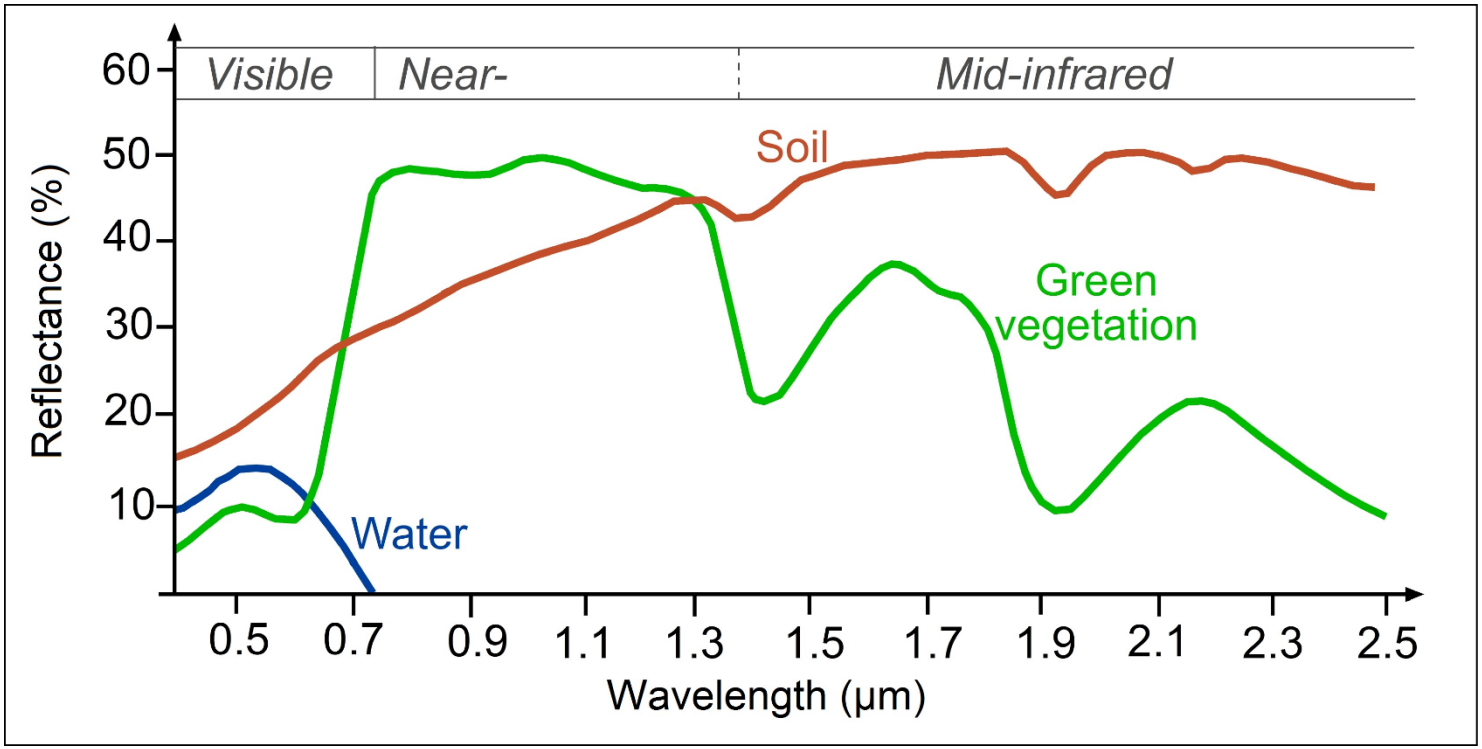






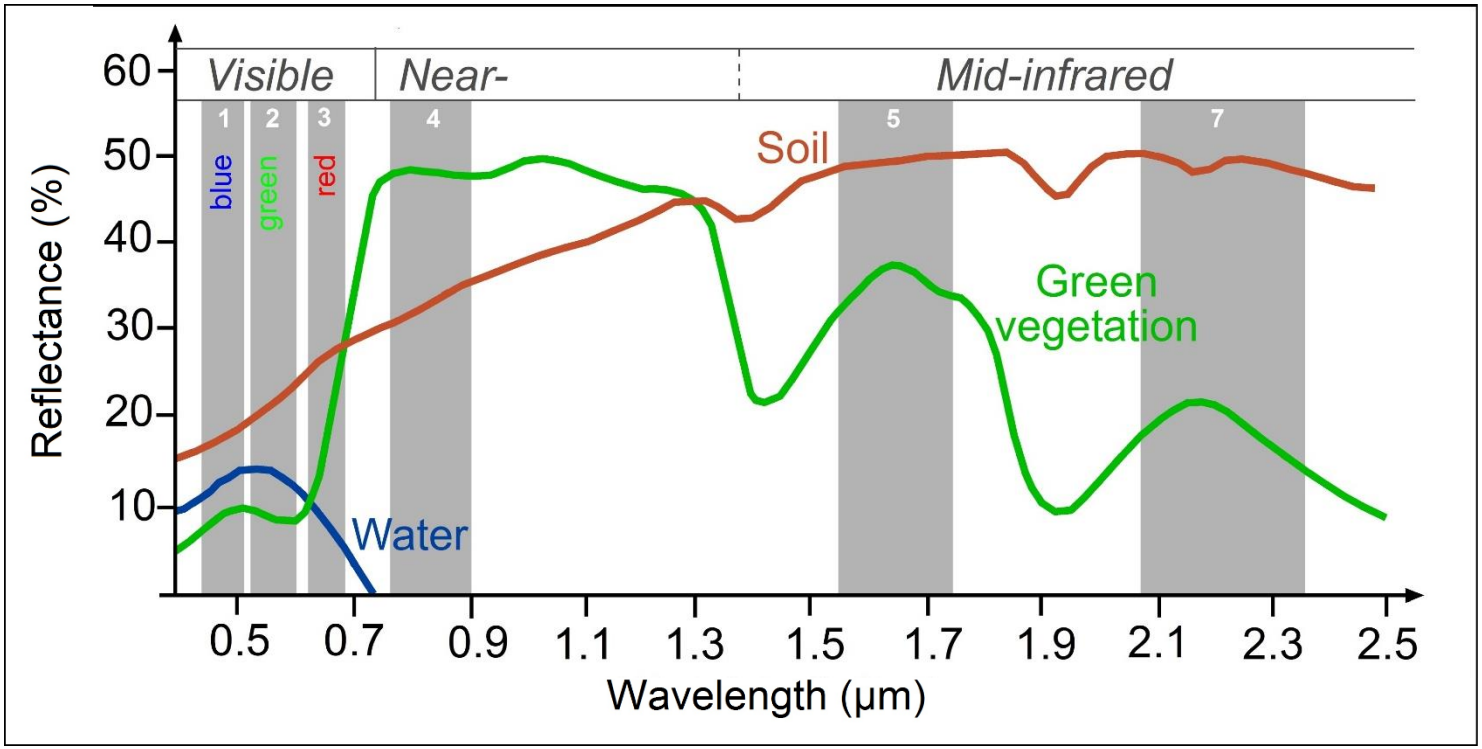
Remote Sensing

s
c
o
r
p
a
n



Spectral signatures of soil, vegetation and water.
Source: Siegmund, Menz 2005 with modifications

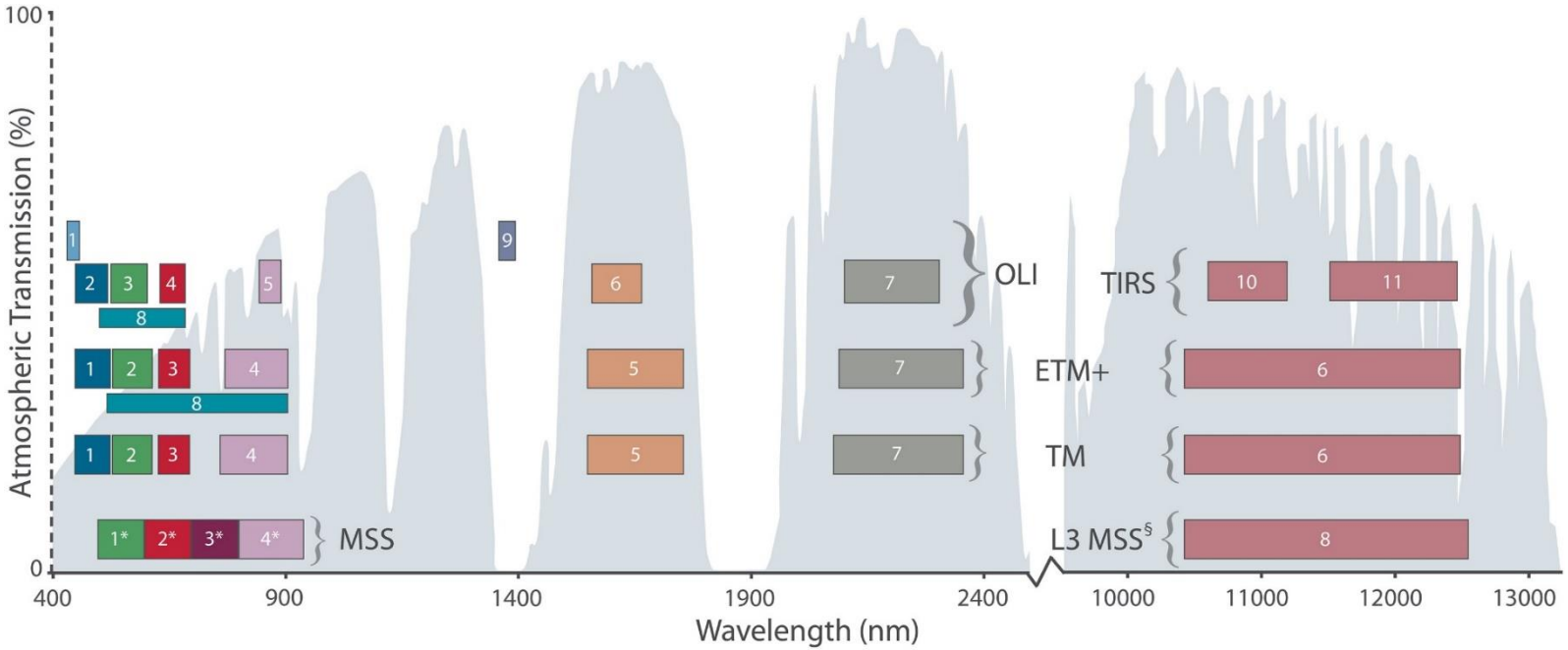
Remote Sensing



Spectral signatures of soil, vegetation and water, and spectral bands of LANDSAT 7.
Source: Siegmund, Menz 2005 with modifications

Remote Sensing: Landsat

s c o r p a n



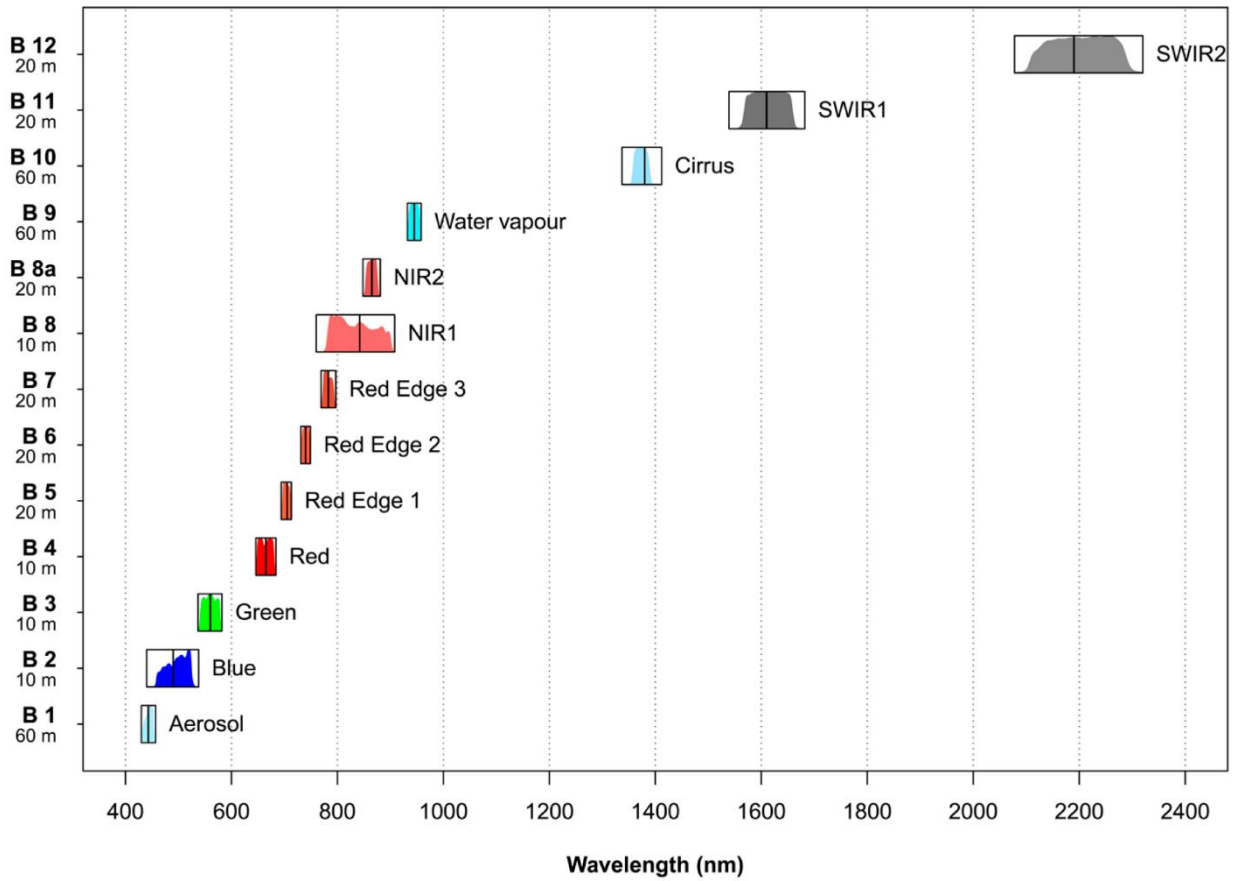
* MSS bands 1–4 were known as bands 4–7, respectively, on Landsats 1–3

§ The 240 m thermal band on Landsat 3 was out of spec within three weeks of launch and turned off in March 1979

Landsat sensors, spectral channels and band-passes, superimposed on atmospheric transmission percentage (grey background). MSS: Landsat-1 through -5; TM: Landsat-4 and -5; ETM+: Landsat-7; OLI and TIRS: Landsat-8. (Source: NASA/Landsat Legacy Project Team and American Society for Photogrammetry and Remote Sensing). Wulder et al., 2018

Remote Sensing: Sentinel 2

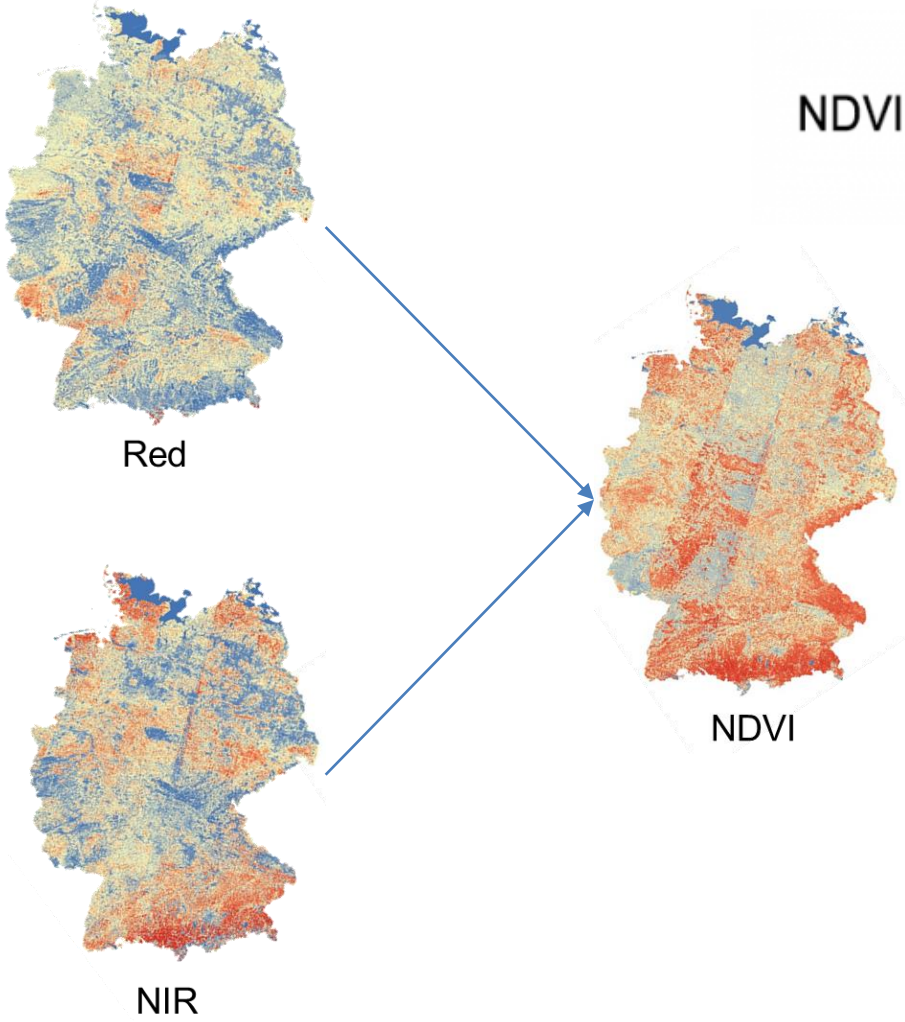
s
c
o
r
p
a
n



Characteristics of the Multi Spectral Instrument (MSI) on board Sentinel-2. The spectral response functions are shown in colors with the central wavelength in black. The band names and the corresponding spatial resolutions (in meters) are also indicated. Immitzer et al., 2016

Remote Sensing: band ratio

s c o r p a n

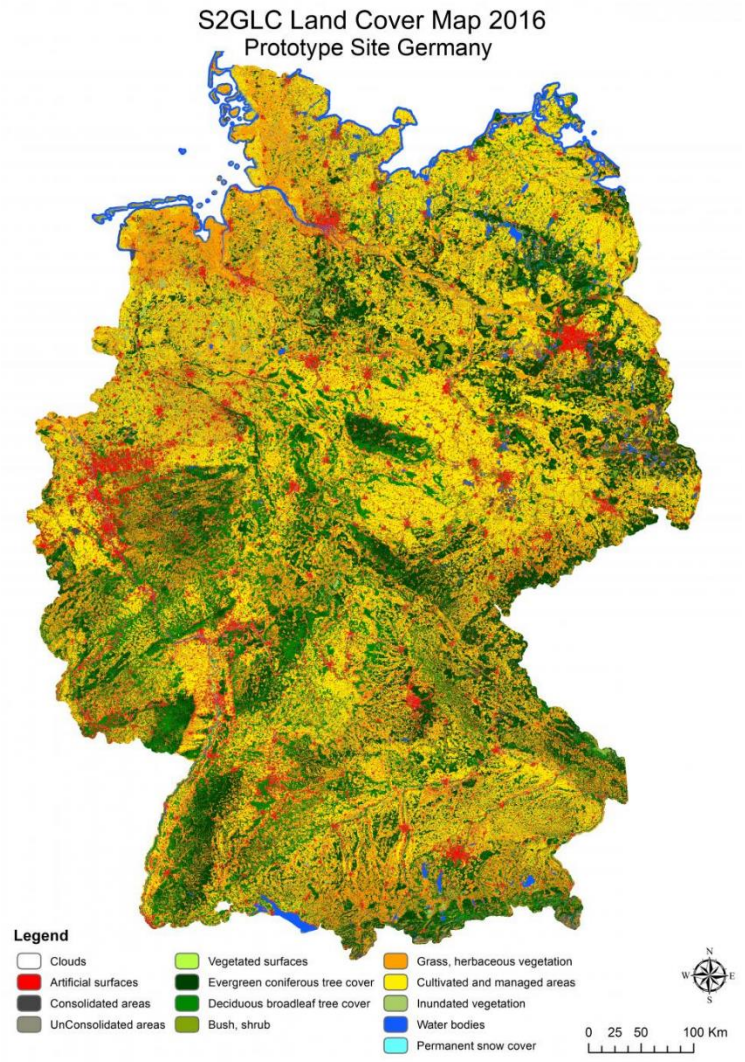


$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

Red and NIR bands of Landsat-8 for Germany

Remote Sensing: Land use/cover

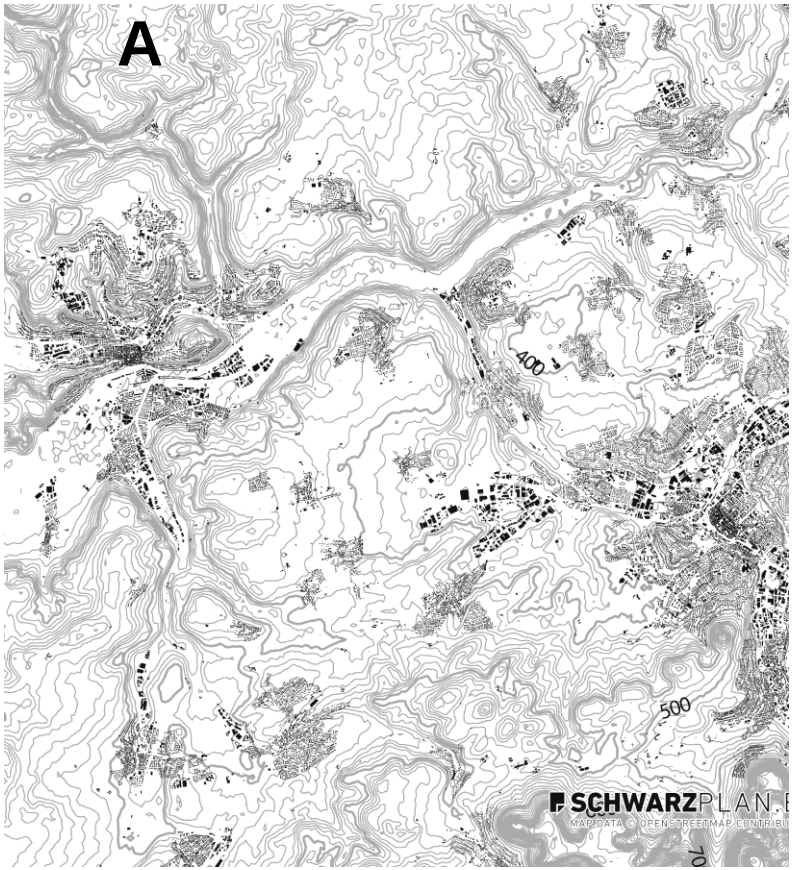
s
c
o
r
p
a
n

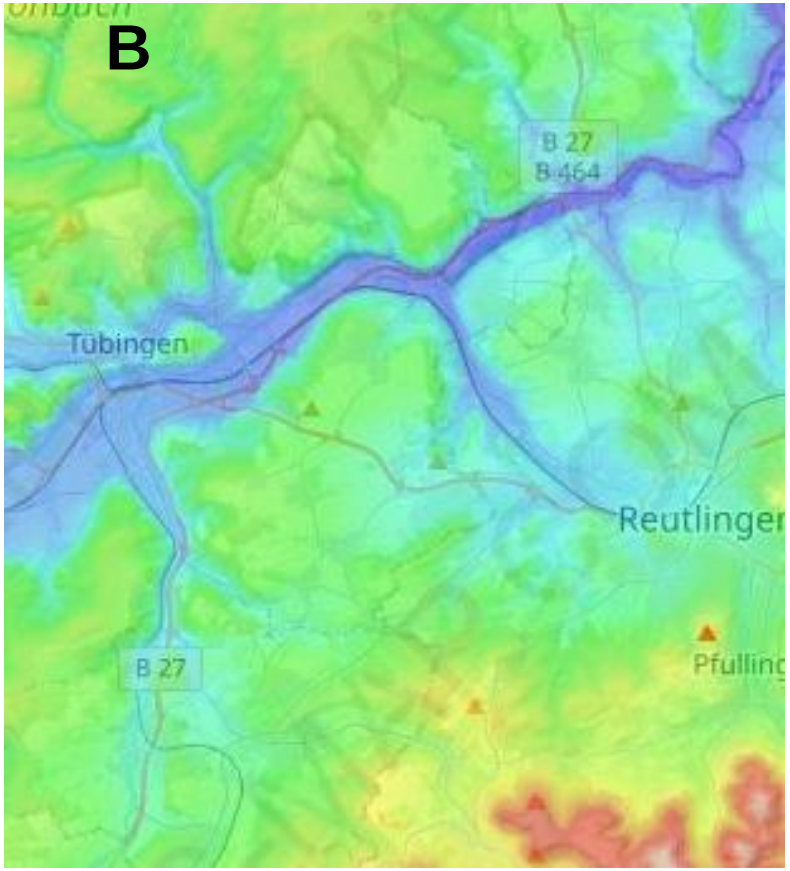


Digital Elevation Model (DEM)

s c o r p a n

- Digital elevation models: is a 3D computer graphics representation of elevation data to represent terrain.

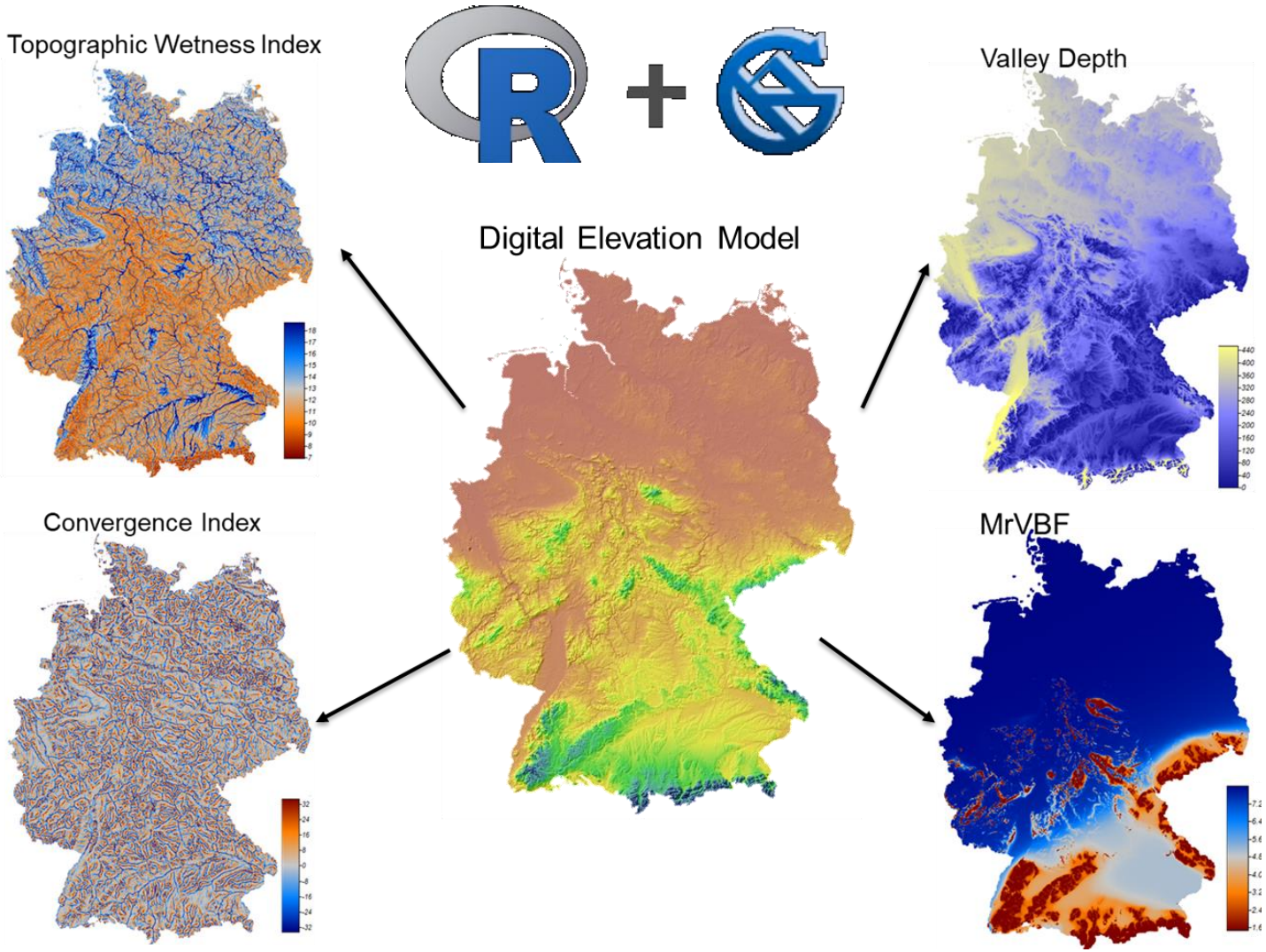




Elevation maps of Tuebinegn, https://en-gb.topographic-map.com/

DEM: Terrain Attributes

s c o r p a n

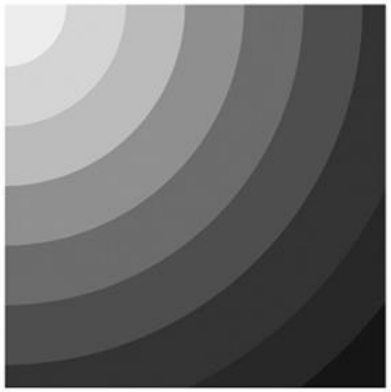


Basic Terrain Derivatives of Germany

Distance layers

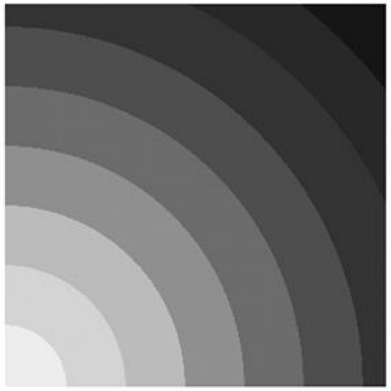
s c o r p a n

Upper left (C1)

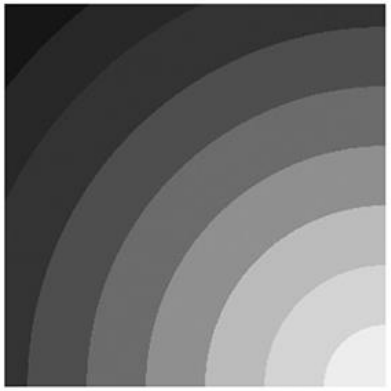


Upper right (C2)

Lower left (C3)



Lower right (C4)



Centre (CC)

European Journal of Soil Science

European Journal of Soil Science, 2018 doi: 10.1111/ejss.12687

Spatial modelling with Euclidean distance fields and machine learning

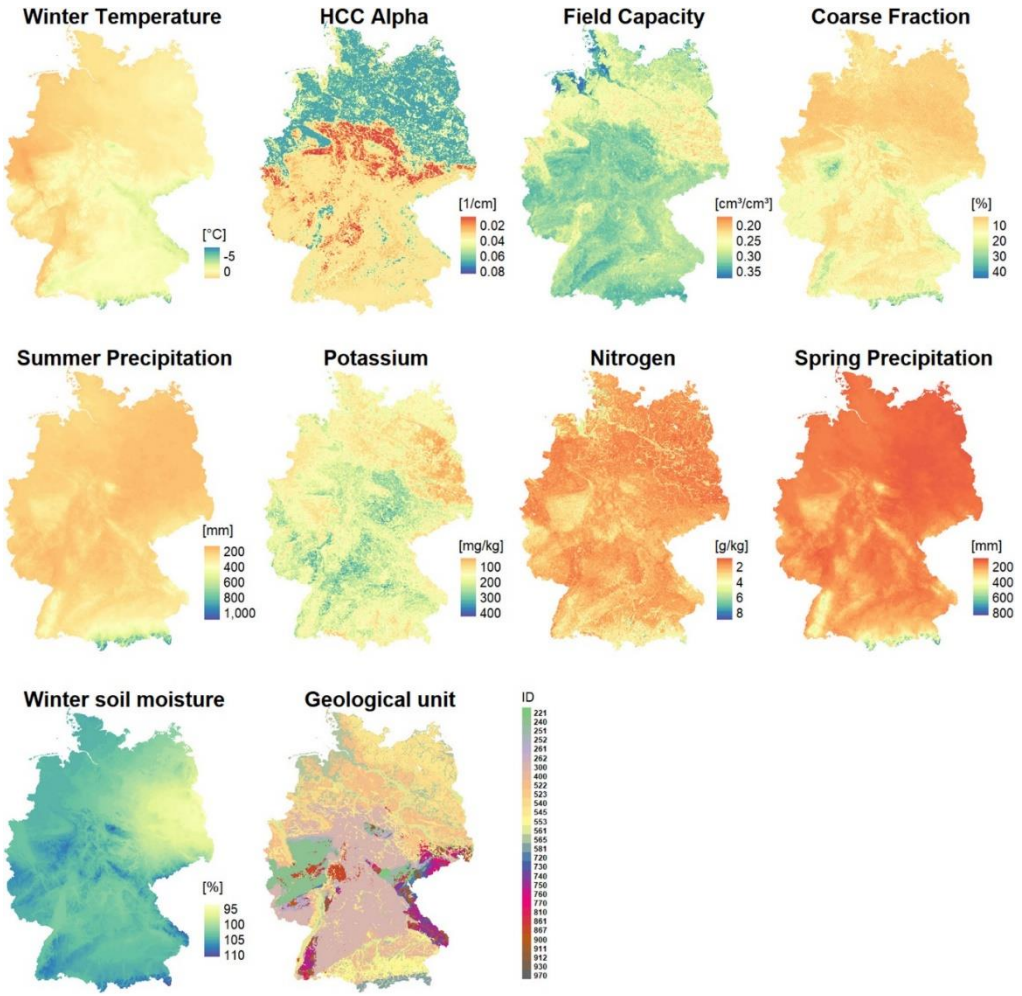
T. Behrens[a], K. Schmidt[a], R. A. Viscarra Rossel[b], P. Gries[a], T. Scholten[a] & R. A. MacMillan[c]

[a]*Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, Rümelinstraße 19-23, 72074, Tübingen, Germany,* [b]*CSIRO Land and Water, GPO Box 1700, Canberra ACT 2601, Australia, and* [c]*LandMapper Environmental Solutions Inc., 7415 118 A Street NW, Edmonton, Alberta Canada*

Visualization of the Euclidean distance fields in machine learning (EDM) predictors.

Other Covariates

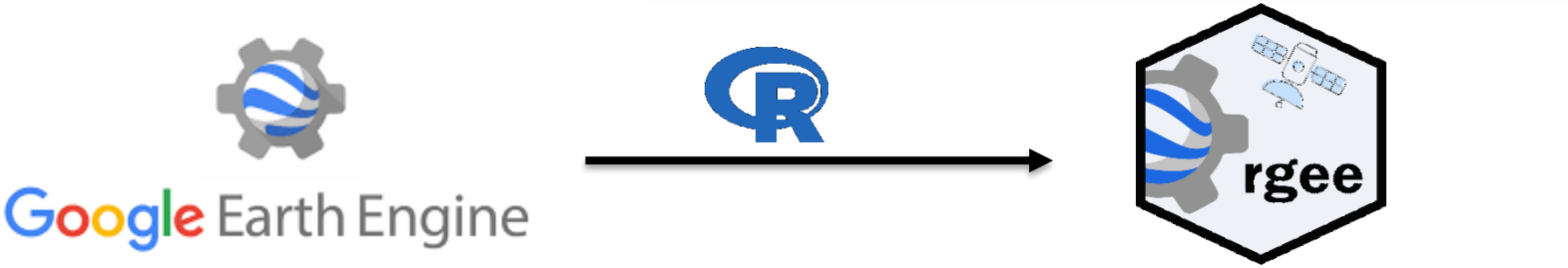
s c o r p a n

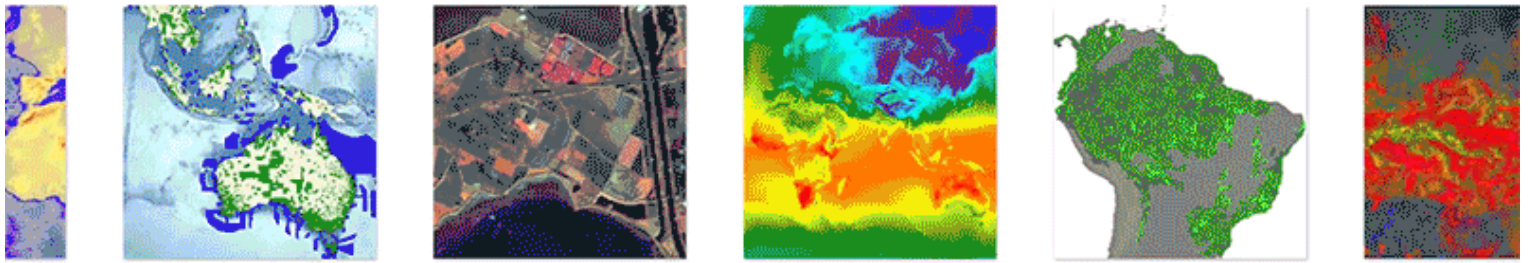


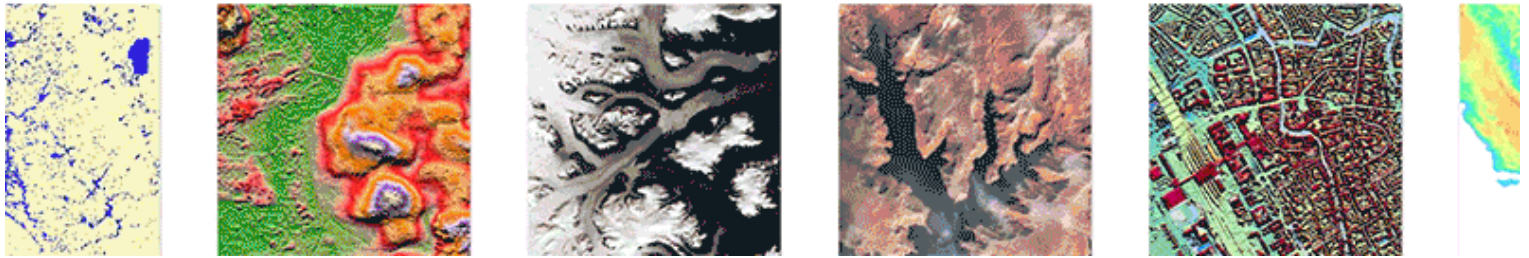
Other Geospatial Data of Germany (Petermann et al., 2020)

Where Can I Obtain the Covariates?

s c o r p a n







Choosing Covariates

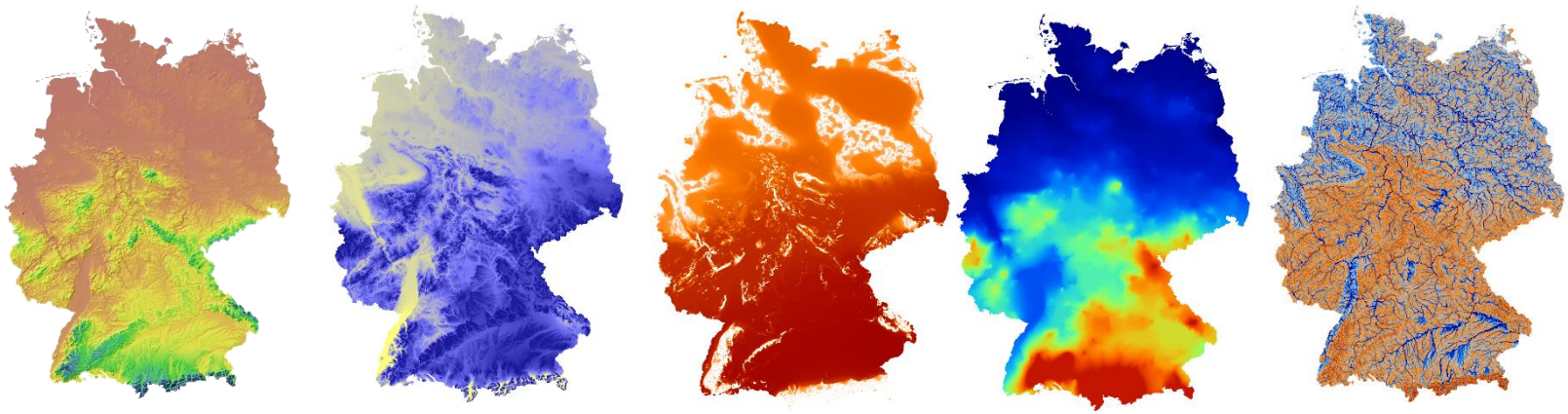
s
c
o
r
p
a
n

- What factors influence soil formation?

  --Link to pedological knowledge

- What data do I have?

- Data should represent the multiple SORPAN covariates that influence soil development in the area

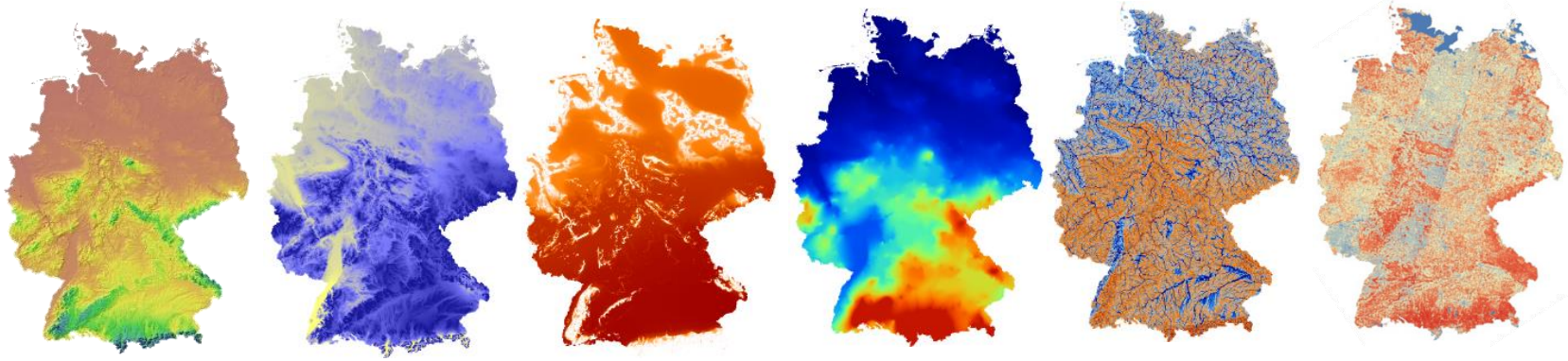


Technical and Practical Notes

s c o r p a n

Covariate data sources

- SRTM digital elevation model (30 m resolution)
- WorldDEM digital elevation model (12 m resolution)
- Landsat-8 satellite images (30 m resolution)
- Sentinel-2 satellite images (10, 20, and 60 m resolution)
- MODIS satellite images (250, 500, and 1000 m resolution)
- Global Land Cover maps (30 m resolution)
- JAXA's ALOS radar images (20 m resolution)
- Monthly precipitation images (1000 m resolution)
- Geology maps (polygon)



Technical and Practical Notes

s
c
o
r
p
a
n

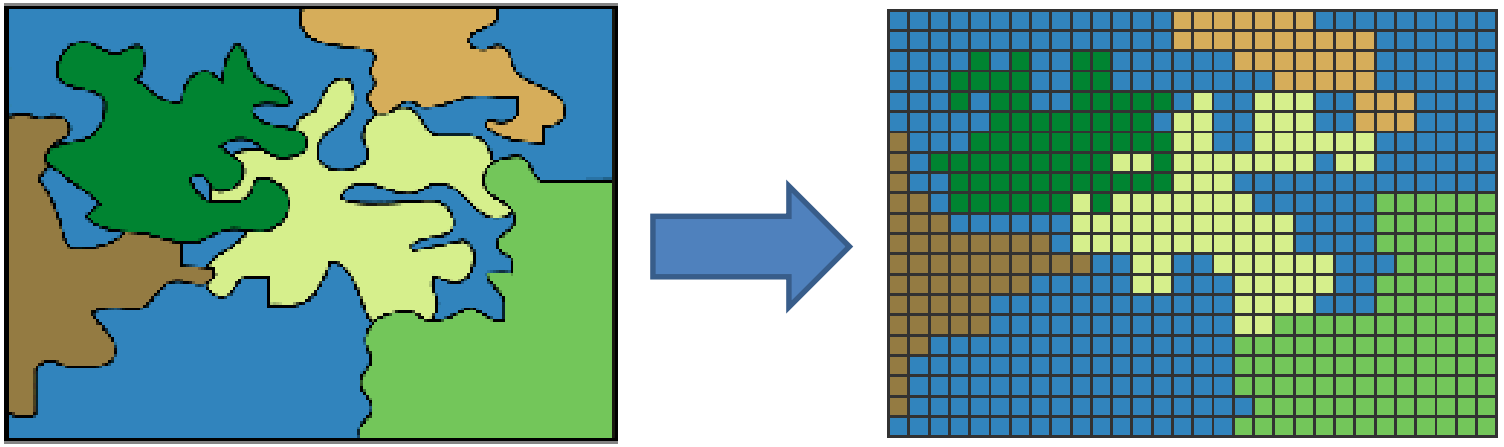
Preparing covariate layers

1. Converting polygon maps to rasters
2. Downscaling or upscaling rasters to a common resolution,
3. Filtering out missing pixels/reducing noise and multicolinearity problems,
4. Overlaying raster stacks and points

Technical and Practical Notes

s
c
o
r
p
a
n

1. Converting polygon maps to rasters



Polygon features

Raster polygon features

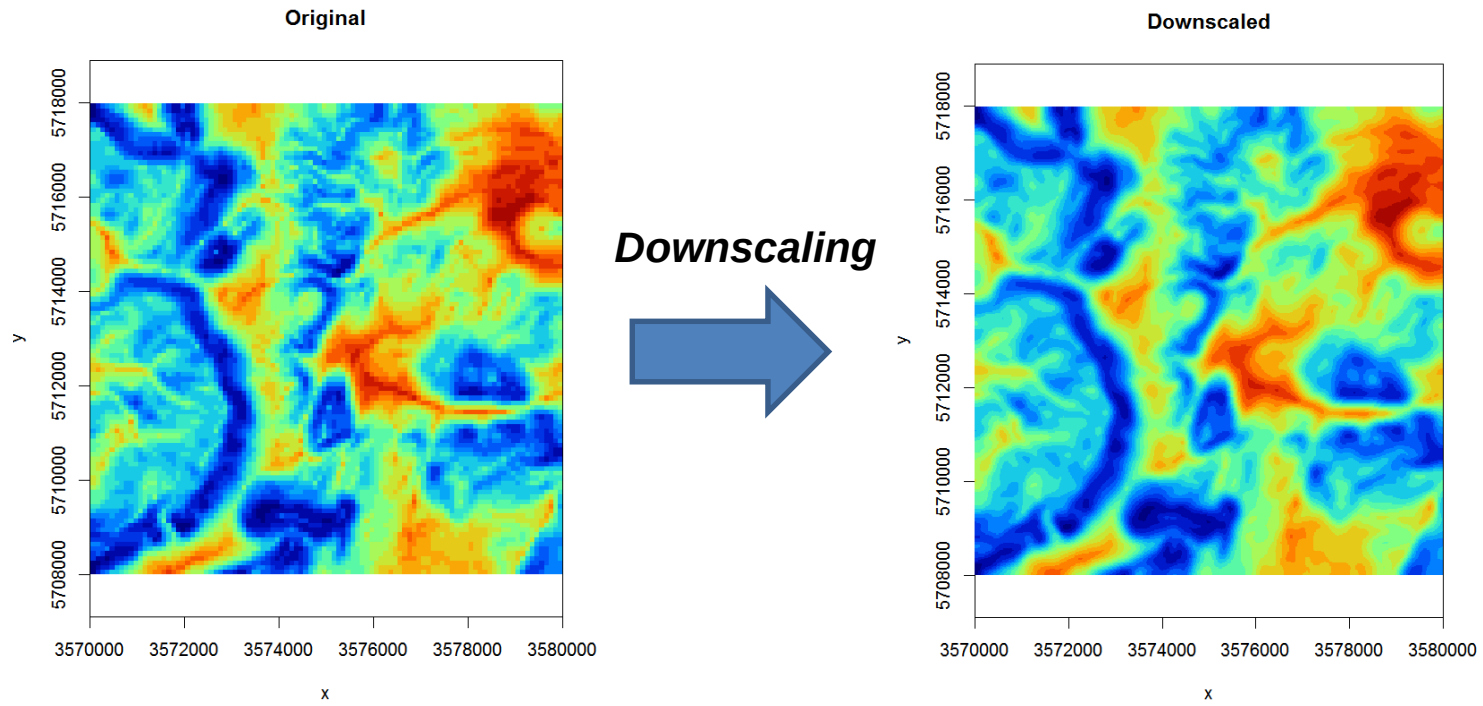
soilmapper.org
desktop.arcgis.com

Technical and Practical Notes

s c o r p a n

2. Downscaling or upscaling rasters

- To adjust the resolution of some covariates that have either too coarse or too fine a resolution compared to the target resolution
- The process of bringing raster layers to a common grid resolution is also known as resampling

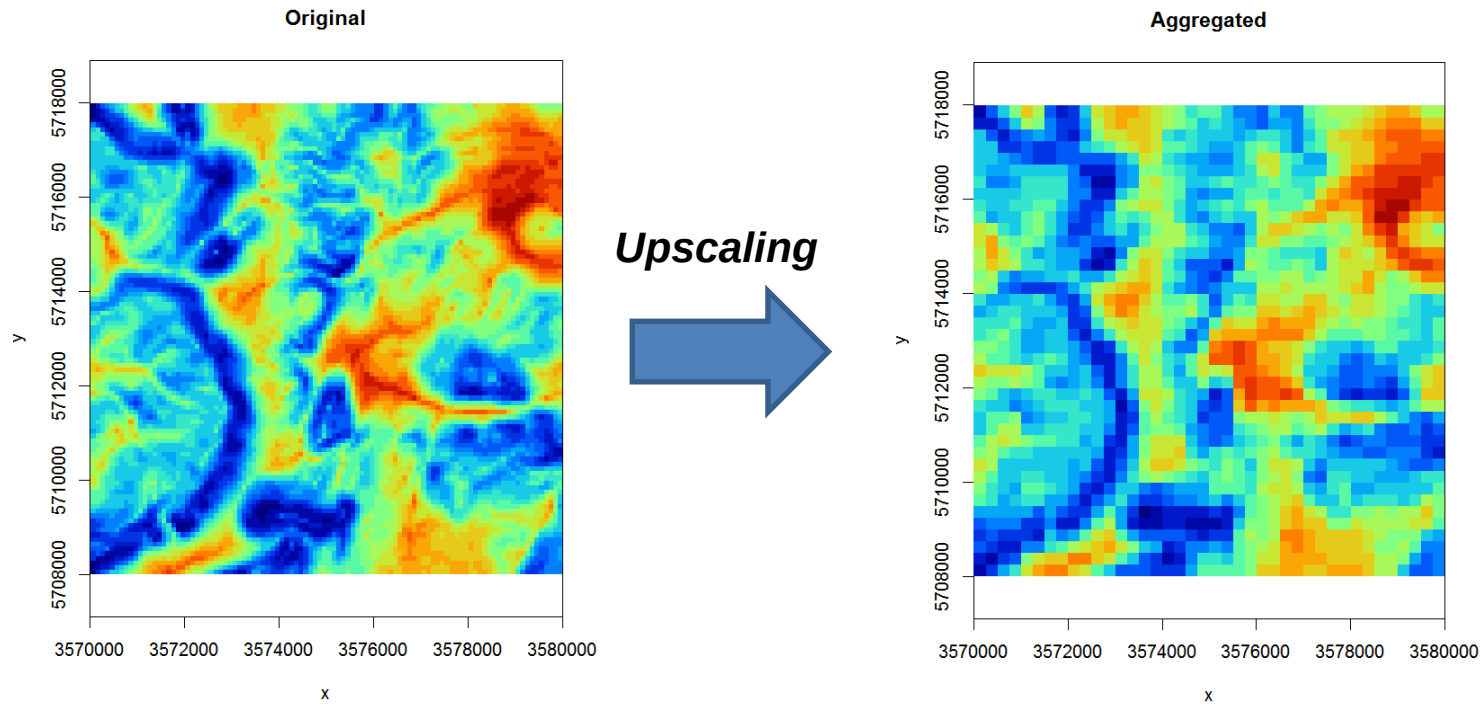


Technical and Practical Notes

s c o r p a n

2. Downscaling or upscaling rasters

- To adjust the resolution of some covariates that have either too coarse or too fine a resolution compared to the target resolution
- The process of bringing raster layers to a common grid resolution is also known as resampling



Technical and Practical Notes

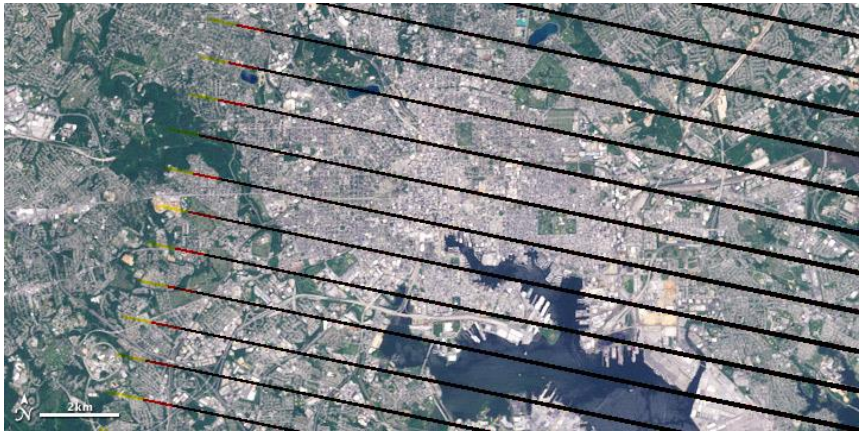
s c o r p a n

3. Filtering out missing pixels and artifacts

- could cause serious problems for producing soil maps as the missing pixels and artifacts would propagate to predictions: if only one layer in the raster stack misses values then predictive models might drop whole rows in the predictions even though data is available for 95% of rows.

- Missing pixels can be efficiently filtered by using for example the gap filling functionality available in the SAGA GIS

- Another way to filter the missing pixels, to reduce noise and to reduce data overlap is to use Principal Components transformation of original data

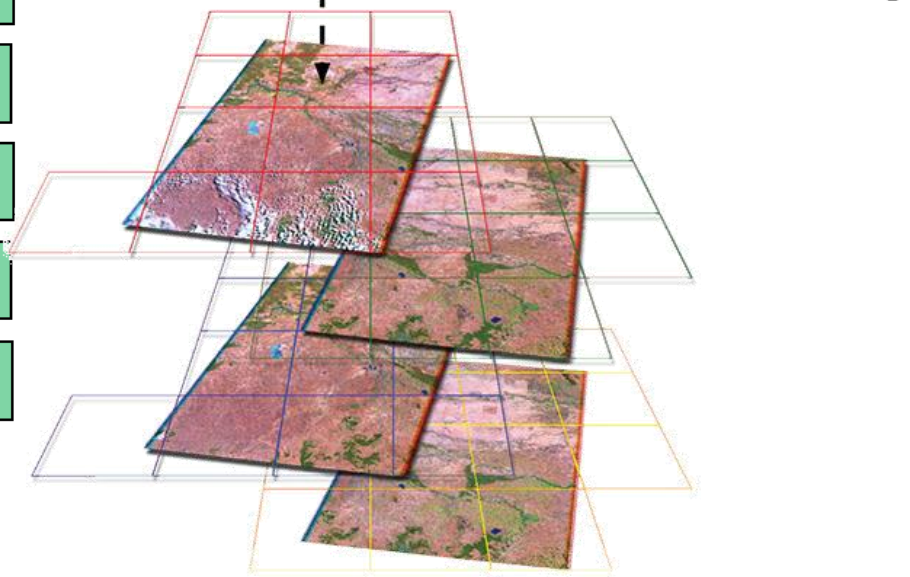


Technical and Practical Notes

s c o r p a n

Raster stacks

Resampling

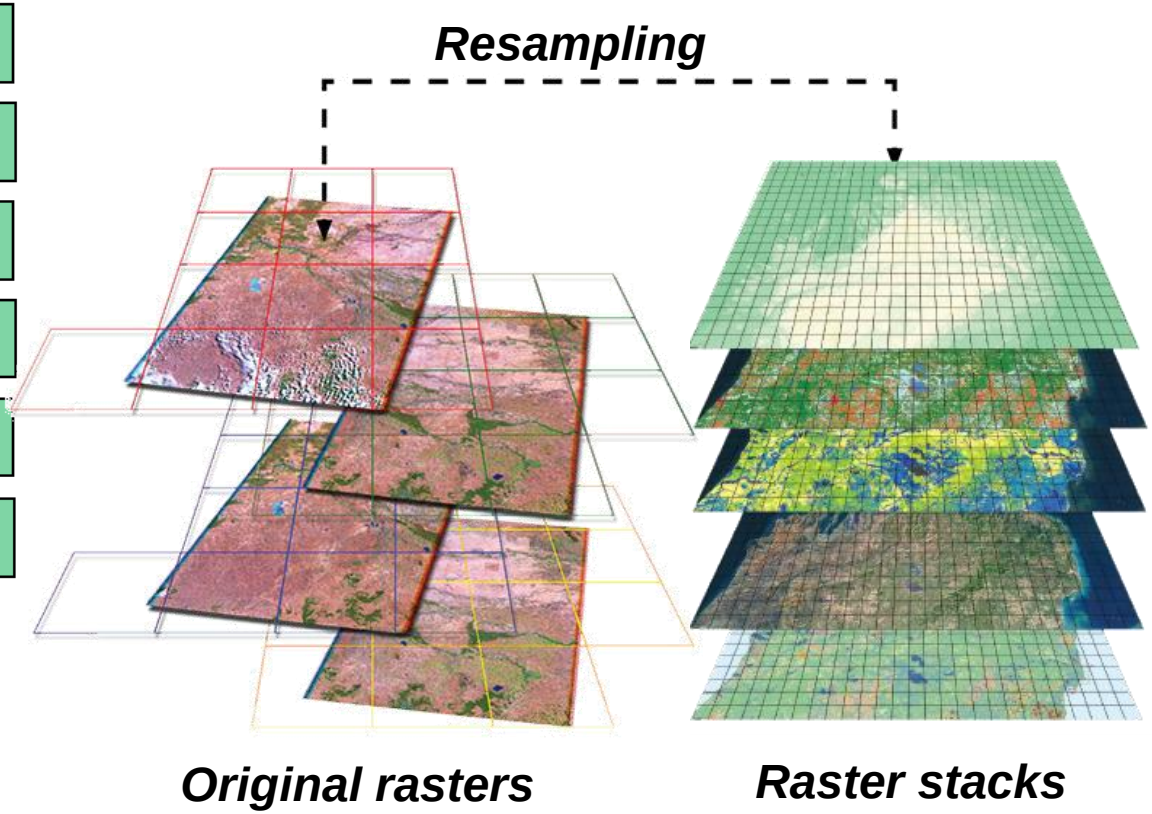


Original rasters

Technical and Practical Notes

s
c
o
r
p
a
n

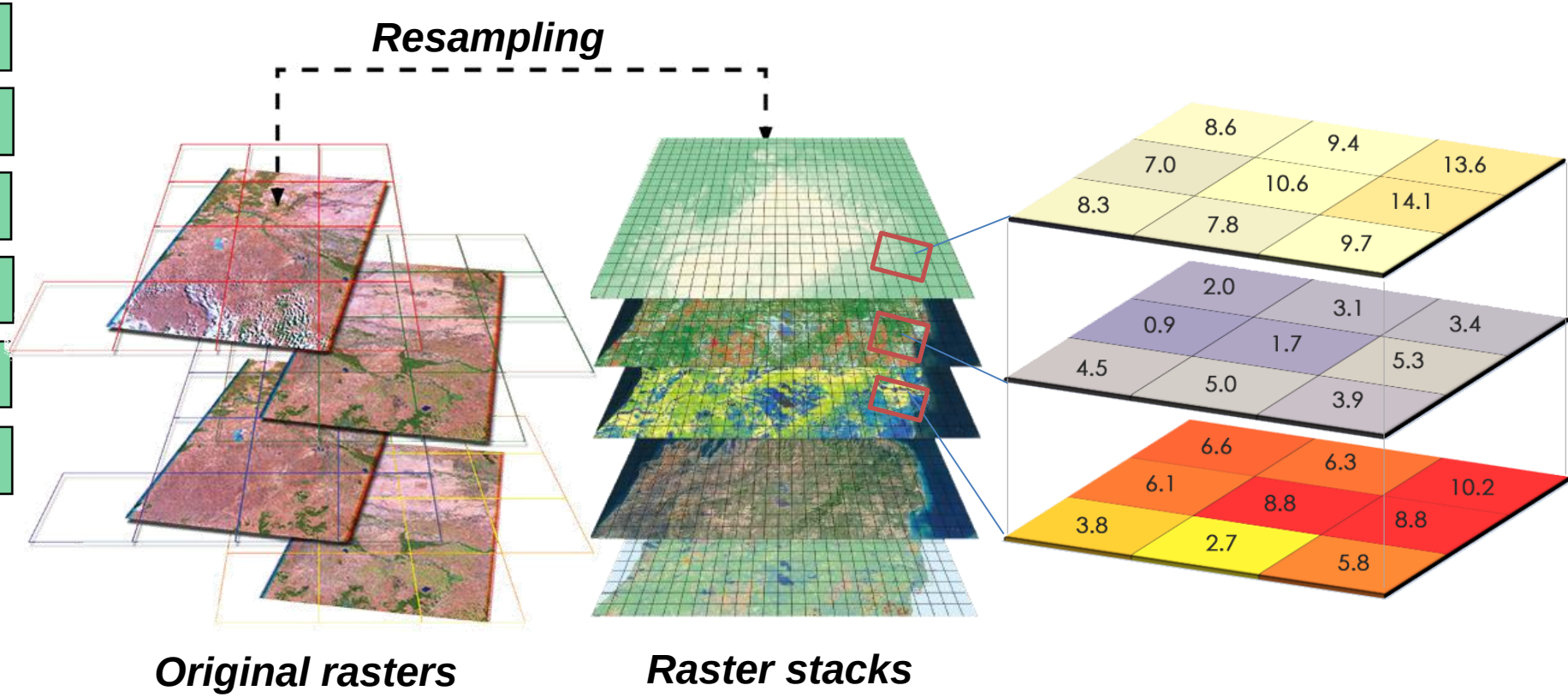
Raster stacks



Technical and Practical Notes

s
c
o
r
p
a
n

Raster stacks

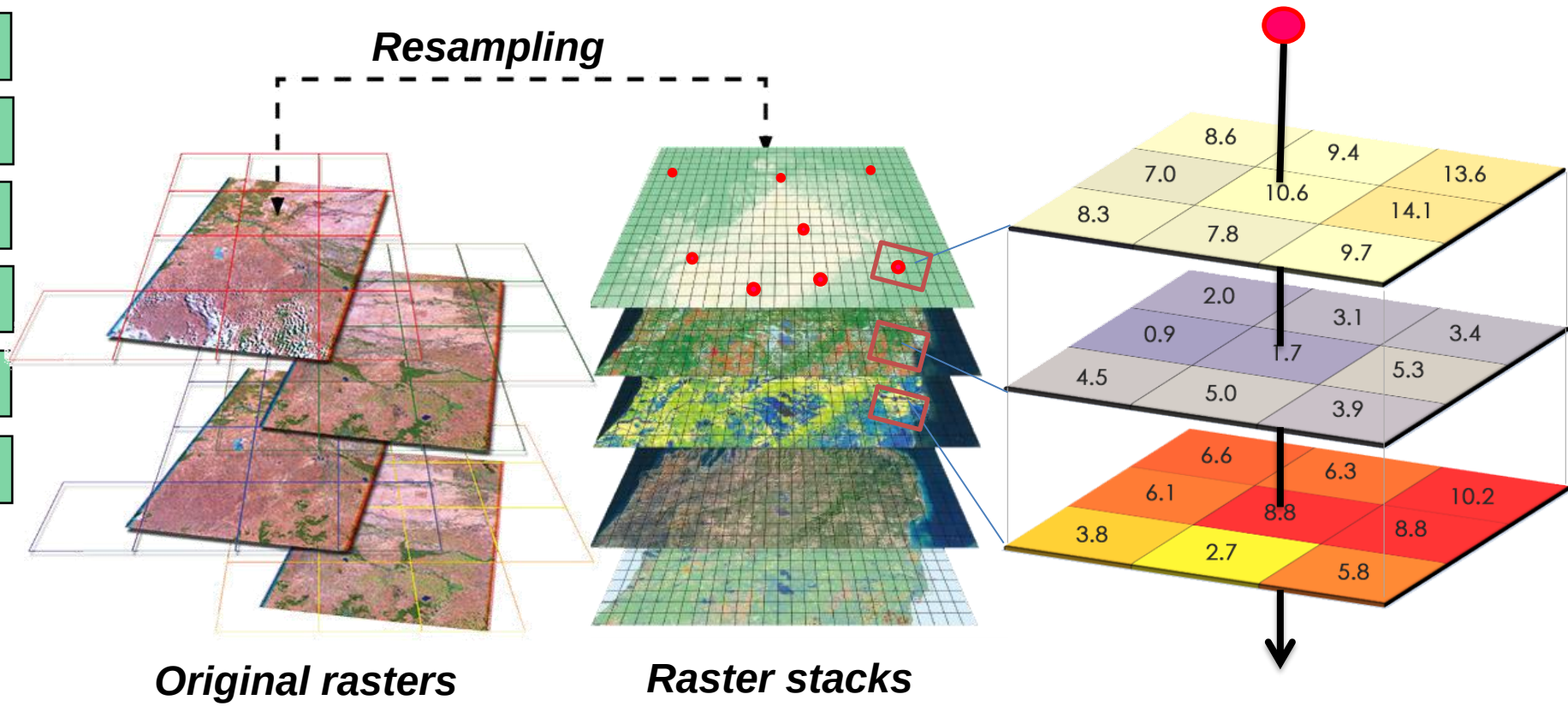


Original rasters ***Raster stacks***

Technical and Practical Notes

s c o r p a n

4. Overlaying raster stacks and points



Original rasters ***Raster stacks***

Technical and Practical Notes

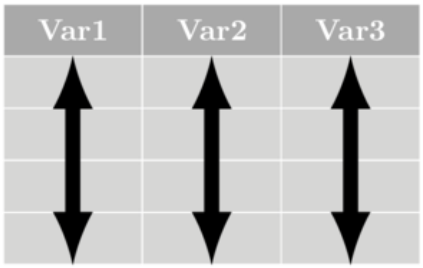
Geo-database
Tidy data
(data.frame)

Predictors
Covariates
Independent variables
Features

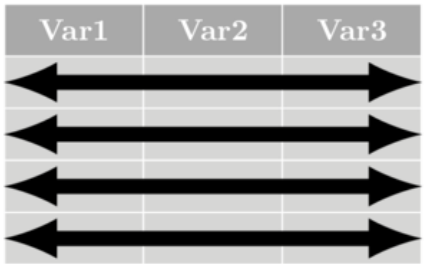
X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	Class
-0.60	3.06	-5.77	0.69	1.46	1.22	1.94	0.61	-0.07	-2.02	-0.87	0.71	0.89	-0.91	0.21	a
0.61	1.61	0.21	2.74	0.46	1.29	0.13	0.55	-0.60	-0.53	0.70	-0.41	-0.62	1.17	-0.32	b
3.81	0.99	-1.21	0.30	0.11	-0.34	-0.29	0.15	-0.09	1.30	0.08	0.66	-0.44	0.45	-0.13	a
0.50	0.14	-0.30	0.64	-1.28	0.28	0.76	-0.61	0.10	-0.07	0.41	-0.53	-0.62	0.06	-0.52	c
1.46	0.23	0.17	0.53	-0.94	0.37	0.64	-0.83	0.20	-0.06	0.18	-0.35	-0.64	0.39	-0.22	a
3.07	-0.19	1.31	0.84	-0.66	1.05	0.94	-1.12	-0.23	-0.31	0.45	-0.54	-0.16	0.38	0.53	a

Response variable
Target variable
Dependent variable
(Soil clay or soil types)

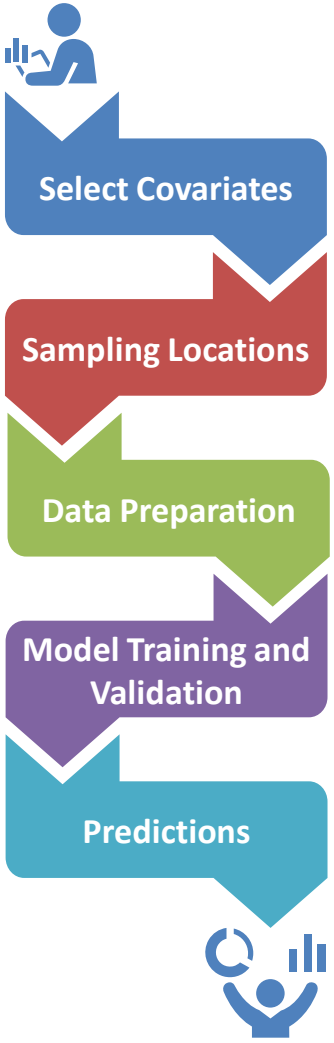
Variables in Columns



Observations in Rows

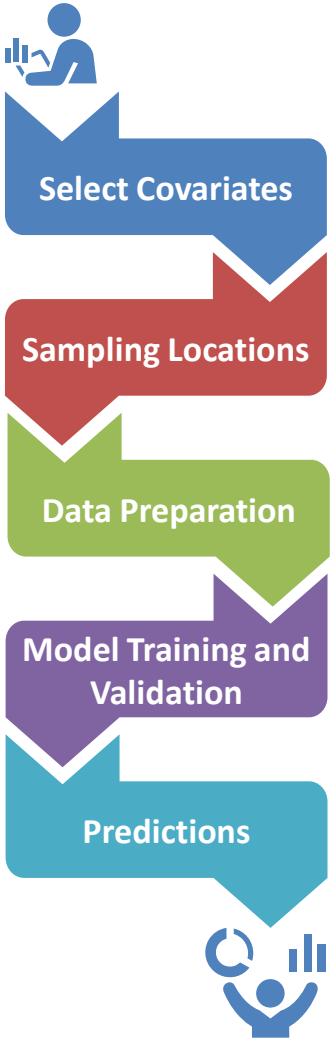


Take-Home Message



1. Environmental covariates, relevant as predictors of soil property/class, are derived from remote sensing, digital elevation, climatic datasets, …
2. Soil samples are collected at the specified locations (e.g., Latin hypercube sampling) and soil property is measured in the laboratory.
3. Intersecting the covariates with the soil point observations.
4. Machine learning models (e.g., random forest) are trained using training data, and accuracy assessment is carried out using the test data set.
5. The ML models are applied to the entire study area in order to produce a soil property/class map.

Take-Home Message



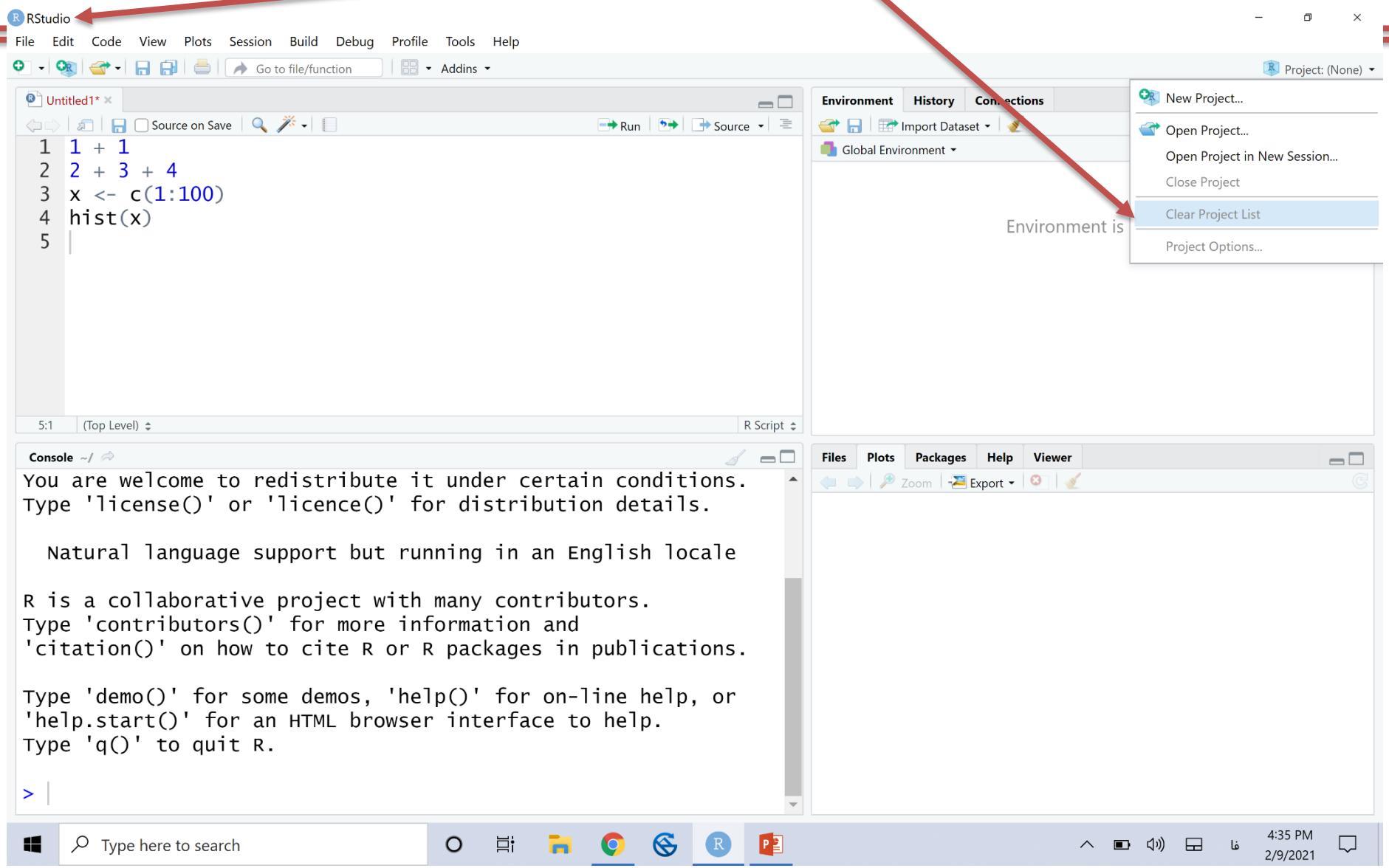
1. **Environmental covariates, relevant as predictors of soil property/class, are derived from remote sensing, digital elevation, climatic datasets, …**
2. Soil samples are collected at the specified locations (e.g., Latin hypercube sampling) and soil property is measured in the laboratory.
3. Intersecting the covariates with the soil point observations.
4. Machine learning models (e.g., random forest) are trained using training data, and accuracy assessment is carried out using the test data set.
5. The ML models are applied to the entire study area in order to produce a soil property/class map.

Please install R and RStudio

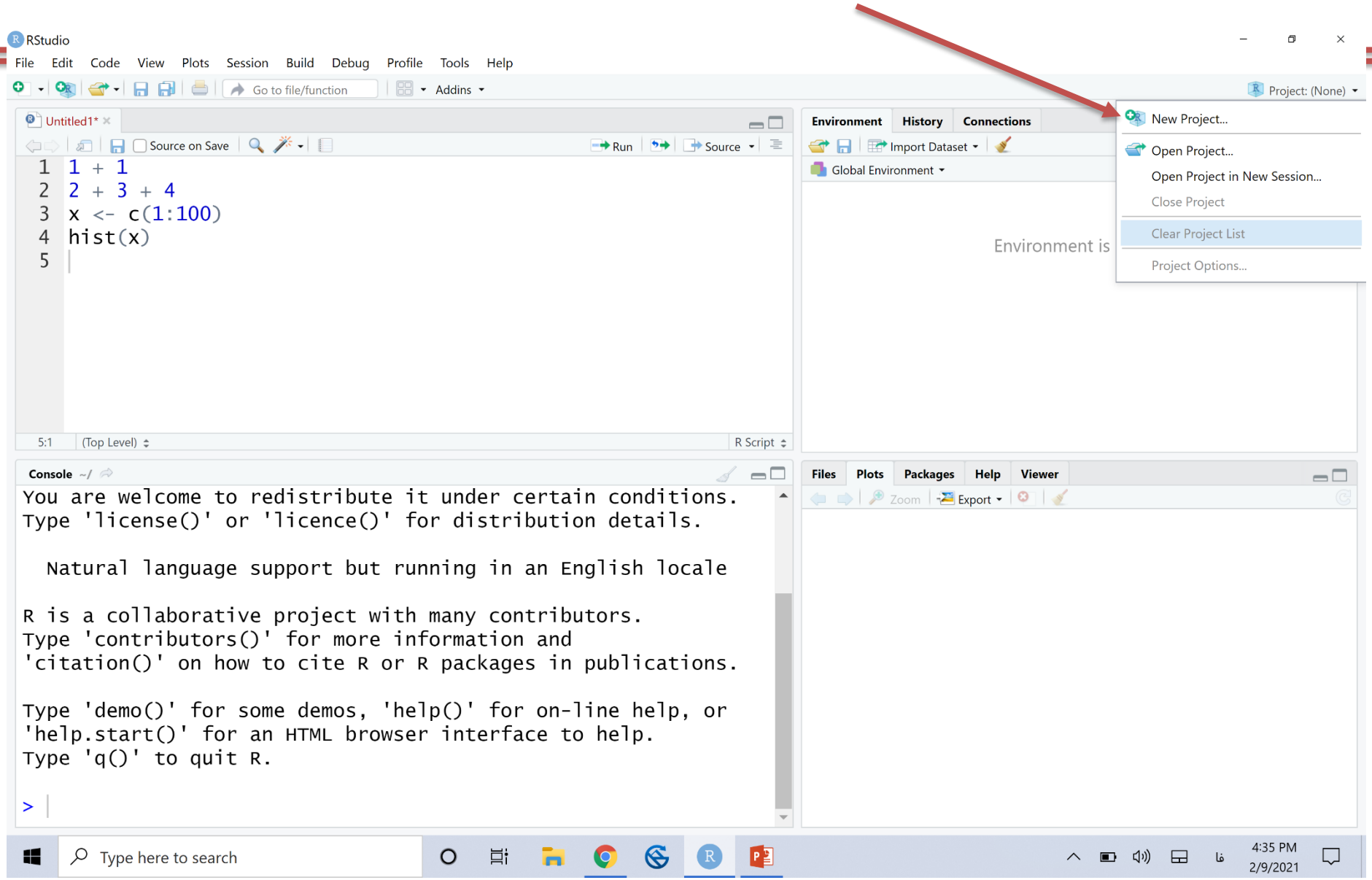
Please install SAGA and Google Earth

1. Open Rstudio

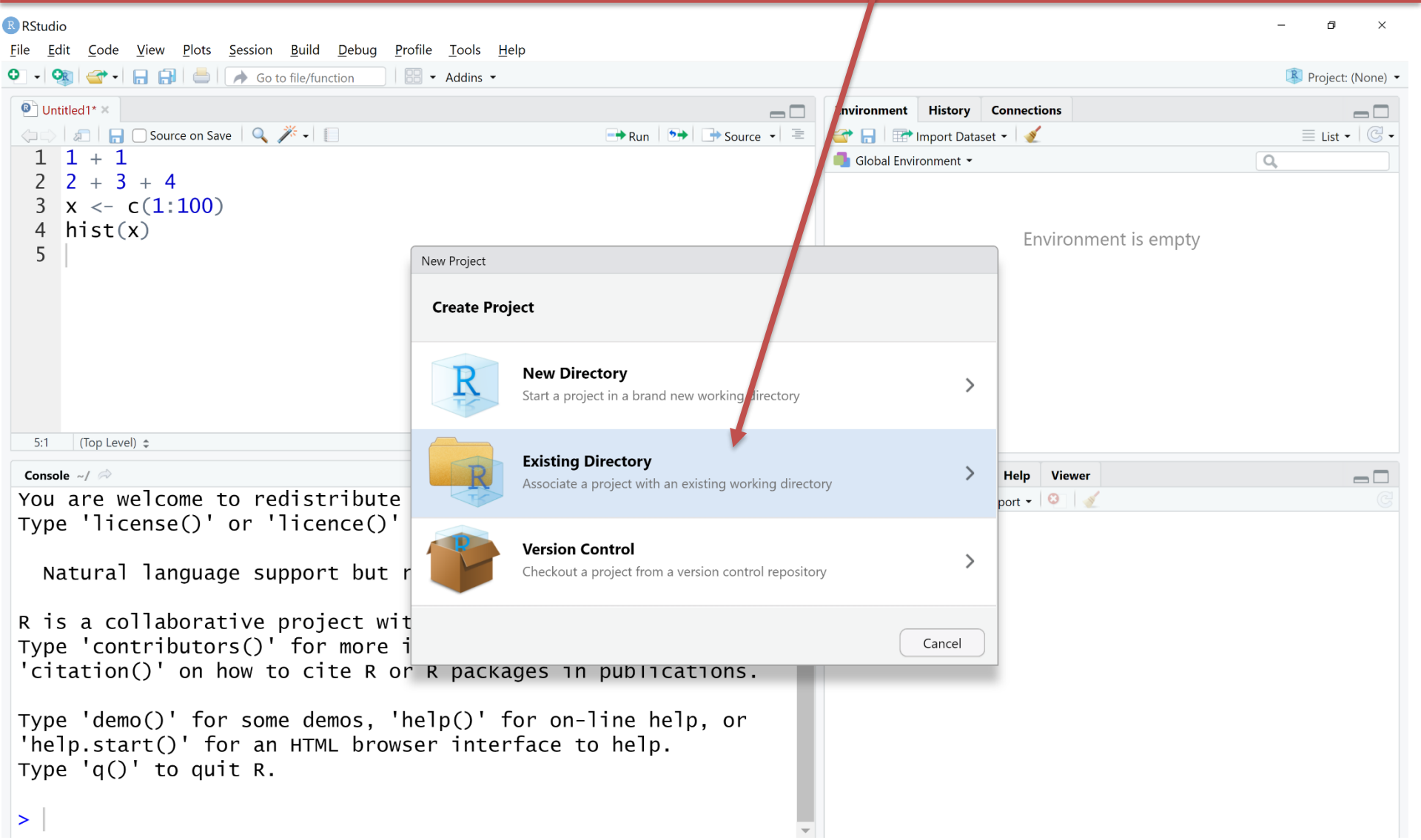
2. Clear Project List



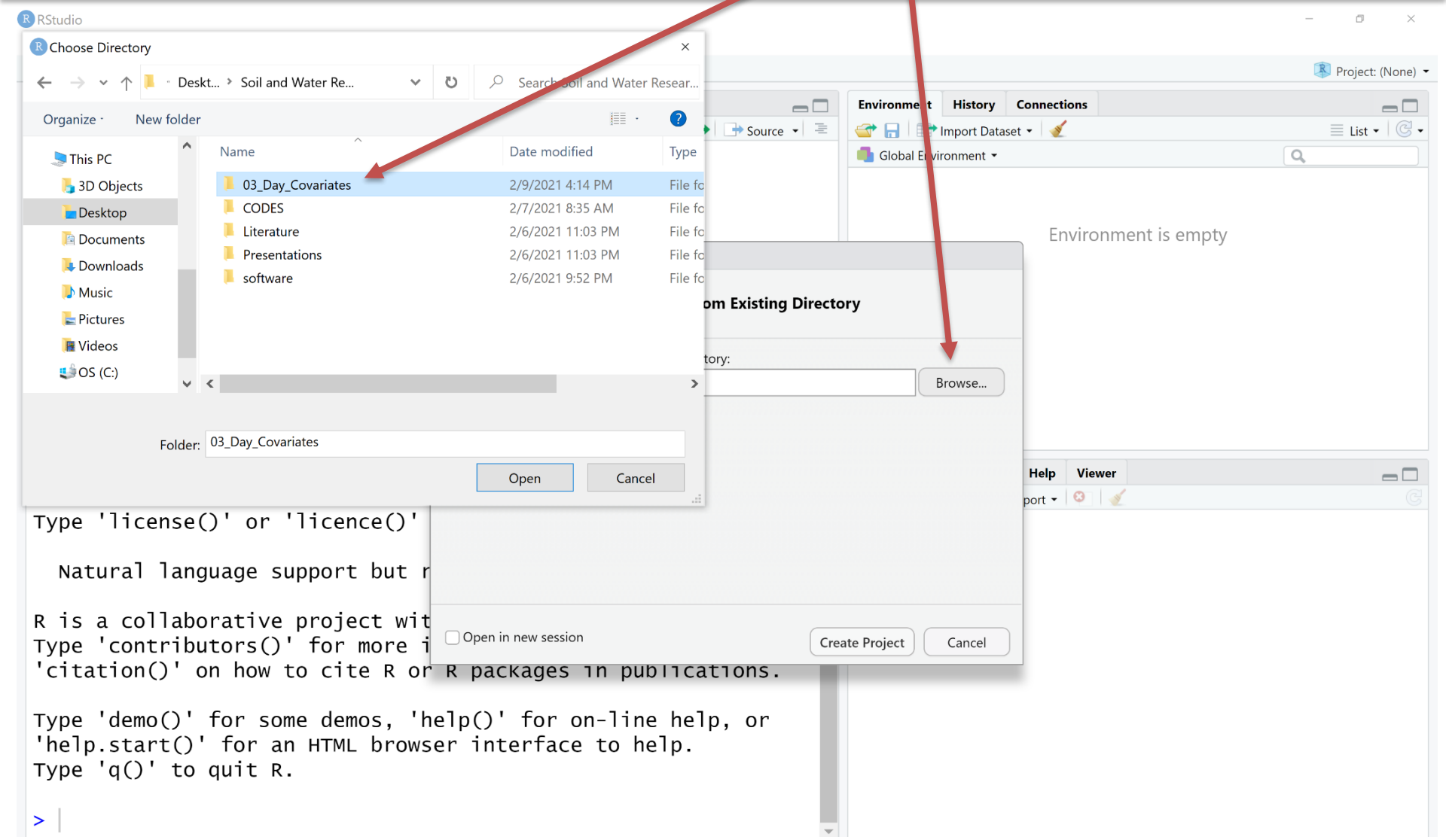
3. New Project



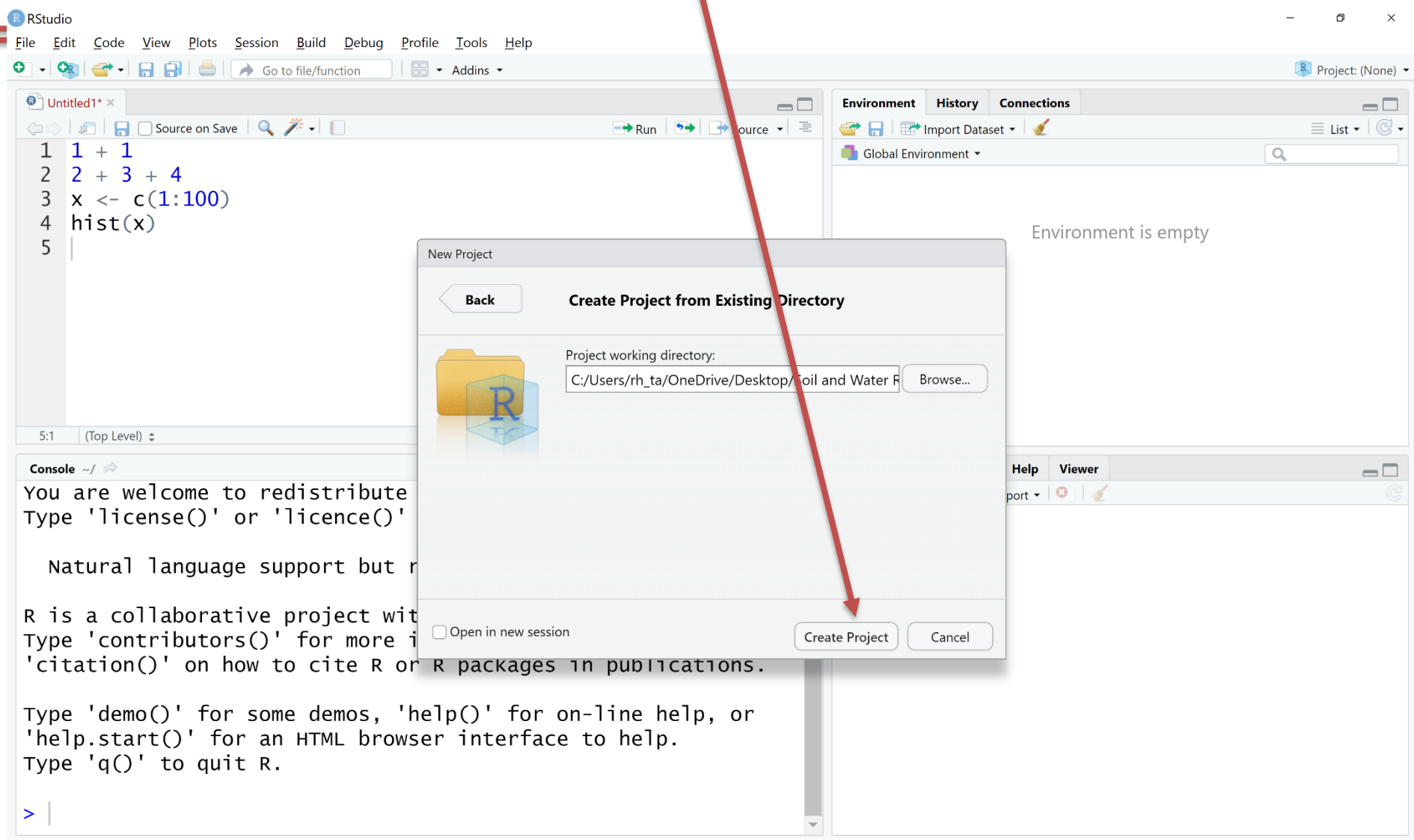
4. Existing Directory



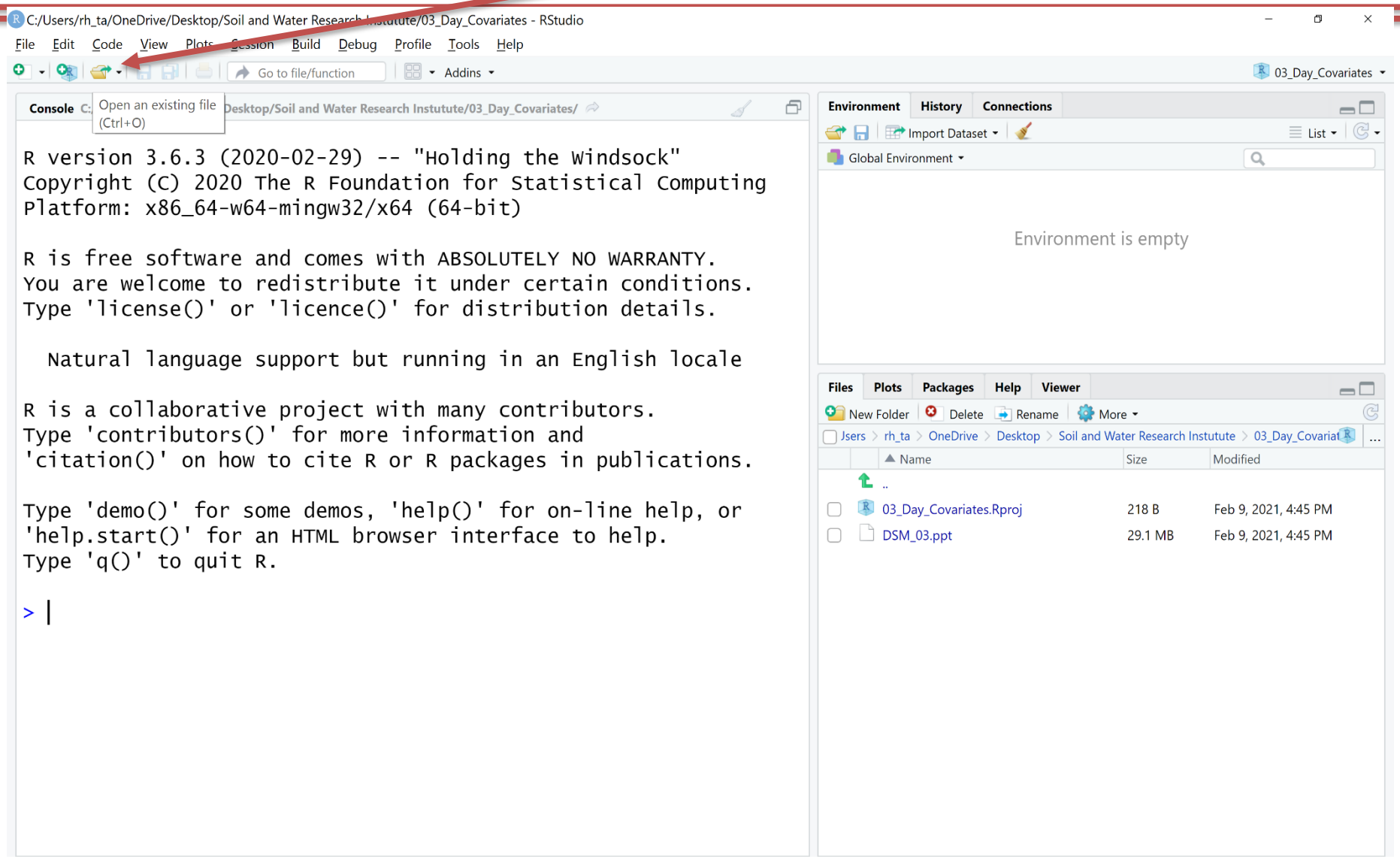
5. Find the folder and find **03_Day_Covariates** and open



6. Create project



7. Open an existing file and find **03_R_Cov.R**



8. Open **03_R_Cov.R**