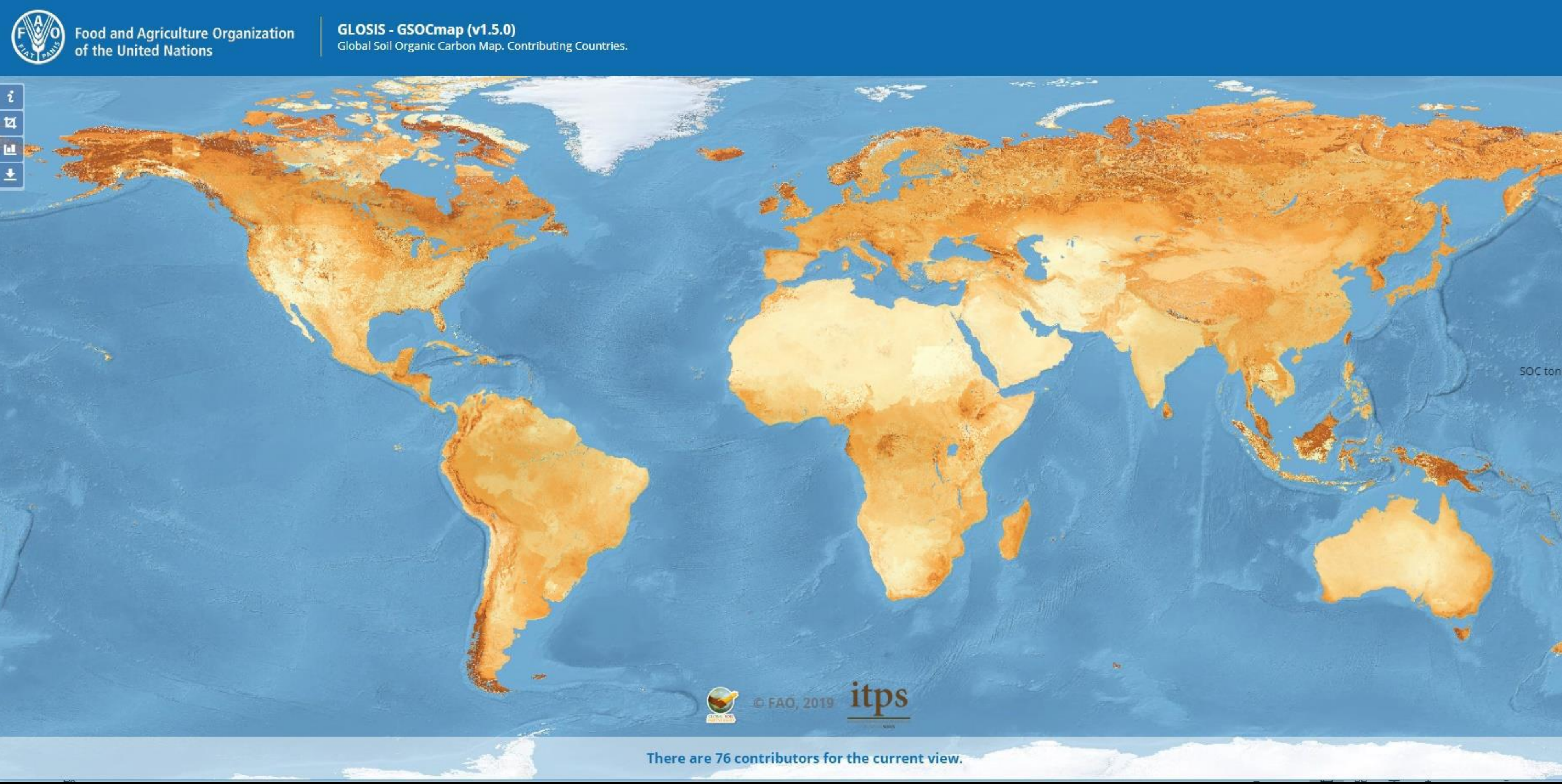


Geostatistics

Ruhollah Taghizadeh
Soil Science and Geomorphology

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

Content

- **Part 1**
  - **Geostatistics**

- **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

<u>From Jenny's Equation</u>

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

<u>Additions:</u>

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

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

McBratney et al., 2003

SCORPAN model

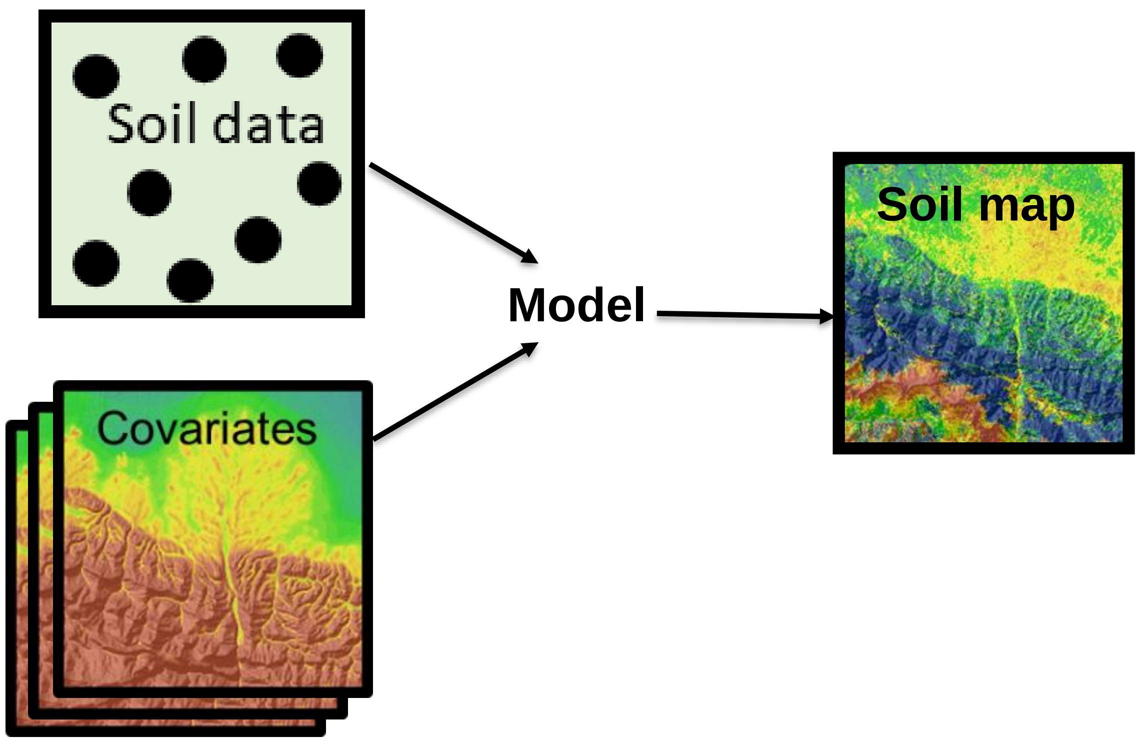
$$\boxed{S} = \boxed{f}\boxed{(s,\ c,\ o,\ r,\ p,\ a,\ n)} + \varepsilon$$

Soil data

Model

Covariates

spatially dependent residuals



SCORPAN Model

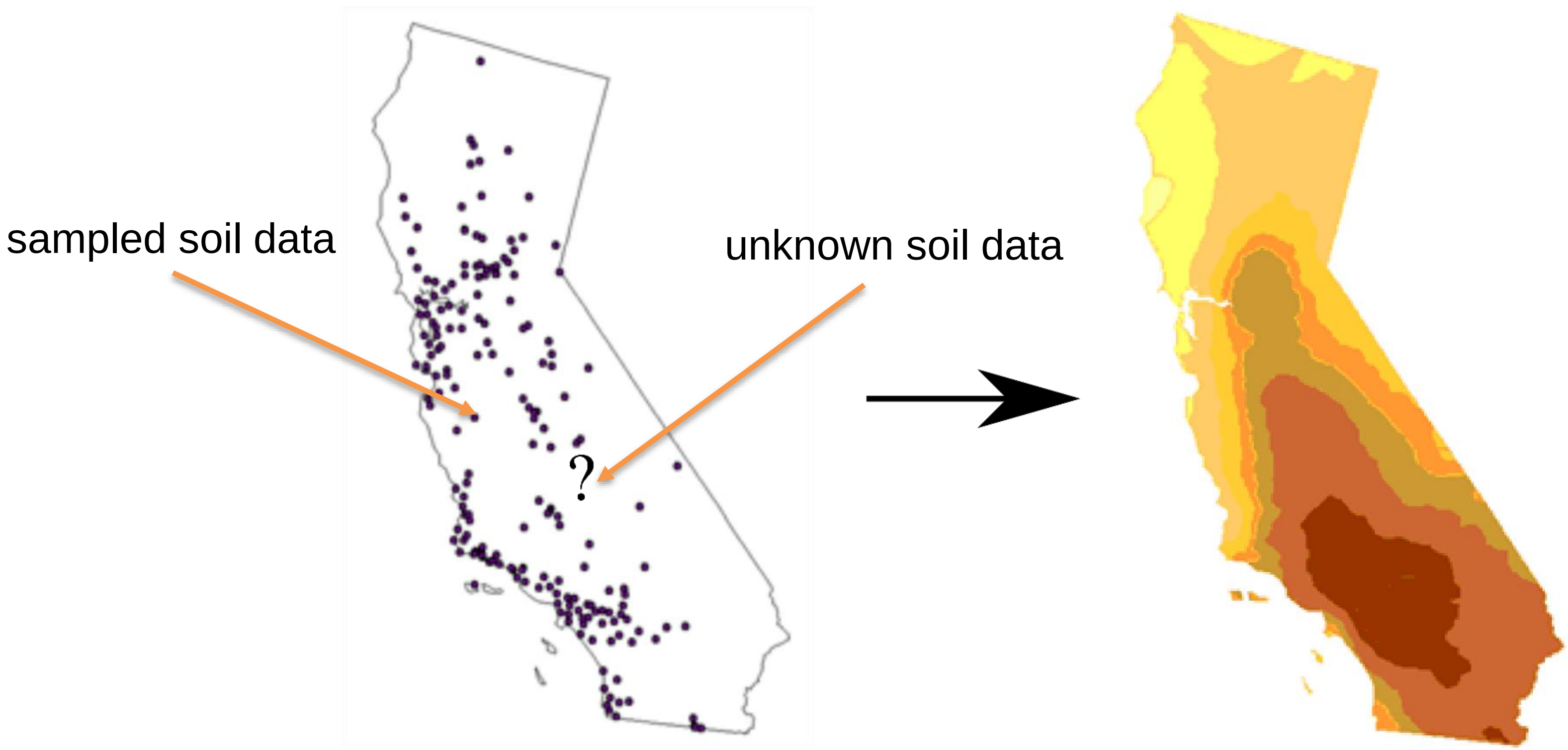
$$S = f(s, c, o, r, p, a, n) + \varepsilon$$

Machine learning

f() : Quantitative function *f* linking S to ***s**co**r**pa**n*** factors

Spatial Interpolation

- Spatial prediction from point samples is one of the main practical applications of geostatistics
- We know the value of some attribute at some observation points, but we need to know it over an entire area - i.e. we want to map it.



Spatial interpolation (Source: http://desktop.arcgis.com)

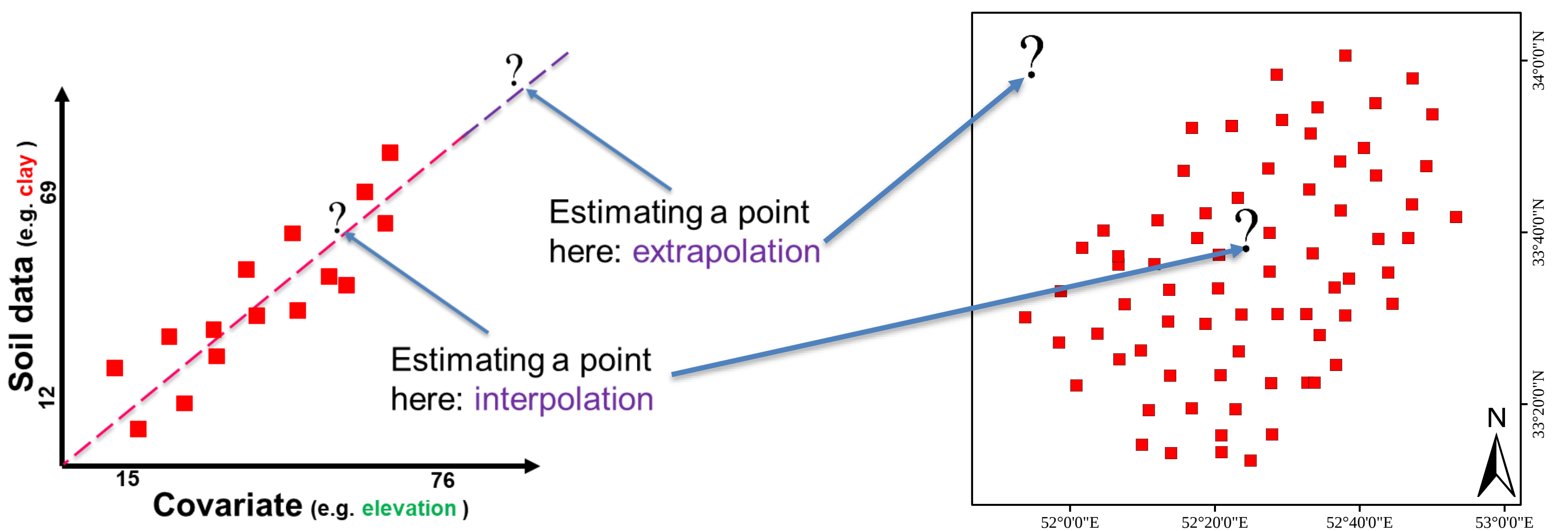
Interpolation vs. Extrapolation

Interpolation

- Estimating the attribute values of locations that are within the range of available data using known data values

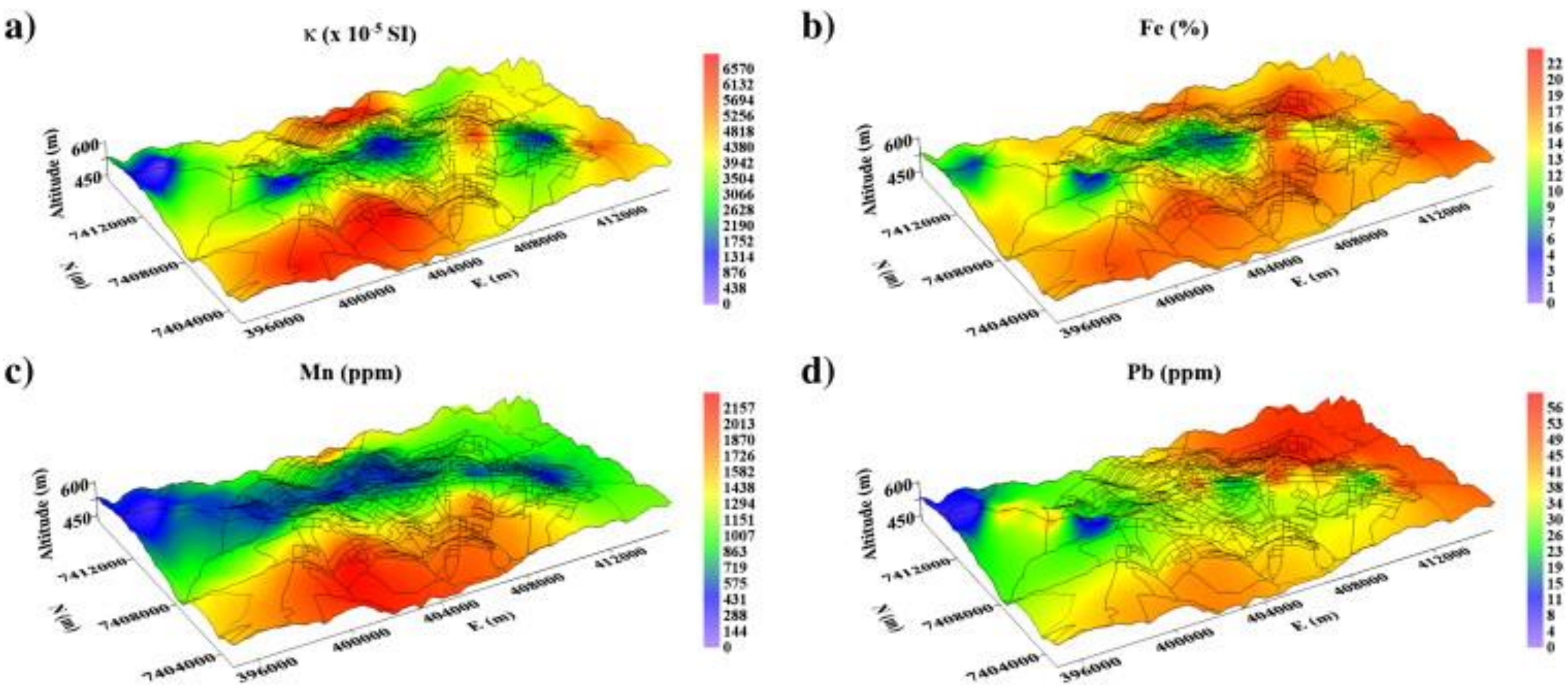
Extrapolation

- Estimating the attribute values of locations outside the range of available data using known data values



We Know,

- Soils and soil properties vary in space



Cervi et al., 2014

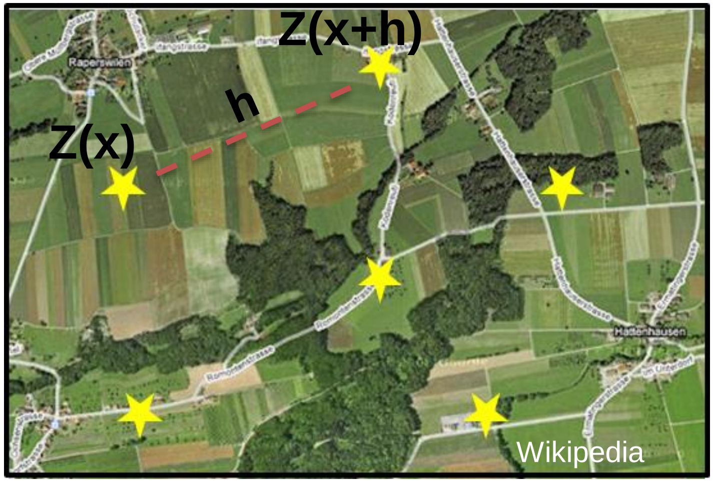
The Variogram

The Central Tool of Geostatistics

- Spatial variation can be quantified using the so-called **semi-variance**

- Semi-variance is calculated according to:

$$\gamma(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{\underbrace{z(x_i)}_{\text{measurement at location x}} - \underbrace{z(x_i + h)}_{\text{measurement at location x+h}}\}^2$$

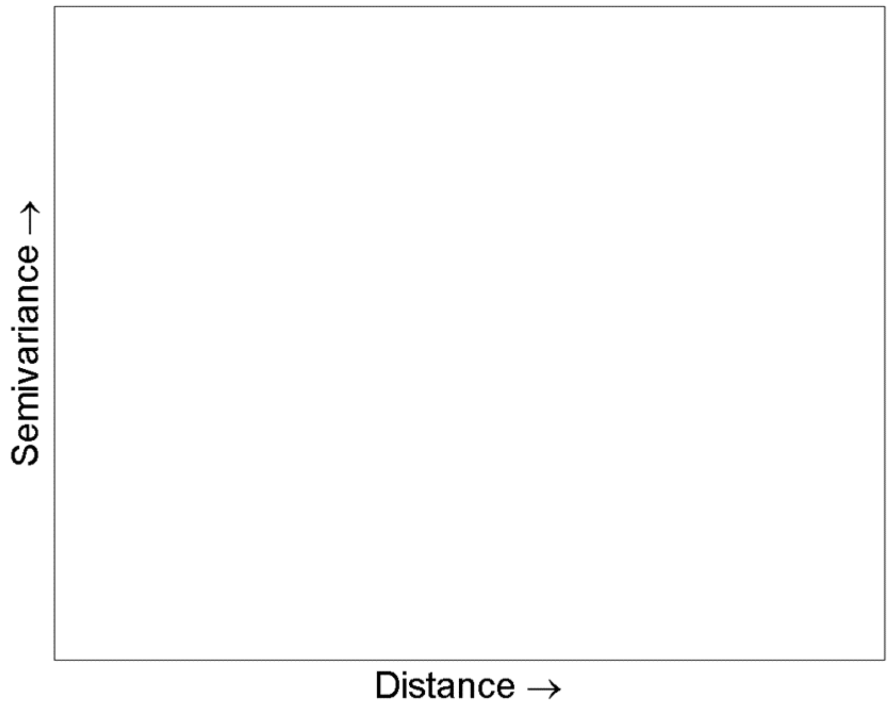


(Source: Heuvelink, 2017, ISRIC)

The Variogram

The Central Tool of Geostatistics

- Plot of semi-variance as a function of the distance is called a semi-variogram

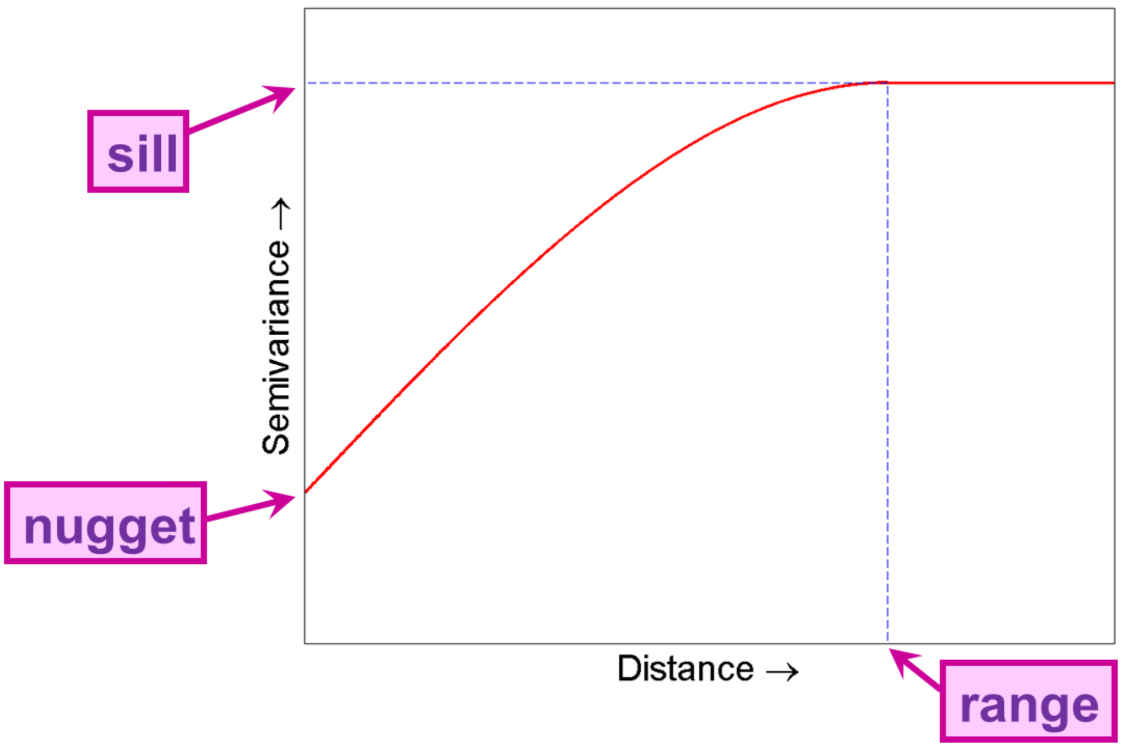


(Source: Heuvelink, 2017, ISRIC)

The Variogram

The Central Tool of Geostatistics

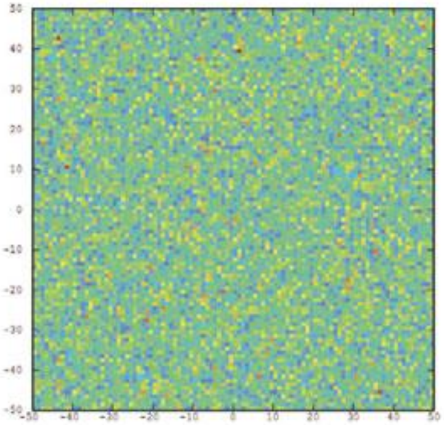
- Typical shape of the semi-variogram, with parameters
  - **Nugget** = measurement errors and short-distance spatial variation
  - **Sill** = variance of the variable of interest
  - **Range** = distance up to which there is spatial correlation

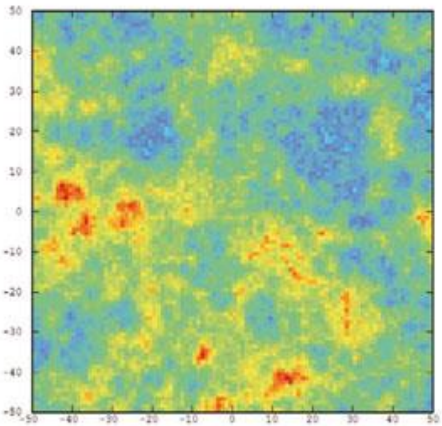


(Source: Heuvelink, 2017, ISRIC)

The Variogram

The Central Tool of Geostatistics

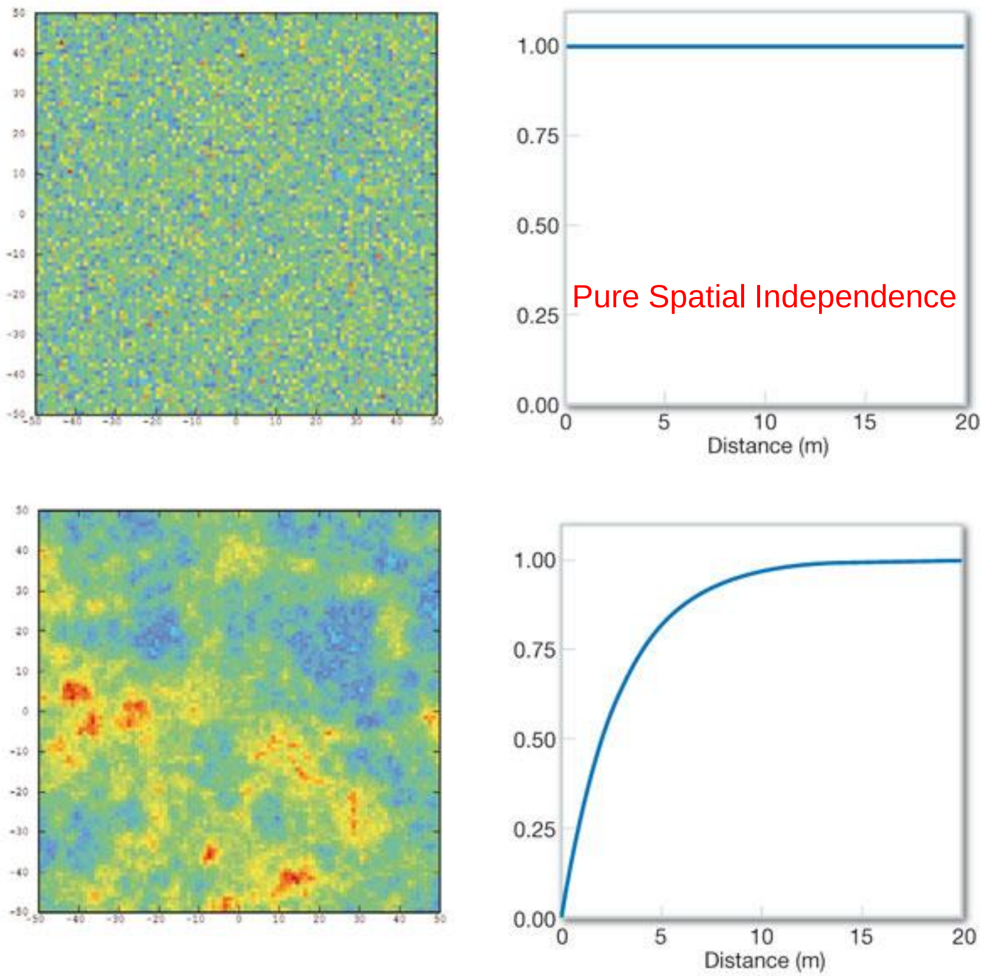




Two phenomena with the same statistical distribution (at left) but with significantly different spatial structures (corresponding variograms at right). Source: https://www.neimagazine.com/

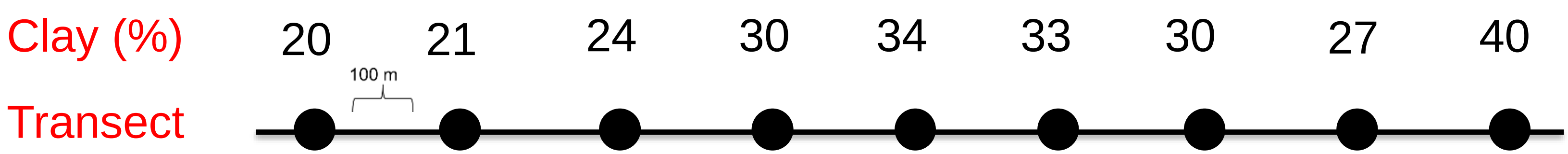
The Variogram

The Central Tool of Geostatistics

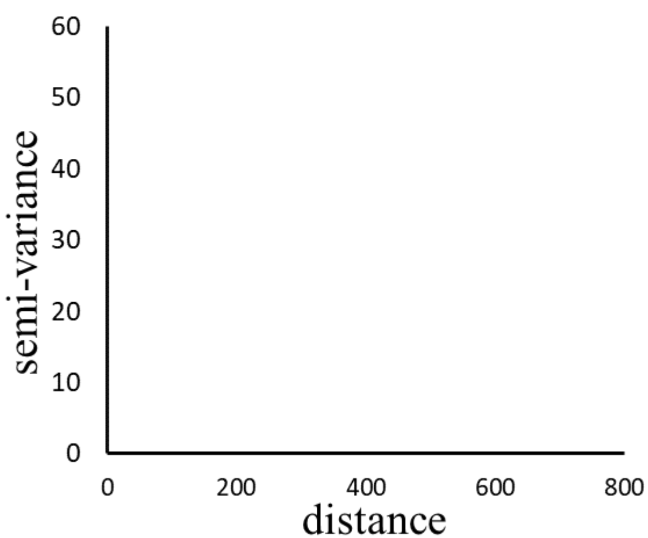


Two phenomena with the same statistical distribution (at left) but with significantly different spatial structures (corresponding variograms at right). Source: https://www.neimagazine.com/

Calculate Variogram: example



$$\gamma(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{\underbrace{z(x_i)}_{\text{measurement at location x}} - \underbrace{z(x_i + h)}_{\text{measurement at location x+h}}\}^2$$

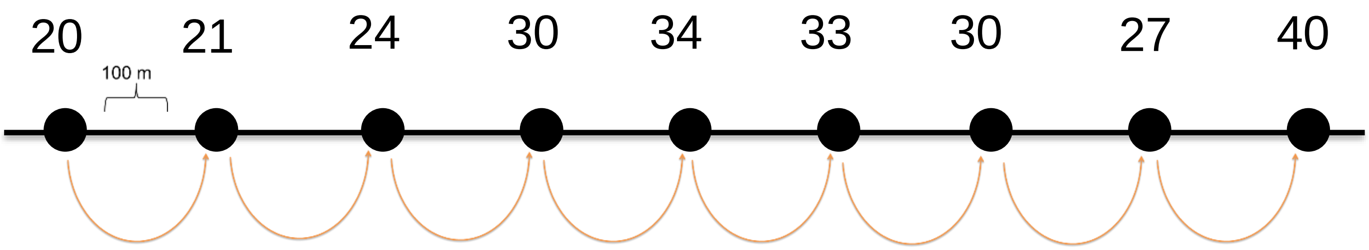


Calculate Variogram: example

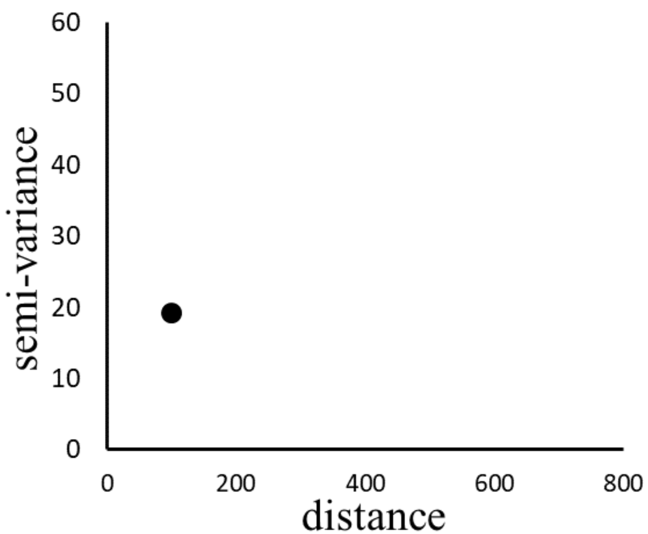
Clay (%)

Transect

d= 100m



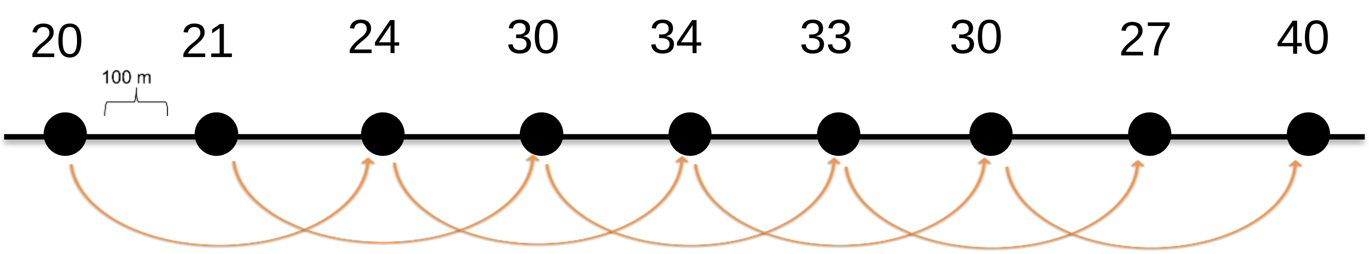
$$\gamma(100) = \frac{1}{16}\left[\begin{array}{l}(20-21)^2+(21-24)^2+(24-30)^2+(30-34)^2 \\ +(34-33)^2+(33-30)^2+(30-27)^2+(27-40)^2\end{array}\right] = 19.12$$



Calculate Variogram: example

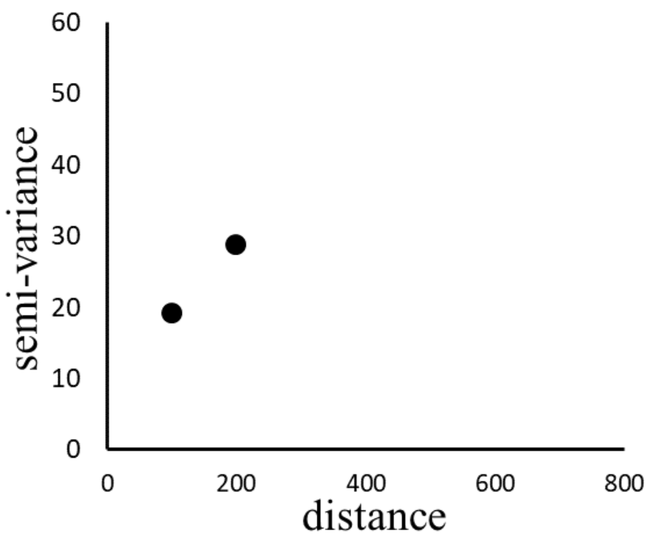
Clay (%)

Transect



d= 200m

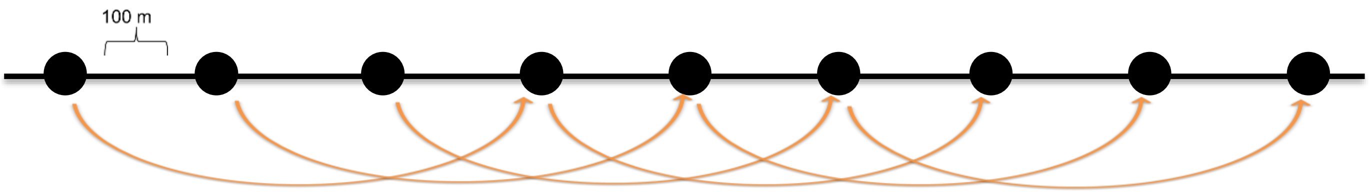
$$\gamma(200) = \frac{1}{14}\left[\begin{array}{c}(20-24)^2 + (21-30)^2 + (24-34)^2 + (30-33)^2 \\ +(34-30)^2 + (33-27)^2 + (30-40)^2\end{array}\right] = 28.71$$



Calculate Variogram: example

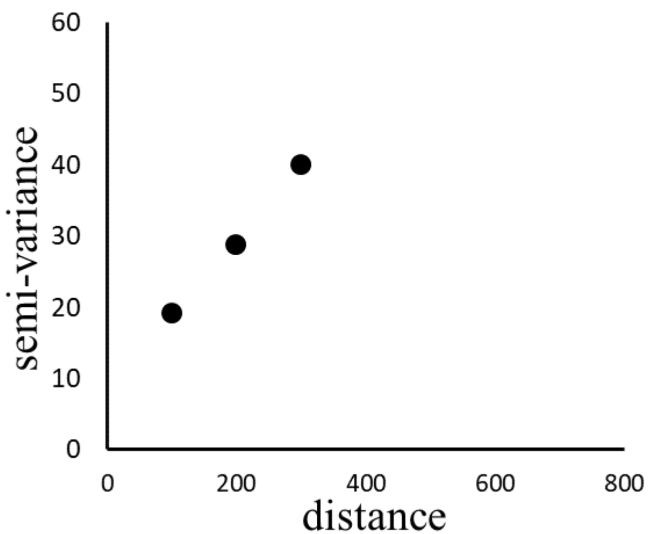
Clay (%) 20 21 24 30 34 33 30 27 40

Transect



d= 300m

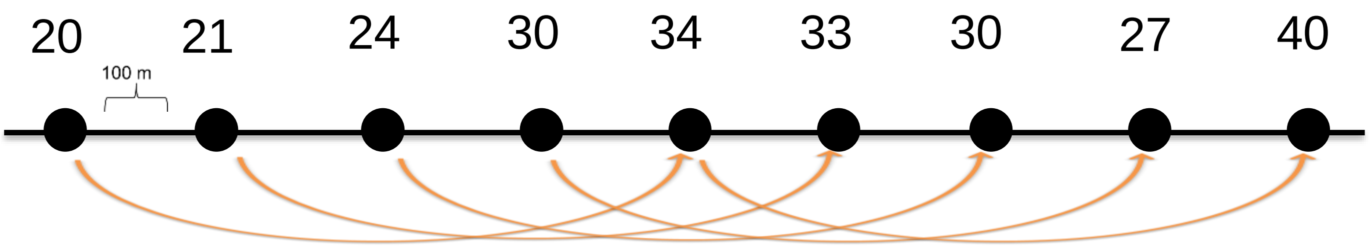
$$\gamma(300) = \frac{1}{12}\left[\begin{array}{c}(20-30)^2 + (21-34)^2 + (24-33)^2 + (30-30)^2 \\ +(34-27)^2 + (33-40)^2\end{array}\right] = 40.00$$



Calculate Variogram: example

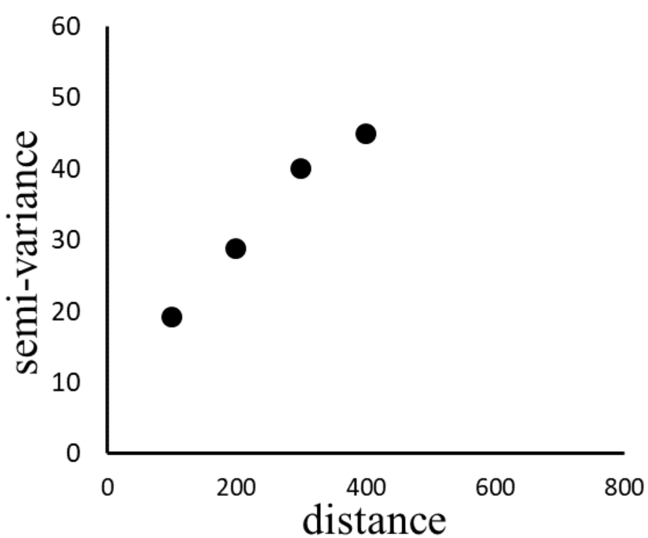
Clay (%)

Transect



d= 400m

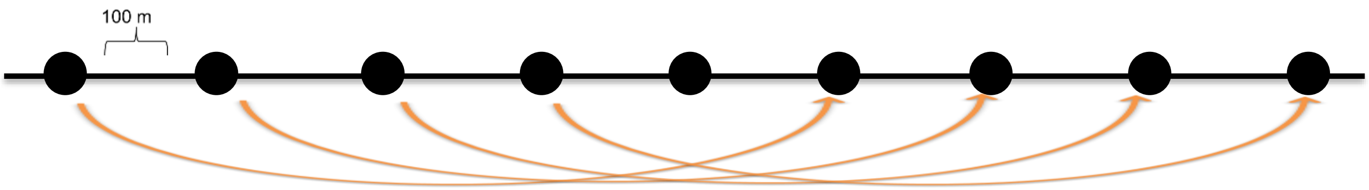
$$\gamma(400) = \frac{1}{10}\left[\begin{array}{c}(20-34)^2 + (21-33)^2 + (24-30)^2 + (30-27)^2 \\ +(34-40)^2\end{array}\right] = 44.90$$



Calculate Variogram: example

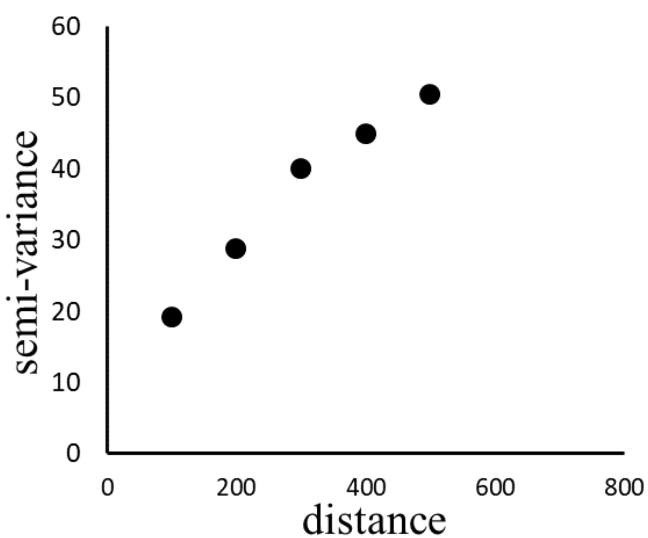
Clay (%) 20 21 24 30 34 33 30 27 40

Transect



d= 500m

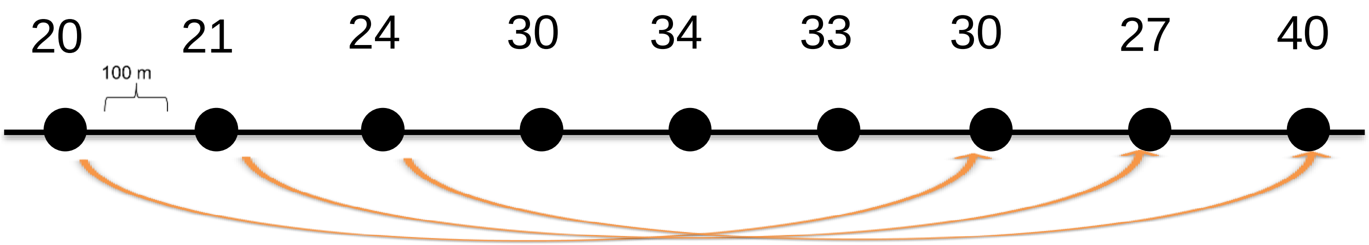
$$\gamma(500) = \frac{1}{8}\left[(20-33)^2 + (21-30)^2 + (24-27)^2 + (30-40)^2\right] = 50.37$$



Calculate Variogram: example

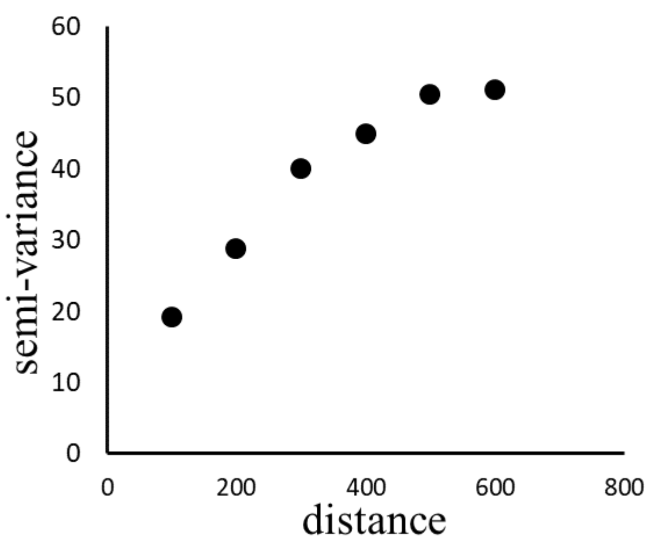
Clay (%)

Transect



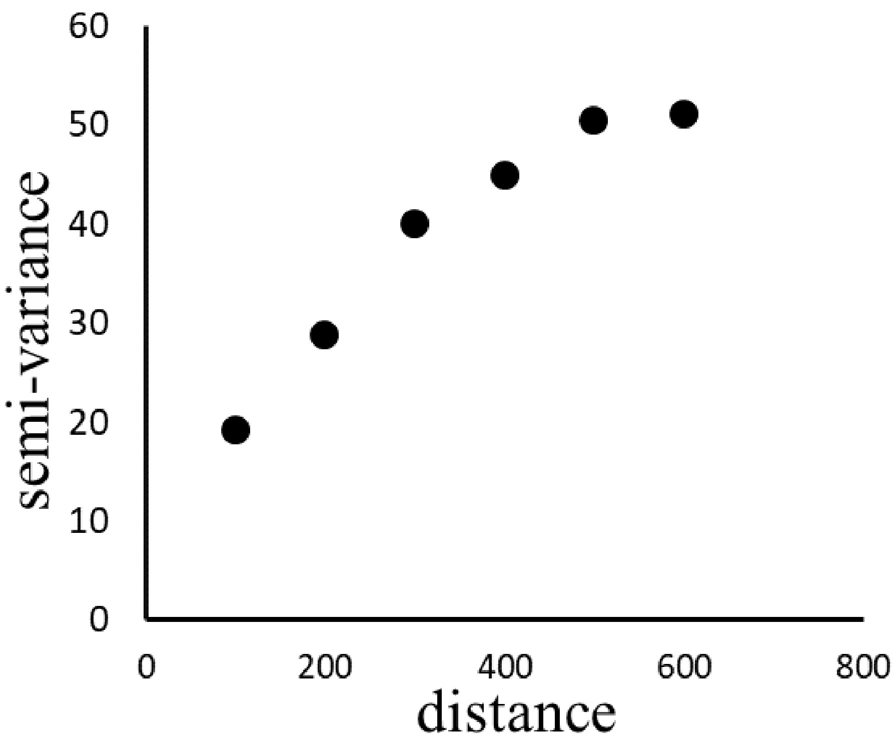
d= 600m

$$\gamma(600) = \frac{1}{6}\left[(20-30)^2 + (21-27)^2 + (24-40)^2\right] = 50.37$$



Calculate Variogram: example

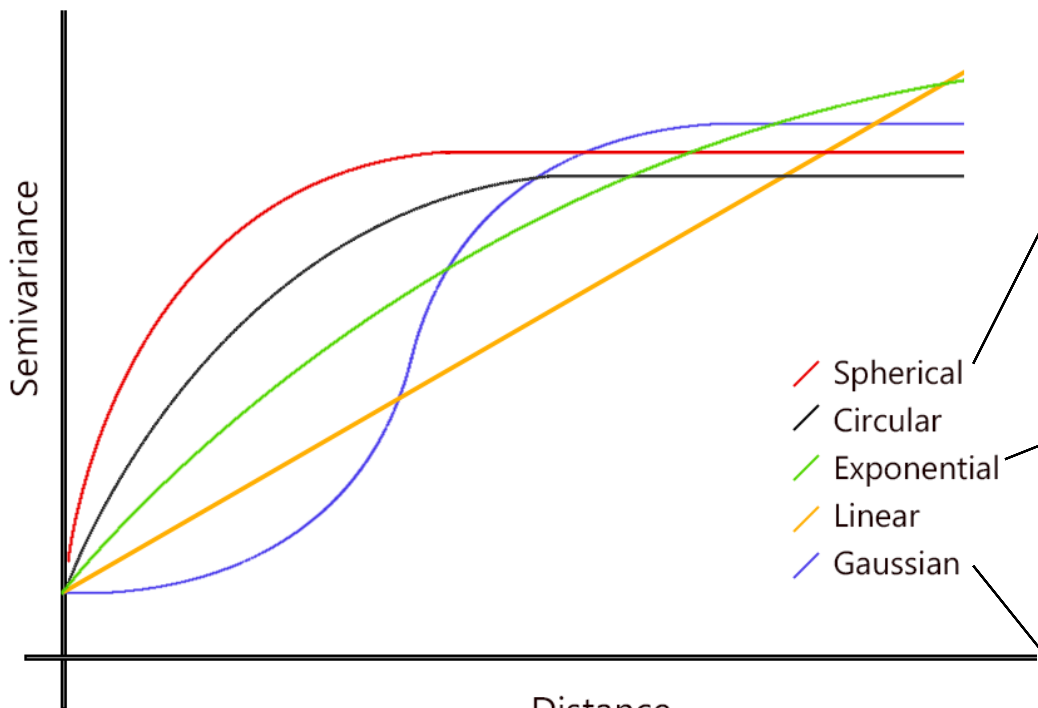
- Experimental semi-variogram is a plot of semi-variance of observations at a number of lag distances
- This is modelled with a continuous function



experimental semi-variogram

Calculate Variogram: example

- There are a number of authorized models that can be used e.g. spherical, exponential, Gaussian
  1. Choose a function shape
  2. Estimate parameters of the chosen shape (e.g. using weighted least squares fitting)



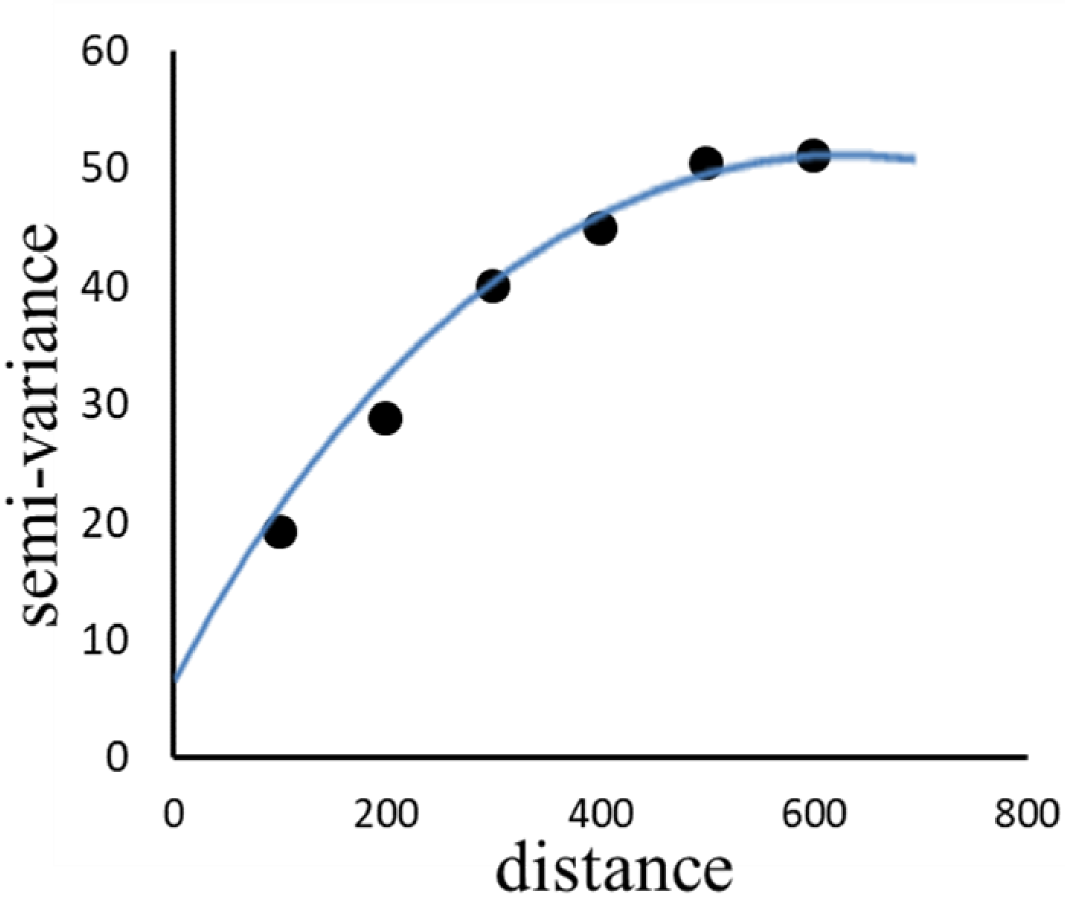
$$\gamma(h) = 0 \quad \text{for } h = 0$$
$$\gamma(h) = c_0 + c \cdot \{\frac{3}{2} \cdot \frac{h}{a} - \frac{1}{2} \cdot \left(\frac{h}{a}\right)^3\} \quad \text{for } 0 < h \leq a$$
$$\gamma(h) = c_0 + c \quad \text{for } h > a$$

$$\gamma(h) = 0 \quad \text{for } h = 0$$
$$\gamma(h) = c_0 + c \cdot (1 - e^{-\frac{h}{a}}) \quad \text{for } h > 0$$

$$\gamma(h) = 0 \quad \text{for } h = 0$$
$$\gamma(h) = c_0 + c \cdot (1 - e^{-\left(\frac{h}{a}\right)^2}) \quad \text{for } h > 0$$

Calculate Variogram: example

- Semi-variogram model for clay example.



Assumption: stationarity and isotropy

- Semi-variogram model
- *Assumption: stationarity and isotropy*

  - The semi-variance of **Z(x)** and **Z(x+h)** only depends on the distance **h** and not on the locations **x** and **x+h** (**stationarity**)

  - The semi-variance is a function of the **length of h** , not of its direction (**isotropy**)

  - *These assumptions are not always realistic*

Geostatistical Interpolation: Kriging

- Kriging is a spatial prediction method
- Kriging is named after the South African engineer, D.G. Krige, who first developed the method.
- Prediction at a location is a linear combination of observations nearby (a weighted average as for inverse distance)
- The weight that is given to each observation depends on the degree of (spatial)correlation: the semi-variogram plays an important role
- Many kriging methods have been developed for different prediction purposes, e.g., block kriging, universal kriging, cokrigin, etc. Here we will only concentrate on the most basic one: ordinary kriging, and also regression kriging

Danie Gerhardus Krige, the inventor of Kriging.

Cressie, N. 1990. The origins of kriging. *Mathematical Geology* 22:239-252.

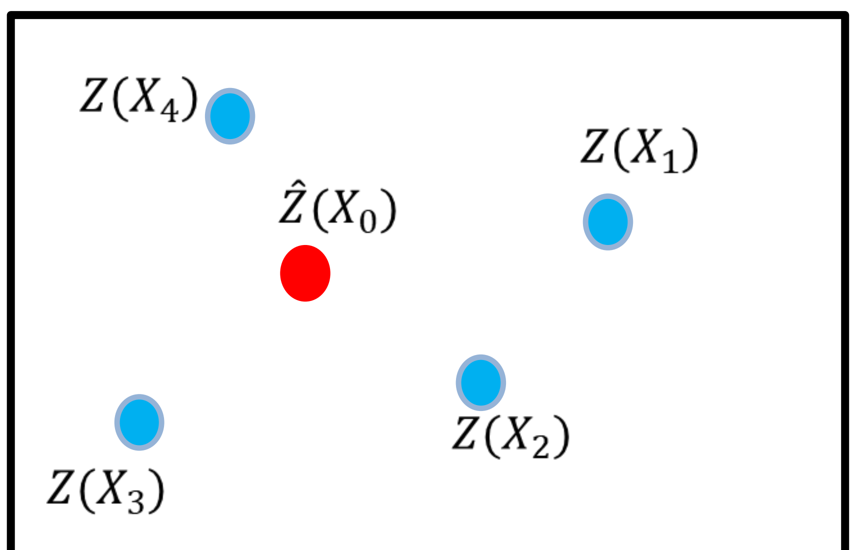
Ordinary Kriging

- The estimated value $\hat{Z}$ at a point X_0 is predicted as the weighted average of the values at all sample points X_i:

$$\hat{Z}(X_0) = \sum_{i=1}^{n} \lambda_i \, . \, Z(X_i)$$

X_0
unknown value to be estimated

X_i
a set of known measurements



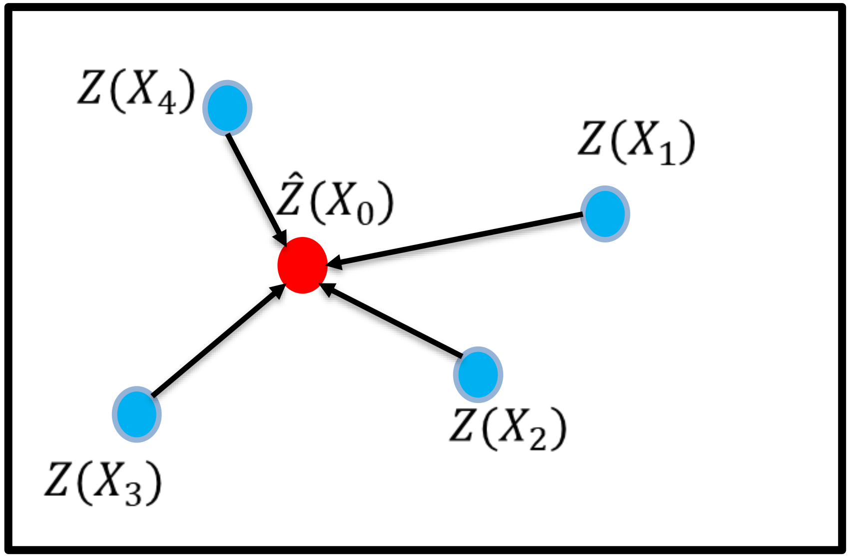
Ordinary Kriging

- The estimated value $\hat{Z}$ at a point X_0 is predicted as the weighted average of the values at all sample points X_i:

IDW

$$\hat{Z}(X_0) = \sum_{i=1}^{n} \lambda_i \, . \, Z(X_i)$$

$$\lambda_i = \frac{1/d_i}{\sum_{i=1}^{n}(1/d_i)}$$



X_0 unknown value to be estimated

X_i a set of known measurements

Ordinary Kriging

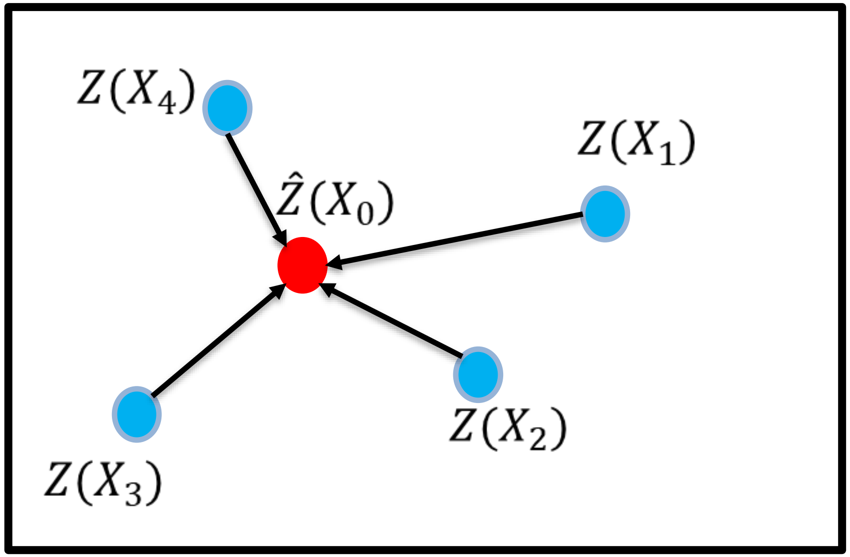
- The estimated value $\hat{Z}$ at a point X_0 is predicted as the weighted average of the values at all sample points X_i:

$$\hat{Z}(X_0) = \sum_{i=1}^{n} \lambda_i . Z(X_i)$$

Kriging weight

X_0 unknown value to be estimated

X_i a set of known measurements



Ordinary Kriging

- kriging tries to choose the optimal weights that produce the minimum estimation error.

$$E\left[\left(\hat{Z}(X_0) - Z(X_0)\right)^2\right]$$

- under the condition:

$$\sum_{i=1}^{n} \lambda_i = 1$$

Ordinary Kriging

- kriging tries to choose the optimal weights that produce the minimum estimation error.

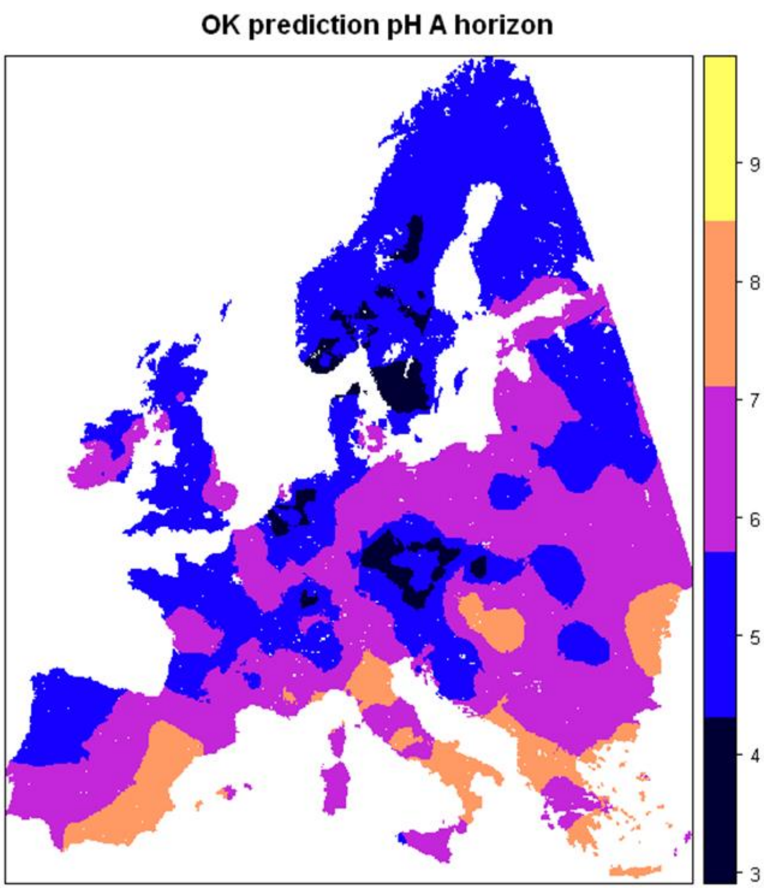
$$E\left[(\hat{Z}(X_0) - Z(X_0))^2\right]$$

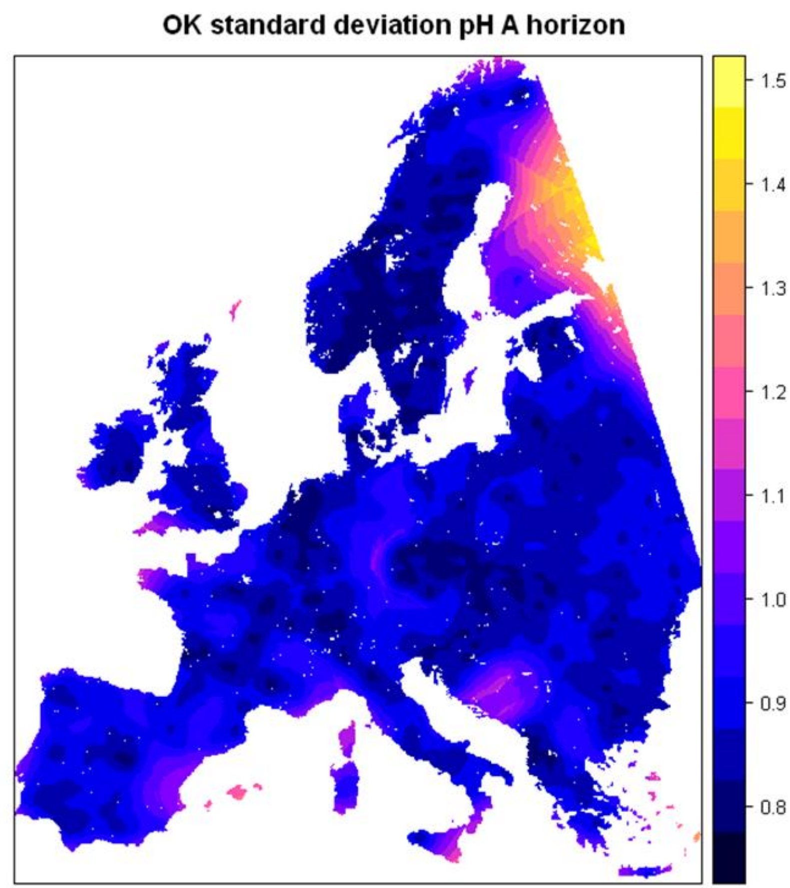
- under the condition:

$$\sum_{i=1}^{n} \lambda_i = 1$$

- The kriging variance at each point is automatically generated as part of the process of computing the weights.

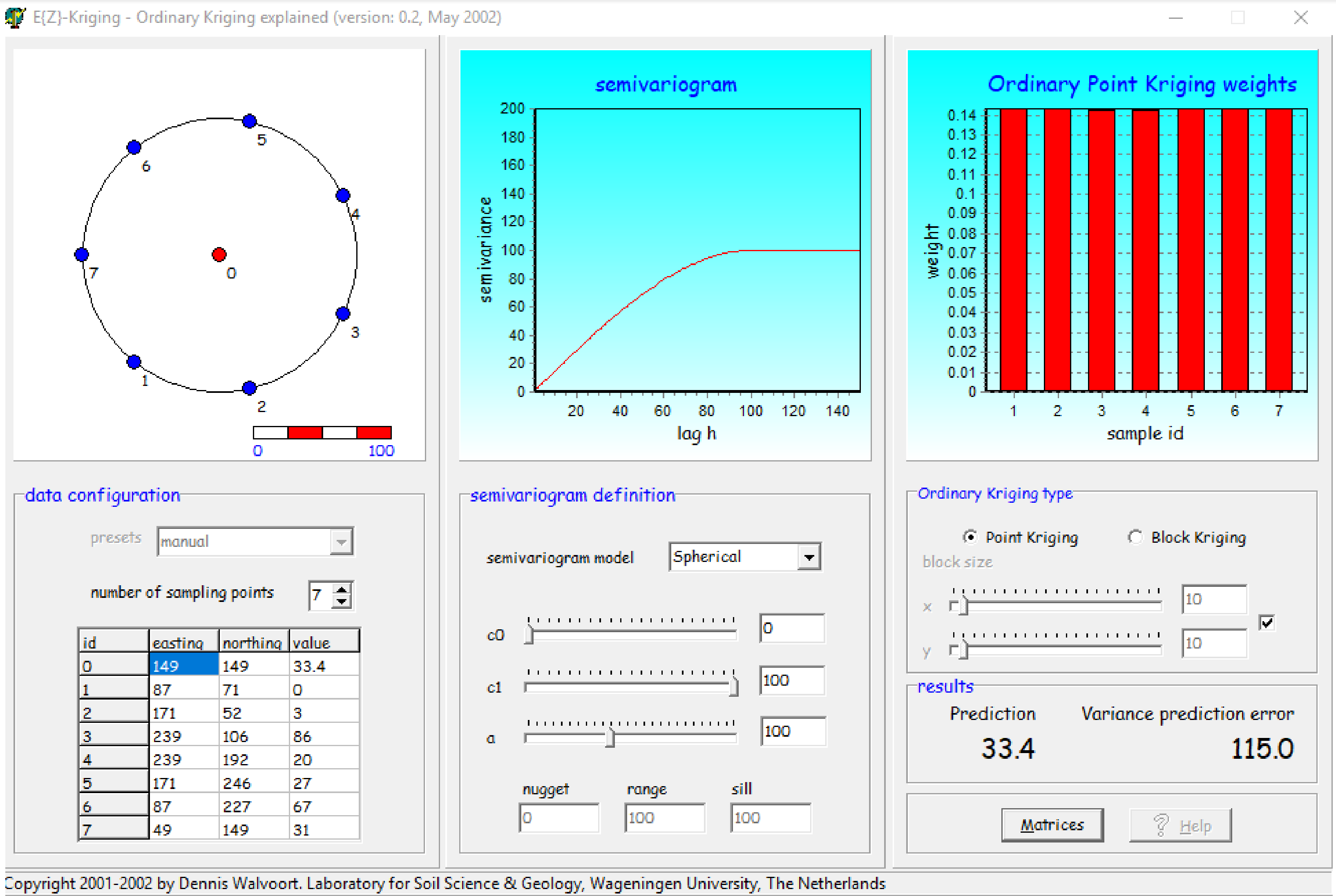
Ordinary Kriging: example



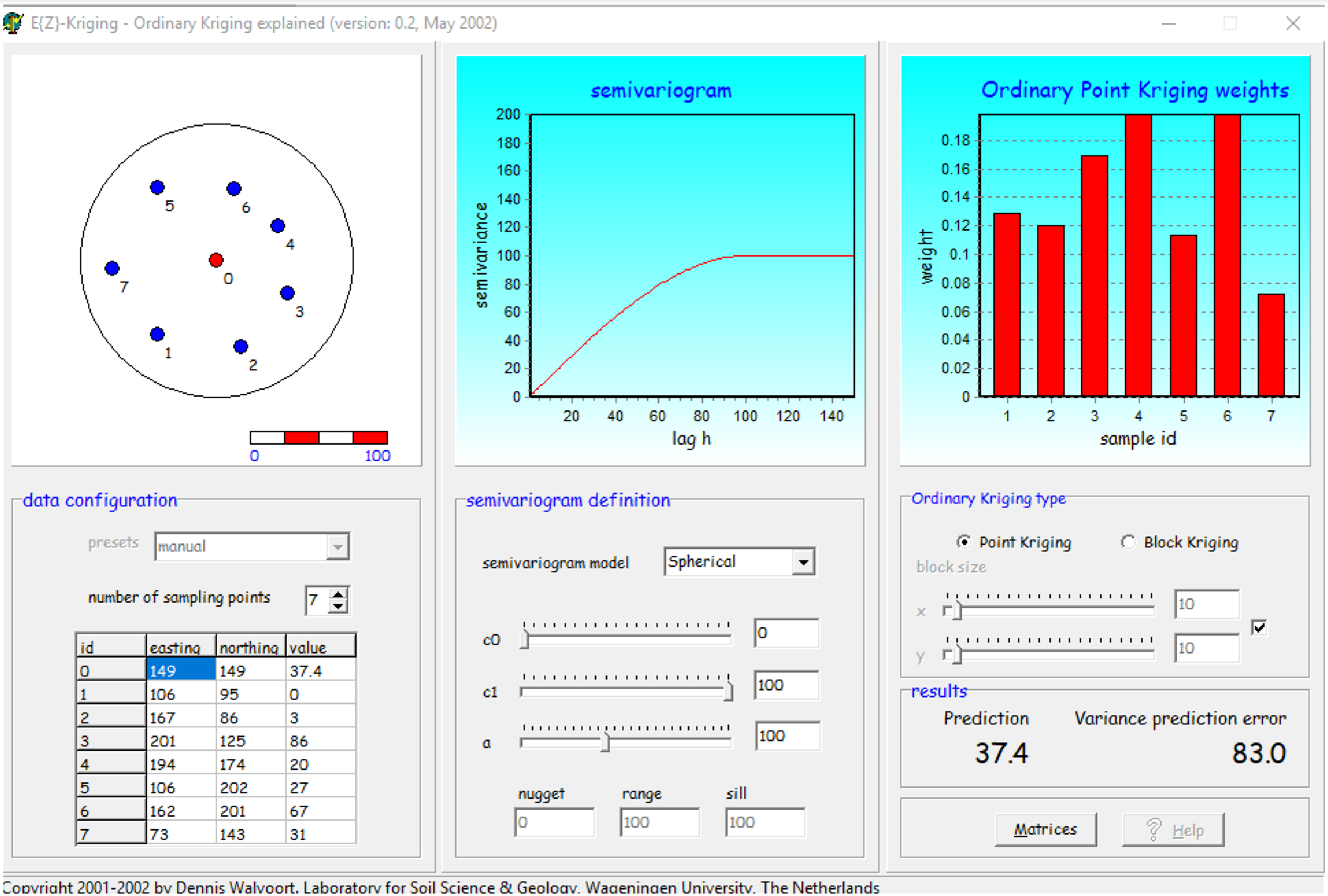


Estimation (left) and standard deviation (right) maps of pH by ordinary kriging.
(Source: Heuvelink, 2017, ISRIC)

E{Z}-Kriging


E{Z}-Kriging - Ordinary Kriging explained (version: 0.2, May 2002)
semivariogram
semivariance
lag h
Ordinary Point Kriging weights
weight
sample id
data configuration
presets manual
number of sampling points 7
id easting northing value
0 149 149 33.4
1 87 71 0
2 171 52 3
3 239 106 86
4 239 192 20
5 171 246 27
6 87 227 67
7 49 149 31
semivariogram definition
semivariogram model Spherical
c0 0
c1 100
a 100
nugget 0
range 100
sill 100
Ordinary Kriging type
Point Kriging
Block Kriging
block size
x 10
y 10
results
Prediction 33.4
Variance prediction error 115.0
Matrices
Help
Copyright 2001-2002 by Dennis Walvoort. Laboratory for Soil Science & Geology, Wageningen University, The Netherlands

E{Z}-Kriging



E{Z}-Kriging - Ordinary Kriging explained (version: 0.2, May 2002)
0
100
semivariogram
semivariance
lag h
Ordinary Point Kriging weights
weight
sample id
data configuration
presets
manual
number of sampling points
7
id
easting
northing
value
0 149 149 37.4
1 106 95 0
2 167 86 3
3 201 125 86
4 194 174 20
5 106 202 27
6 162 201 67
7 73 143 31
semivariogram definition
semivariogram model
Spherical
c0
c1
a
nugget
range
sill
Ordinary Kriging type
Point Kriging
Block Kriging
block size
results
Prediction
Variance prediction error
37.4
83.0
Matrices
Help
Copyright 2001-2002 by Dennis Walvoort, Laboratory for Soil Science & Geology, Wageningen University, The Netherlands

Regression Kriging

- Regression + kriging of residuals: is a hybrid spatial interpolation technique

- Linear regression is used to make a relationship between soil data and covariates

- Kriging is used to model the residuals

Regression Kriging

- Regression + kriging of residuals: is a hybrid spatial interpolation technique

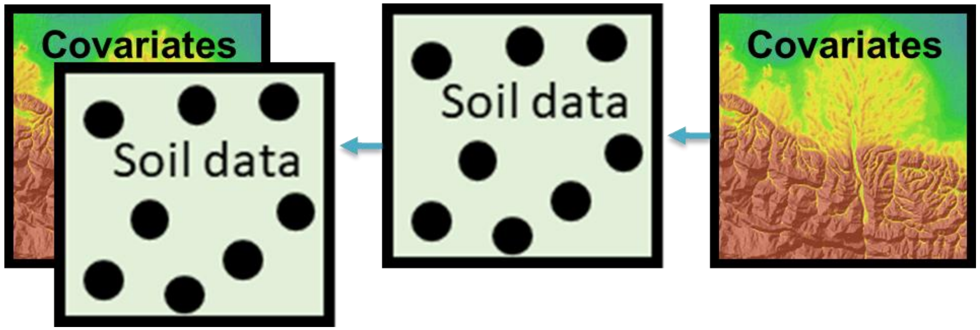
- Linear regression is used to make a relationship between soil data and covariates

- Kriging is used to model the residuals

$$\underbrace{\boxed{S}}_{\text{Soil data}} = \underbrace{\boxed{f}}_{\text{Model}}\underbrace{\boxed{(s,\ c,\ o,\ r,\ p,\ a,\ n)}}_{\text{Covariates}} + \underbrace{\varepsilon}_{\text{spatially dependent residuals}}$$

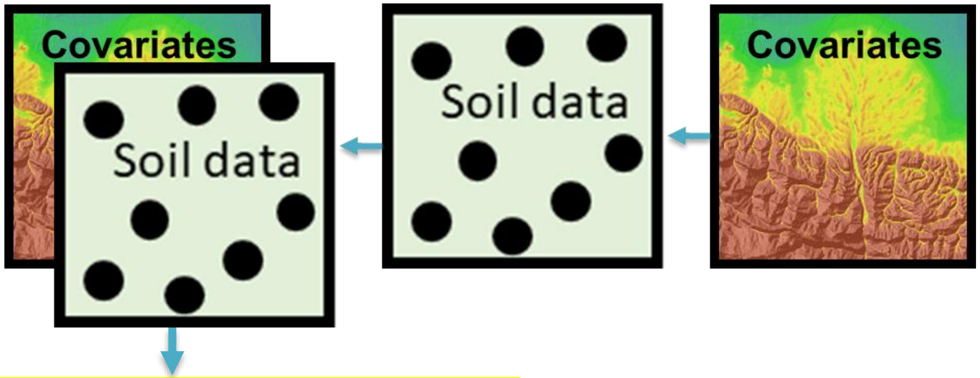
Regression Kriging: workflow

Overlay soil data and covariates



Regression Kriging: workflow

Overlay soil data and covariates



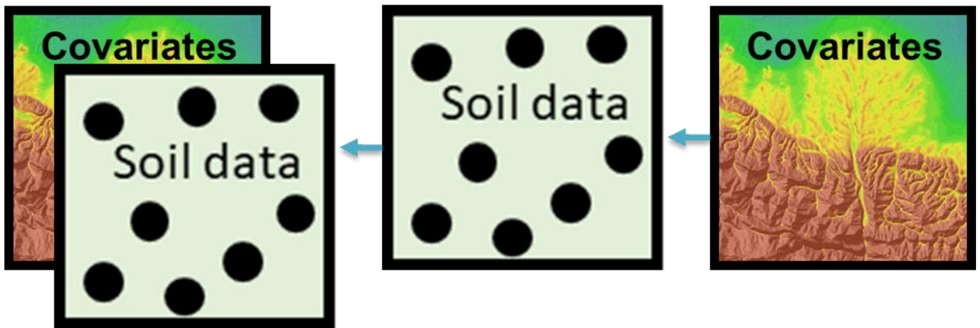
Samples	X	Y	Clay	Elevation
1	765342	2658345	**69**	**76**
2	765269	2695483	**12**	**15**

Tabulate soil data and covariates
and build a linear regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Regression Kriging: workflow

Overlay soil data and covariates



Samples	X	Y	Clay	Elevation
1	765342	2658345	**69**	**76**
2	765269	2695483	**12**	**15**

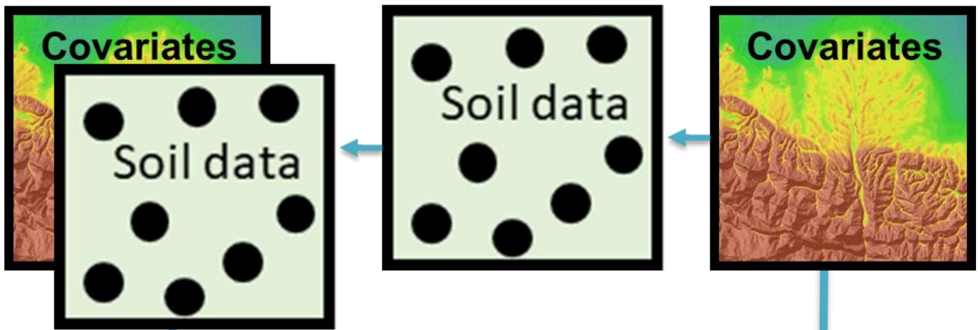
Soil clay = 0.024 + 0.83 × elevation

Tabulate soil data and covariates
and build a linear regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Regression Kriging: workflow

Overlay soil data and covariates



Samples	X	Y	Clay	Elevation
1	765342	2658345	**69**	**76**
2	765269	2695483	**12**	**15**

Tabulate soil data and covariates
and build a linear regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Soil clay = 0.024 + 0.83 × elevation

apply the model to all unobserved locations

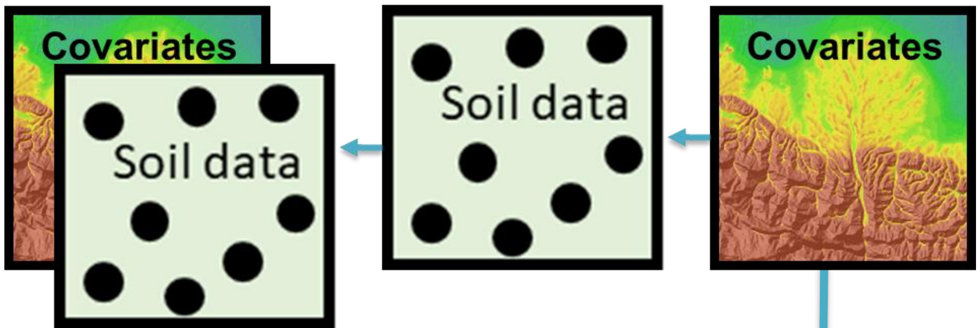
Kriging of residuals

Map 2

Map 1

Regression Kriging: workflow

Overlay soil data and covariates



Samples	X	Y	Clay	Elevation
1	765342	2658345	**69**	**76**
2	765269	2695483	**12**	**15**

Tabulate soil data and covariates
and build a linear regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_1 X_1 + \beta_2 X_2 \ + \cdots \ + \beta_n X_n$$

Soil clay = 0.024 + 0.83 × elevation

apply the model to all unobserved locations

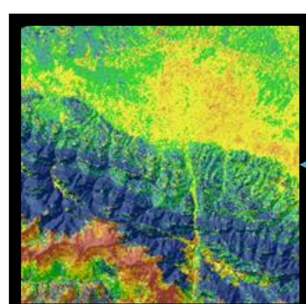
Kriging of residuals

Map 2

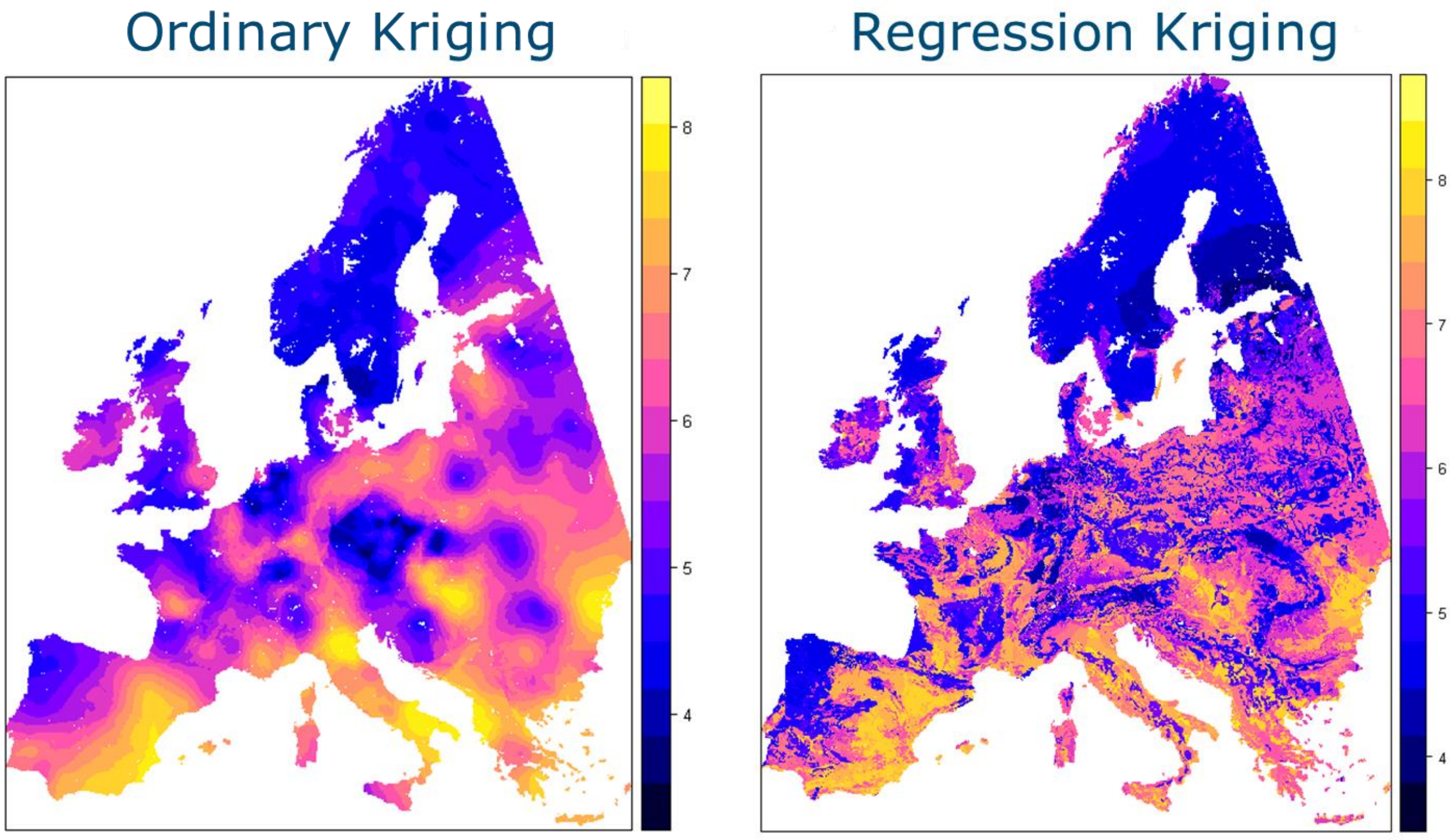
Map 1

add up the two maps

Regression Kriging map



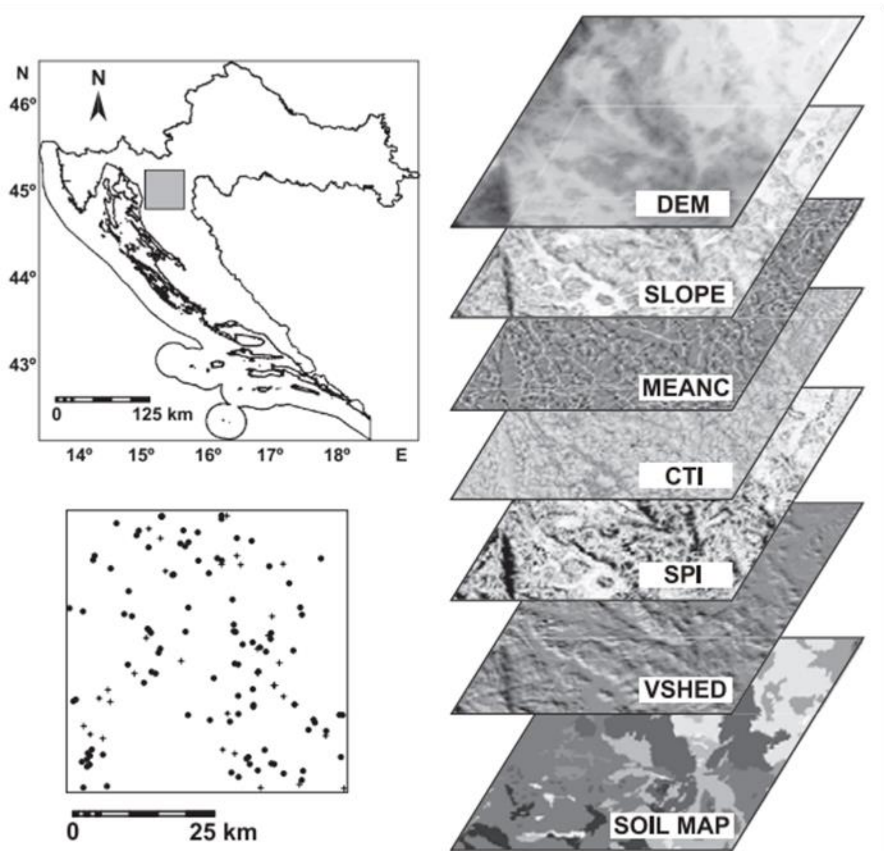
Regression Kriging: example I



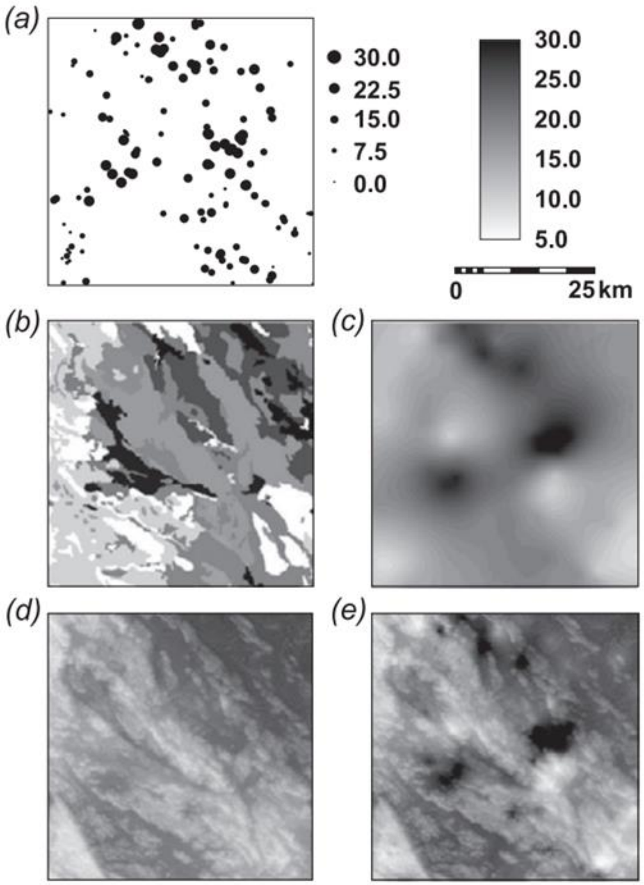
Estimation maps of pH by ordinary kriging (left) and regression kriging (right) .
(Source: Heuvelink, 2017, ISRIC)

Regression Kriging: example II

Hengl et al., 2004



Location of the study area (upper-left), profiles used for interpolation (●) and validation (+) (lower-left) and maps of predictors (right).



Topsoil thickness (DEPTH) measured at 135 locations (a) comparison of predictions made by using: soil map only (b), ordinary kriging (c), plain regression (d) and regression-kriging (e).

Practice

Import point data

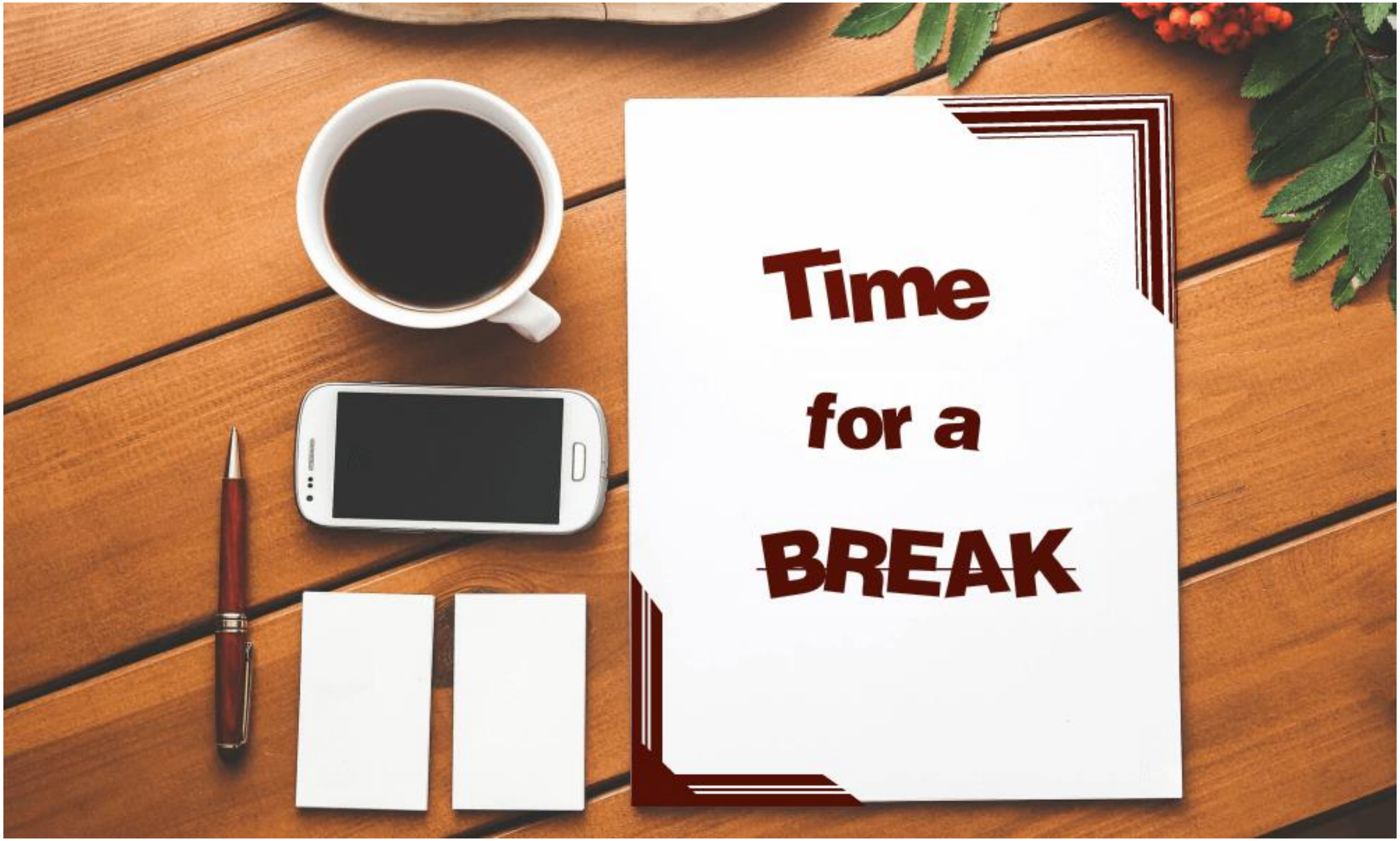
Variogram analysis

Kriging predicting

IDW prediction

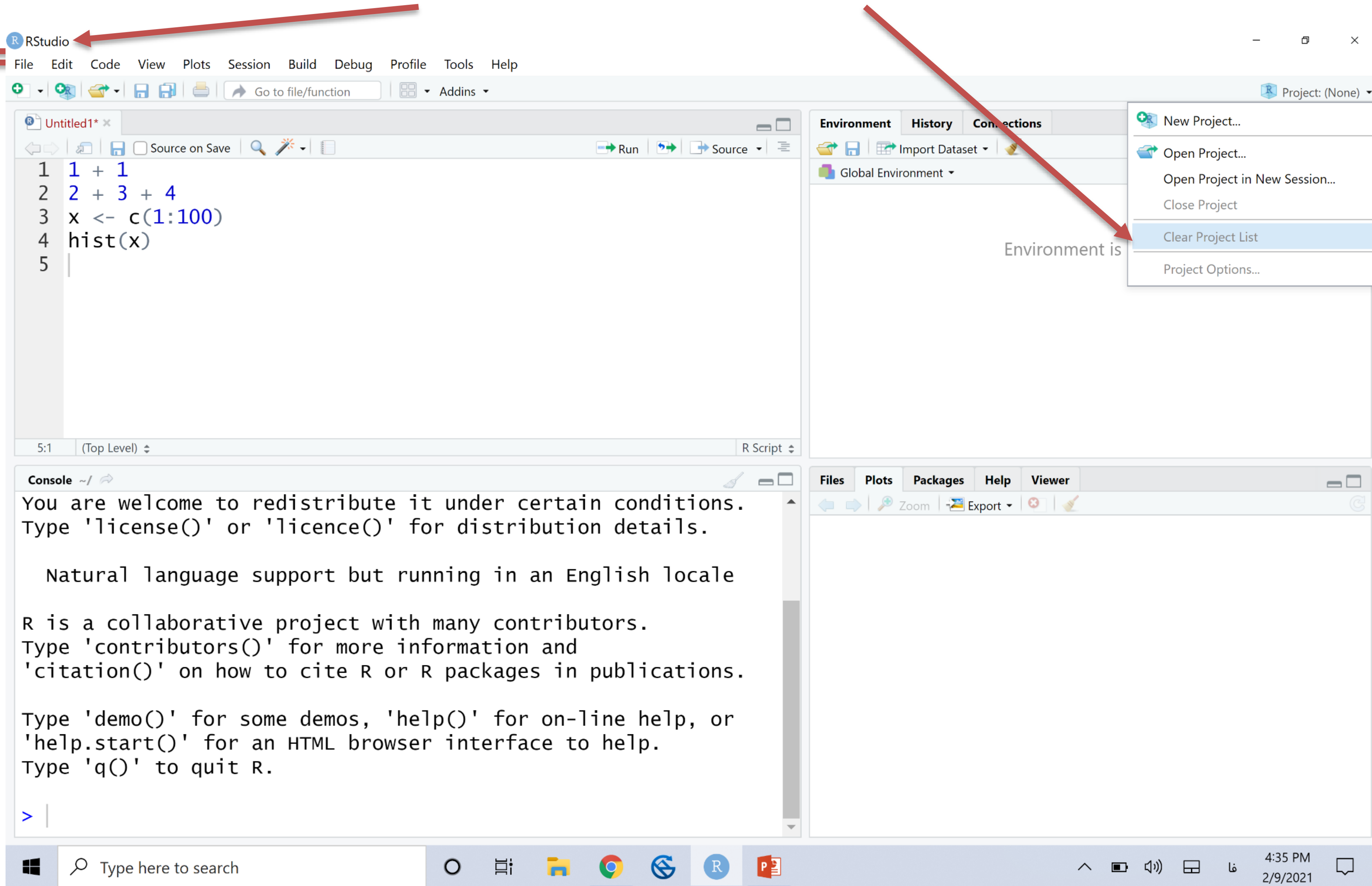
Regression Kriging=random forest + res kriged

spatial stochastic simulations

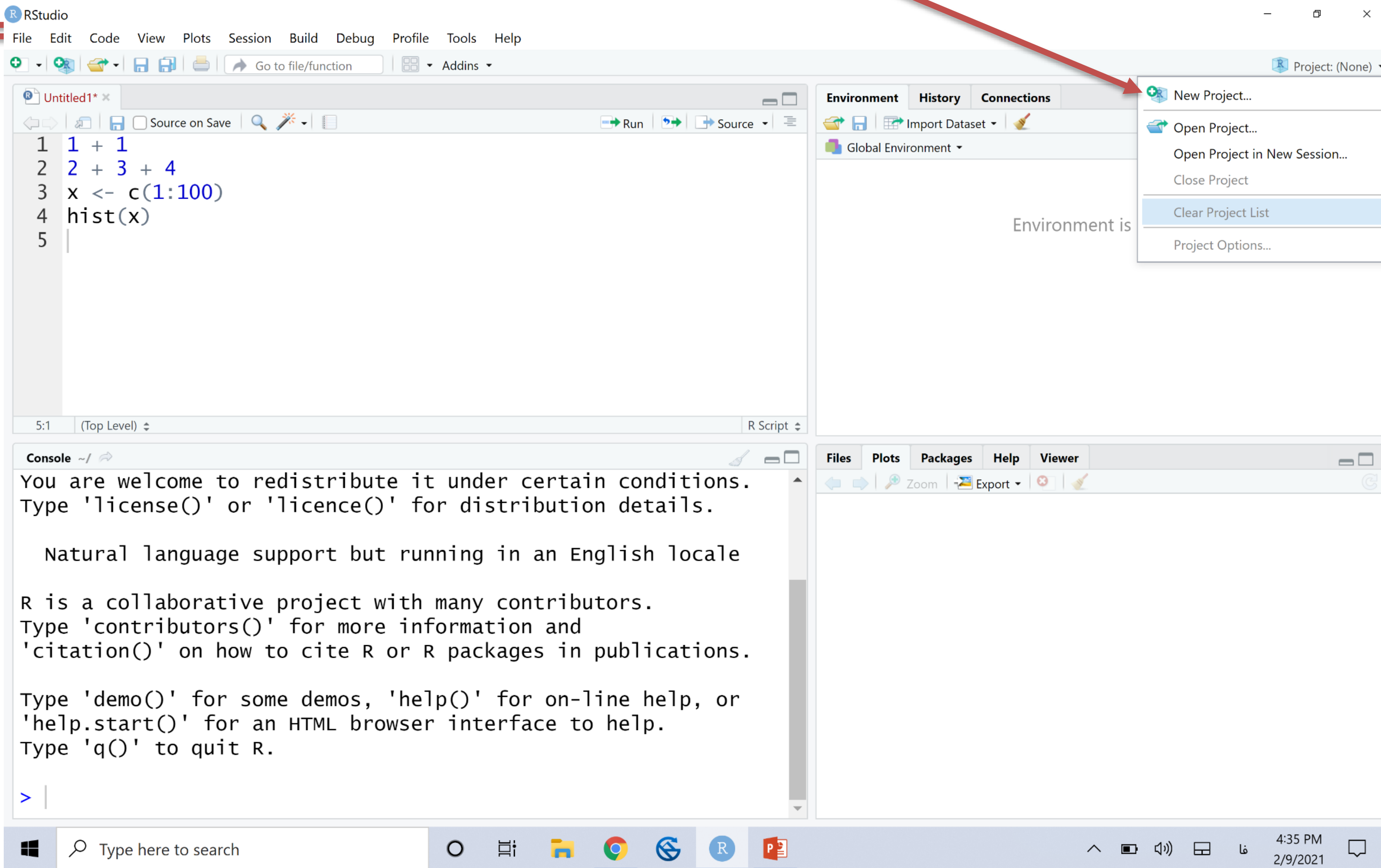

Time
for a
BREAK

1. Open Rstudio

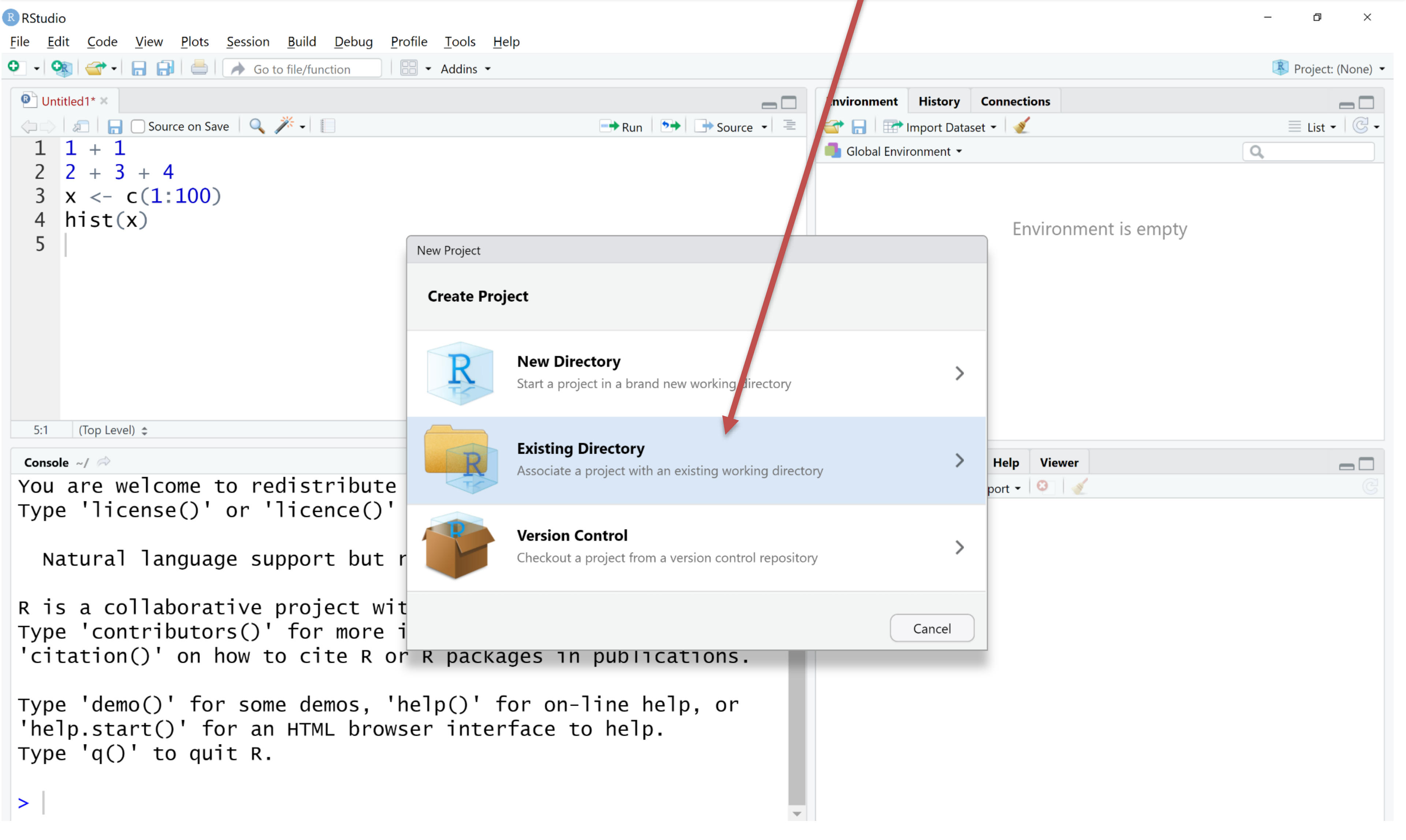
2. Clear Project List



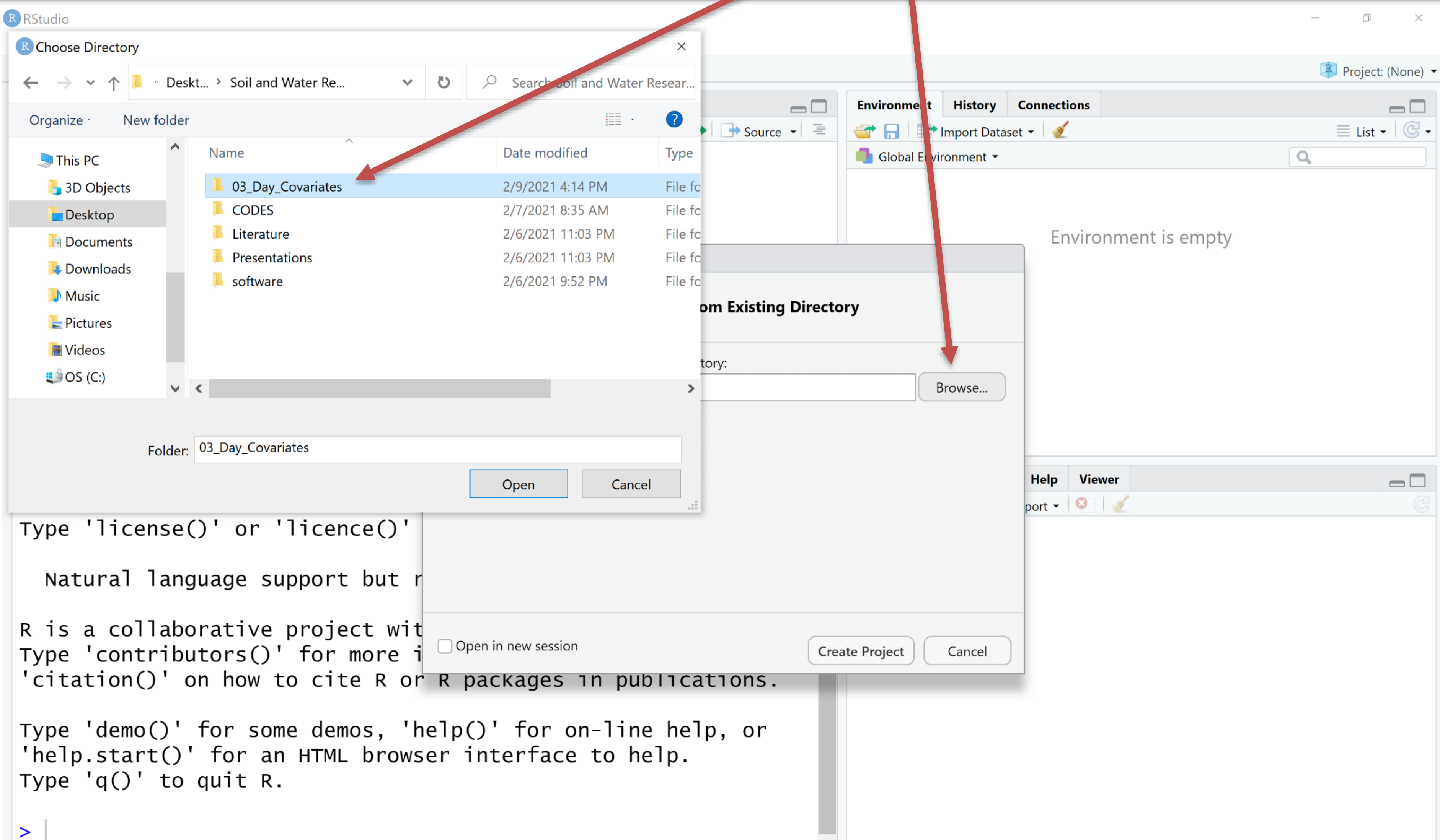
3. New Project



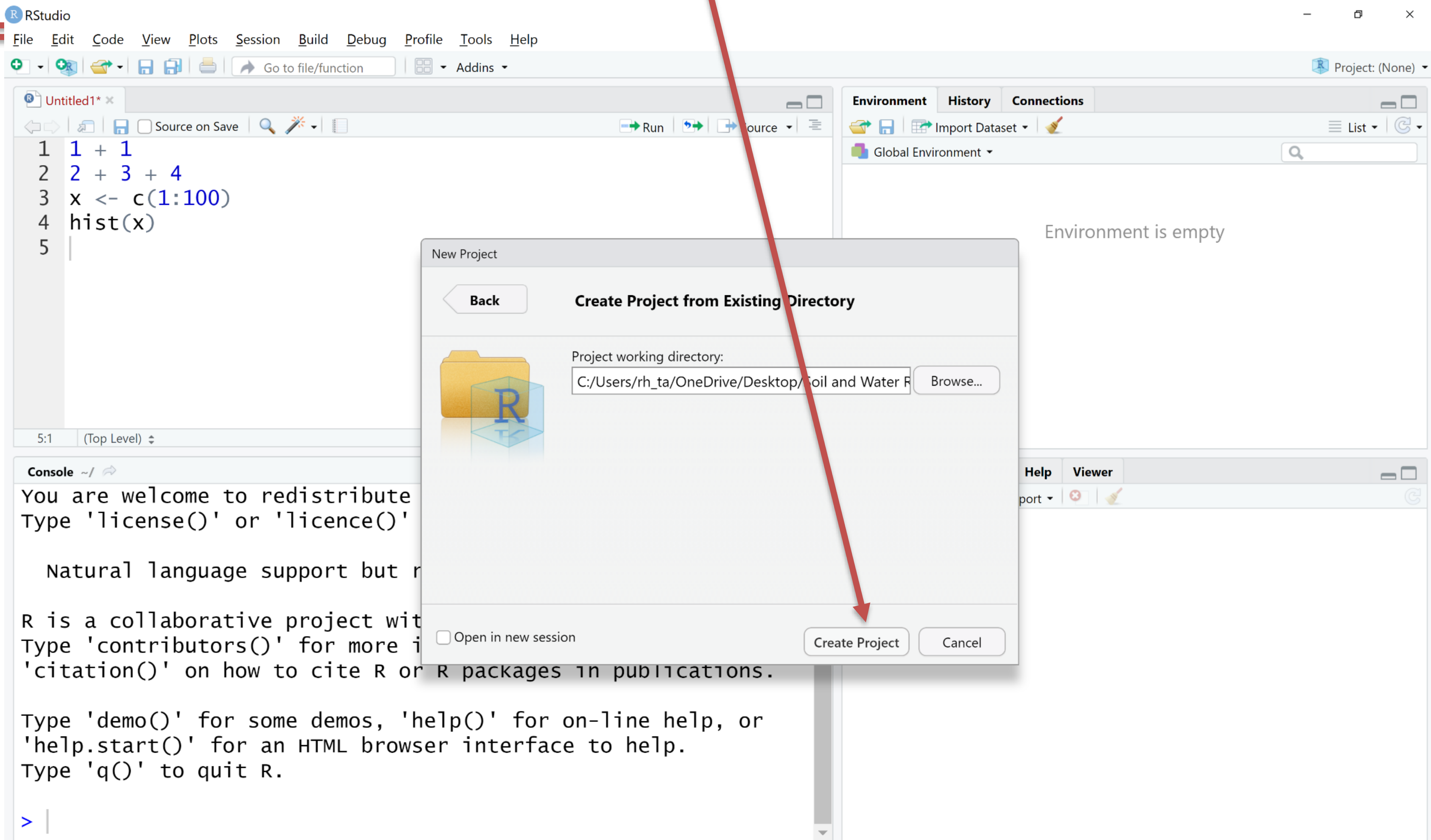
4. Existing Directory



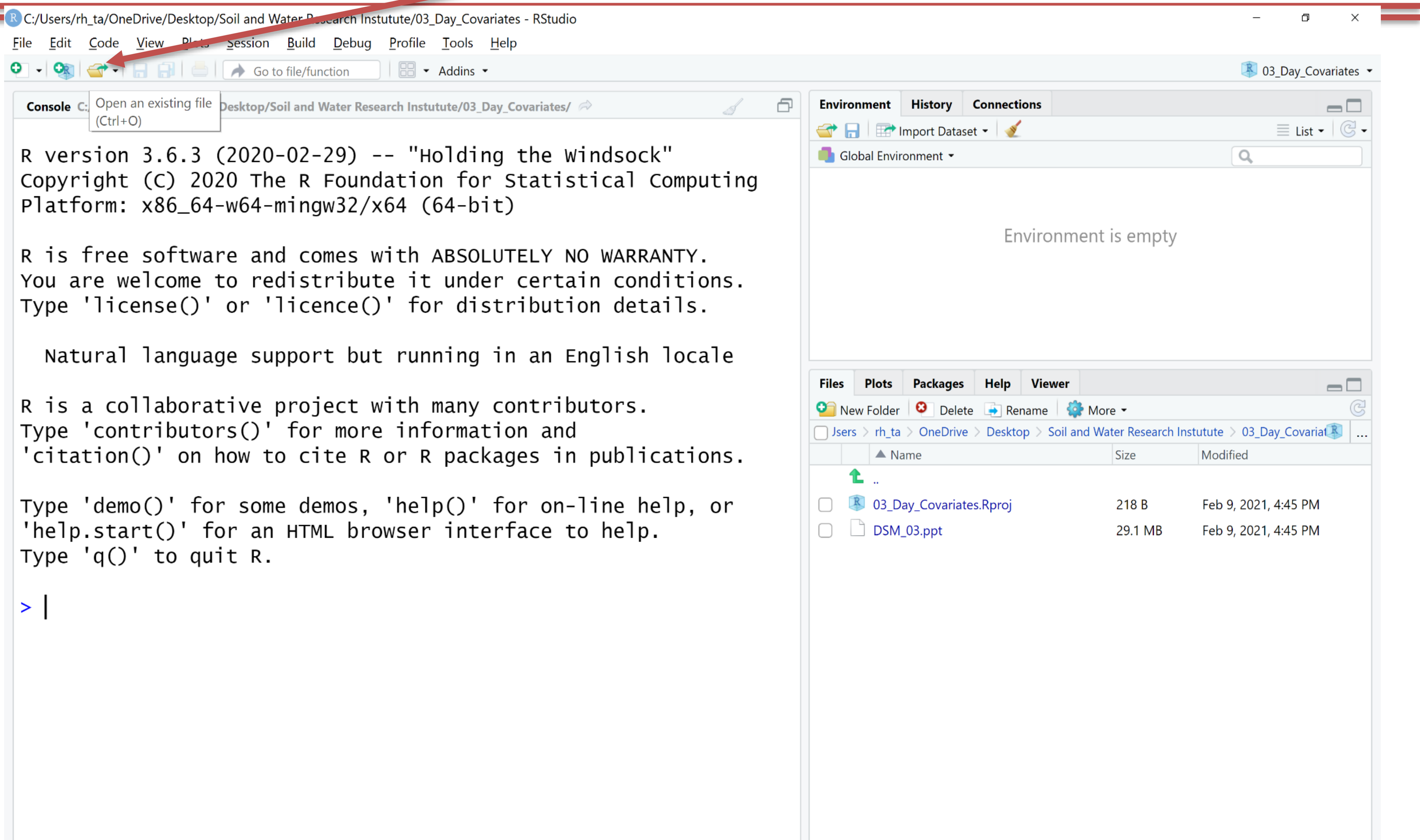
5. Find the folder and find **Day_06_Geostatistics** and open



6. Create project



7. Open an existing file and find **06_R_Cov.R**



8. Open **06_R_Cov.R**