

There are 76 contributors for the current view.

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**

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;

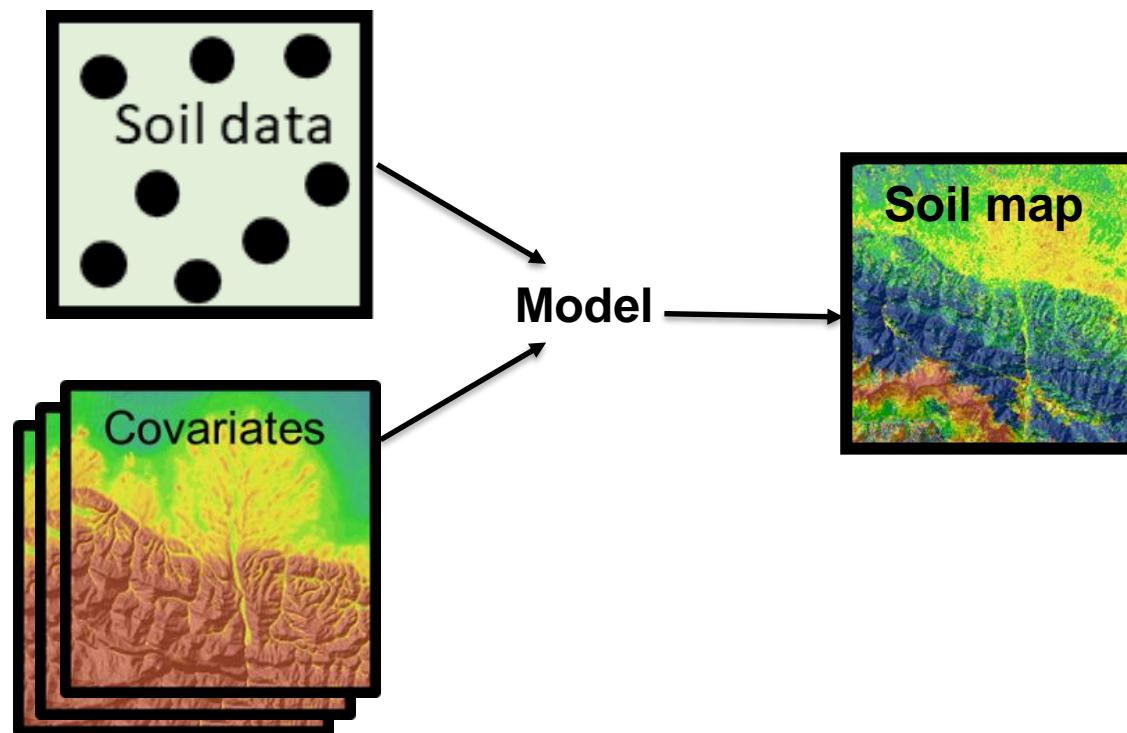
ε : auto-correlated random spatial variation.

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

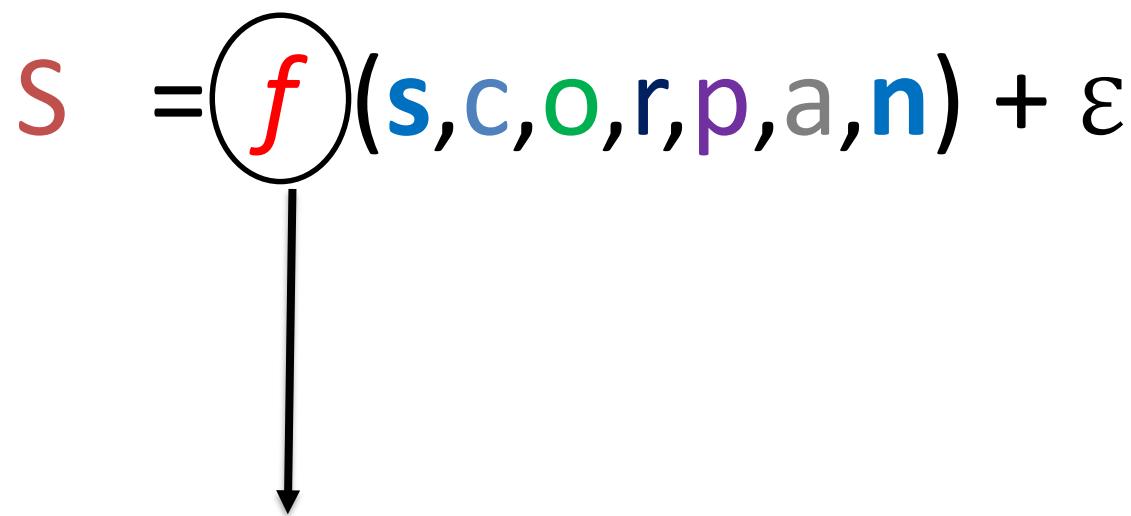
SCORPAN model

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

Soil data Model Covariates spatially dependent residuals



SCORPAN Model

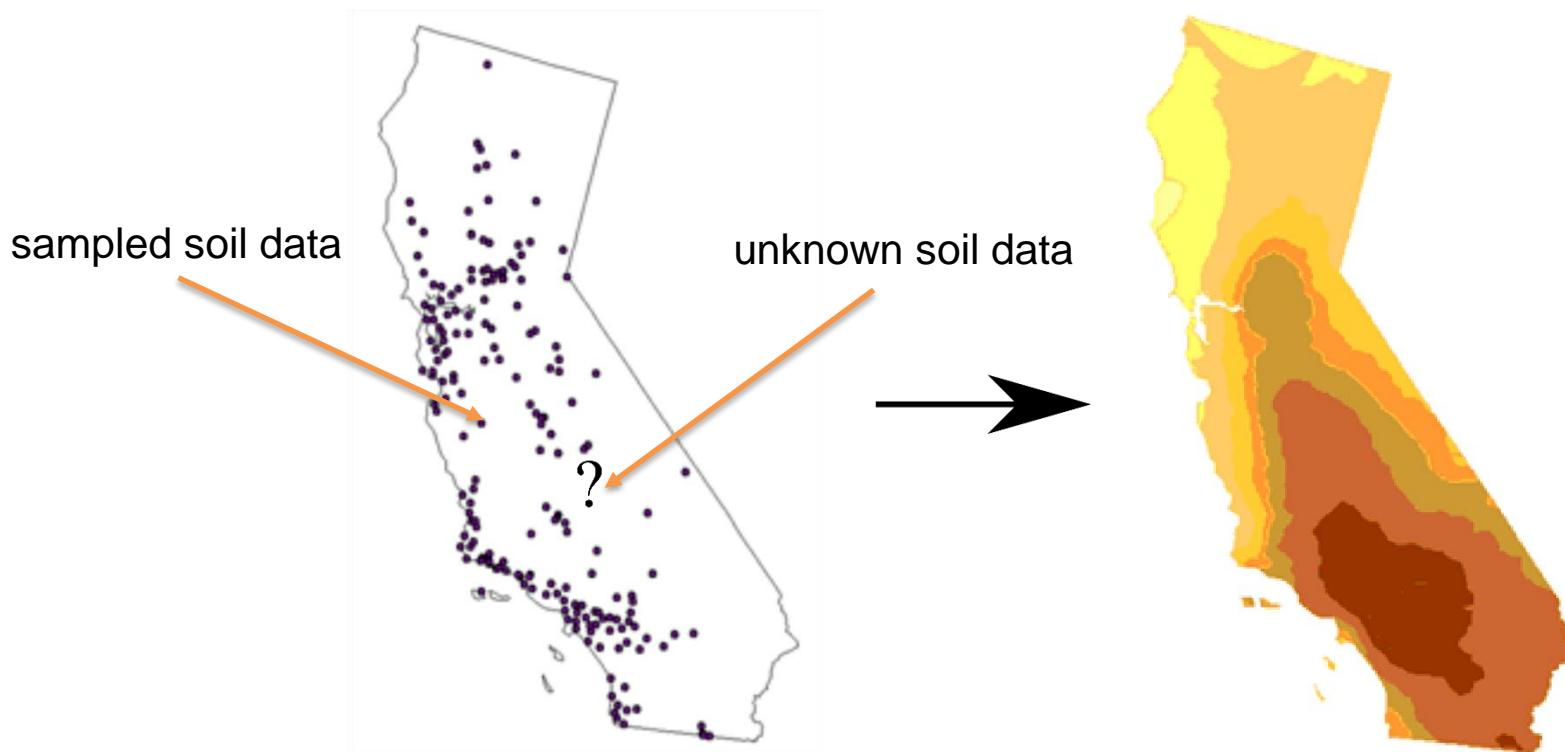
$$S = f(s, c, o, r, p, a, n) + \varepsilon$$


Machine learning

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

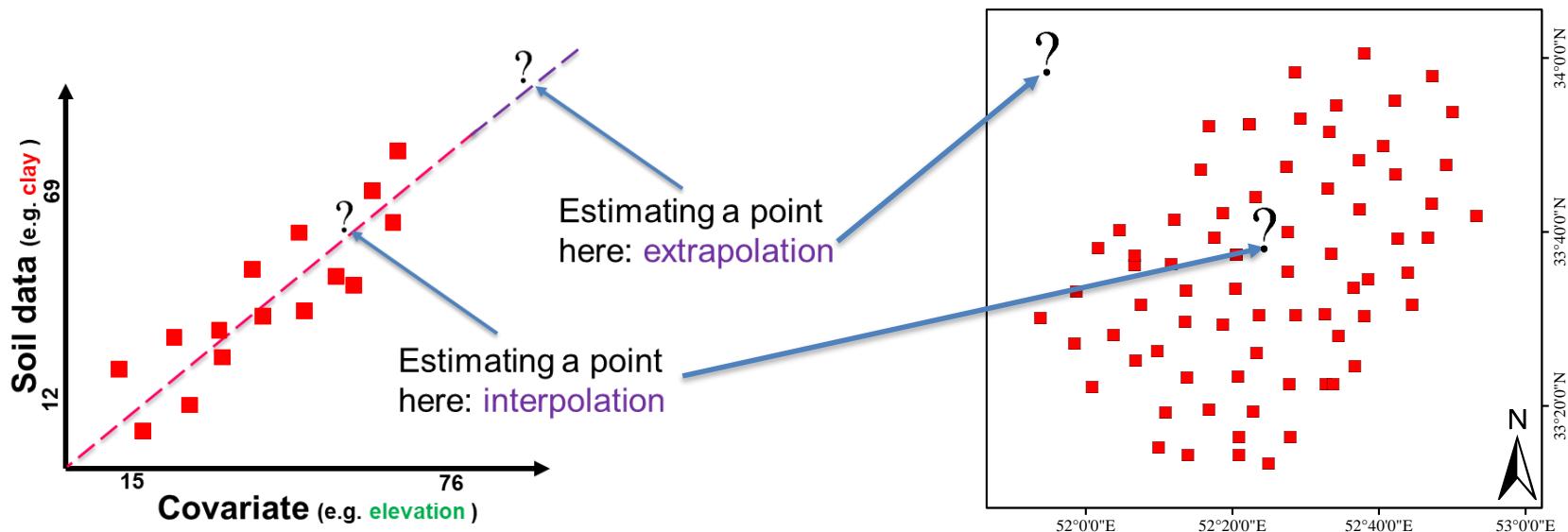
Interpolation vs. Extrapolation

Interpolation

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

Extrapolation

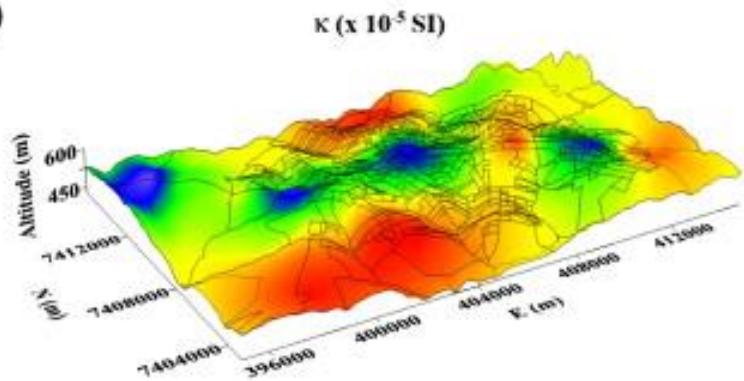
- Estimating the attribute values of locations outside the range of available data using known data values



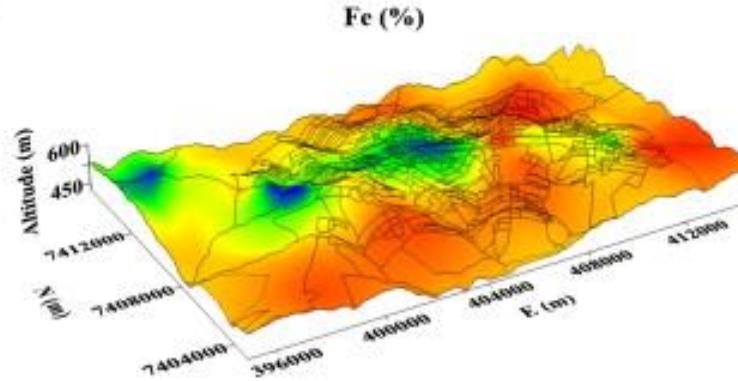
We Know,

- Soils and soil properties vary in space

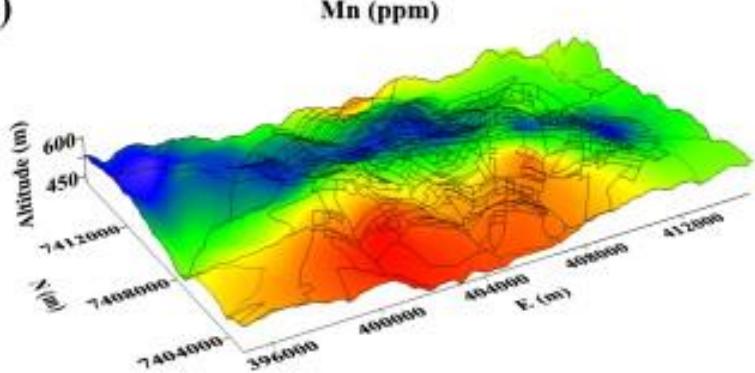
a)



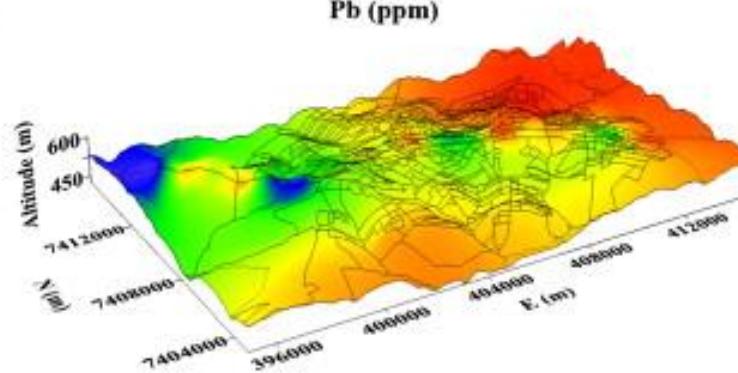
b)



c)



d)



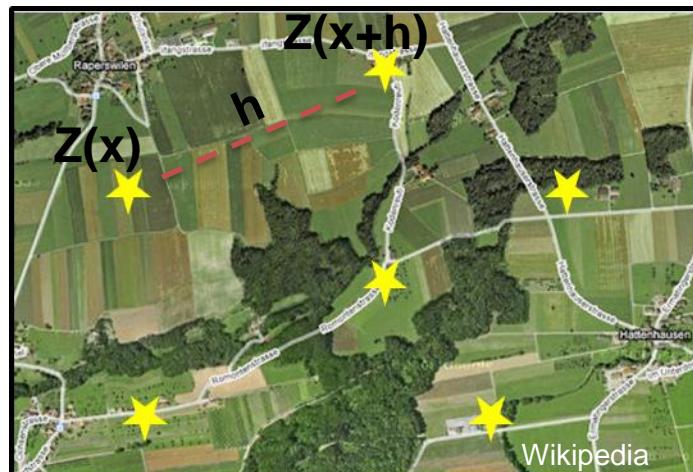
The Variogram

The Central Tool of Geostatistics

- Spatial variation can be quantified using the so-called **semi-variance**
- Semi-variance is calculated according to:

$$\gamma(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{z(x_i) - z(x_i + h)\}^2$$

measurement at location x
measurement at location $x+h$



The Variogram

The Central Tool of Geostatistics

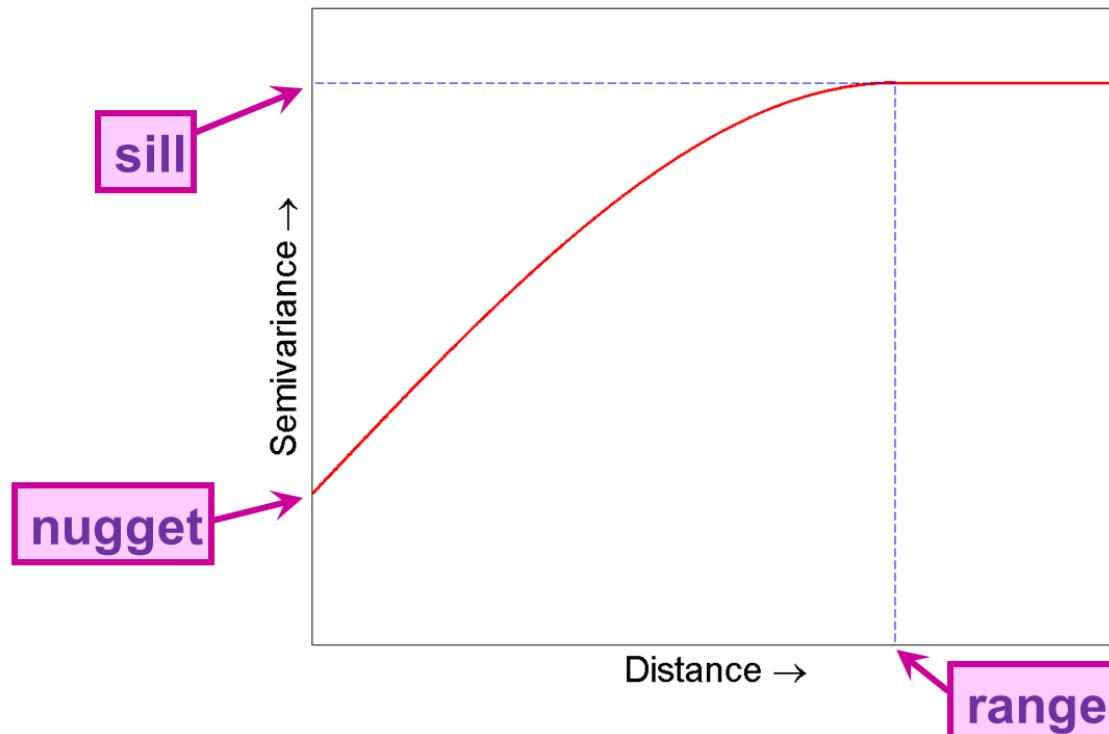
- Plot of semi-variance as a function of the distance is called a semi-variogram



The Variogram

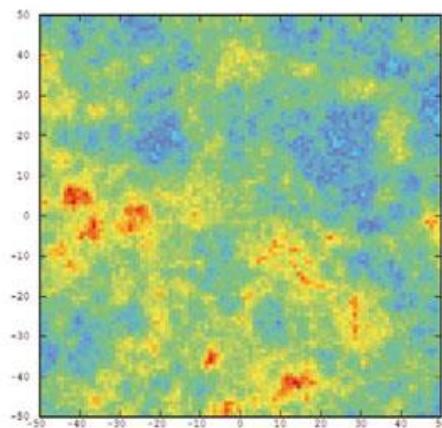
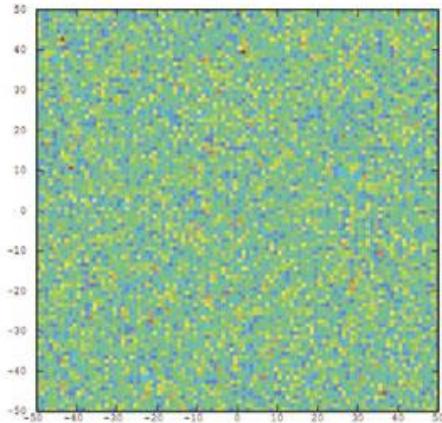
The Central Tool of Geostatistics

- 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



The Variogram

The Central Tool of Geostatistics

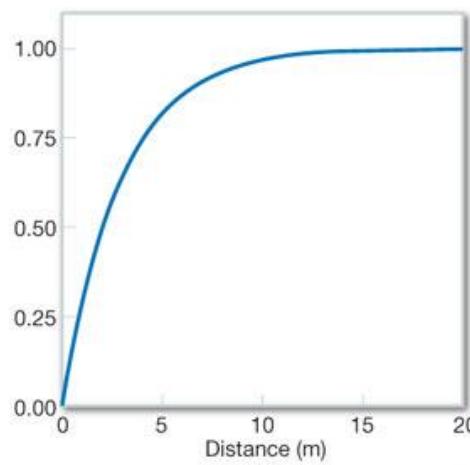
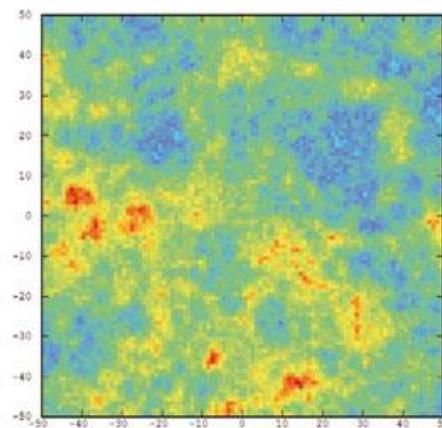
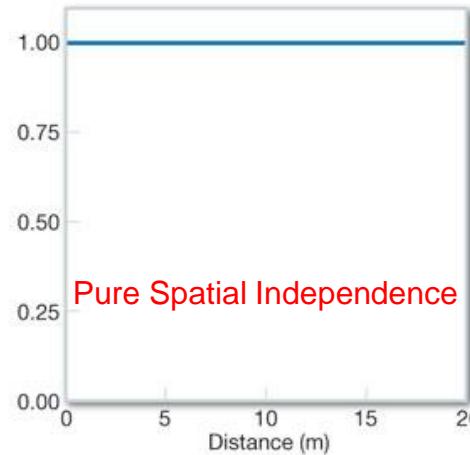
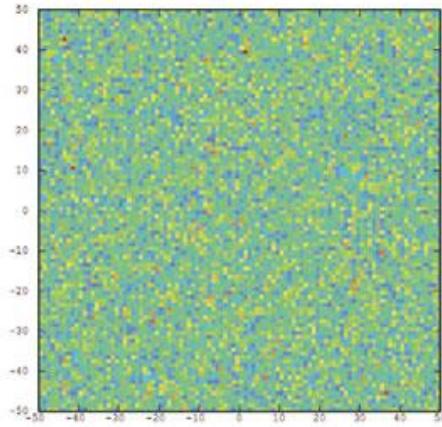


Two phenomena with the same statistical distribution (at left) but with significantly different spatial structures (corresponding variograms at right). Source:

<https://www.neimagazine.com/>

The Variogram

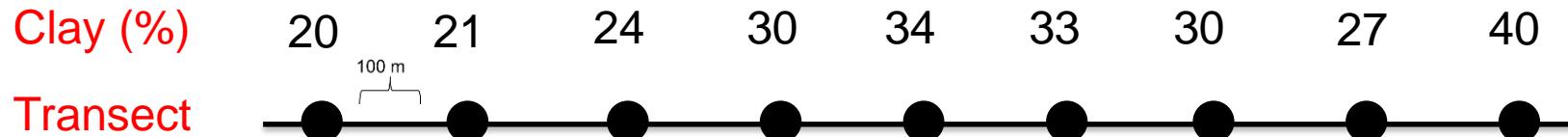
The Central Tool of Geostatistics



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

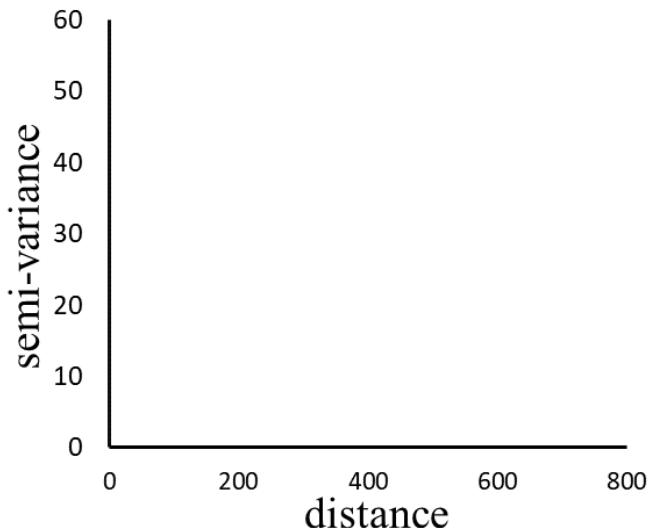
<https://www.neimagazine.com/>

Calculate Variogram: example

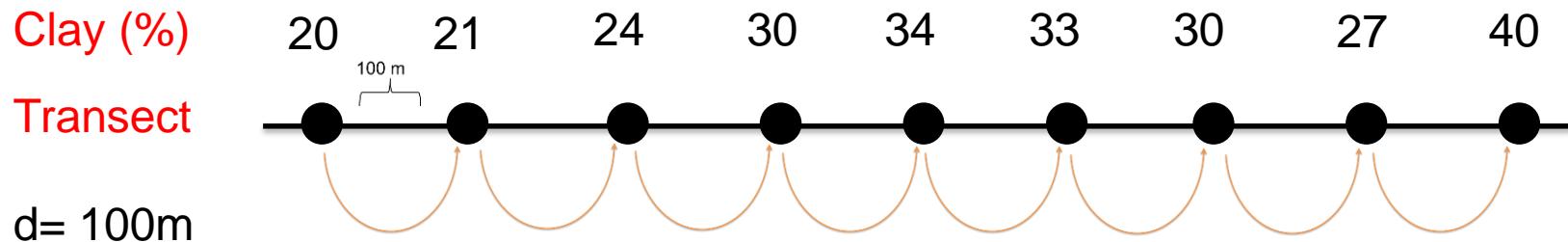


$$\gamma(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{z(x_i) - z(x_i + h)\}^2$$

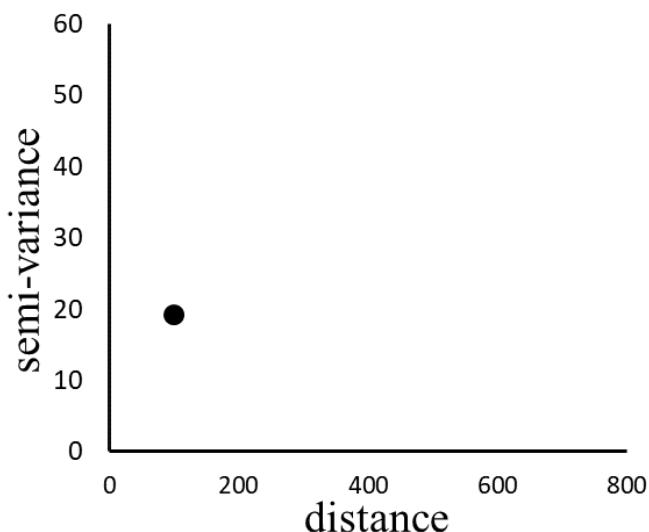
measurement at location x
measurement at location $x+h$



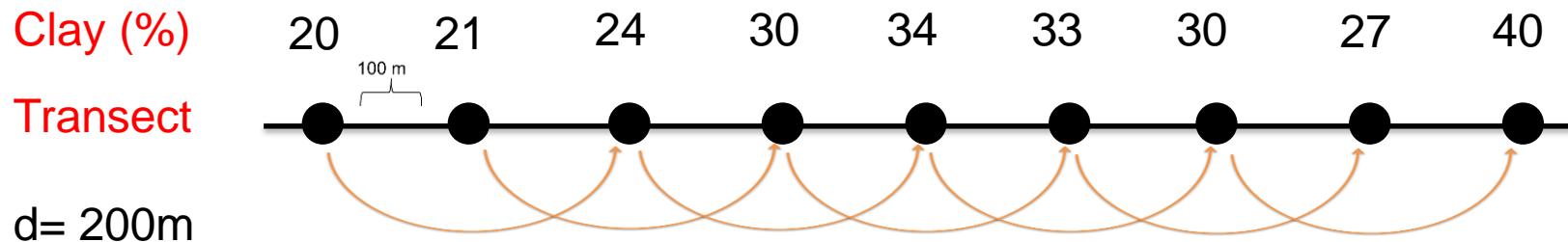
Calculate Variogram: example



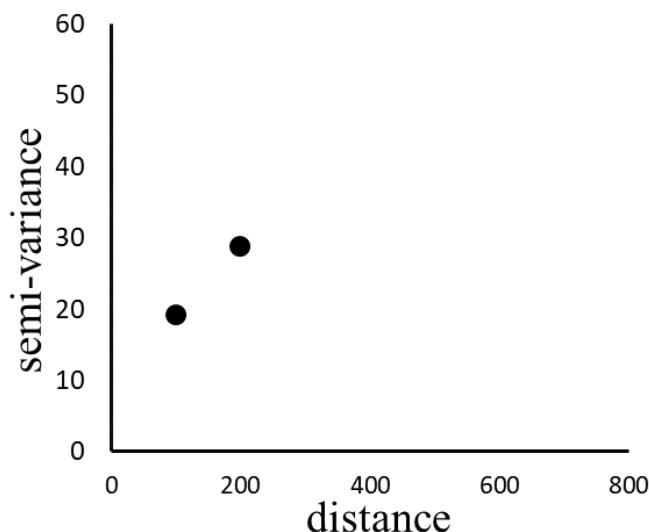
$$\gamma(100) = \frac{1}{16} \left[(20 - 21)^2 + (21 - 24)^2 + (24 - 30)^2 + (30 - 34)^2 + (34 - 33)^2 + (33 - 30)^2 + (30 - 27)^2 + (27 - 40)^2 \right] = 19.12$$



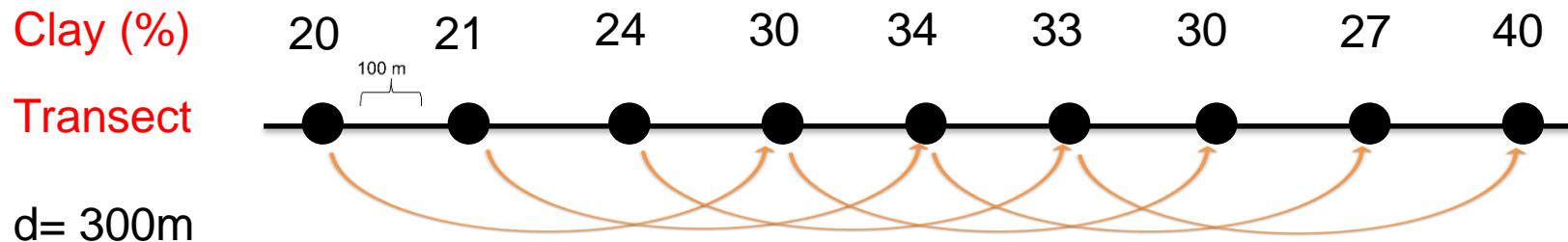
Calculate Variogram: example



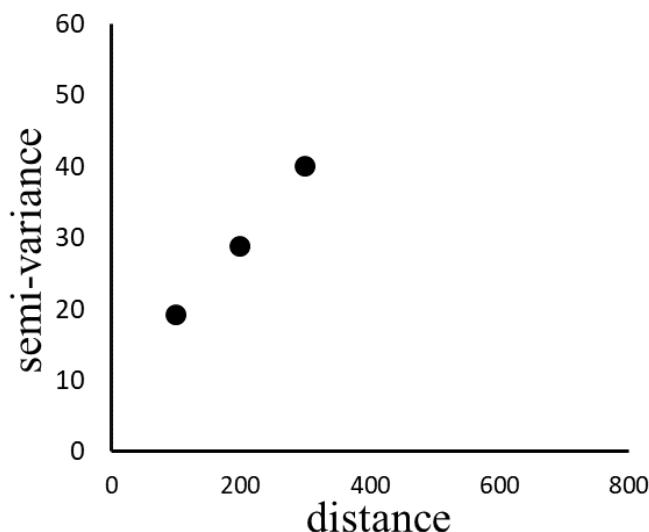
$$\gamma(200) = \frac{1}{14} \left[(20 - 24)^2 + (21 - 30)^2 + (24 - 34)^2 + (30 - 33)^2 + (34 - 30)^2 + (33 - 27)^2 + (30 - 40)^2 \right] = 28.71$$



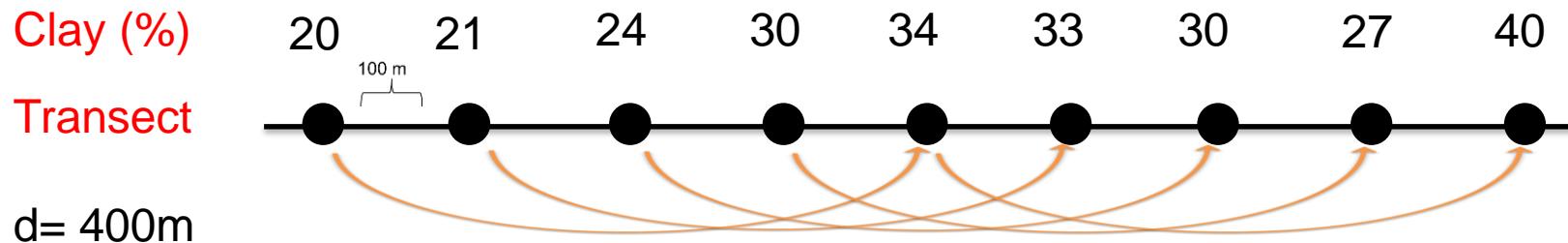
Calculate Variogram: example



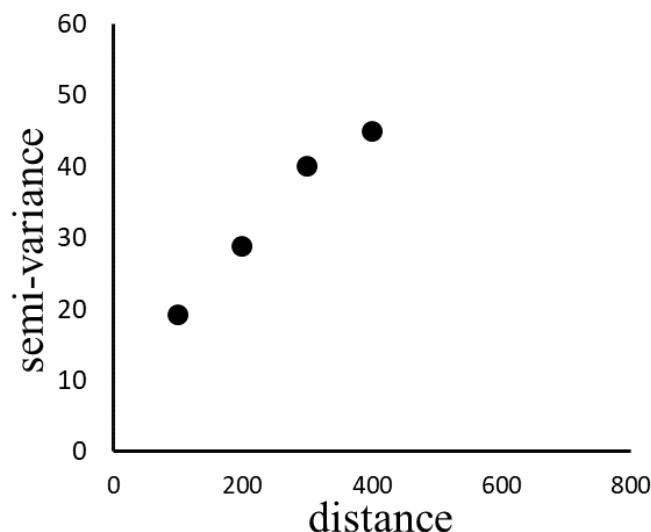
$$\gamma(300) = \frac{1}{12} \left[(20 - 30)^2 + (21 - 34)^2 + (24 - 33)^2 + (30 - 30)^2 + (34 - 27)^2 + (33 - 40)^2 \right] = 40.00$$



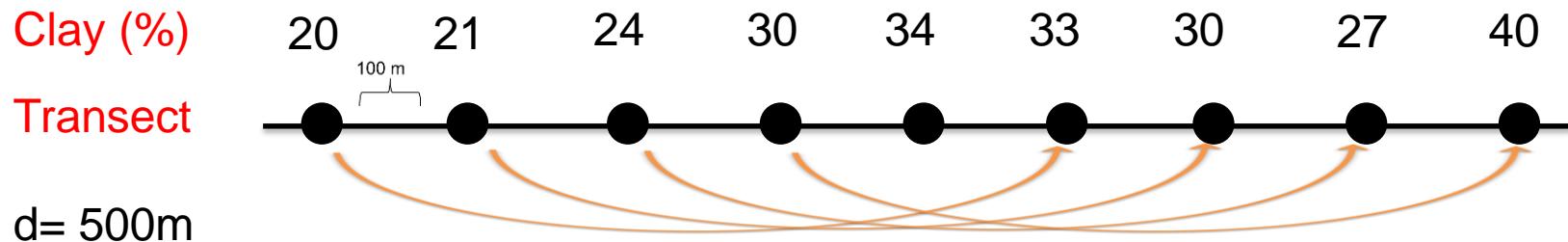
Calculate Variogram: example



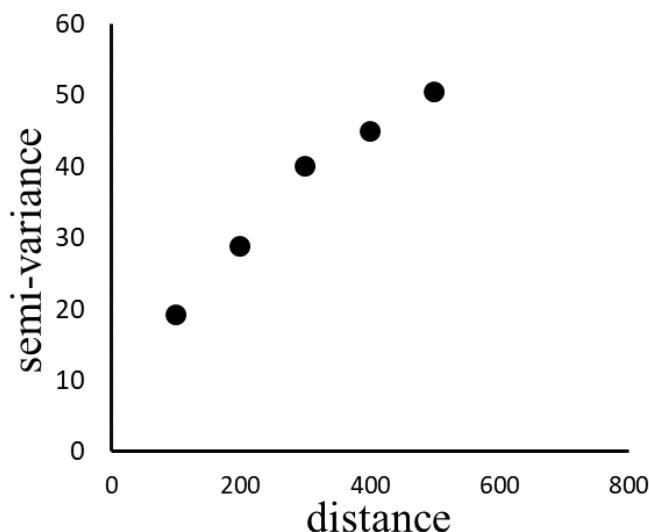
$$\gamma(400) = \frac{1}{10} \left[(20 - 34)^2 + (21 - 33)^2 + (24 - 30)^2 + (30 - 27)^2 + (34 - 40)^2 \right] = 44.90$$



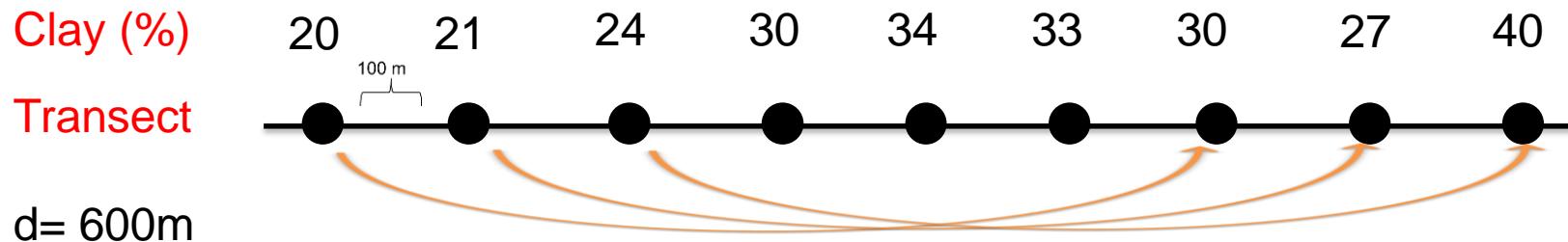
Calculate Variogram: example



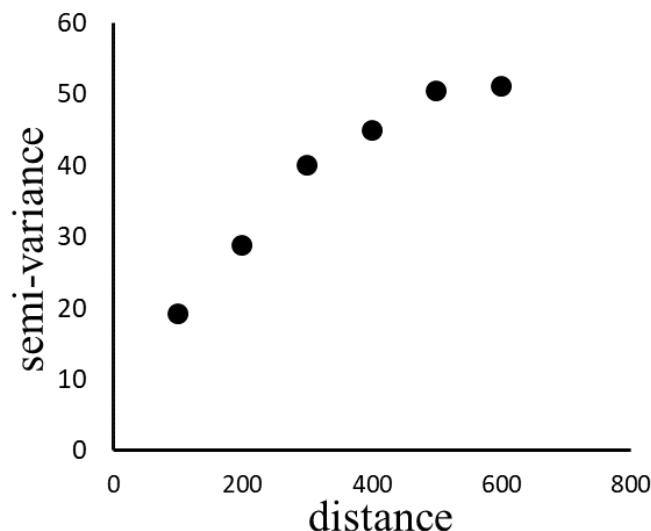
$$\gamma(500) = \frac{1}{8} [(20 - 33)^2 + (21 - 30)^2 + (24 - 27)^2 + (30 - 40)^2] = 50.37$$



Calculate Variogram: example

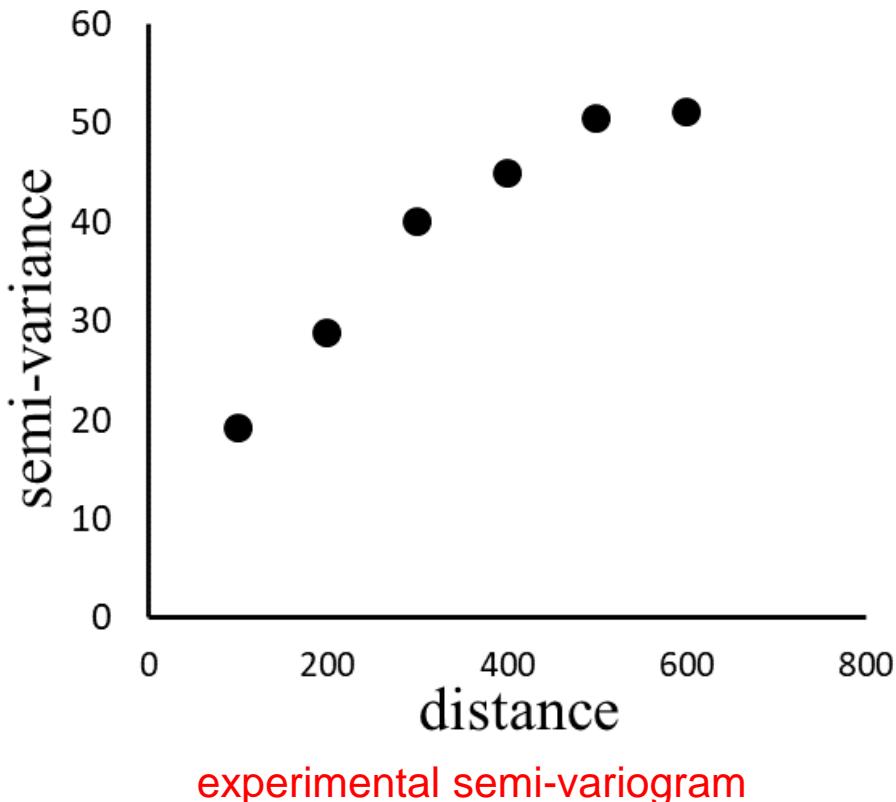


$$\gamma(600) = \frac{1}{6} [(20 - 30)^2 + (21 - 27)^2 + (24 - 40)^2] = 50.37$$



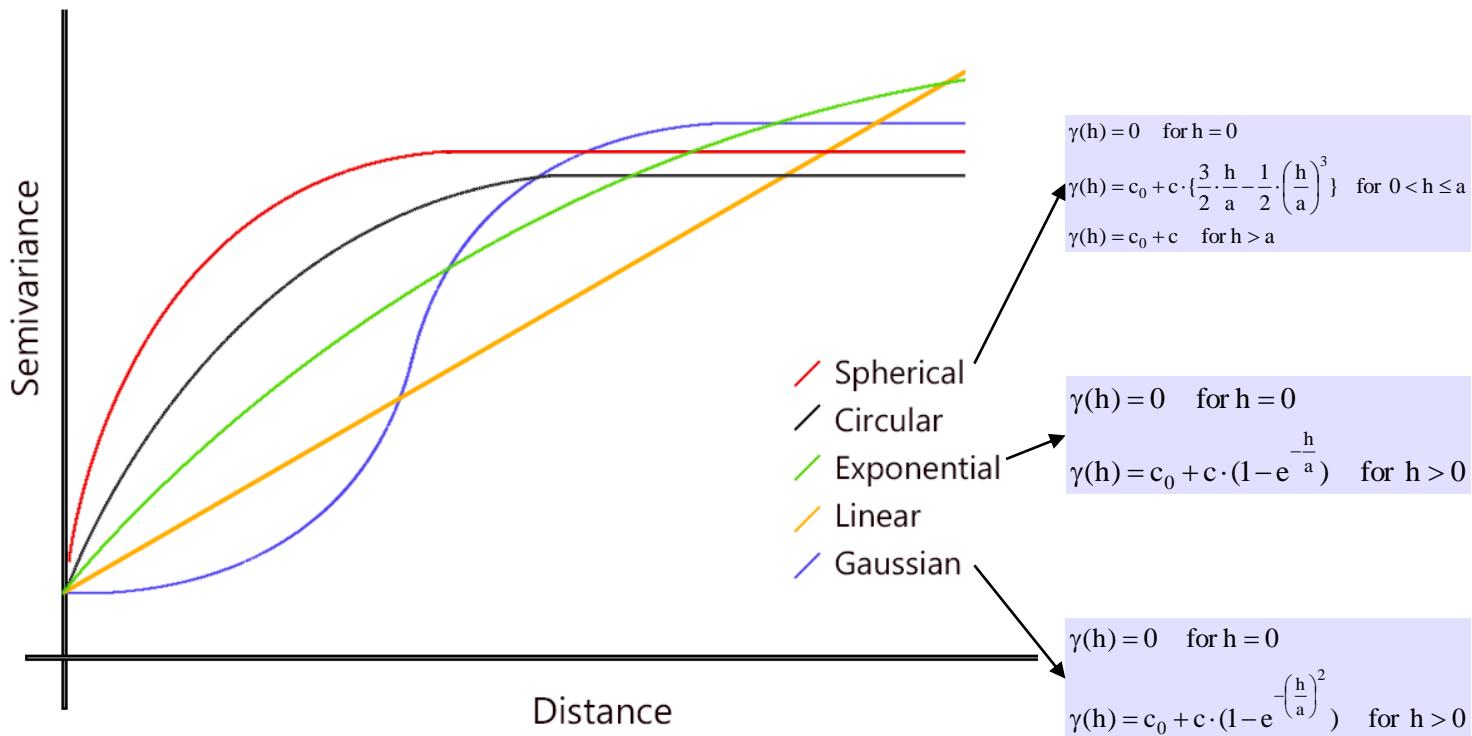
Calculate Variogram: example

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



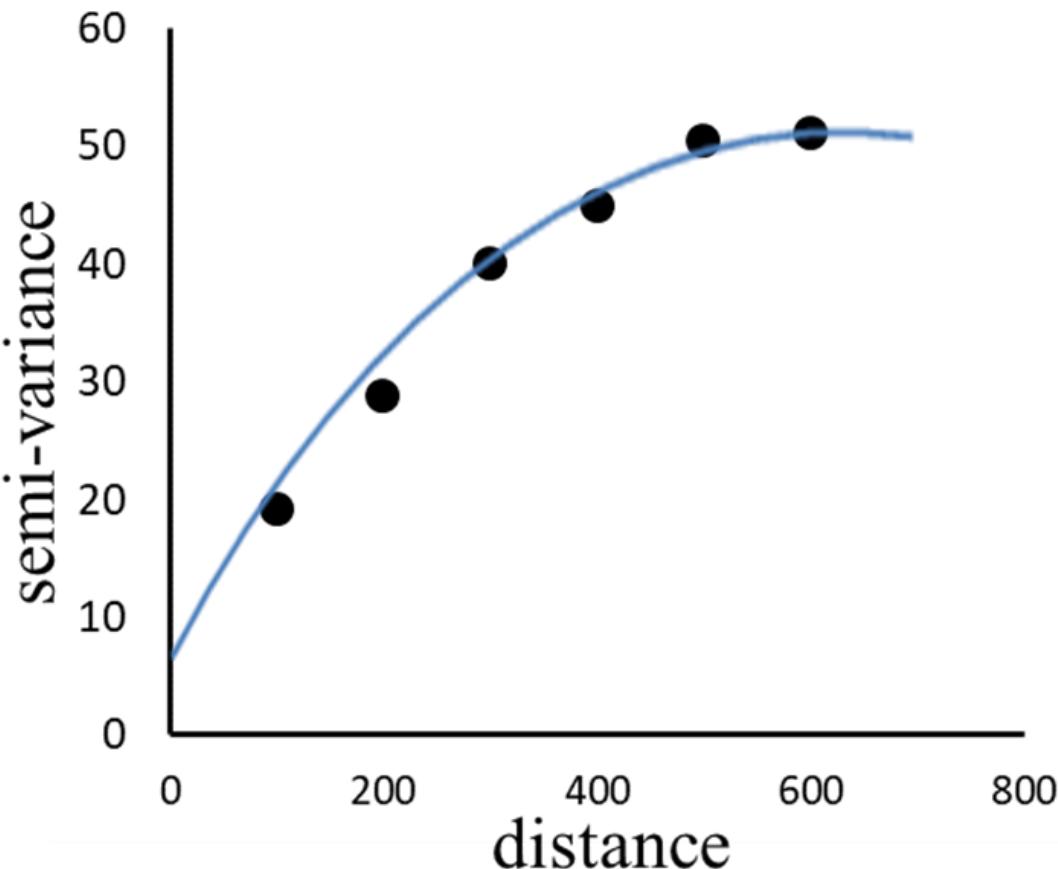
Calculate Variogram: example

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



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, cokriging, etc. Here we will only concentrate on the most basic one: **ordinary kriging**, and also regression kriging



Danie Gerhardus Krige,
the inventor of Kriging.

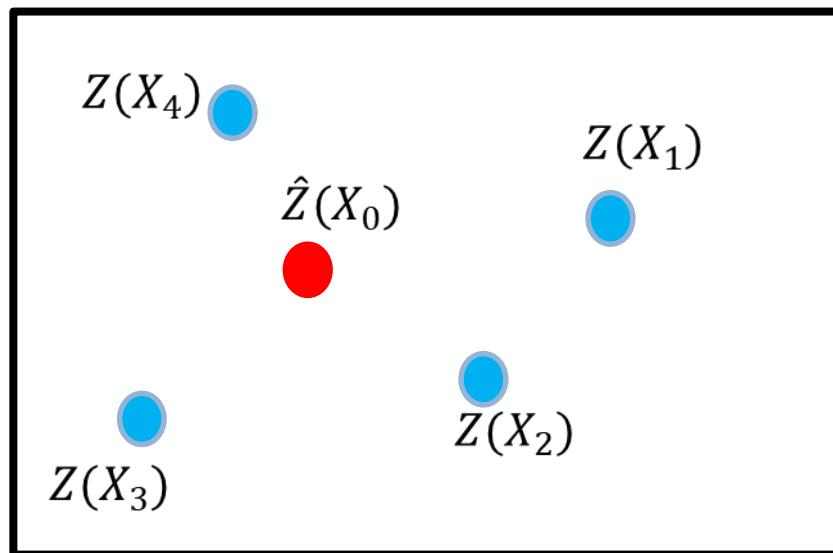
Ordinary Kriging

- 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 :

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

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

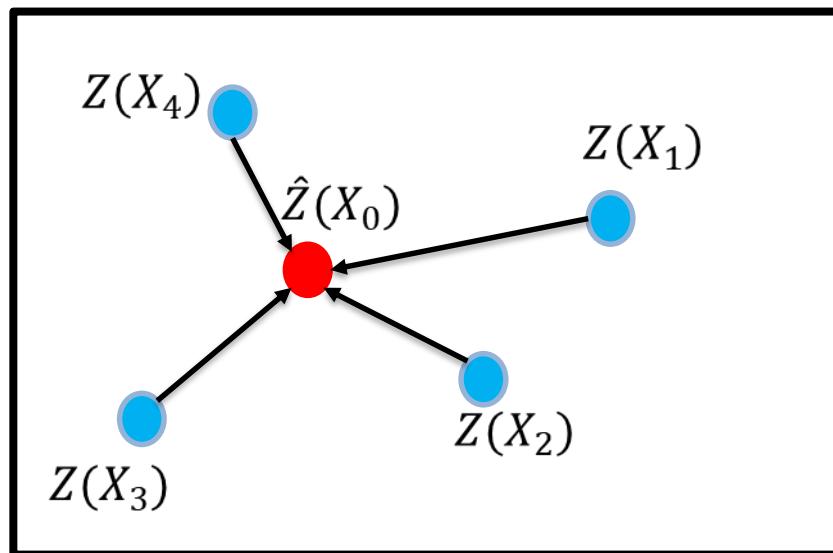


$$\lambda_i = \frac{1/d_i}{\sum_{i=1}^n (1/d_i)}$$

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

X_0 ●
unknown value to be estimated

X_i ●
a set of known measurements

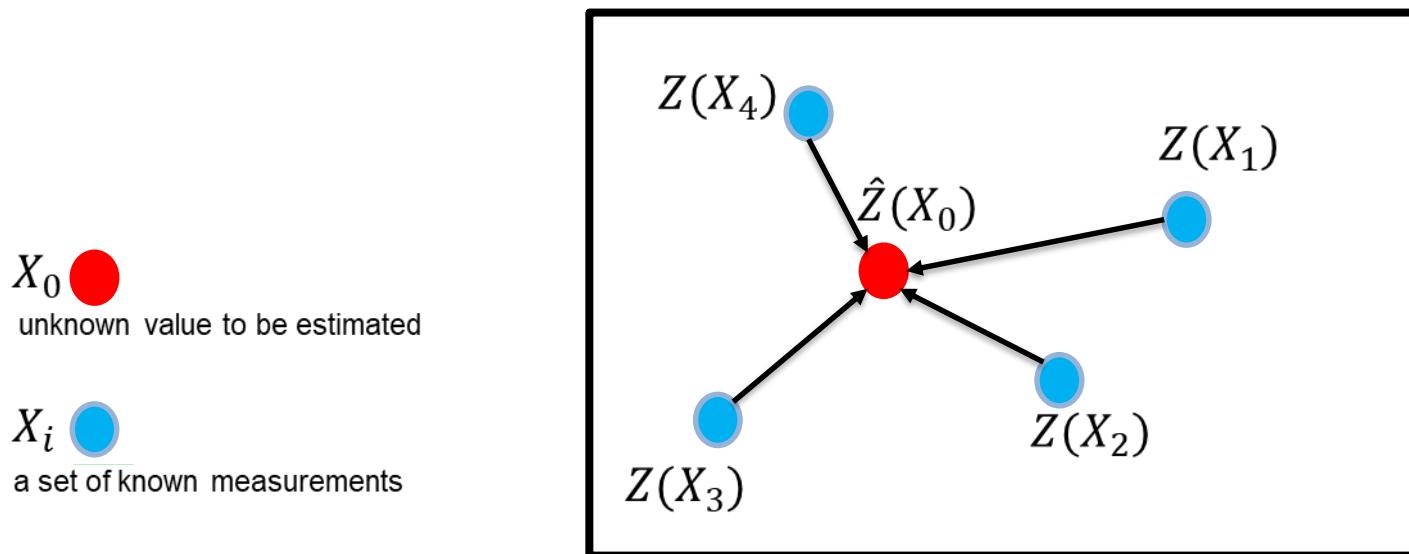


Ordinary Kriging

- 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 :

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

Kriging weight



Ordinary Kriging

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

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

- 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.

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

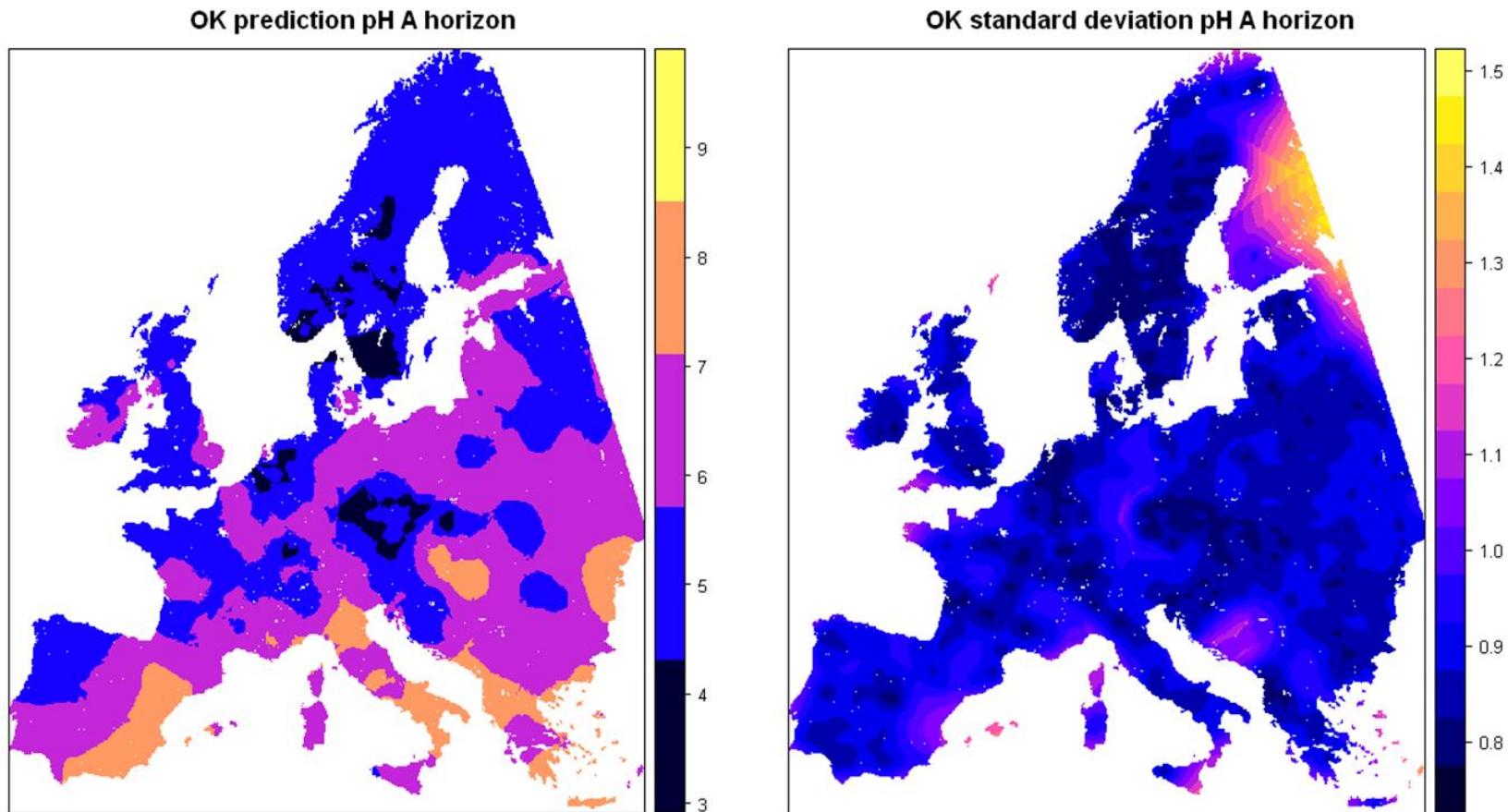
- under the condition:

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



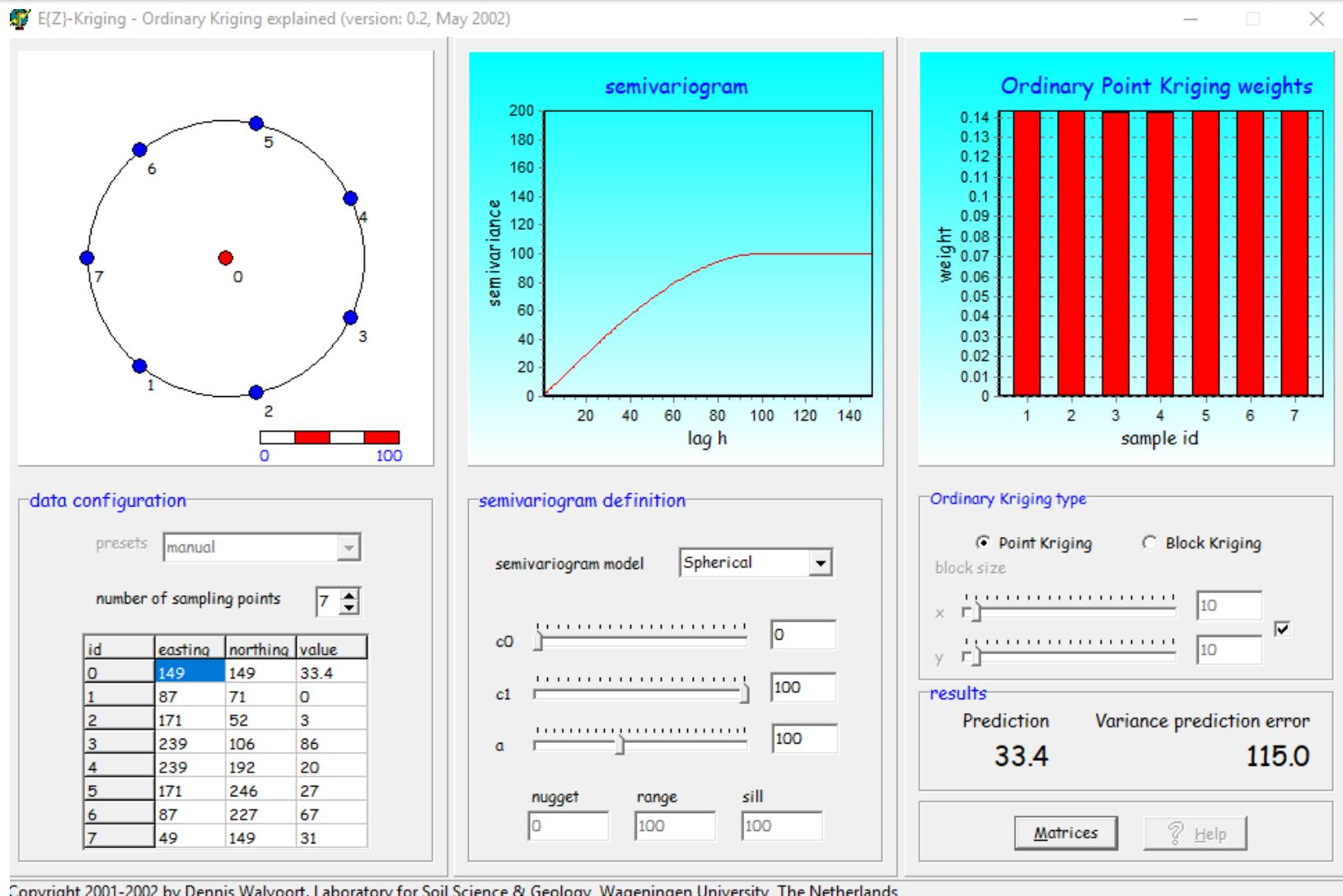
- The kriging variance at each point is automatically generated as part of the process of computing the weights.

Ordinary Kriging: example

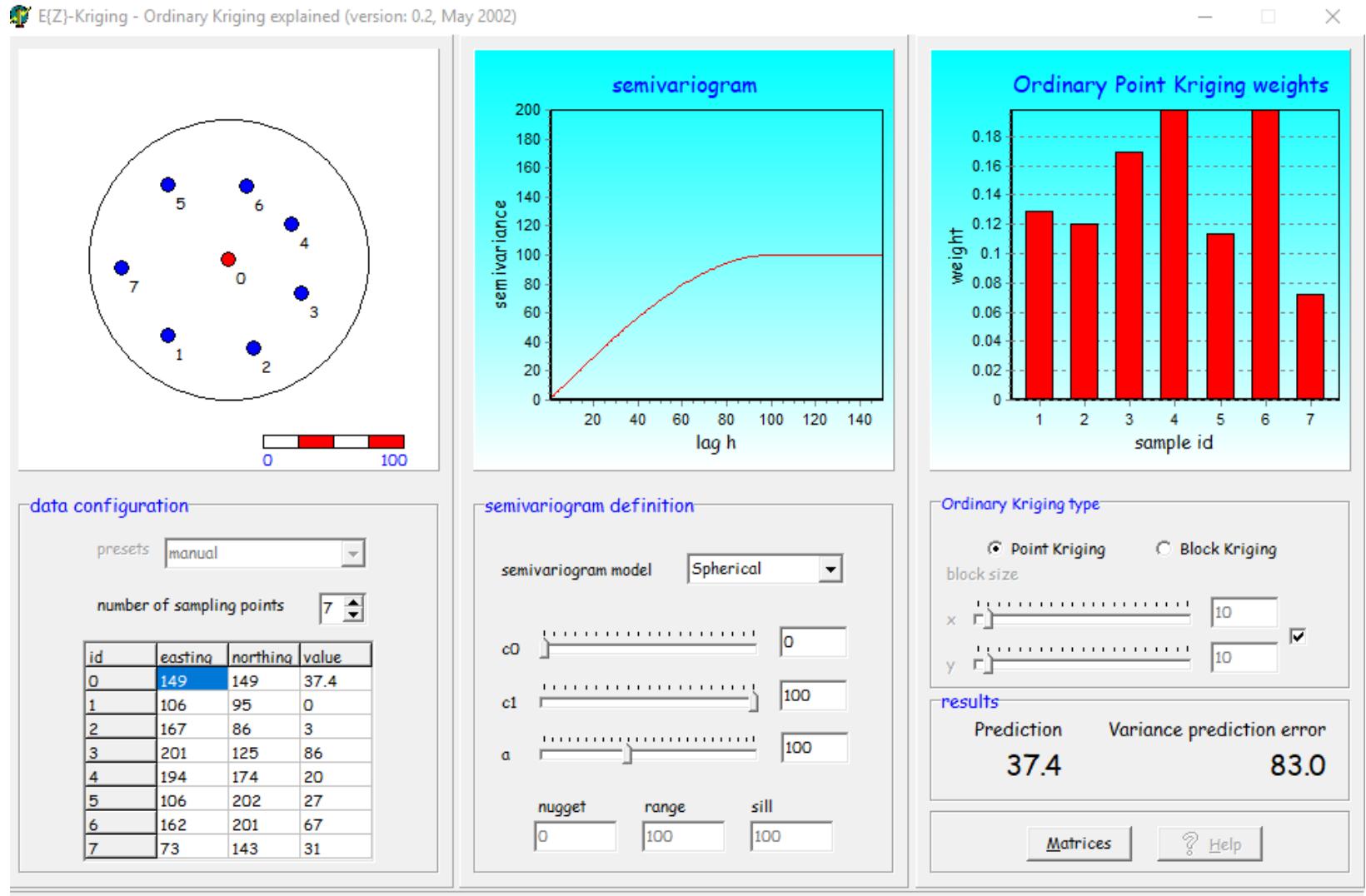


Estimation (left) and standard deviation (right) maps of pH by ordinary kriging.
(Source: Heuvelink, 2017, ISRIC)

$E\{Z\}$ -Kriging



$E\{Z\}$ -Kriging

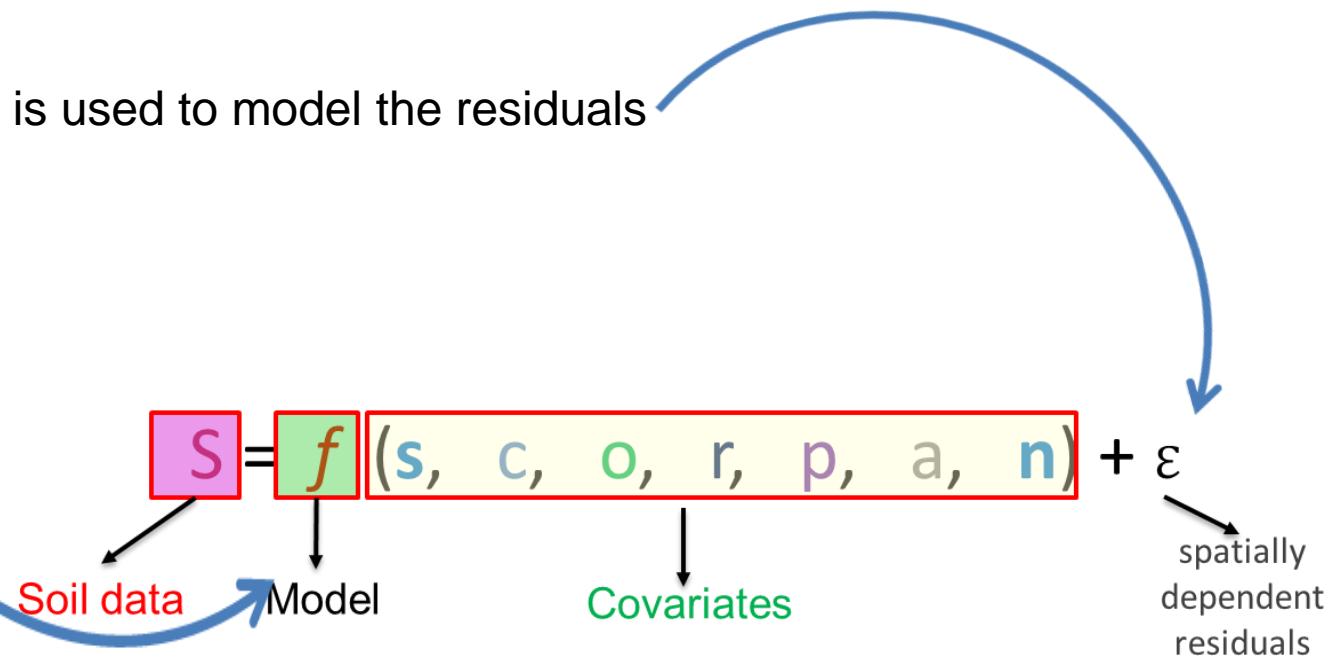


Regression Kriging

- 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

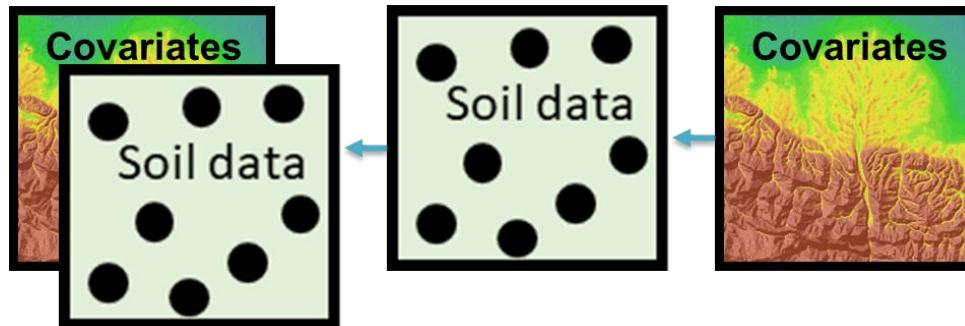
Regression Kriging

- 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



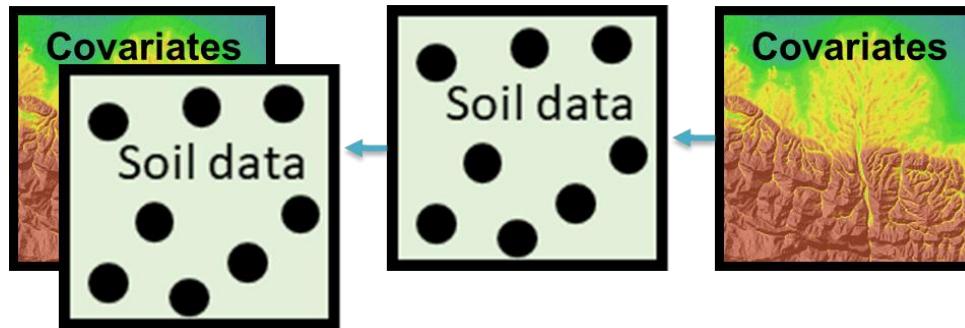
Regression Kriging: workflow

Overlay soil data and covariates



Regression Kriging: workflow

Overlay soil data and covariates



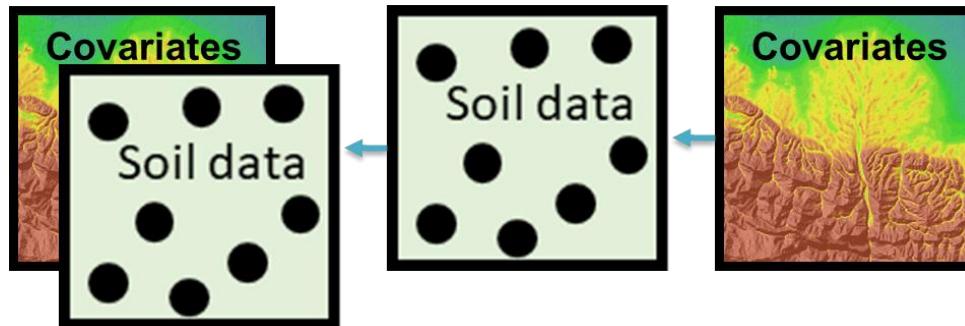
Samples	X	Y	Clay	Elevation
1	765342	2658345	69	76
2	765269	2695483	12	15

Tabulate soil data and covariates
and build a linear regression

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

Regression Kriging: workflow

Overlay soil data and covariates



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

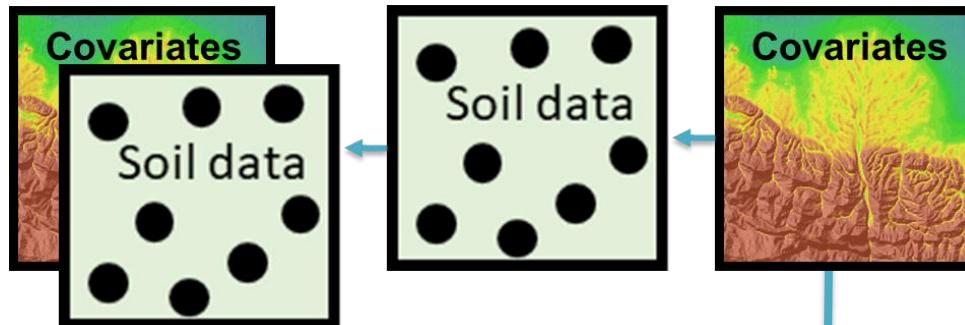
$$\text{Soil clay} = 0.024 + 0.83 \times \text{elevation}$$

Tabulate soil data and covariates
and build a linear regression

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

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

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

$$\text{Soil clay} = 0.024 + 0.83 \times \text{elevation}$$

apply the model to all
unobserved locations

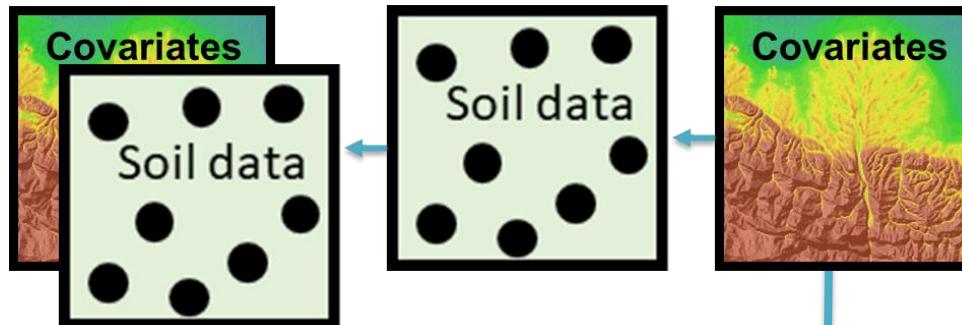
Map 2

Kriging of residuals

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

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

$$\text{Soil clay} = 0.024 + 0.83 \times \text{elevation}$$

apply the model to all
unobserved locations

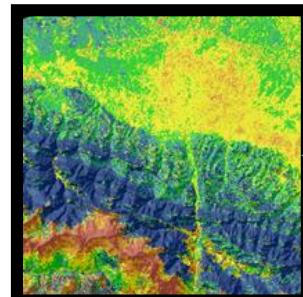
Kriging of residuals

Map 2

Map 1

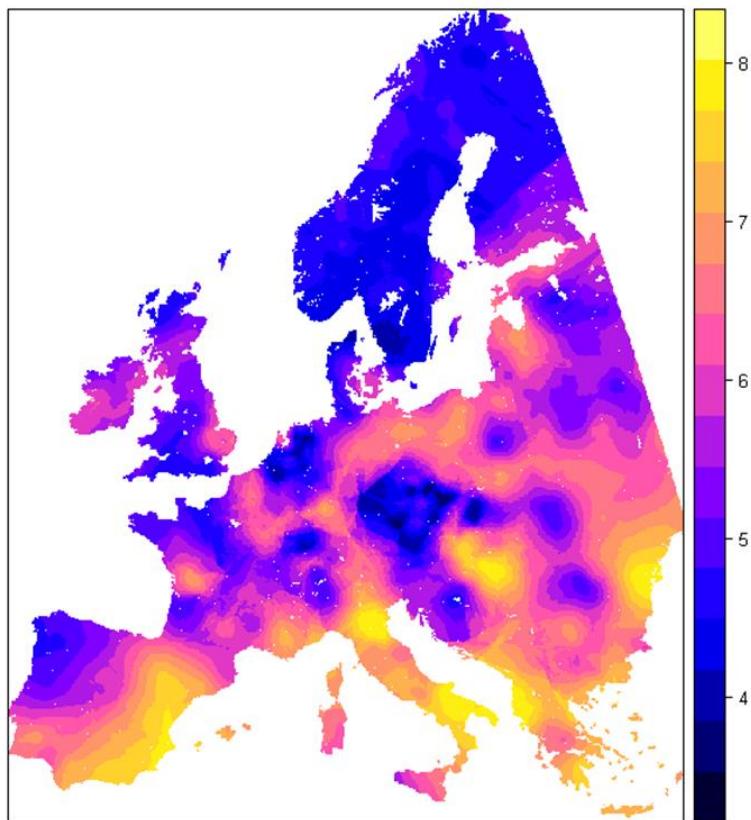
add up the two maps

Regression Kriging map

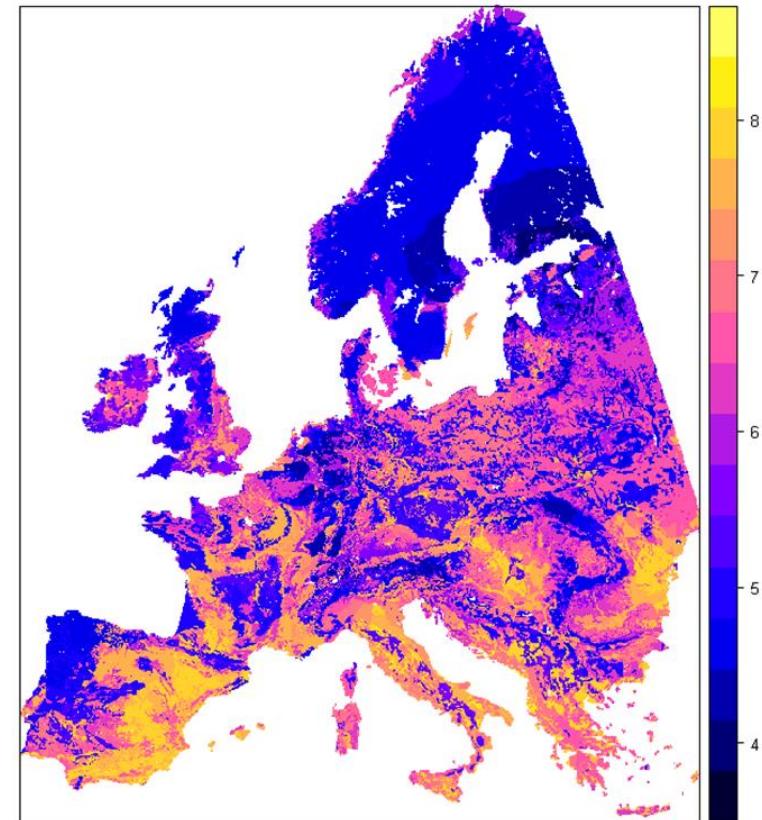


Regression Kriging: example I

Ordinary Kriging



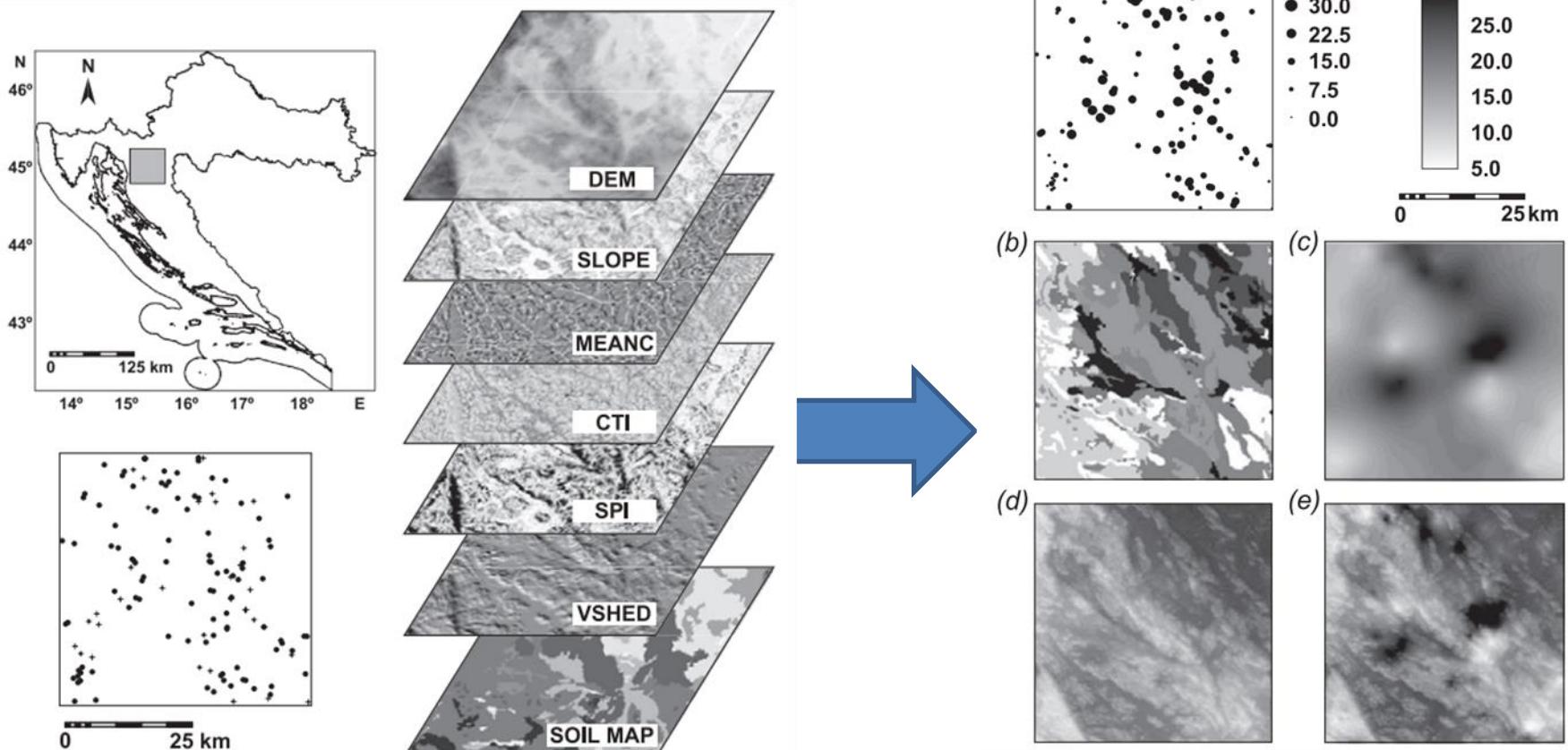
Regression Kriging



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



1. Open Rstudio

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with various icons. The main workspace contains a script editor tab titled "Untitled1*" with the following R code:

```
1 1 + 1
2 2 + 3 + 4
3 x <- c(1:100)
4 hist(x)
5
```

The bottom panel is the R Console, which displays the standard R welcome message and information about its natural language support and collaborative nature.

2. Clear Project List

A red arrow points from the text "Environment is" to the "Project List" menu in the top right corner of the RStudio window. The "Project List" menu is open, showing the following options:

- New Project...
- Open Project...
- Open Project in New Session...
- Close Project
- Clear Project List** (highlighted)
- Project Options...

3. New Project

The screenshot shows the RStudio interface. In the top right corner, a red arrow points to the 'Project' dropdown menu, which is open to show options like 'New Project...', 'Open Project...', and 'Clear Project List'. The 'Environment' tab is selected in the main workspace. On the left, there's an 'Untitled1*' script editor with some R code. Below it is a 'Console' window displaying the standard R welcome message. At the bottom, there's a taskbar with various icons.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Untitled1* Go to file/function Addins

1 1 + 1
2 2 + 3 + 4
3 x <- c(1:100)
4 hist(x)
5

5:1 (Top Level) R Script

Console ~/
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Type here to search

Project: (None)

New Project...
Open Project...
Open Project in New Session...
Close Project
Clear Project List
Project Options...

Environment History Connections

Import Dataset

Global Environment

Files Plots Packages Help Viewer

4:35 PM 2/9/2021

4. Existing Directory

The screenshot shows the RStudio interface with a 'Create Project' dialog box overlaid. The dialog has three options: 'New Directory', 'Existing Directory', and 'Version Control'. A large red arrow points to the 'Existing Directory' option. The RStudio environment includes a top menu bar, a code editor with some R code, a console window displaying a welcome message, and an empty 'Environment' tab.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Untitled1* Go to file/function Addins

Project: (None)

1 1 + 1
2 2 + 3 + 4
3 x <- c(1:100)
4 hist(x)
5

5:1 (Top Level) ◆

Console ~/ ◆
You are welcome to redistribute
Type 'license()' or 'licence()'
Natural language support but r
R is a collaborative project wit
Type 'contributors()' for more i
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

New Project

Create Project

New Directory Start a project in a brand new working directory >

Existing Directory Associate a project with an existing working directory > ↓

Version Control Checkout a project from a version control repository >

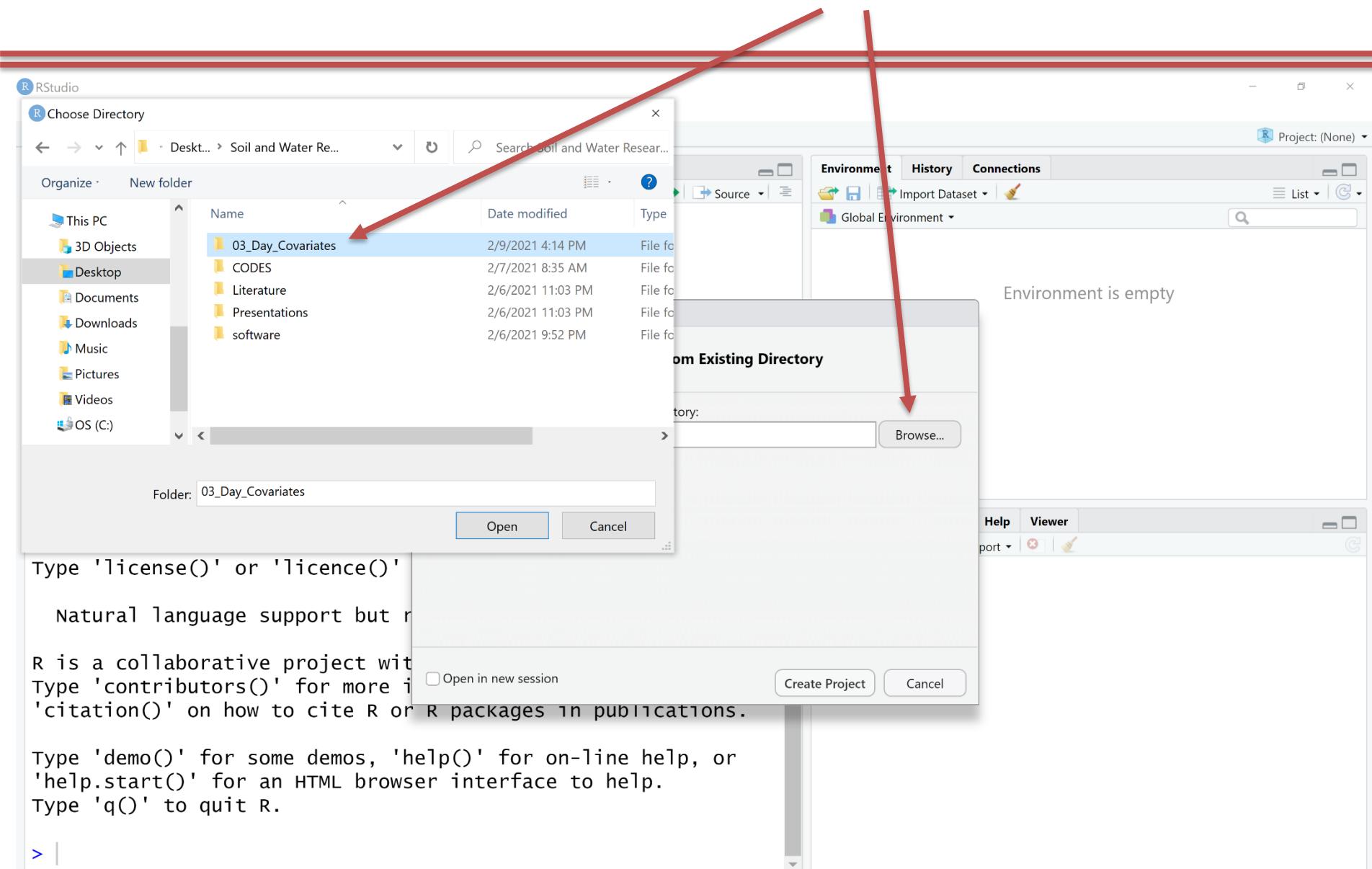
Environment History Connections

Global Environment

Environment is empty

Help Viewer

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



6. Create project

The screenshot shows the RStudio interface with a 'New Project' dialog box open in the foreground. The dialog is titled 'Create Project from Existing Directory' and displays a working directory path: 'C:/Users/rh_ta/OneDrive/Desktop/oil and Water R'. A red arrow points to the 'Create Project' button at the bottom right of the dialog. The RStudio environment includes a code editor with R code, a console window displaying R startup messages, and various toolbars and panels.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Untitled1* | Go to file/function | Addins

1 1 + 1
2 2 + 3 + 4
3 x <- c(1:100)
4 hist(x)
5

5:1 (Top Level) ▾

Console ~ /

You are welcome to redistribute
Type 'license()' or 'licence()'
Natural language support but r
R is a collaborative project with
Type 'contributors()' for more i
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

New Project

Back Create Project from Existing Directory

Project working directory:
C:/Users/rh_ta/OneDrive/Desktop/oil and Water R | Browse...

Open in new session

Create Project Cancel

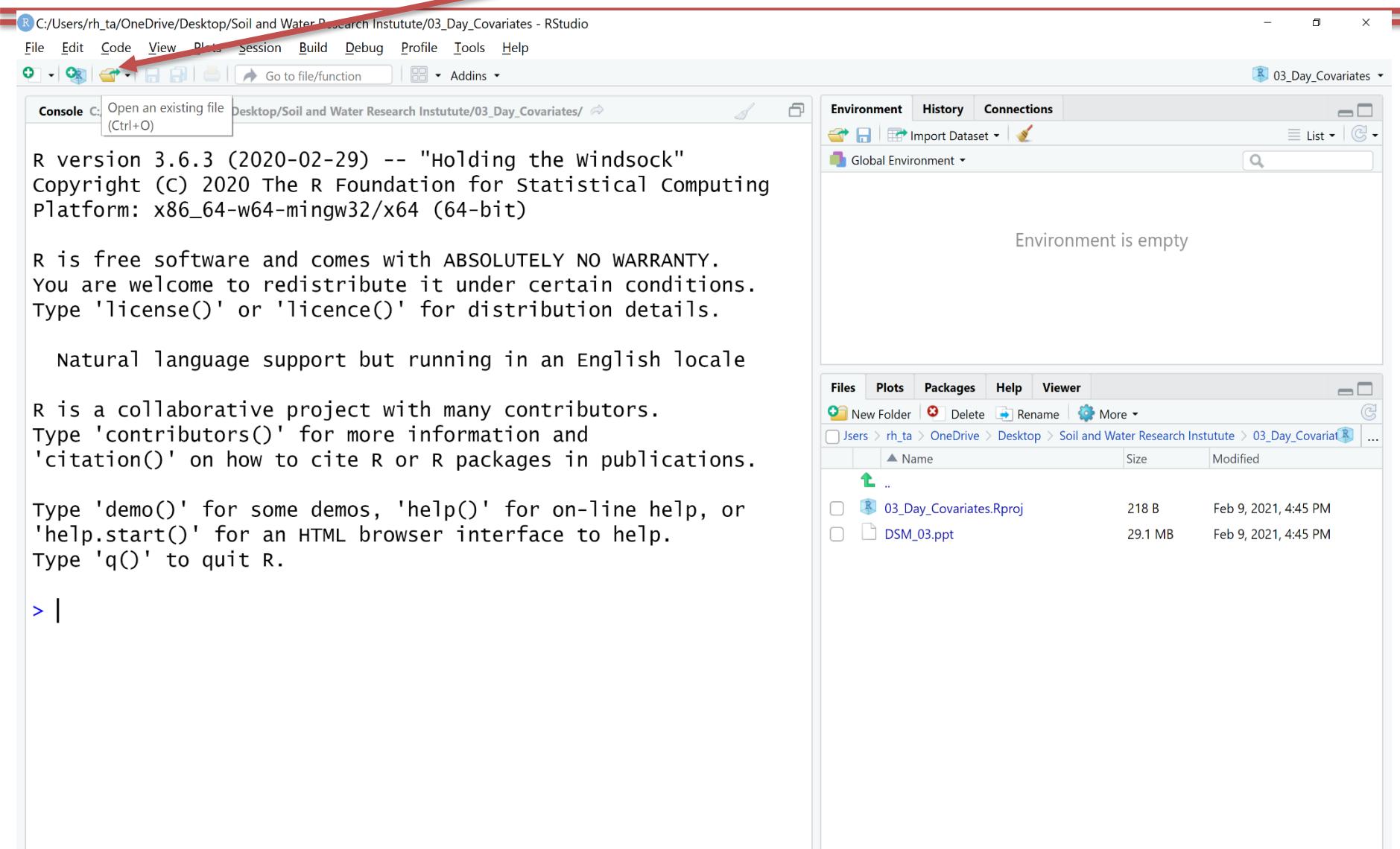
Environment History Connections

Global Environment

Environment is empty

Help Viewer

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



R C:/Users/rh_ta/OneDrive/Desktop/Soil and Water Research Institute/03_Day_Covariates - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

+ - Go to file/function Addins

Console C: Open an existing file Desktop/Soil and Water Research Institute/03_Day_Covariates/ Environment History Connections

R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Jusers > rh_ta > OneDrive > Desktop > Soil and Water Research Institute > 03_Day_Covariates

Name	Size	Modified
03_Day_Covariates.Rproj	218 B	Feb 9, 2021, 4:45 PM
DSM_03.ppt	29.1 MB	Feb 9, 2021, 4:45 PM

8. Open 03_R_Cov.R

