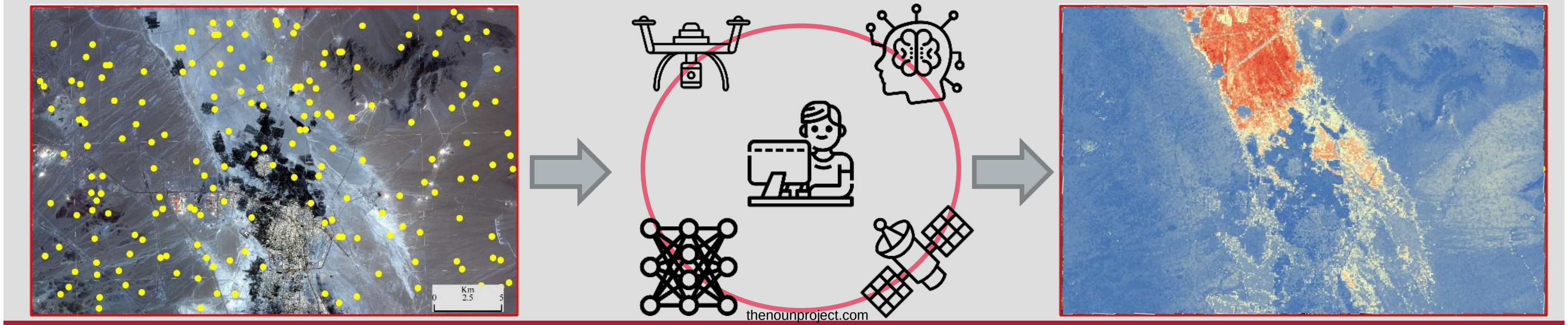
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Mathematisch-Naturwissenschaftliche Fakultät

Department of Geosciences · Soil Science & Geomorphology



Digital Soil Mapping

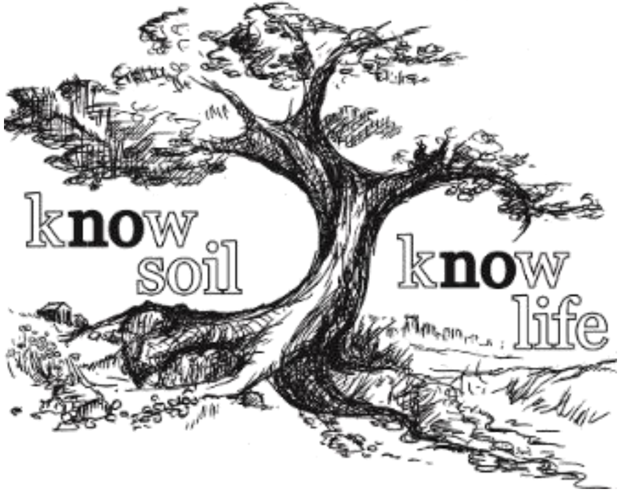
Ruhollah Taghizadeh

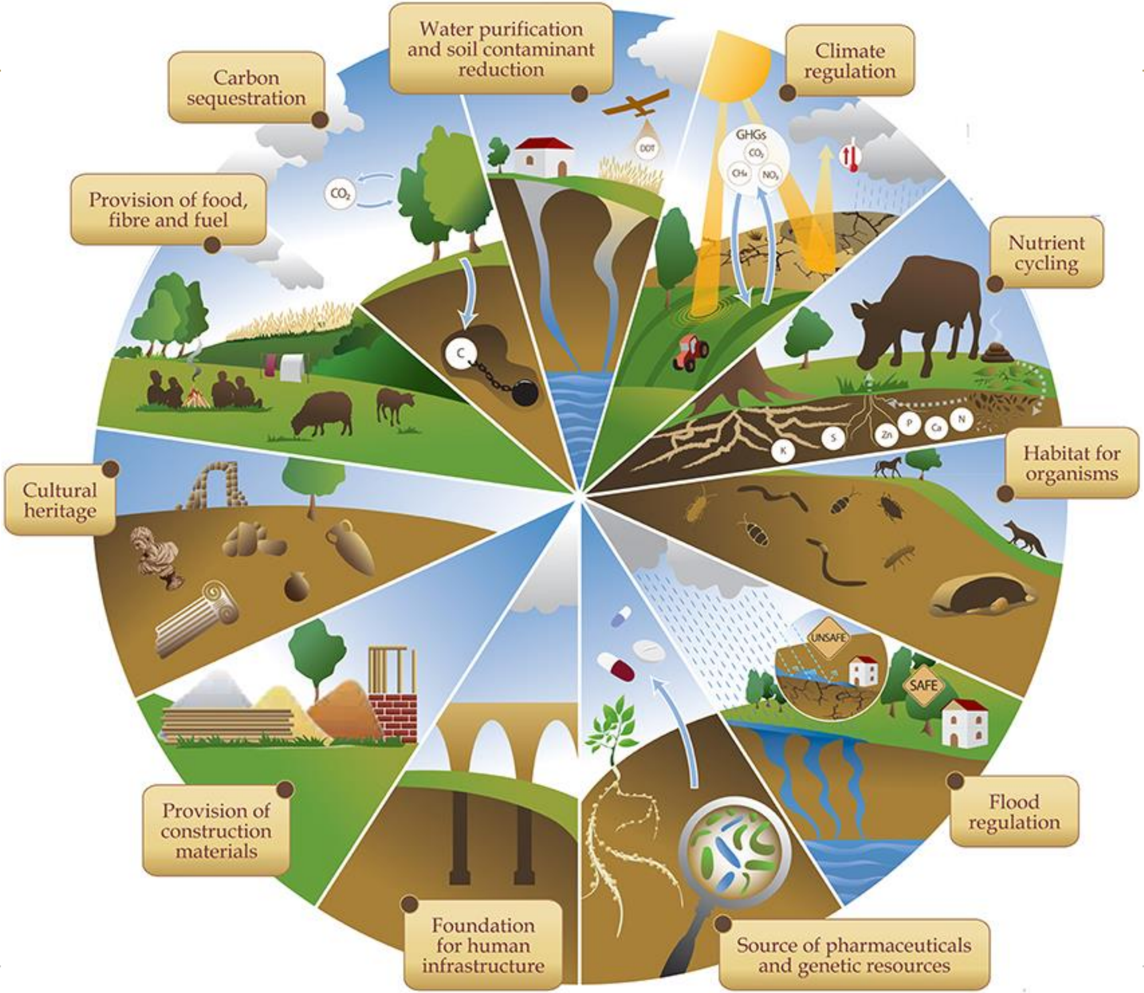
Content

- Conventional Soil Mapping
- Digital Soil Mapping
- Environmental Covariates
- Interpolation (Geostatistical Analysis)

- Machine Learning Methods
- Decision Tree and Random Forest
- Model Validation and Uncertainty Analysis
- Soil Depth Functions and 3D Maps

Why Soil Matters?





Soil Degradation

Soil erosion

Land clearing

Soil salinization

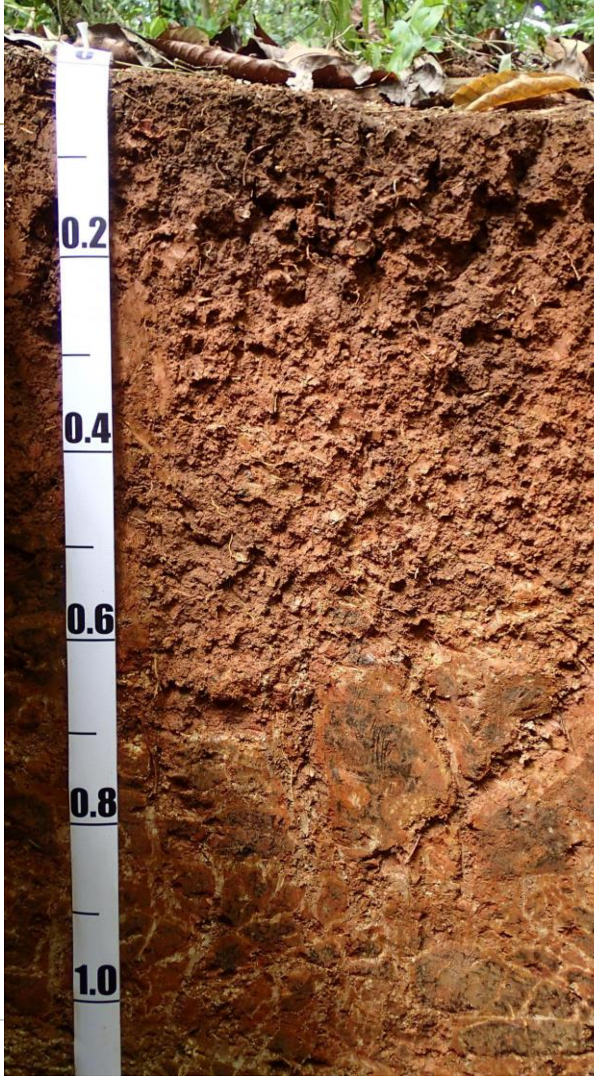
Soil compaction

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Soil Degradation

Soil Functions

Sustainable Soil Management



Soil map

is an essential tool in achieving sustainable use of the land

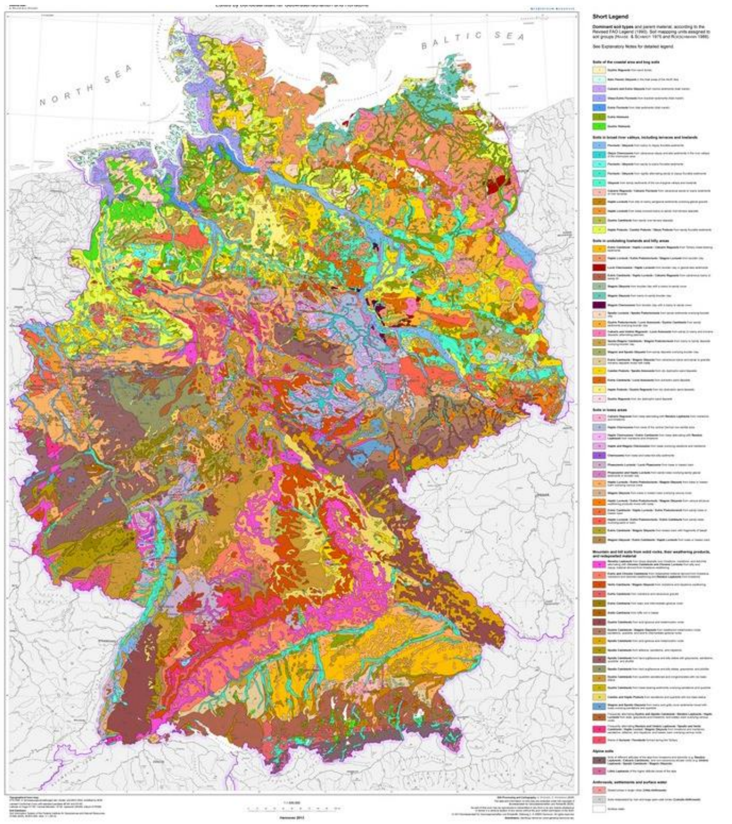
Soil Map

Definition:

Soil map is a geographical representation showing diversity of soil types and/or soil properties (soil pH, textures, organic matter, depths of horizons etc.) in the area of interest.

Users:

Decision makers, researchers/scientists, agronomists, farmers, etc.



Soil Map of the Federal Republic of Germany 1 : 1,000,000

(https://www.bgr.bund.de)

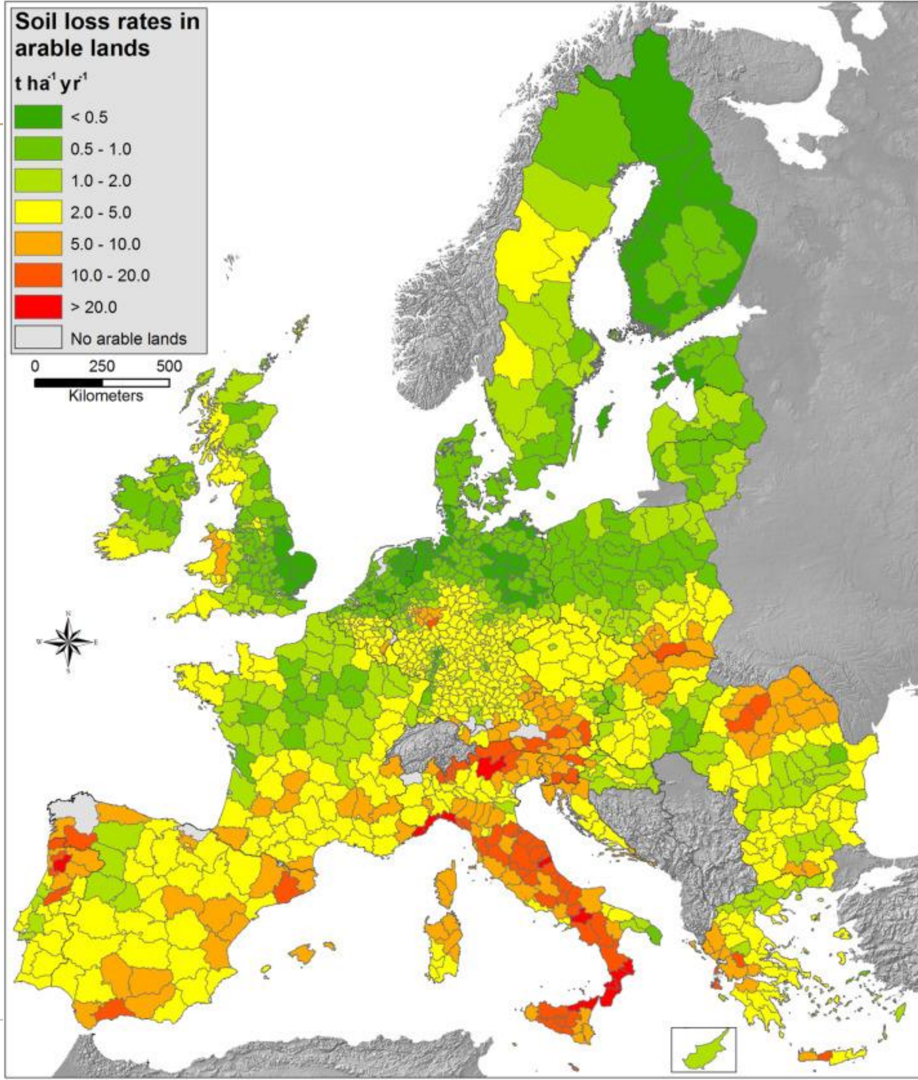
Dent & Young, 1981; Zhu, et al. 2001

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Environmental Science & Policy
Volume 54, December 2015, Pages 438-447

The new assessment of soil loss by water erosion in Europe

Panos Panagos [a], Pasquale Borrelli [a], Jean Poesen [c], Cristiano Ballabio [a], Emanuele Lugato [a], Katrin Meusburger [b], Luca Montanarella [a], Christine Alewell [b]



Article | Peer review | Metrics | Related articles

Original research article

07 Oct 2019

Development of a harmonised soil profile analytical database for Europe: a resource for supporting regional soil management

Jeppe Aagaard Kristensen[1,2,★], Thomas Balstrøm[2], Robert J. A. Jones[3], Arwyn Jones[4], Luca Montanarella[4], Panos Panagos[4], and Henrik Breuning-Madsen[2,★,†]

[1]Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden

[2]Department of Geosciences and Natural Resource Management, University of Copenhagen, 1350 Copenhagen K, Denmark

[3]School of Energy, Environment and AgriFood, Cranfield University, College Road, Cranfield, MK43 0AL, UK

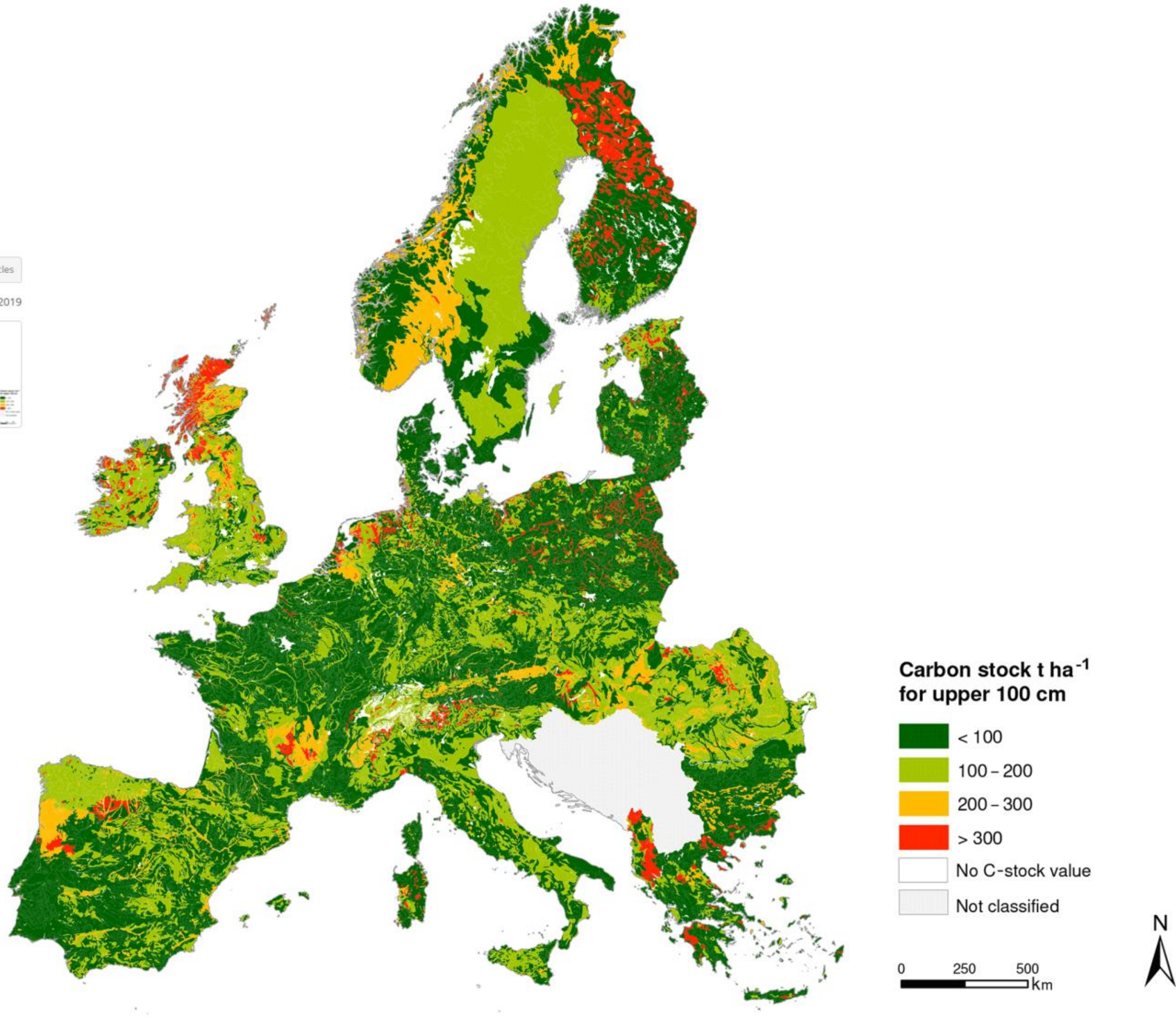
[4]European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra (VA), Italy

★These authors contributed equally to this work.

†deceased

Correspondence: Jeppe Aagaard Kristensen (jeppe.aa.kristensen@gmail.com)

Received: 30 Mar 2019 – Discussion started: 06 May 2019 – Revised: 20 Aug 2019 – Accepted: 28 Aug 2019 – Published: 07 Oct 2019



Article | Peer review | Metrics | Related articles

Original research article

07 Oct 2019

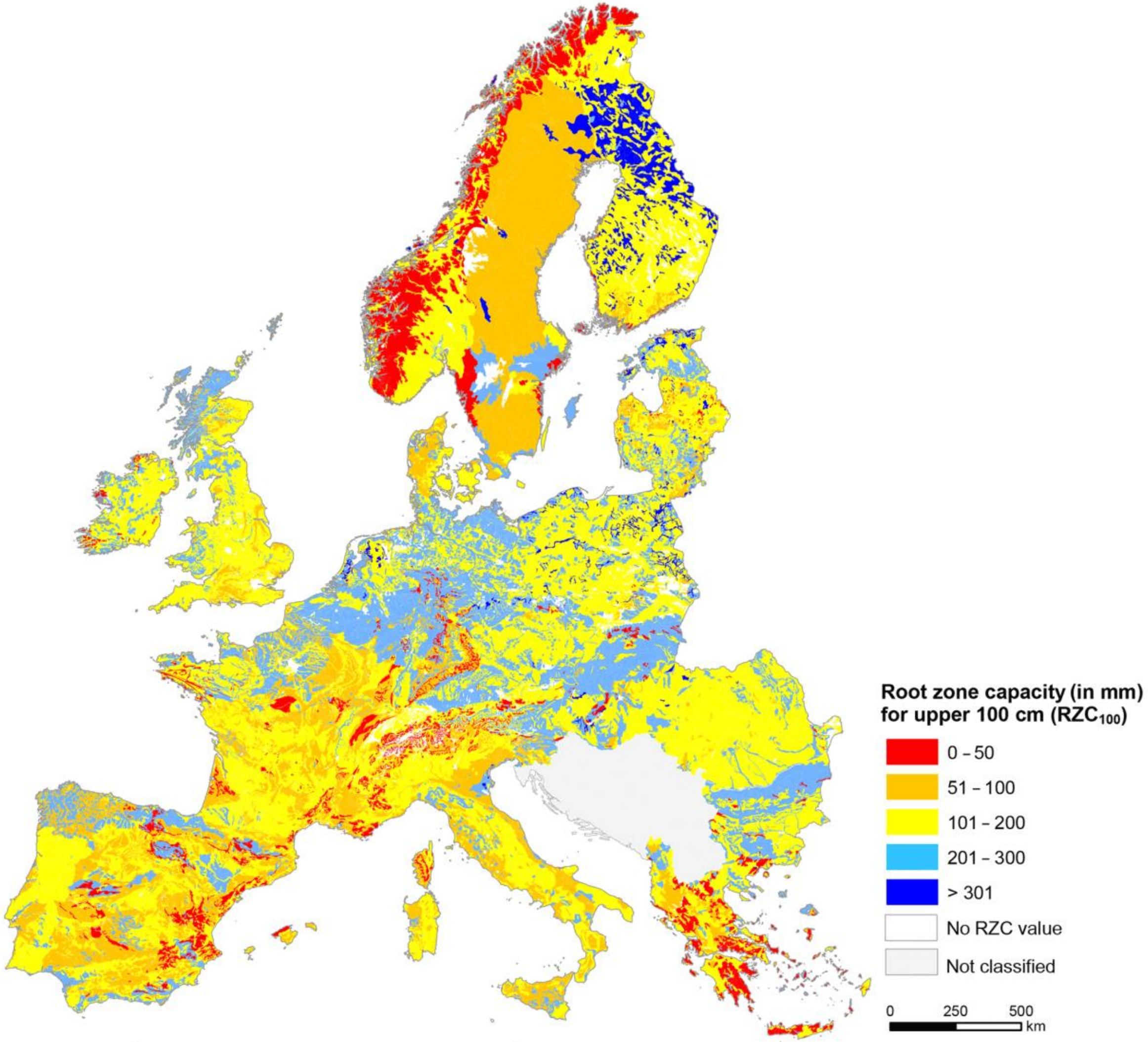
Development of a harmonised soil profile analytical database for Europe: a resource for supporting regional soil management

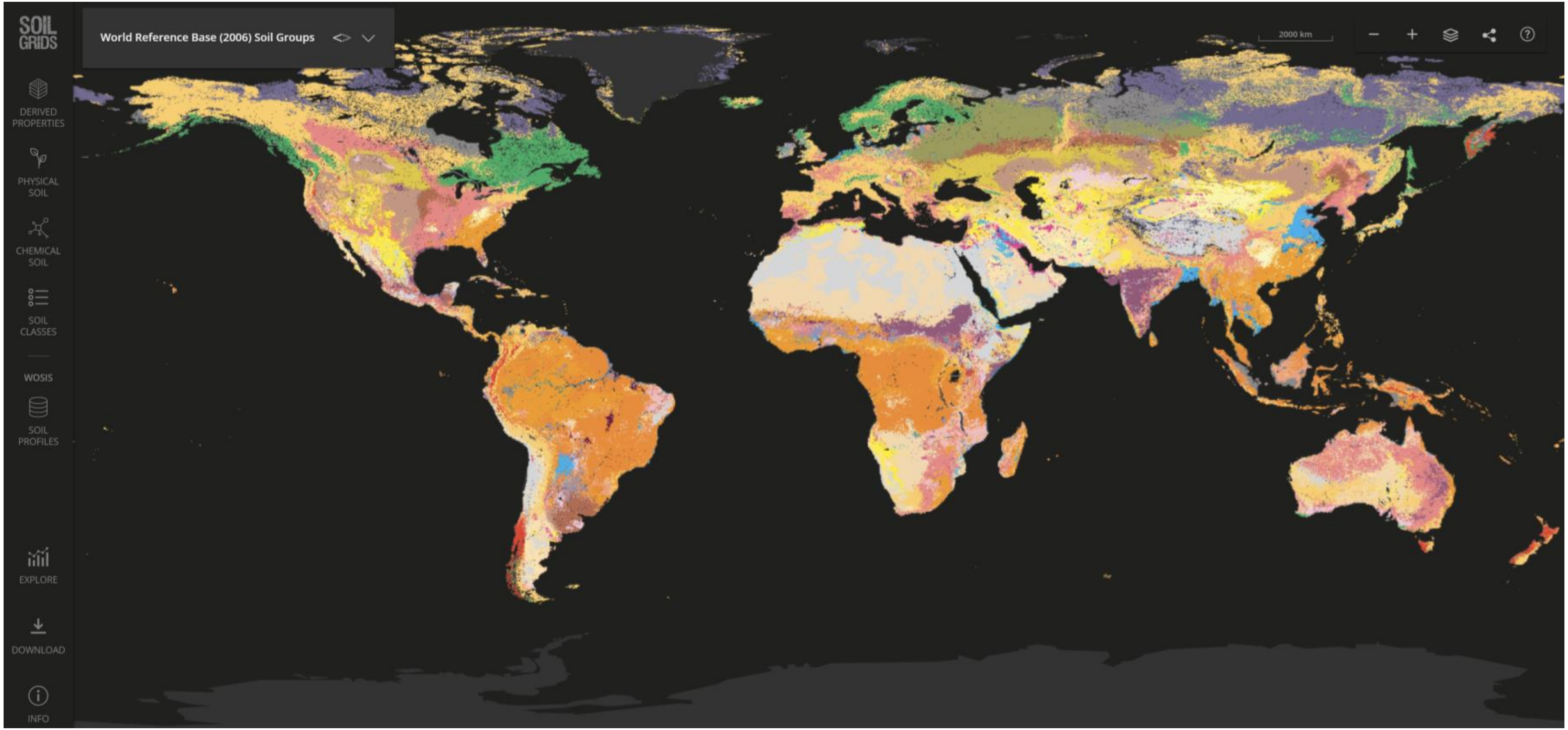
Jeppe Aagaard Kristensen[1,2,★], Thomas Balstrøm[2], Robert J. A. Jones[3], Arwyn Jones[4], Luca Montanarella[4], Panos Panagos[4], and Henrik Breuning-Madsen[2,★,†]

[1]Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 223 62 Lund, Sweden
[2]Department of Geosciences and Natural Resource Management, University of Copenhagen, 1350 Copenhagen K, Denmark
[3]School of Energy, Environment and AgriFood, Cranfield University, College Road, Cranfield, MK43 0AL, UK
[4]European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra (VA), Italy
★These authors contributed equally to this work.
†deceased

Correspondence: Jeppe Aagaard Kristensen (jeppe.aa.kristensen@gmail.com)

Received: 30 Mar 2019 – Discussion started: 06 May 2019 – Revised: 20 Aug 2019 – Accepted: 28 Aug 2019 – Published: 07 Oct 2019

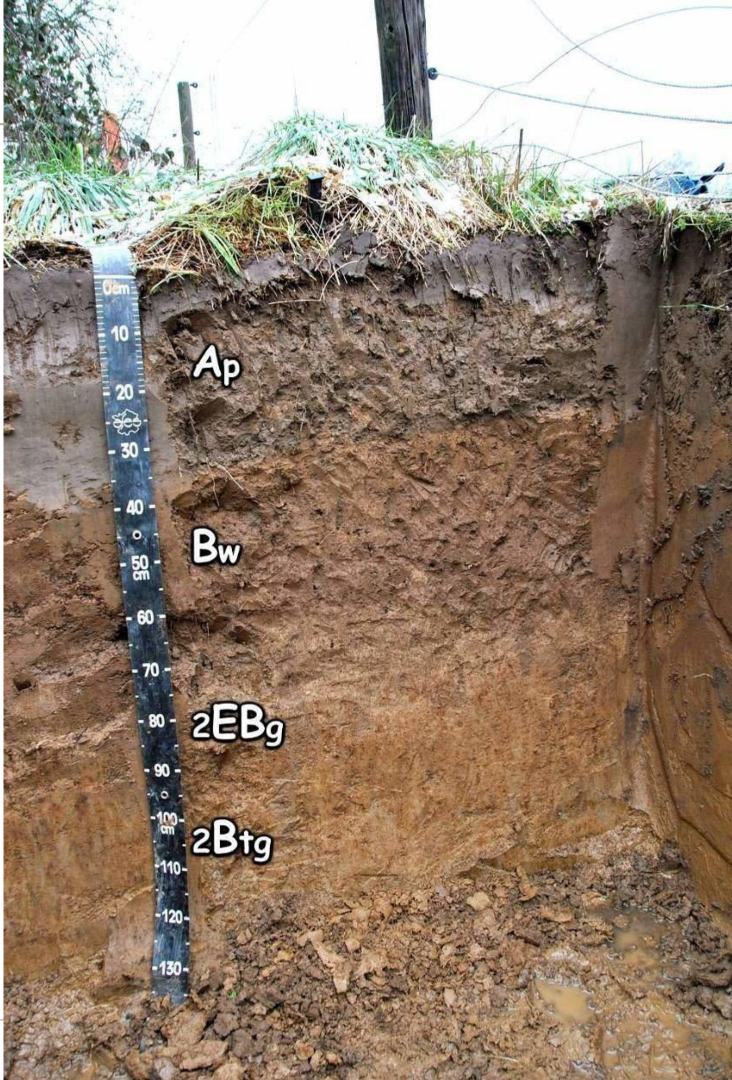




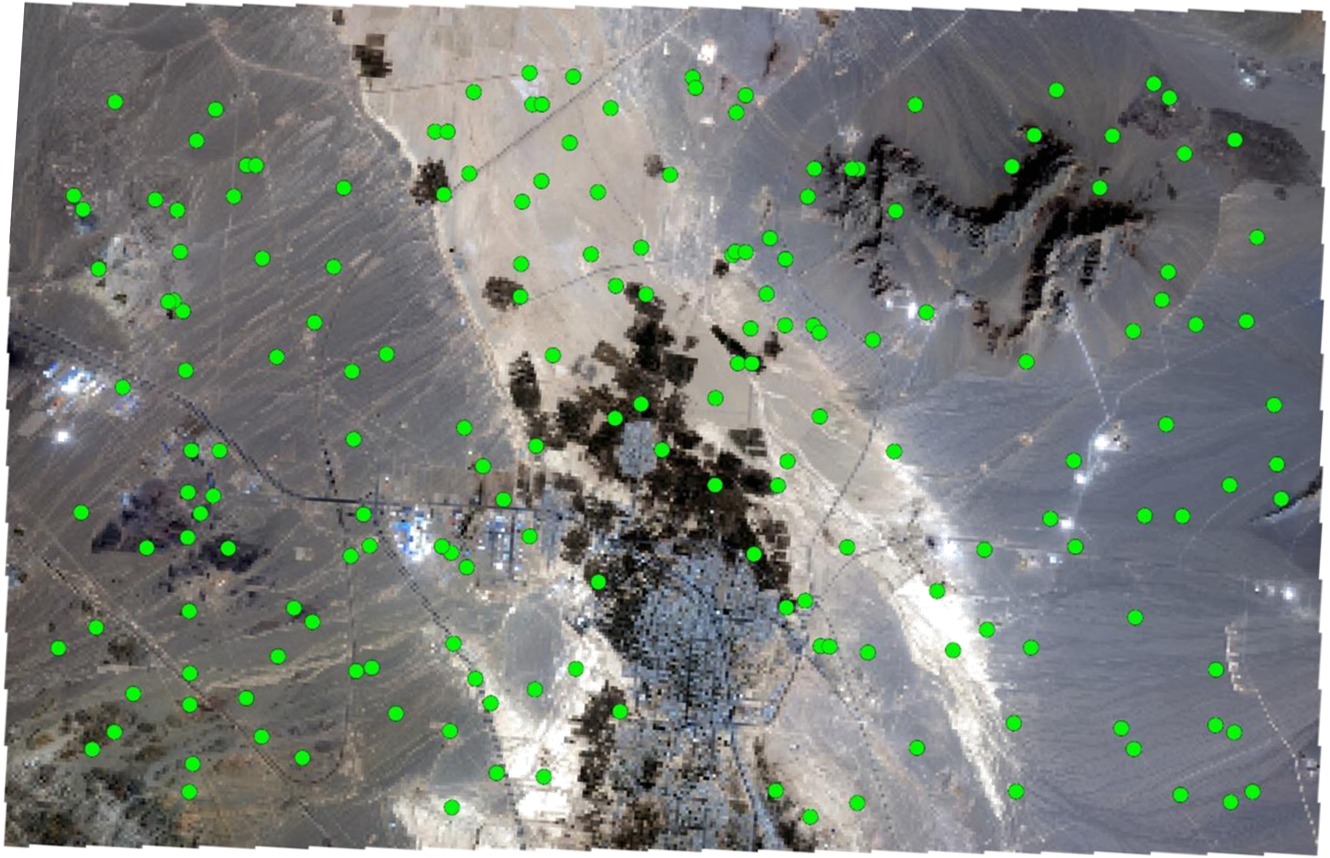
https://soilgrids.org/

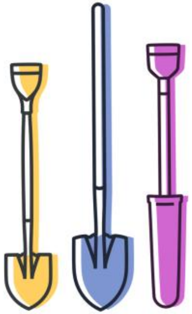
How to Make a Soil Map?

1. study area
2. sampling plan
3. filed data collection
4. soil analysis
5. soil mapping

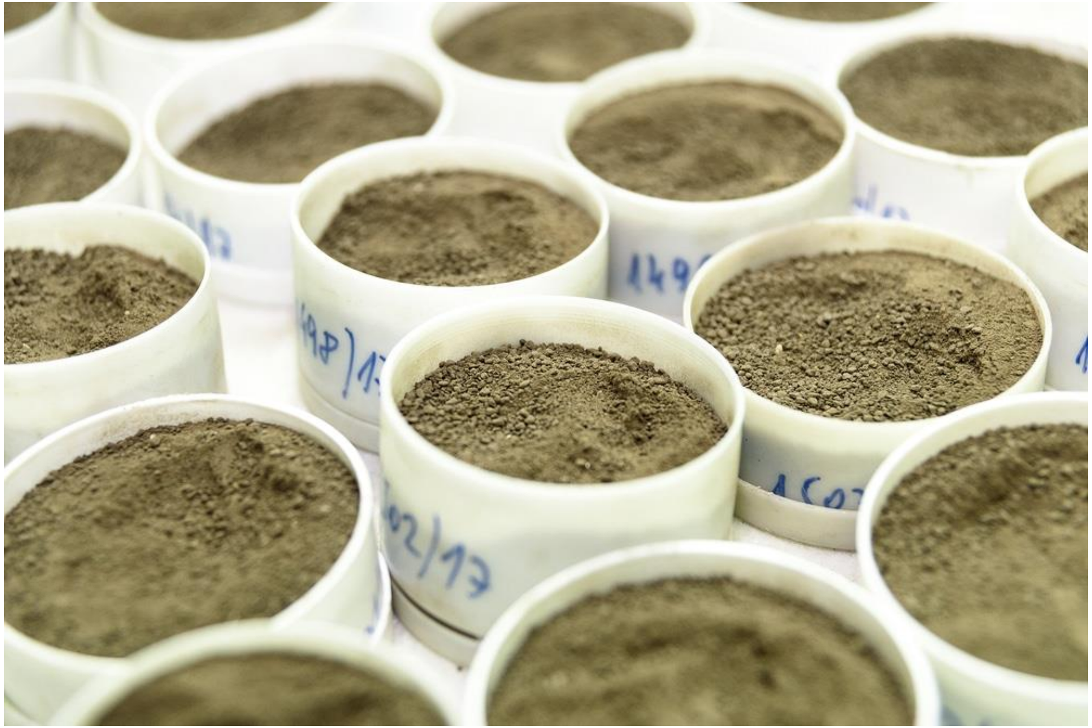


sampling plan and filed data collection

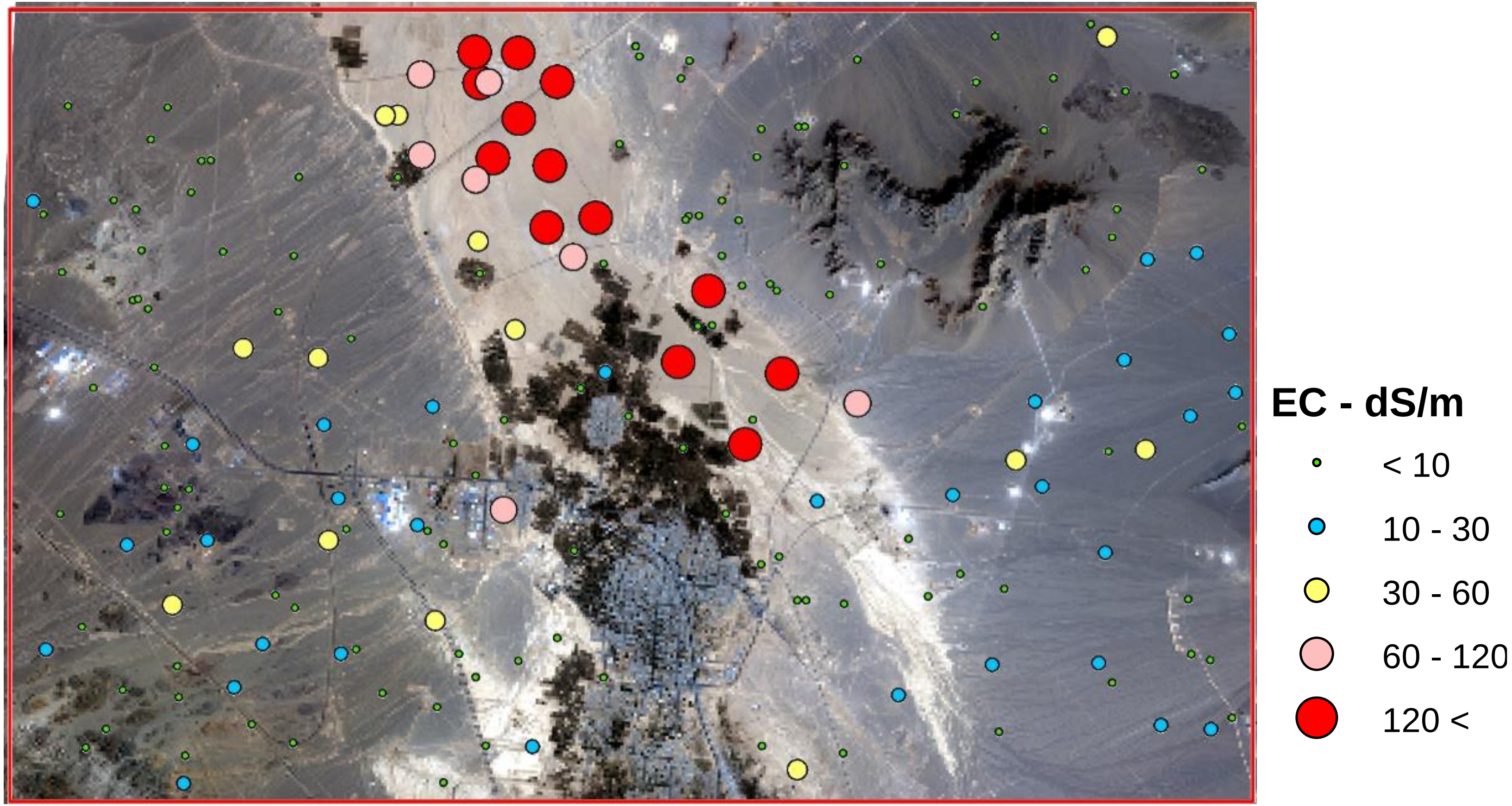


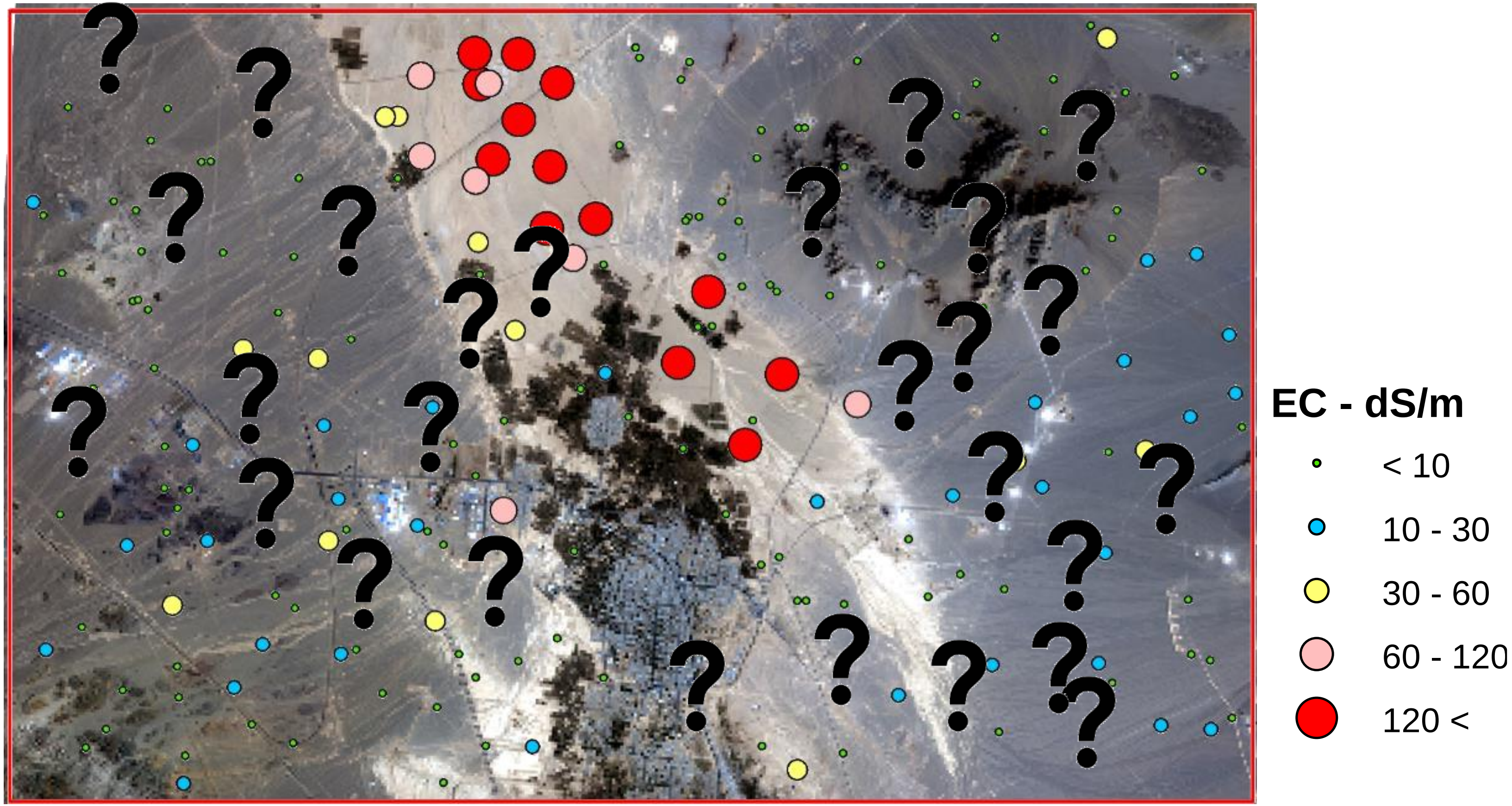


soil analysis



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN




EC - dS/m
< 10
10 - 30
30 - 60
60 - 120
120 <

How to Make a Soil Map?

1. study area
2. sampling plan
3. filed data collection
4. soil analysis

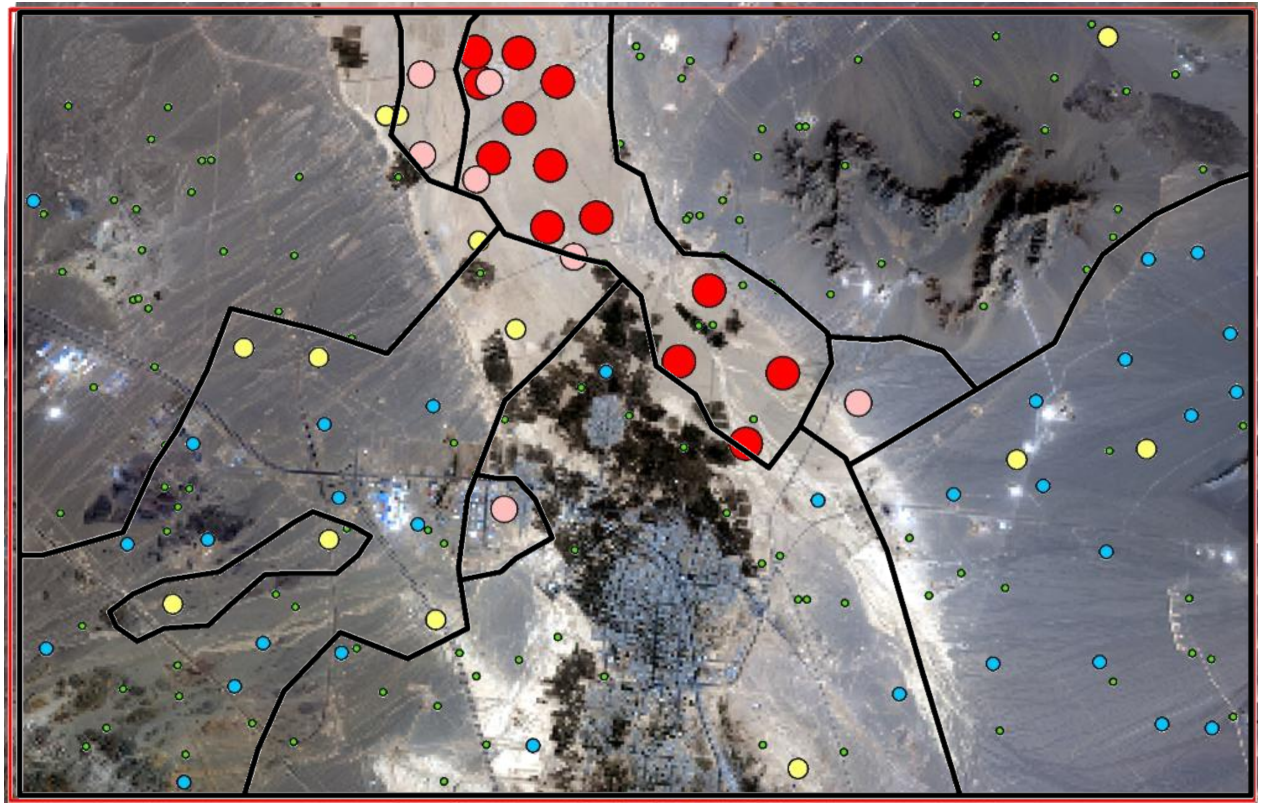
-mean value!

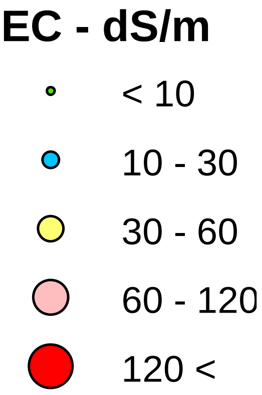
5. soil mapping

1. **C**onventional **S**oil **M**apping

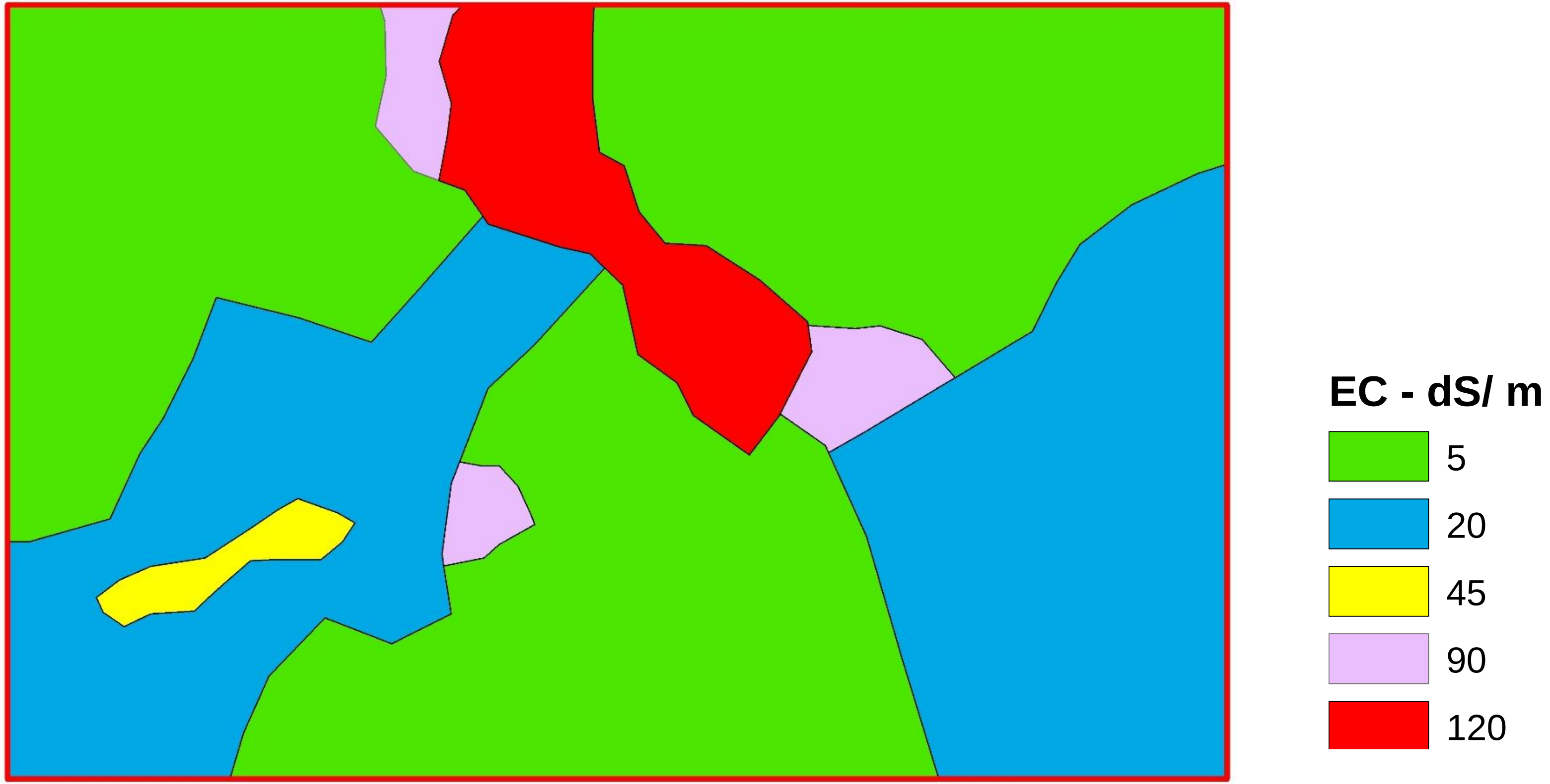
2. **D**igital **S**oil **M**apping

Conventional Soil Mapping

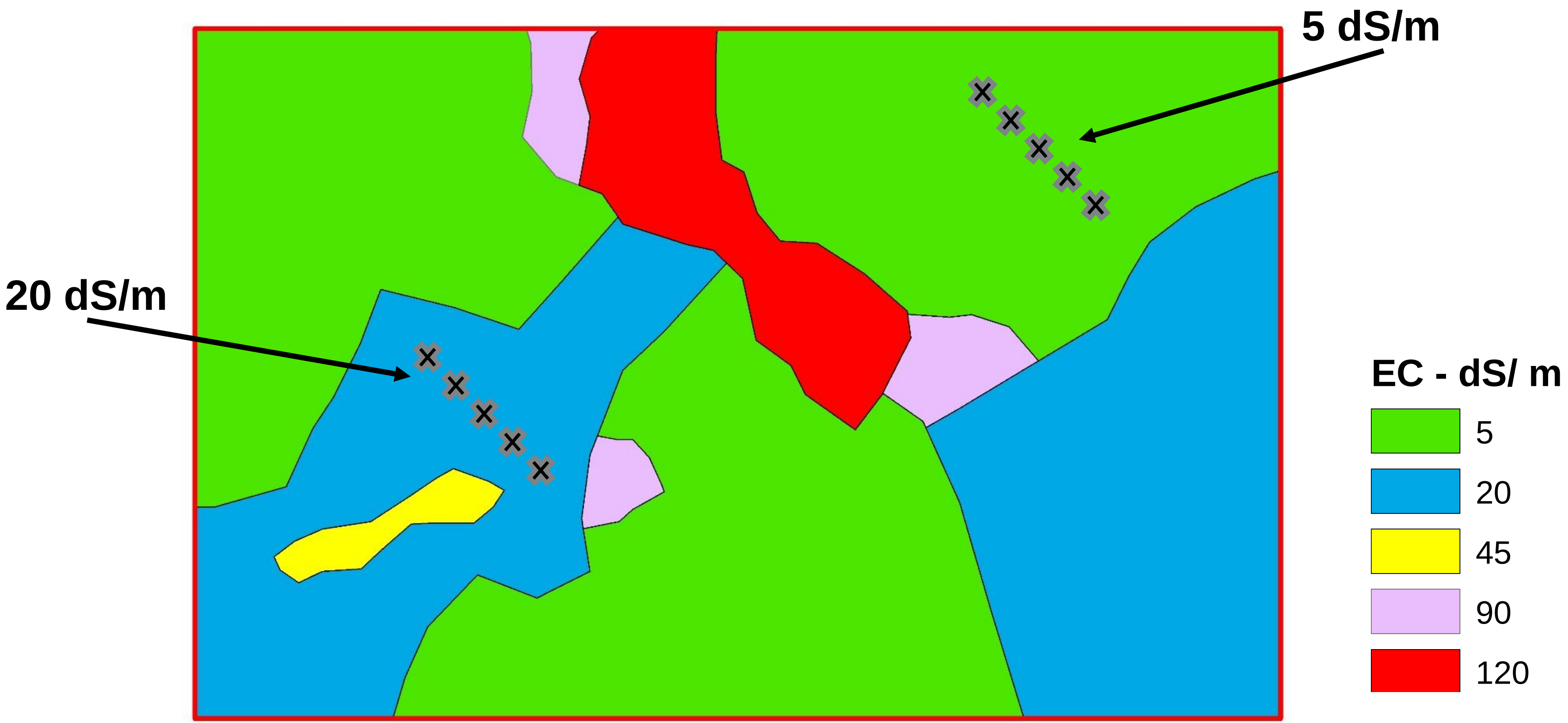




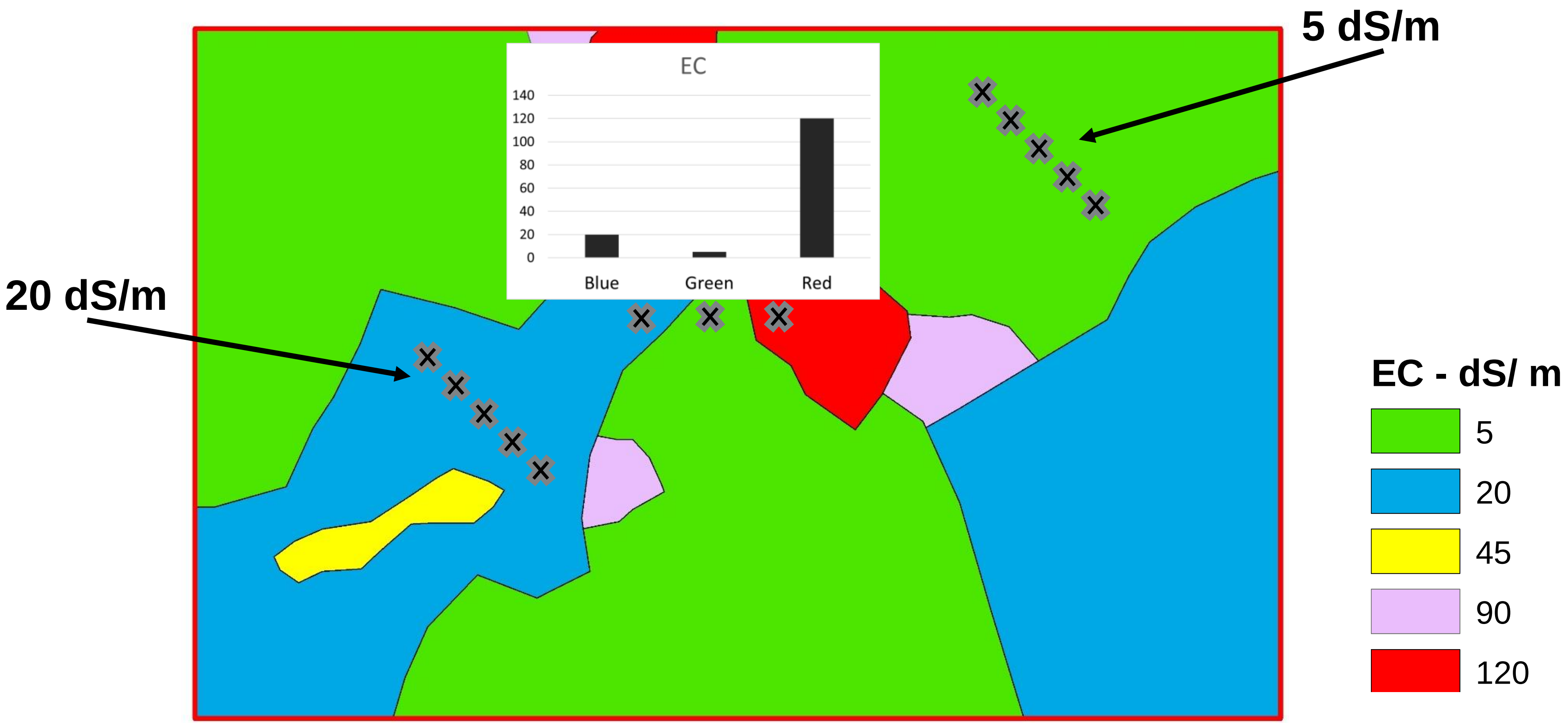
Conventional Soil Mapping



Conventional Soil Mapping



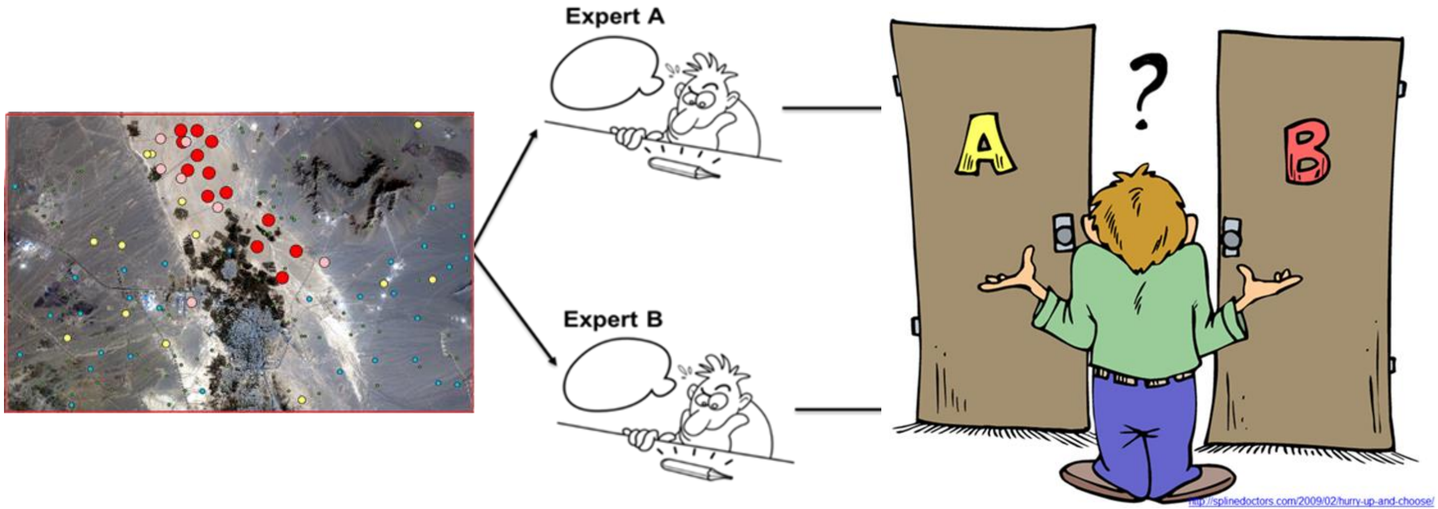
Conventional Soil Mapping



Conventional Soil Mapping

Limitations:

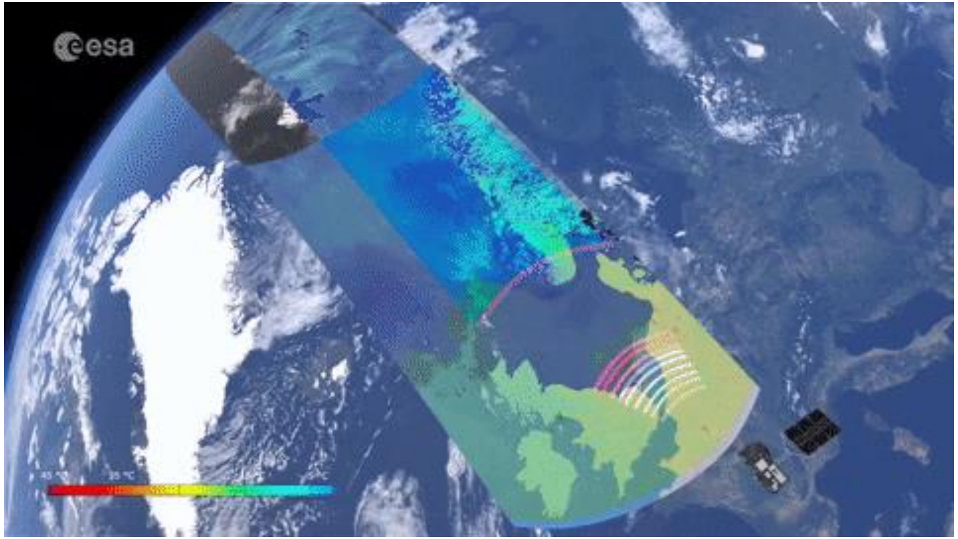
1. Don’t consider the continuous nature of soil variability.
2. Are expensive and time-consuming €
3. Are not suitable for quantitative modelling purposes
4. Depend on the experience and skills of the soil surveyor.
5. Unknown accuracy



http://splinedoctors.com/2009/02/hurry-up-and-choose/

Dent & Young, 1981; Zhu, et al. 2001

Conventional to Digital



Remote sensing

Unmanned aerial vehicle

Dent & Young, 1981; Zhu, et al. 2001

What is digital soil mapping?

- Digital Soil Mapping (DSM), predictive soil mapping, or pedometric mapping = computer assisted production of digital maps of soil type and soil properties, by use of mathematical and statistical **models** that combine information from **soil observations** with information contained in correlated environmental variables (**covariates**).

- DSM = the application of pedometric methods to automate the production of soil maps

- DSM is NOT digitising soil maps

McBratney et al., 2003; Lagacherie& McBratney, 2004

DSM: Some Historical Perspectives

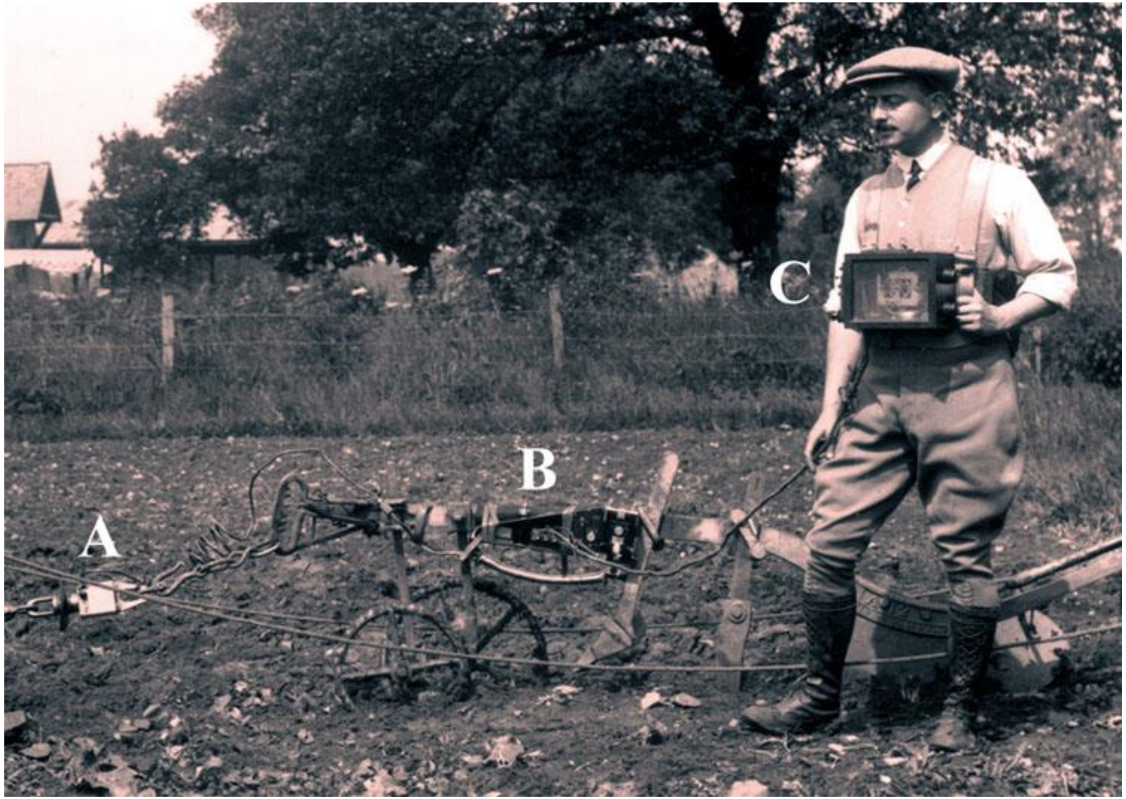
In a broader sense, automatic mapping of soil information.
Achieved in many ways:

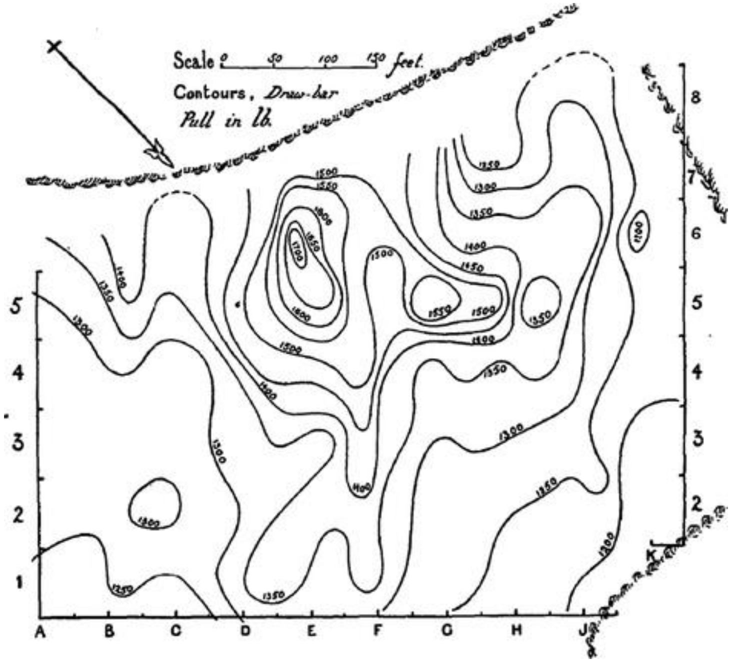
1. Collect dense observation (digital soil cartography)
2. Jenny clorpt functions
3. Interpolation (geostatistics)
4. Combination of clorpt & geostatistics

McBratney et al., 2003; Lagacherie& McBratney, 2004

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Collect Dense Soil Data ‘Automatically’

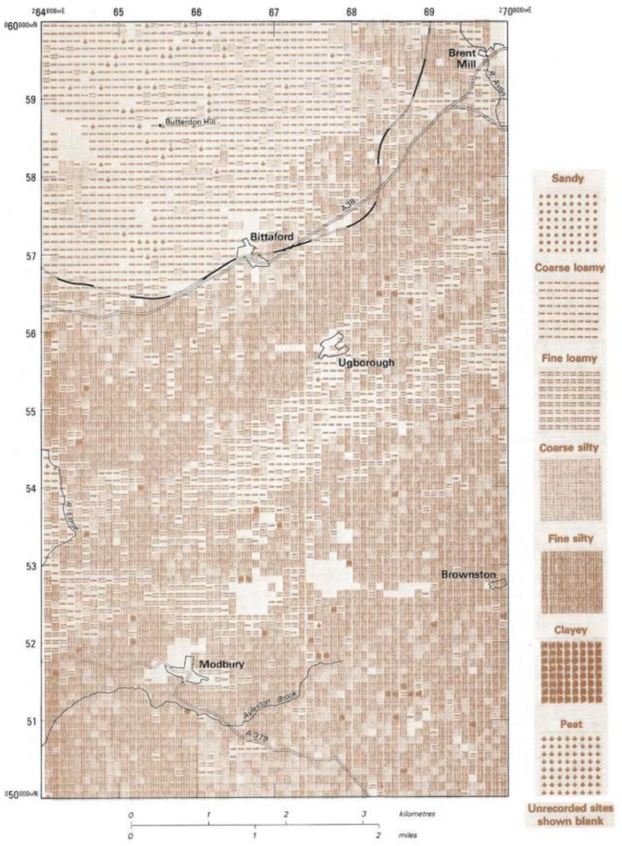




Studies in soil cultivation (Haines & Keen, 1925)

Collect Dense Soil Data ‘Manually’

- Soils of the Ivybridge area in the West of England.
- 60 km^2 inspected from auger borings on a 100-m square grid.
- 60 columns x 101 rows
- At each ‘pixel’ characteristics of the soil profile and site were recorded.



Webster, R., Harrod, T.R., Staines, S.J. & Hogan, D.V. 1979. Grid sampling and computer mapping of the Ivybridgearea, Devon. Soil Survey Technical Monograph No 12, Soil Survey of England and Wales, RothamstedExperimental Station, Harpenden.

Jenny's (clorpt) Function

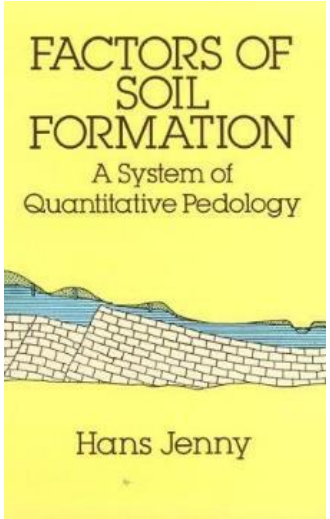
Soil forms under influence of **FIVE** soil forming factors:

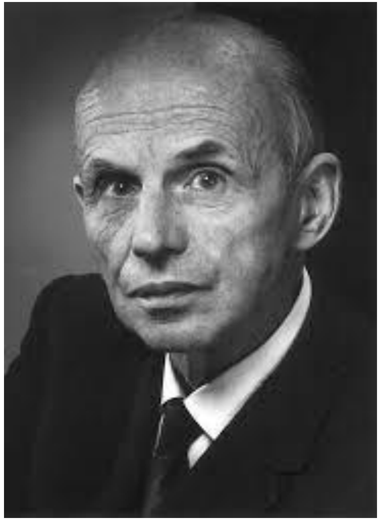
1. climate
2. organisms
3. relief
4. parent material
5. Time

Each factor is an independent variable.

Jenny's (clorpt) equation (1941), soil forming factors:

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





- Since Hans Jenny published his formulation in 1941, it has been used by researchers all over the world as a qualitative list for understanding the factors that may be important for producing the soil pattern within a region (Conventional Soil Mapping).

Jenny’s (clorpt) Function

Soil Science Society of America Journal

Division S-5—Soil Genesis, Morphology & Classification

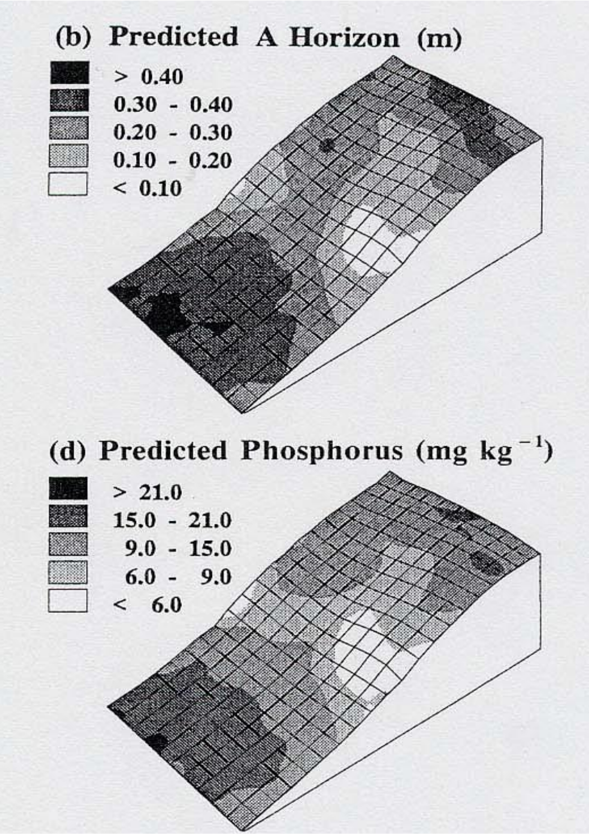
Soil Attribute Prediction Using Terrain Analysis

Correction(s) for this article

I. D. Moore, P. E. Gessler, G. A. Nielsen, G. A. Peterson

First published: 01 March 1993 | https://doi.org/10.2136/sssaj1993.03615995005700020026x |

Moore et al. (1993) gave the first two-dimensional example using a set of terrain attributes derived from a digital elevation model on a 15-m grid to predict continuous soil properties such as A horizon thickness and p for a small catchment in Colorado.



Interpolation (geostatistics)

- Neighbourhood law
- Predict values at unknown locations using values at measured locations
- Many interpolation methods: kriging, IDW, etc

Irrigation Science 1, 197 – 208 (1980)

Irrigation Science
© by Springer-Verlag 1980

Spatial Variability of Soil Sampling for Salinity Studies in Southwest Iran

S. Hajrasuliha, N. Baniabbassi, J. Metthey, and D. R. Nielsen [1]

Received January 21, 1980

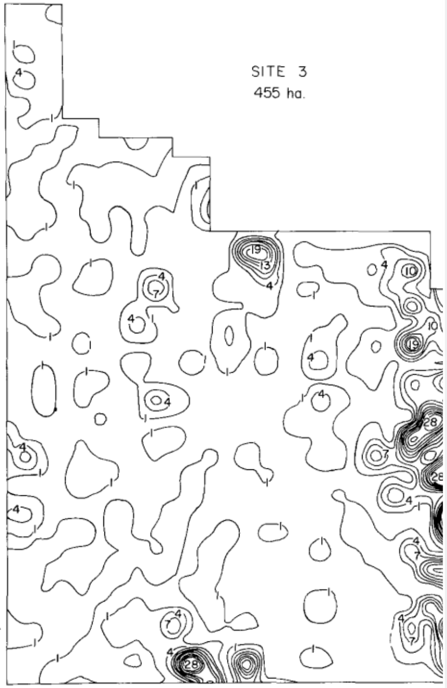


Fig. 5. Isosalinity contour map of Site 3. Numbers refer to EC values in mmho cm^{-1}

Geostatistics + clorpt

- In this approach ‘clorpt’ is used to predict the soil property of interest from environmental variables and kriging is used on the residuals (Regression Kriging).

ELSEVIER

Geoderma 67 (1995) 215–226

GEODERMA

Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging

I.O.A. Odeh [a,b,*], A.B. McBratney [a], D.J. Chittleborough [b]

[a] *Co-operative Research Centre for Sustainable Cotton Production, Department of Agricultural Chemistry and Soil Science, The University of Sydney, Sydney, N.S.W. 2006, Australia*

[b] *The University of Adelaide, Department of Soil Science, Waite Campus, Glen Osmond, S.A. 5064, Australia*

Received 1 February 1994; accepted 18 January 1995

Introducing the SCORPAN Model

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Geoderma 117 (2003) 3–52

GEODERMA

www.elsevier.com/locate/geoderma

On digital soil mapping

A.B. McBratney[a,*], M.L. Mendonça Santos[b], B. Minasny[a]

[a] *Australian Centre for Precision Agriculture, Faculty of Agriculture, Food and Natural Resources, McMillan Building A05, The University of Sydney, Sydney, New South Wales 2006, Australia*

[b] *EMBRAPA-Centro Nacional de Pesquisa de Solos, Rua Jardim Botânico 1024, 22.460-000, Rio de Janeiro, RJ, Brazil*

Received 19 November 2002; received in revised form 14 May 2003; accepted 5 June 2003

Abstract

We review various recent approaches to making digital soil maps based on geographic information systems (GIS) data layers, note some commonalities and propose a generic framework for the future. We discuss the various methods that have been, or could be, used for fitting quantitative relationships between soil properties or classes and their 'environment'. These include generalised linear models, classification and regression trees, neural networks, fuzzy systems and

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

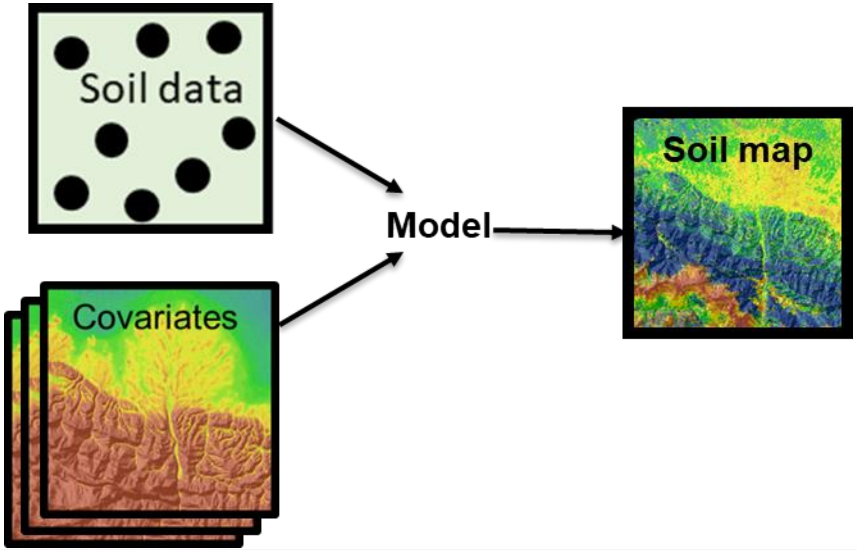
$$S = f(s, c, o, r, p, a, n) + \varepsilon$$

Soil data

Model

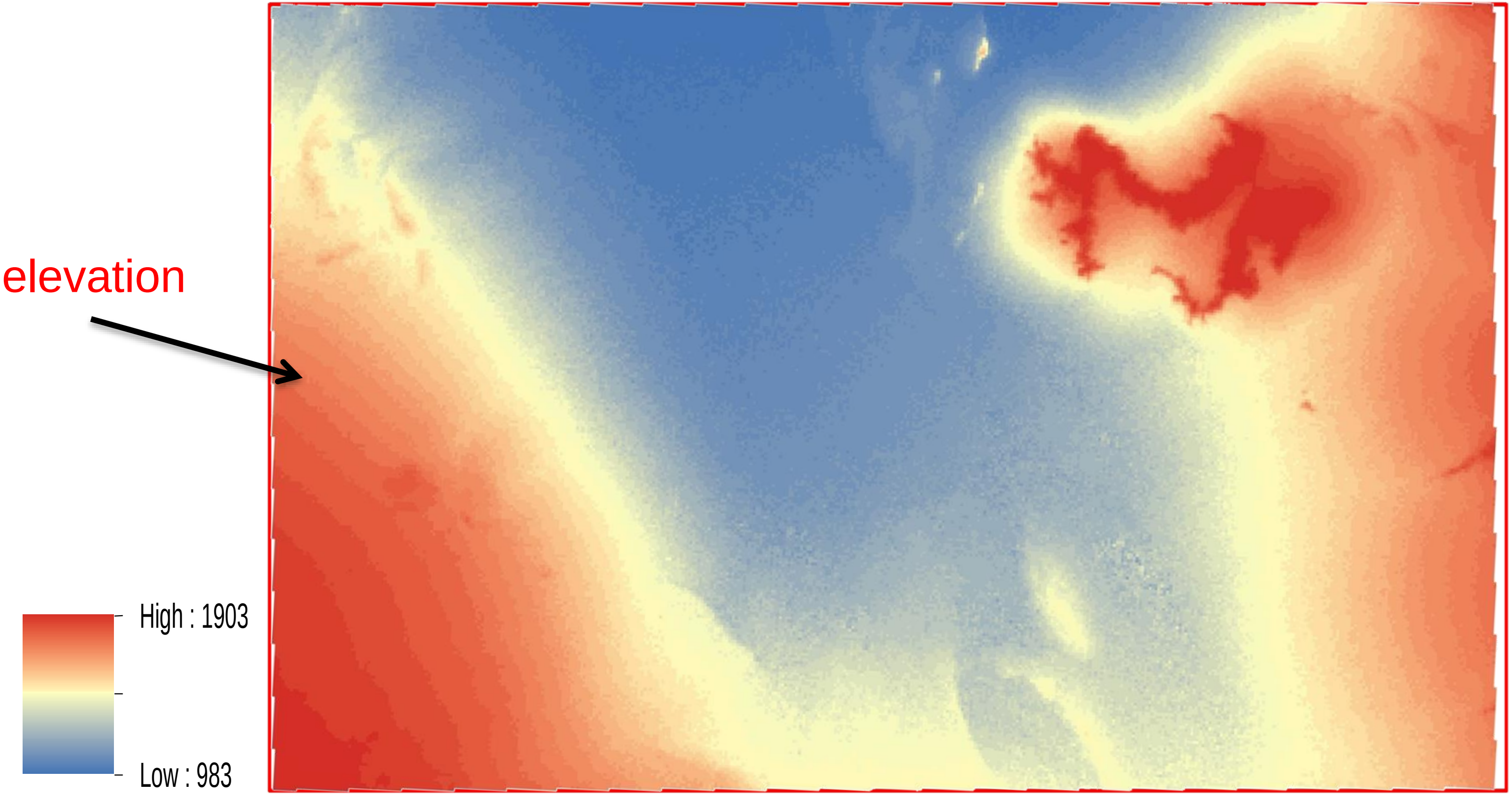
Covariates

spatially dependent residuals



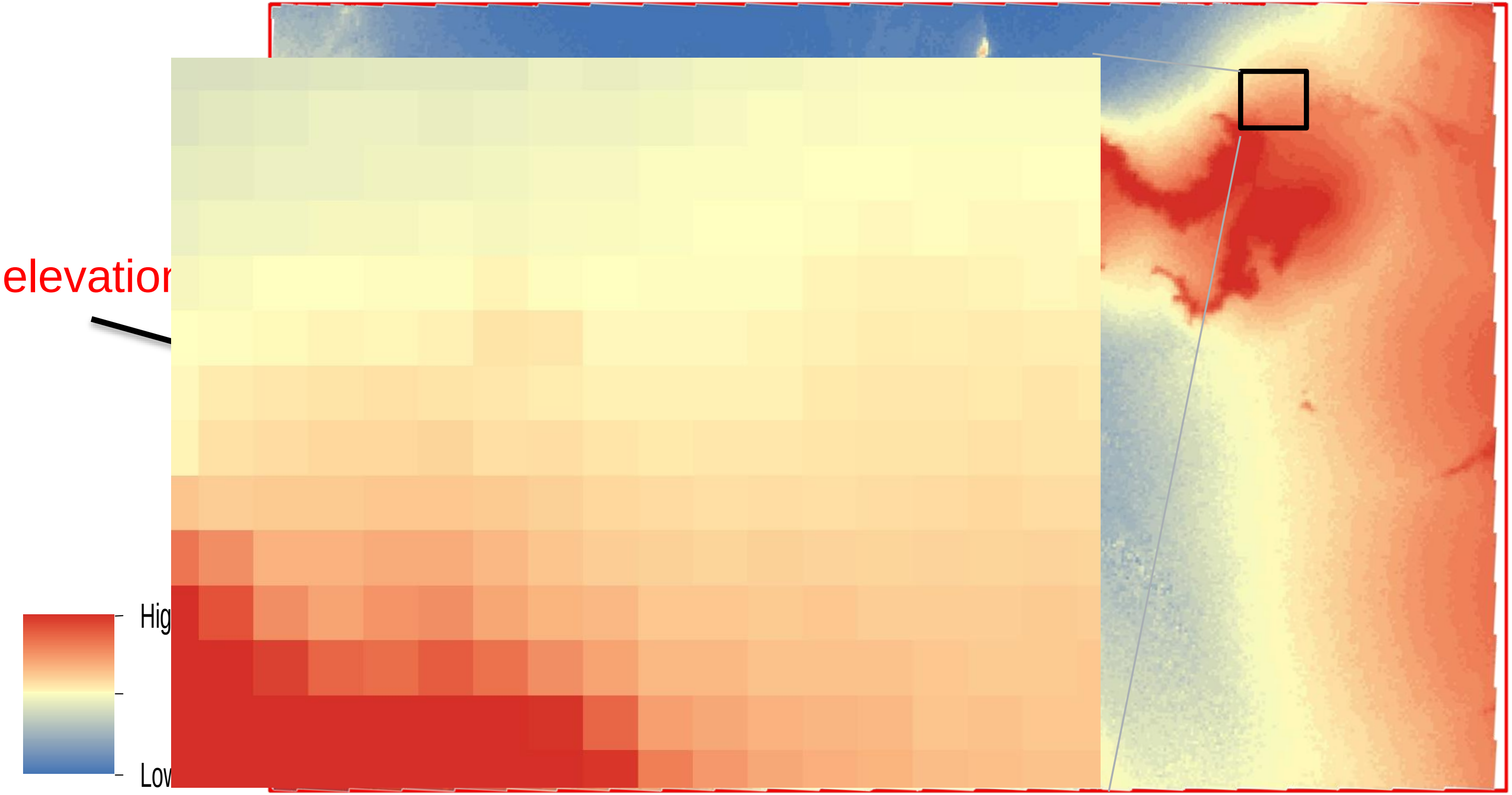
SCORPAN model

elevation



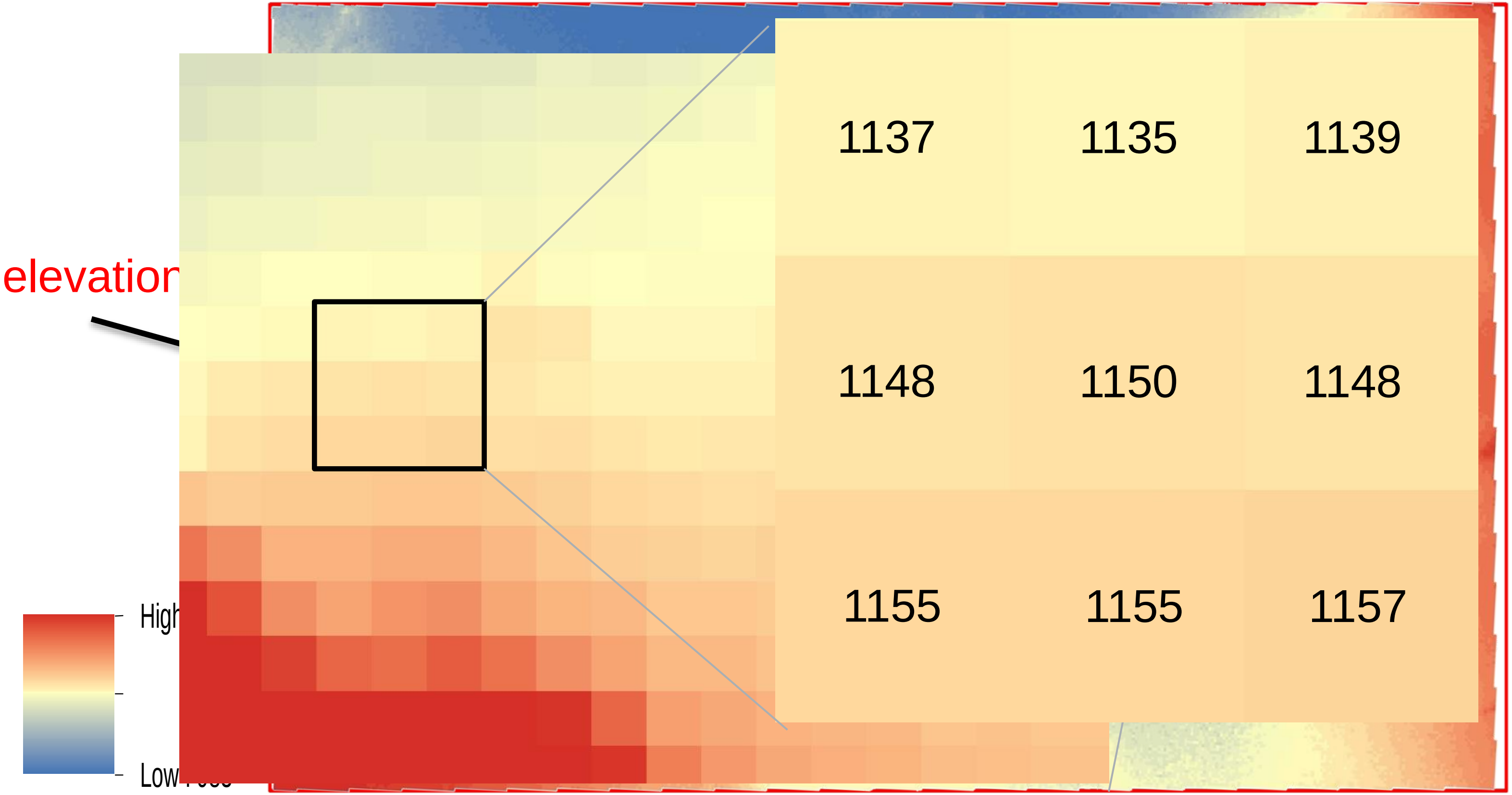
Study area

SCORPAN model



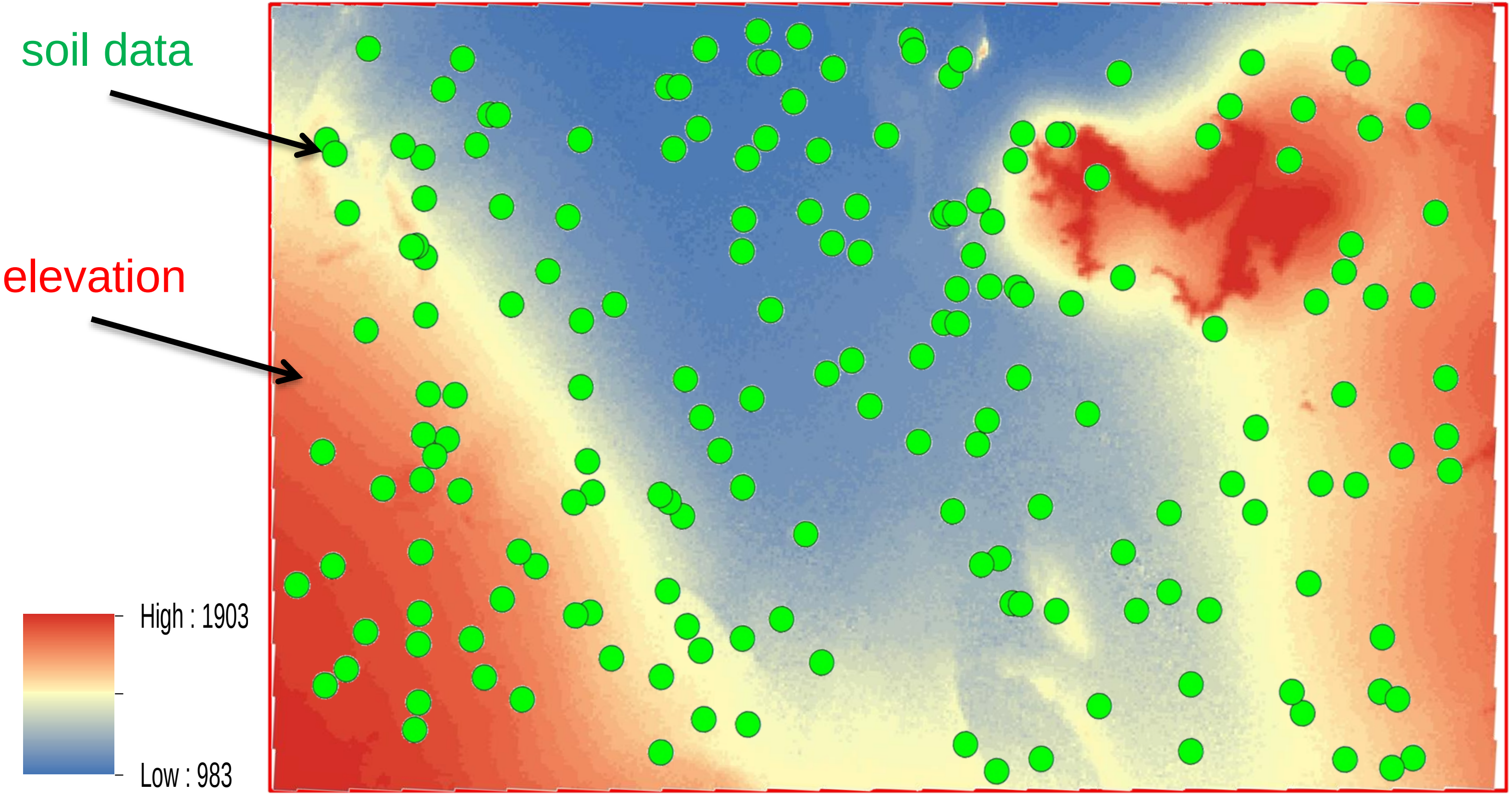
Study area

SCORPAN model



Study area

SCORPAN model

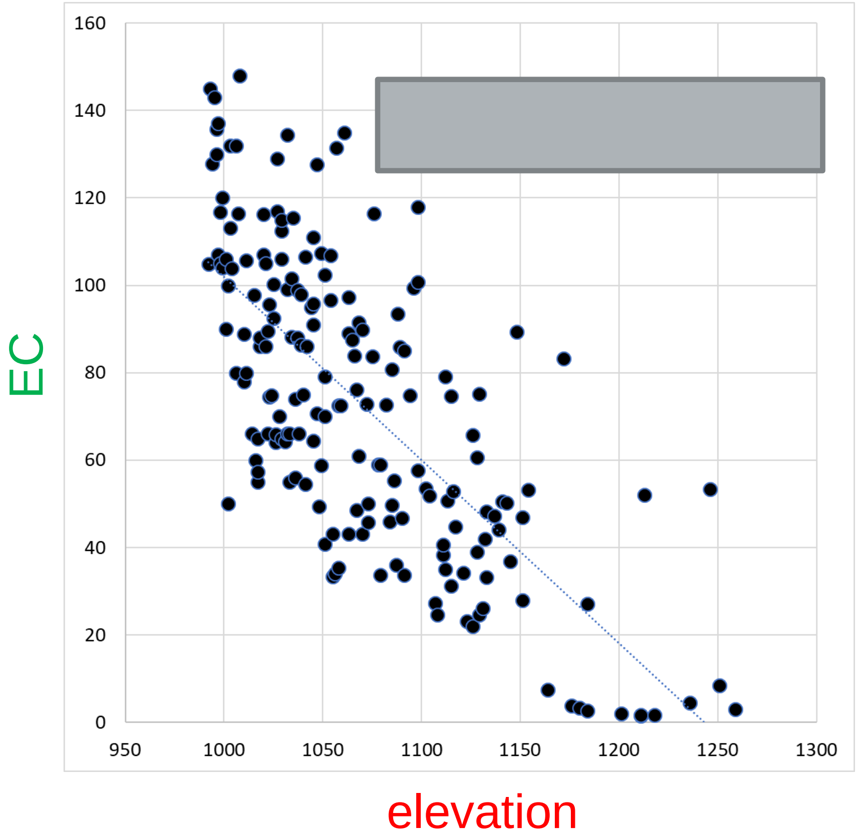


Study area

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

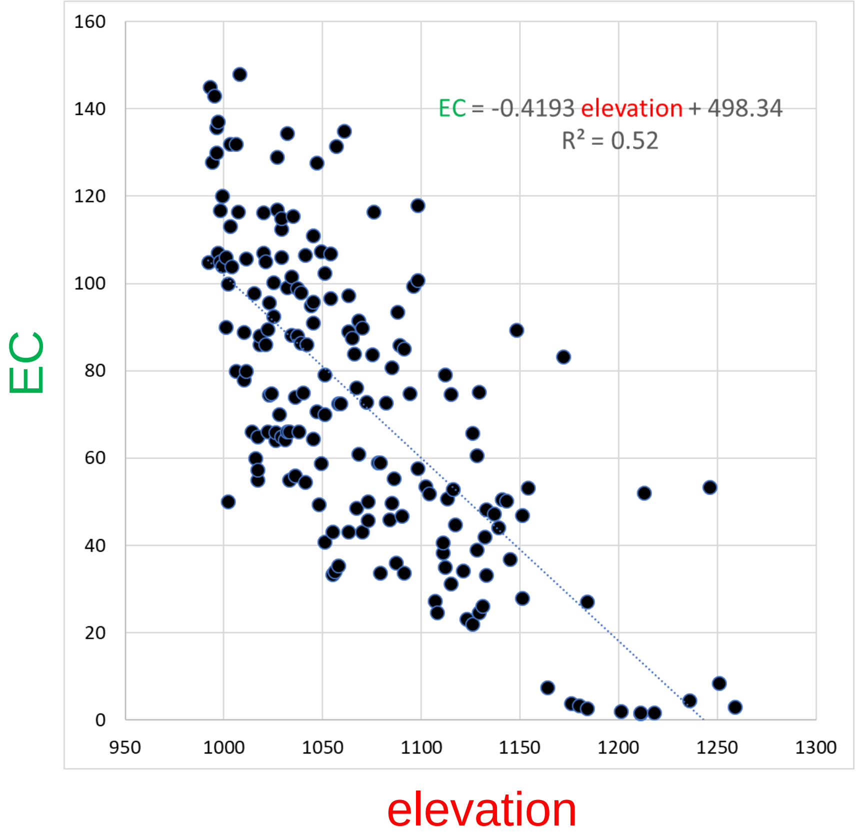
SCORPAN model

No	Elevation	EC
1	1004	138.9
2	1003	232
3	1006	232
4	1008	183
5	1017	5.67
6	1018	34
7	1036	34
8	1045	14.8
9	1029	28.1
10	1026	5.86
11	1037	28.8
12	1033	13.61
13	998	151.8
14	996	135.8
15	997	137.1
16	992	184.8
17	993	209
18	995	227



SCORPAN model

No	Elevation	EC
1	1004	138.9
2	1003	232
3	1006	232
4	1008	183
5	1017	5.67
6	1018	34
7	1036	34
8	1045	14.8
9	1029	28.1
10	1026	5.86
11	1037	28.8
12	1033	13.61
13	998	151.8
14	996	135.8
15	997	137.1
16	992	184.8
17	993	209
18	995	227



EBERHARD KARLS UNIVERSITÄT TÜBINGEN

SCORPAN model

1137	1135	1139
1148	1150	1148
1155	1155	1157

elevation

EC = -0.4193 elevation + 498.34

f()

22	22	21
17	16	17
14	14	13

EC

- Soil Map
- Conventional Soil Map
- Digital Soil Map
- SCORPAN

SCORPAN model

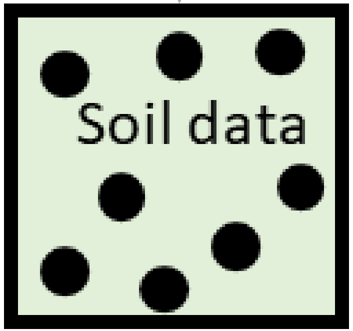
$$\boxed{S} = \boxed{f}\,\boxed{(s,\ c,\ o,\ r,\ p,\ a,\ n)} + \varepsilon$$

Soil data

Model

Covariates

spatially dependent residuals



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

Soil Data (Sampling)

- Sampling is the process of selecting a sample from a population.

- Lack of funds for sampling and analysis is a major problem for any soil mapping project

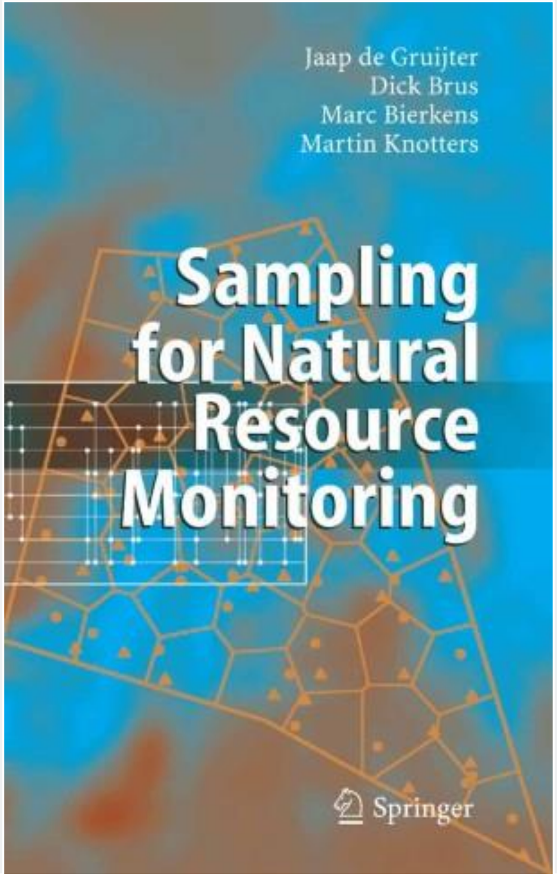
- Soil sampling plan? Optimal?

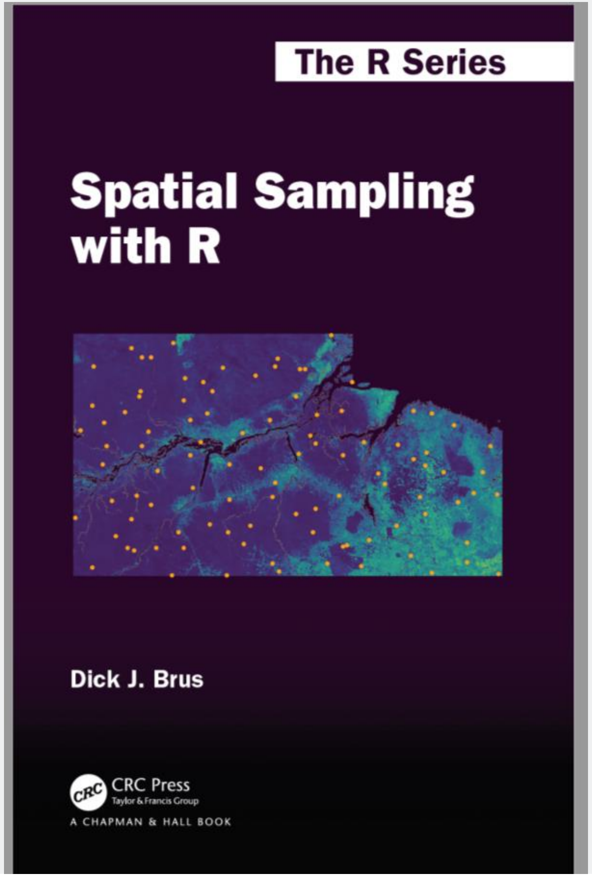
- Design-based
- Model-based

- Probability
- Non-Probability

- Population
- Map

Sampling: Useful Resources





ELSEVIER
Geoderma
Volume 338, 15 March 2019, Pages 464-480

Sampling for digital soil mapping: A tutorial supported by R scripts

D.J. Brus [a, b]

Show more

https://doi.org/10.1016/j.geoderma.2018.07.036

Get rights and content

ELSEVIER
Computers & Geosciences
Volume 32, Issue 9, November 2006, Pages 1378-1388

A conditioned Latin hypercube method for sampling in the presence of ancillary information ☆

Budiman Minasny, Alex B. McBratney

SCORPAN Model

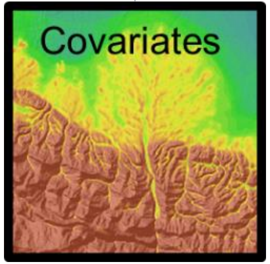
$$\boxed{S} = \boxed{f}\boxed{(s,\ c,\ o,\ r,\ p,\ a,\ n)} + \varepsilon$$

Soil data

Model

Covariates

spatially dependent residuals



Environmental Covariates

s
c
o
r
p
a
n

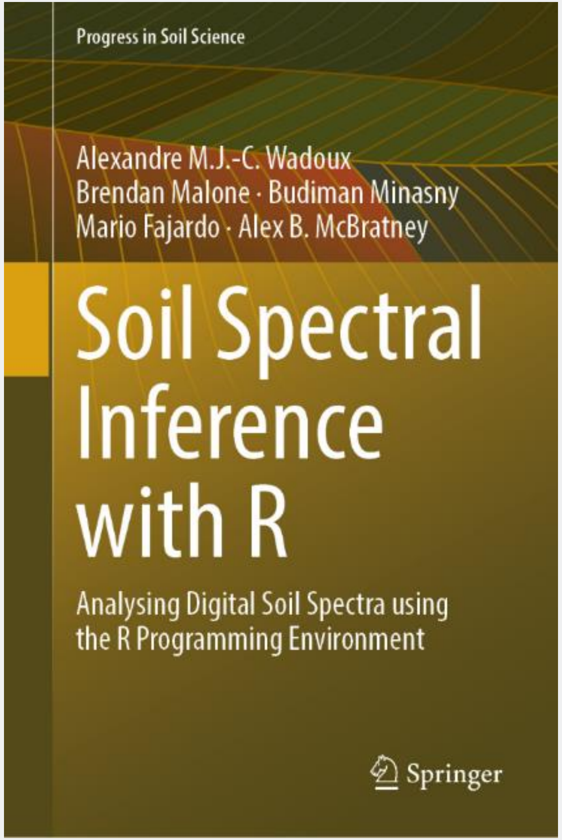
Possible sources of information to represent seven scorpan factors

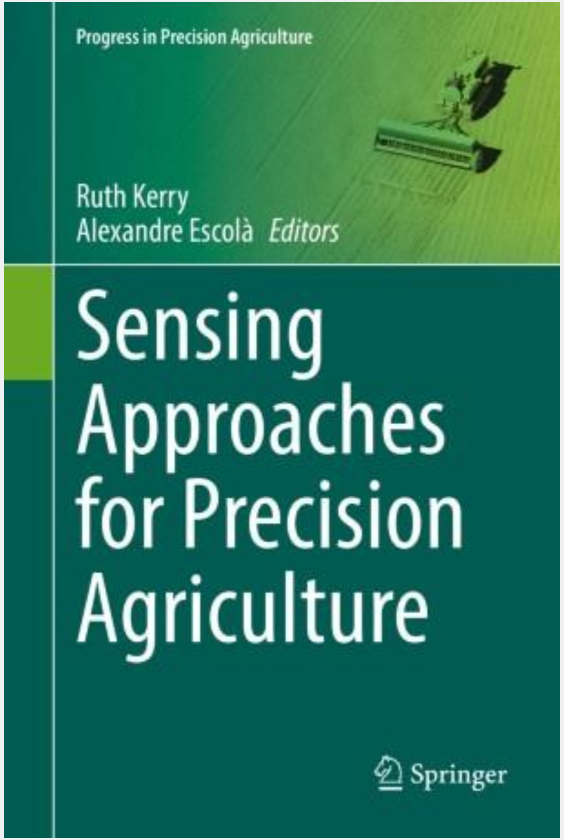
scorpan factor	Possible representatives
s	Legacy soil maps, point observations, expert knowledge
c	Temperature and precipitation records
o	Vegetation maps, species abundance maps, yield maps, land use maps
r	Digital elevation model, terrain attributes
p	Legacy geology maps, gamma radiometric information
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

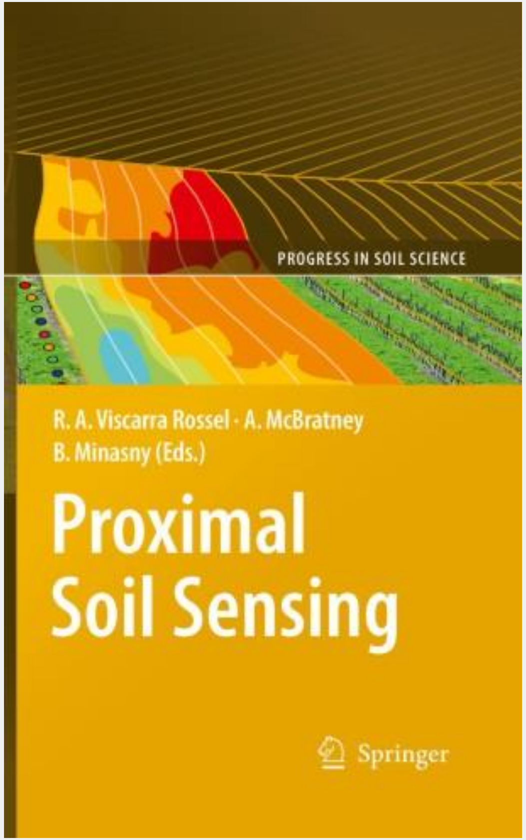
- Proximal sensing, Remote sensing, and digital elevation models are three prominent examples of high-resolution environmental covariates can be used to represent various scorpan factors

McBratney et al., 2003

Soil Sensing: Useful Resources

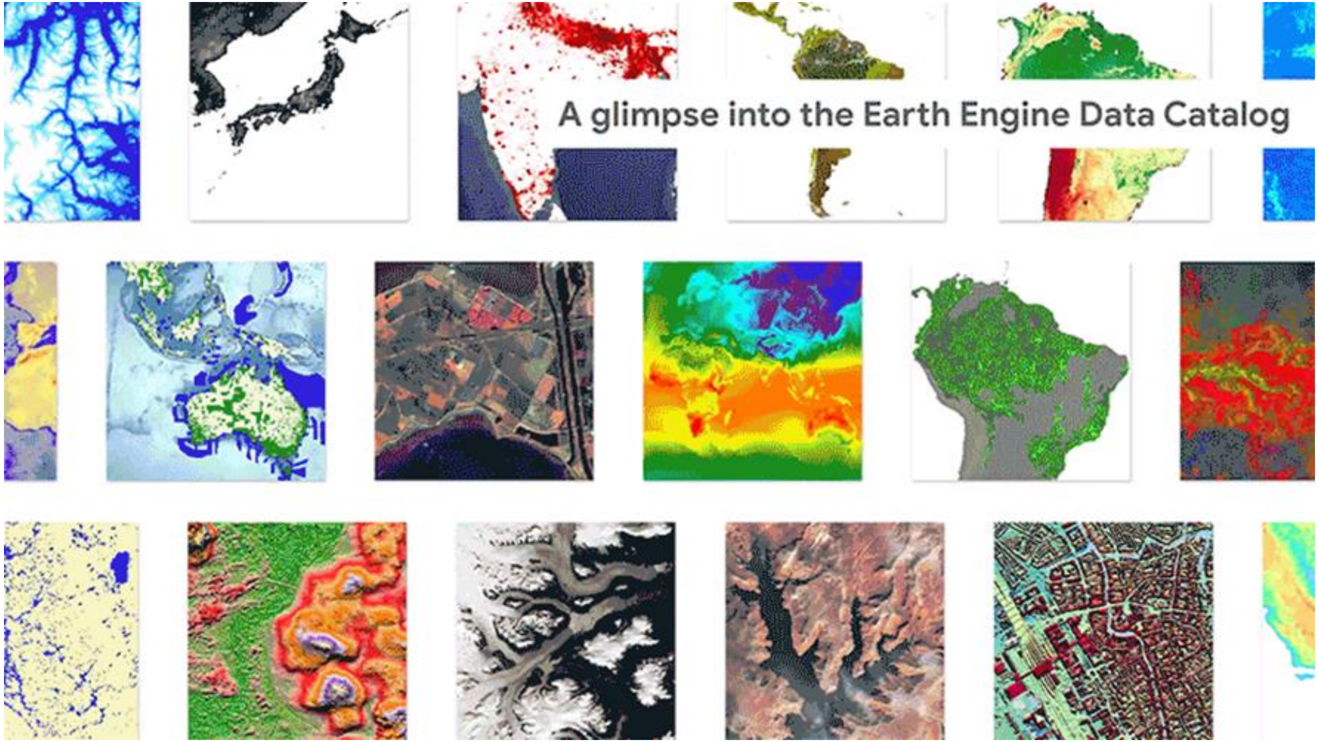






Where Can I Obtain the Covariates?

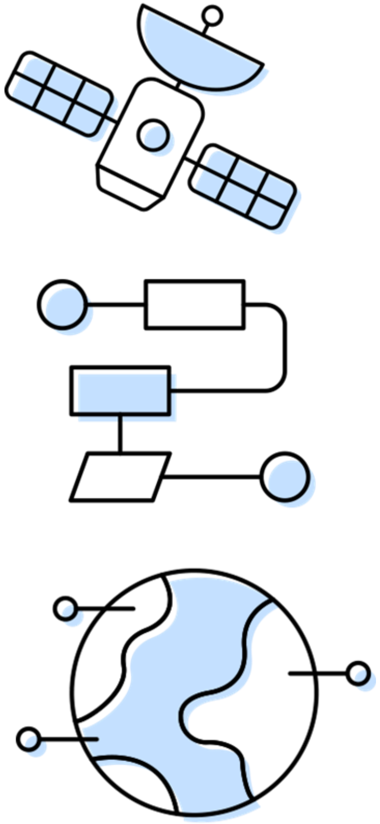
s
c
o
r
p
a
n



- What is Google Earth Engine?

GEE is a

public data catalog,
compute infrastructure,
geospatial APIs,
and an interactive app server



The Earth Engine Public Data Catalog

Landsat and Sentinel
Raw, TOA, SR, ...

MODIS
Daily, NBAR, LST, ...

Terrain
SRTM, GTOPO, NED, ...

Land Cover
GlobCover, NLCD, ...

Atmospheric
NOAA NCEP, OMI, ...

... and many more, updating daily!

> 200 public datasets

> 4000 new images every day

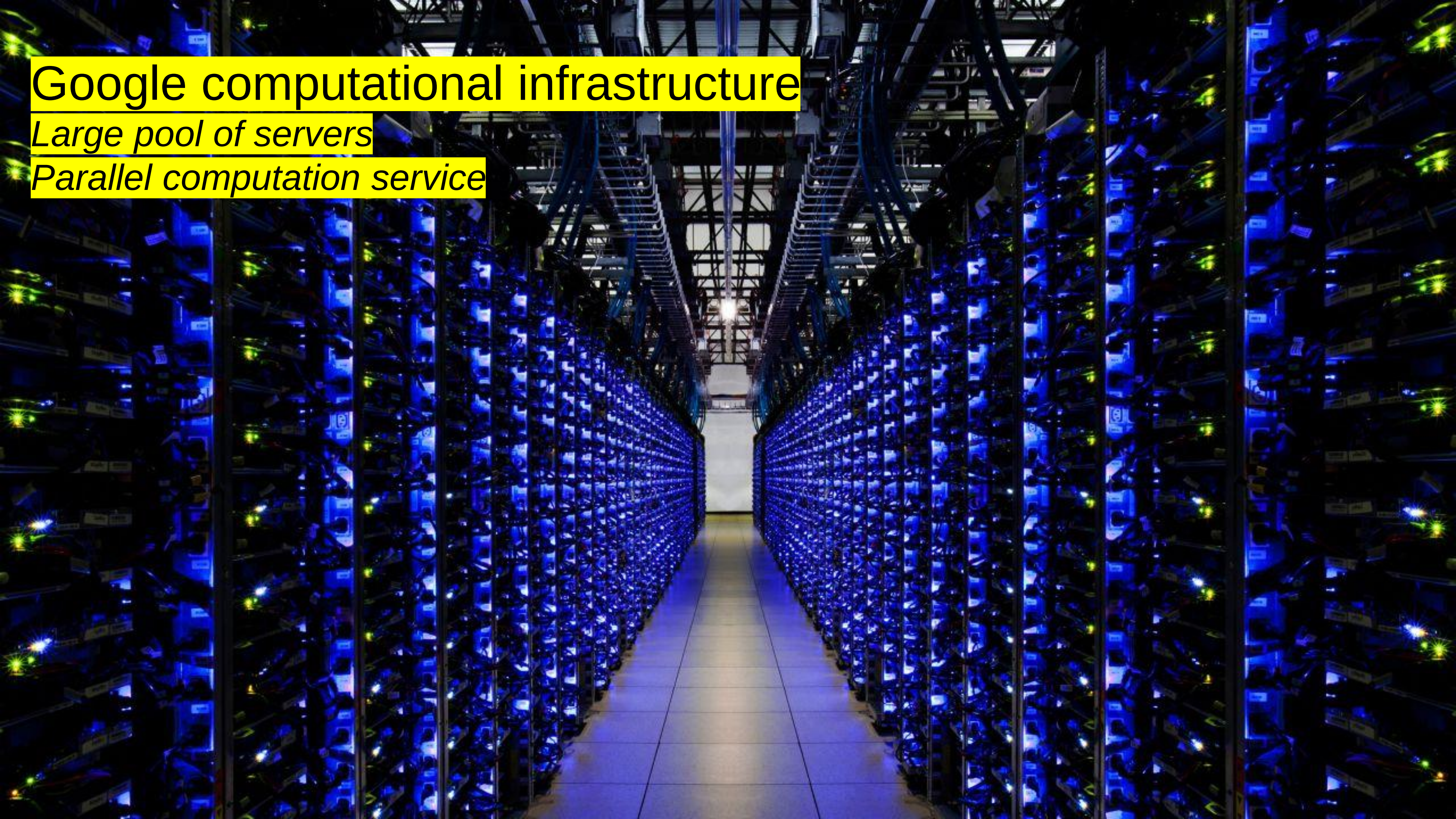
> 5 million images

> 5 petabytes of data

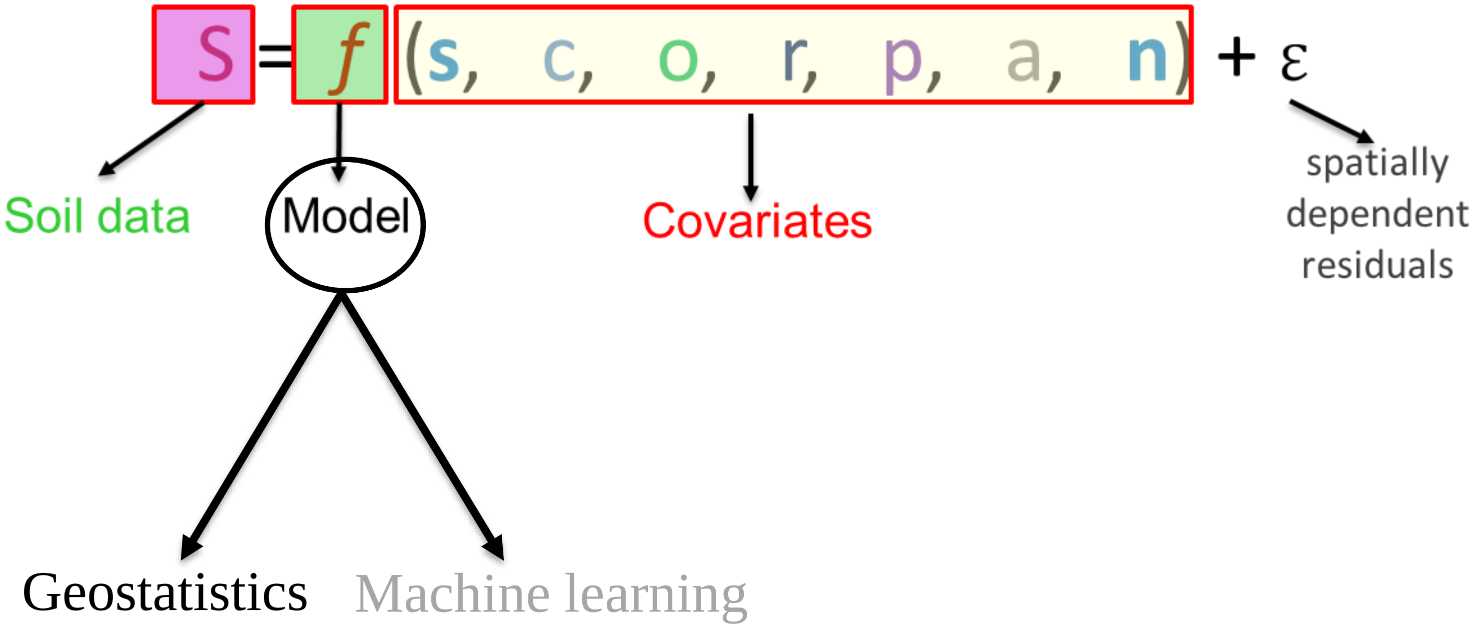
Google computational infrastructure

Large pool of servers

Parallel computation service



SCORPAN Model

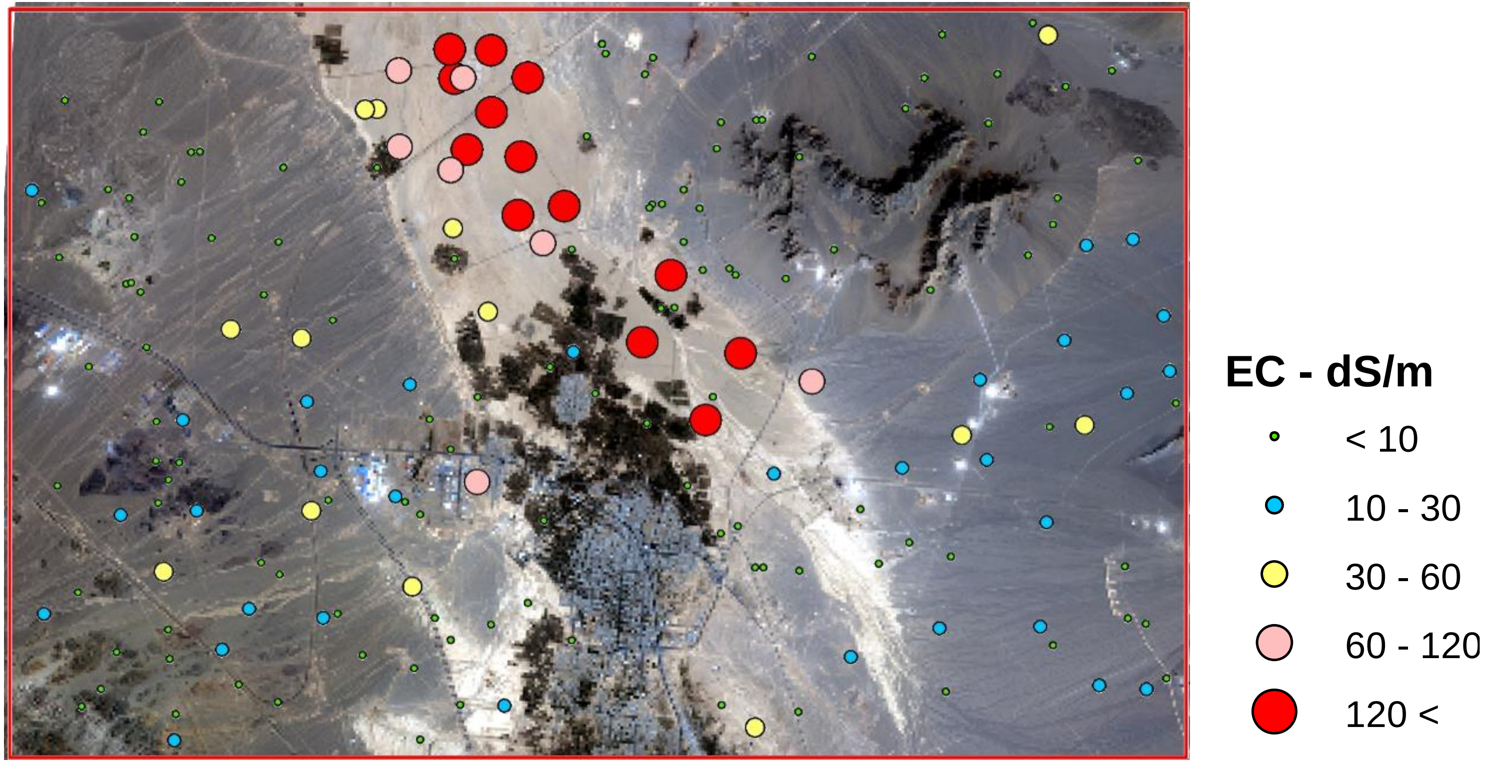


f() : Quantitative function *f* linking S to ***scorpan*** factors

UNIVERSITÄT TÜBINGEN

We Know,

- Soils and soil properties vary in space

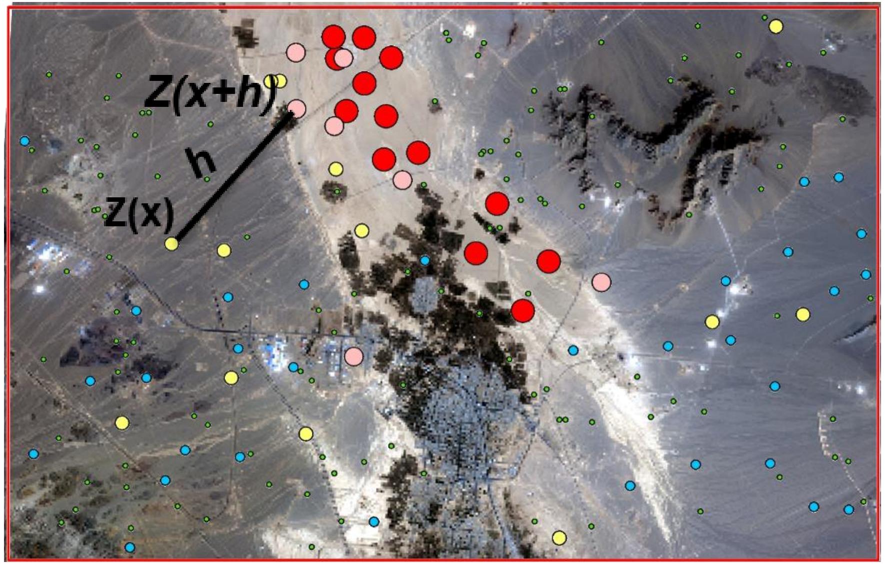


The Variogram: The Central Tool of Geostatistics

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

- Semi-variance is calculated according to:

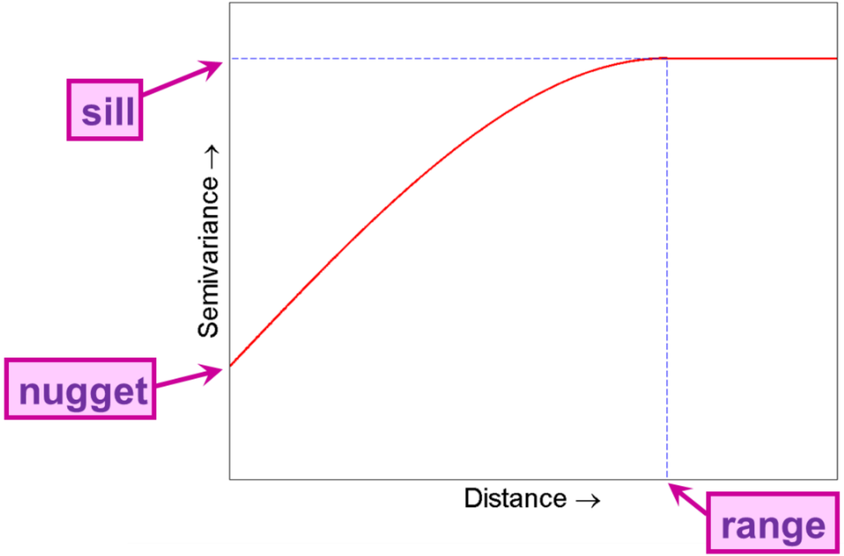
$$\gamma(h)=\frac{1}{2M(h)}\sum_{i=1}^{M(h)}\{\underbrace{z(x_i)}_{\text{measurement at location x}}-\underbrace{z(x_{i}+h)}_{\text{measurement at location x+h}}\}^2$$



(Source: Heuvelink, 2017, ISRIC)

The Variogram: The Central Tool of Geostatistics

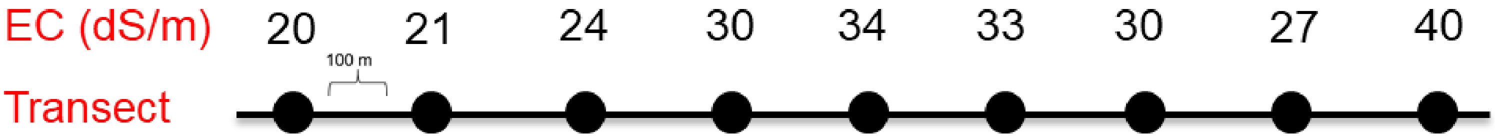
- Plot of semi-variance as a function of the distance is called a semi-variogram



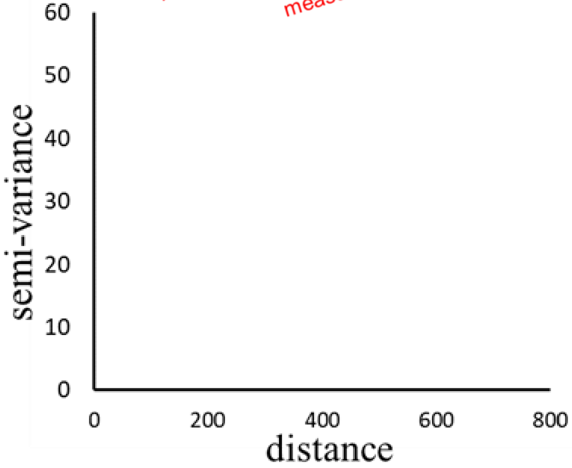
- **Nugget** = measurement errors and short-distance spatial variation
- **Sill** = variance of the variable of interest
- **Range** = distance up to which there is spatial correlation

(Source: Heuvelink, 2017, ISRIC)

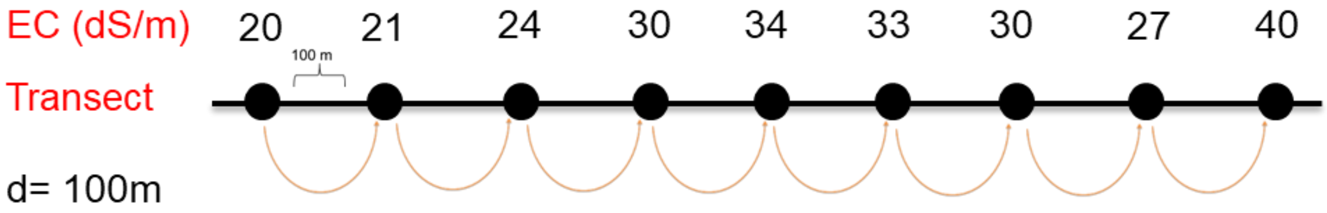
Calculate Variogram: example



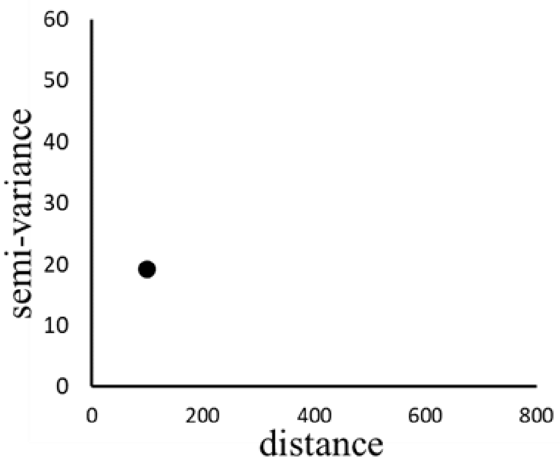
$$\gamma(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{\underbrace{z(x_i)}_{\text{measurement at location x}} - \underbrace{z(x_i + h)}_{\text{measurement at location x+h}}\}^2$$



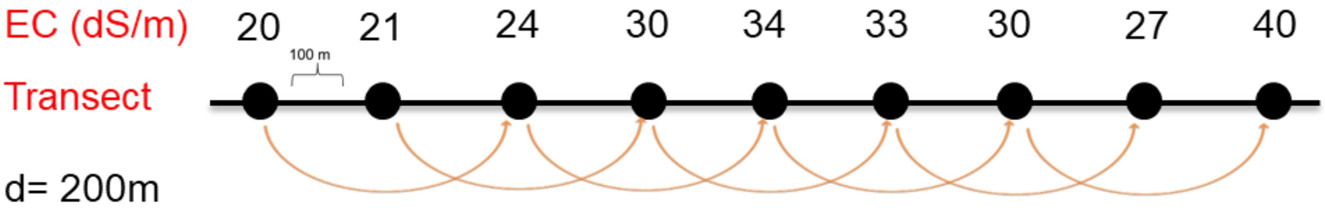
Calculate Variogram: example



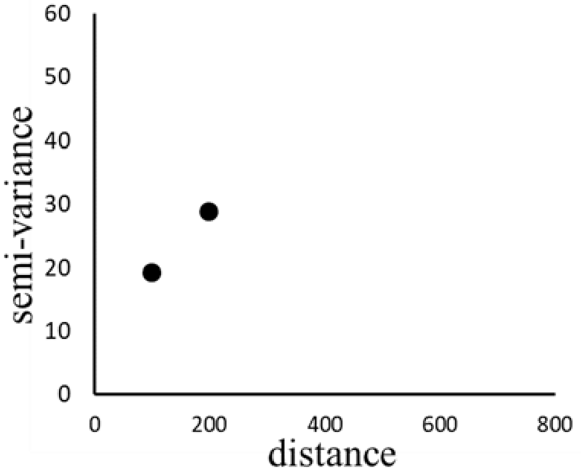
$$\gamma(100) = \frac{1}{16}\begin{bmatrix}(20-21)^2 + (21-24)^2 + (24-30)^2 + (30-34)^2 \\ +(34-33)^2 + (33-30)^2 + (30-27)^2 + (27-40)^2\end{bmatrix} = 19.12$$



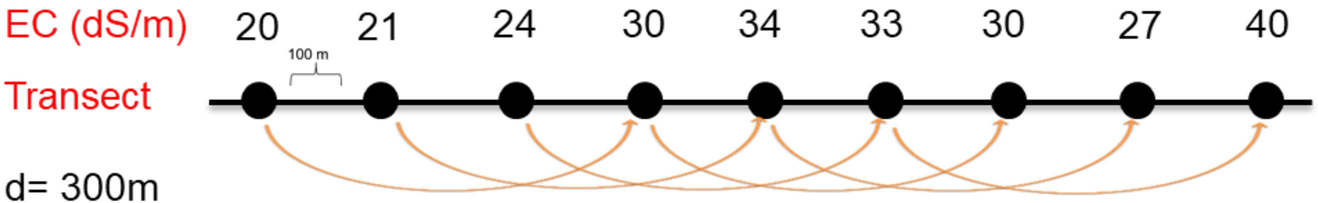
Calculate Variogram: example



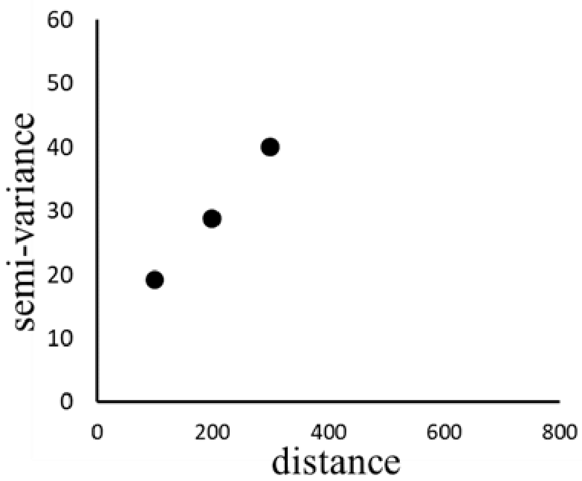
$$\gamma(200) = \frac{1}{14}\left[\begin{array}{c}(20-24)^2 + (21-30)^2 + (24-34)^2 + (30-33)^2 \\ +(34-30)^2 + (33-27)^2 + (30-40)^2\end{array}\right] = 28.71$$



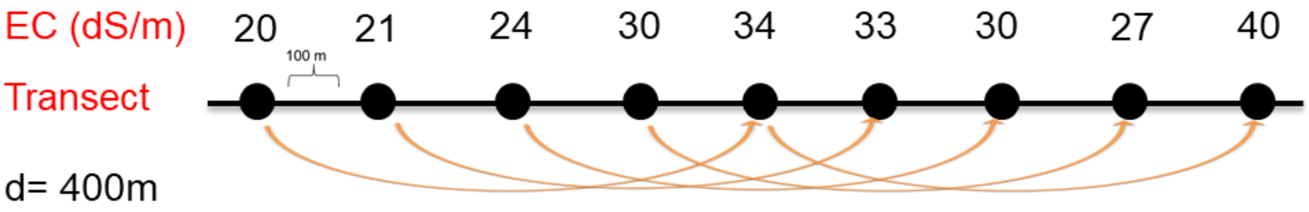
Calculate Variogram: example



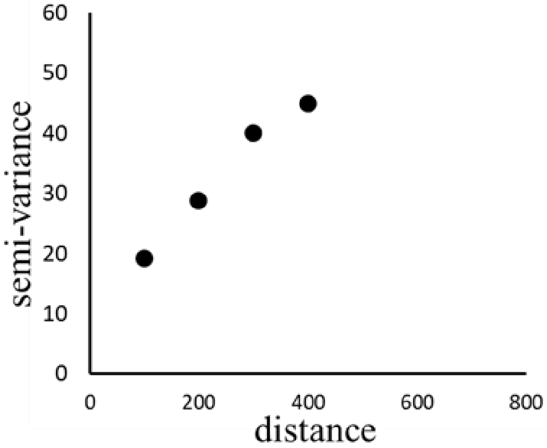
$$\gamma(300) = \frac{1}{12}\left[\begin{array}{c}(20-30)^2 + (21-34)^2 + (24-33)^2 + (30-30)^2 \\ +(34-27)^2 + (33-40)^2\end{array}\right] = 40.00$$



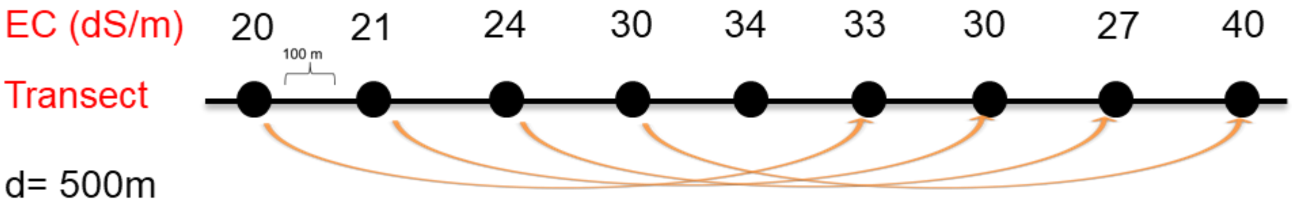
Calculate Variogram: example



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

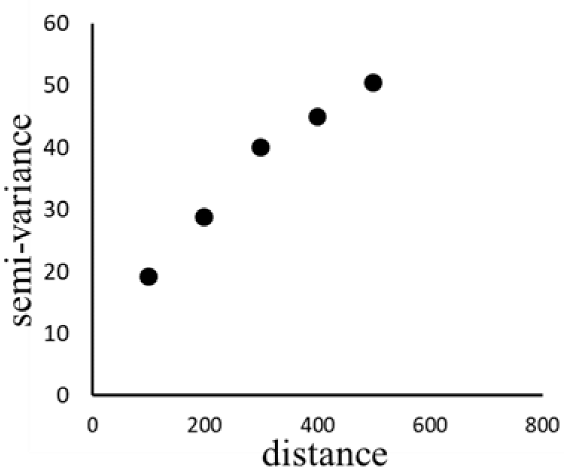


Calculate Variogram: example



d= 500m

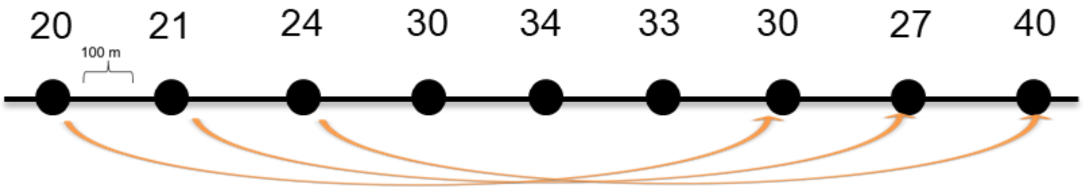
$$\gamma(500) = \frac{1}{8}\left[(20-33)^2 + (21-30)^2 + (24-27)^2 + (30-40)^2\right] = 50.37$$



Calculate Variogram: example

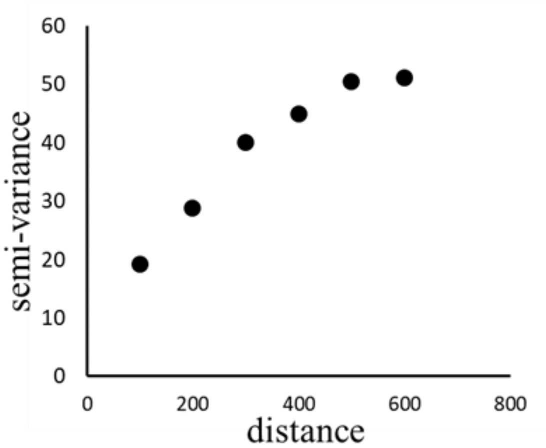
EC (dS/m)

Transect



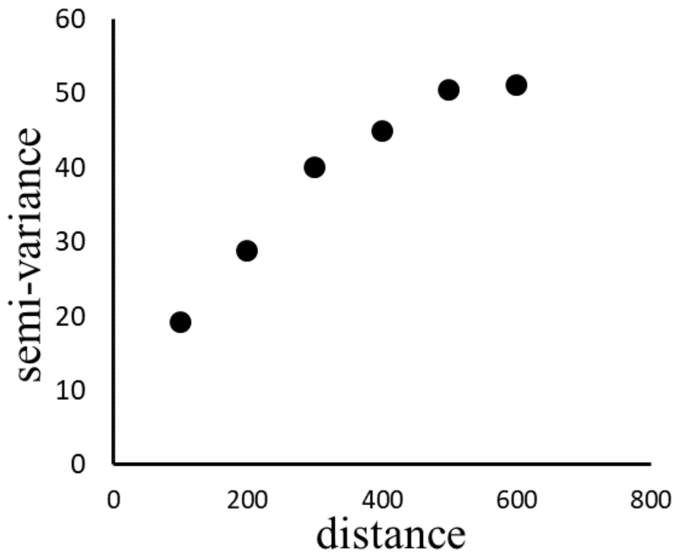
d= 600m

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



Calculate Variogram: example

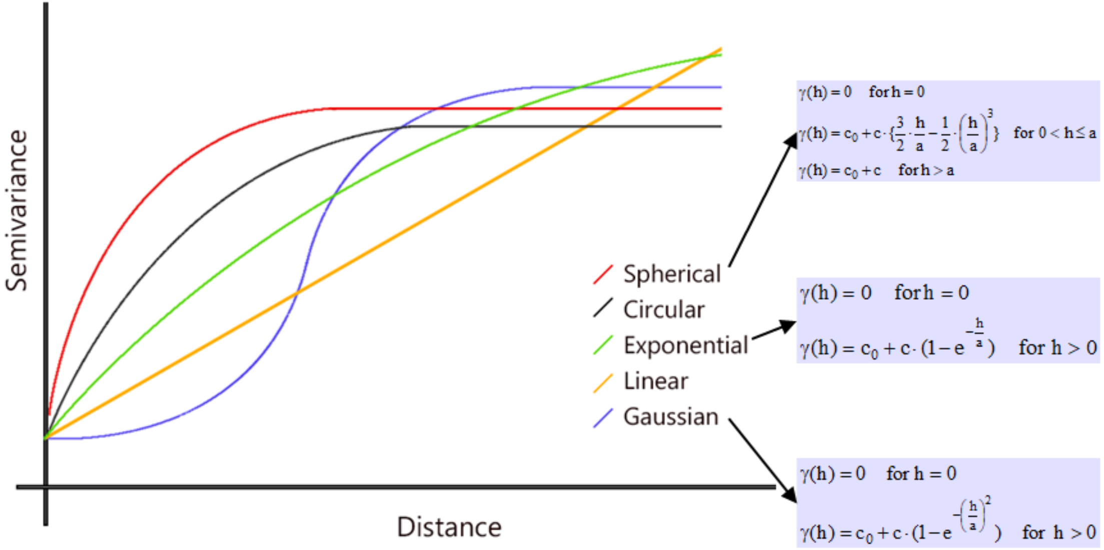
- Experimental semi-variogram is a plot of semi-variance of observations at a number of lag distances
- This is modelled with a continuous function



Calculate Variogram: example

- There are a number of authorized models that can be used e.g. spherical, exponential, Gaussian

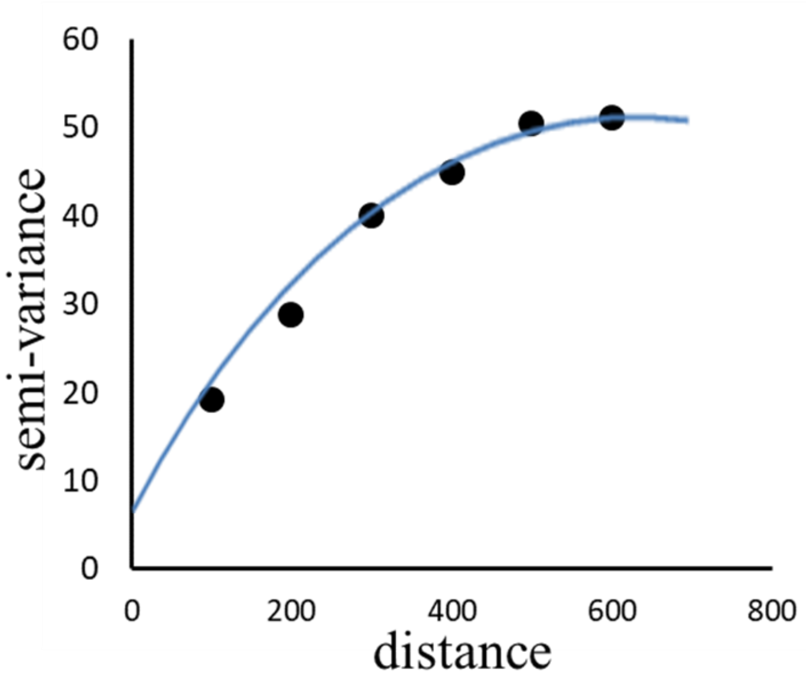
1. Choose a function shape
2. Estimate parameters of the chosen shape (e.g. using weighted least squares fitting)



EBERHARD KARLS UNIVERSITÄT TÜBINGEN

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*

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

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 :

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

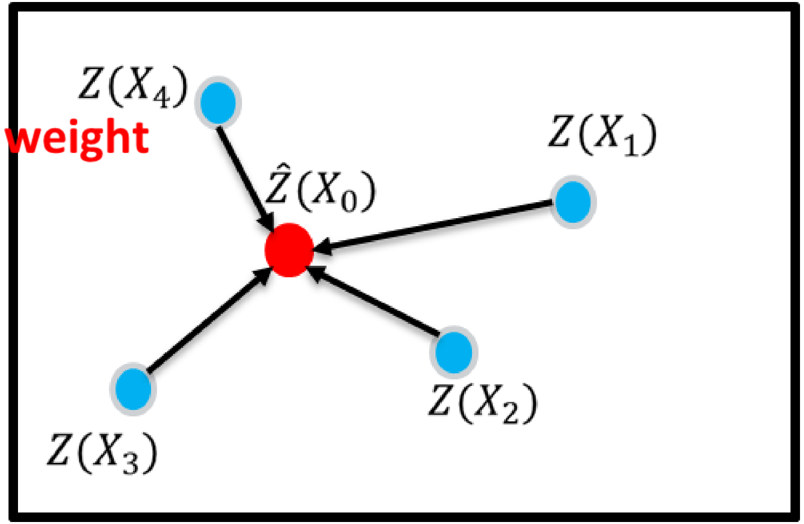
Kriging weight

IDW

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

X_0 unknown value to be estimated

X_i a set of known measurements



Ordinary Kriging

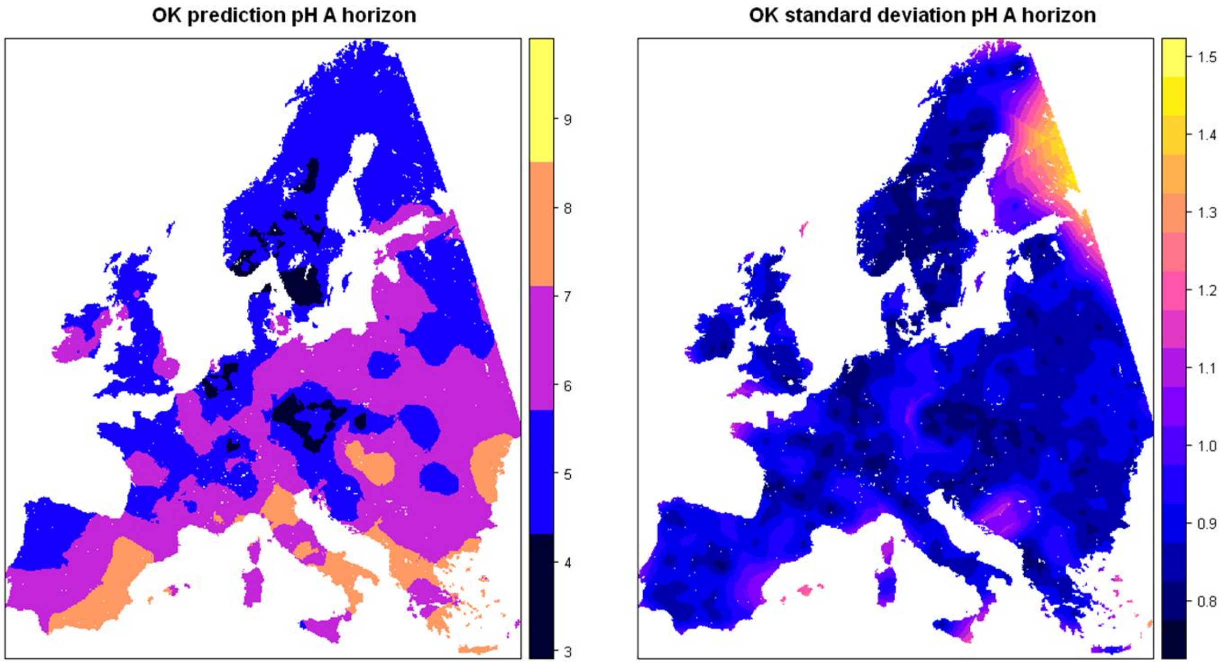
- kriging tries to choose the optimal weights that produce the minimum estimation error.

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

- under the condition:

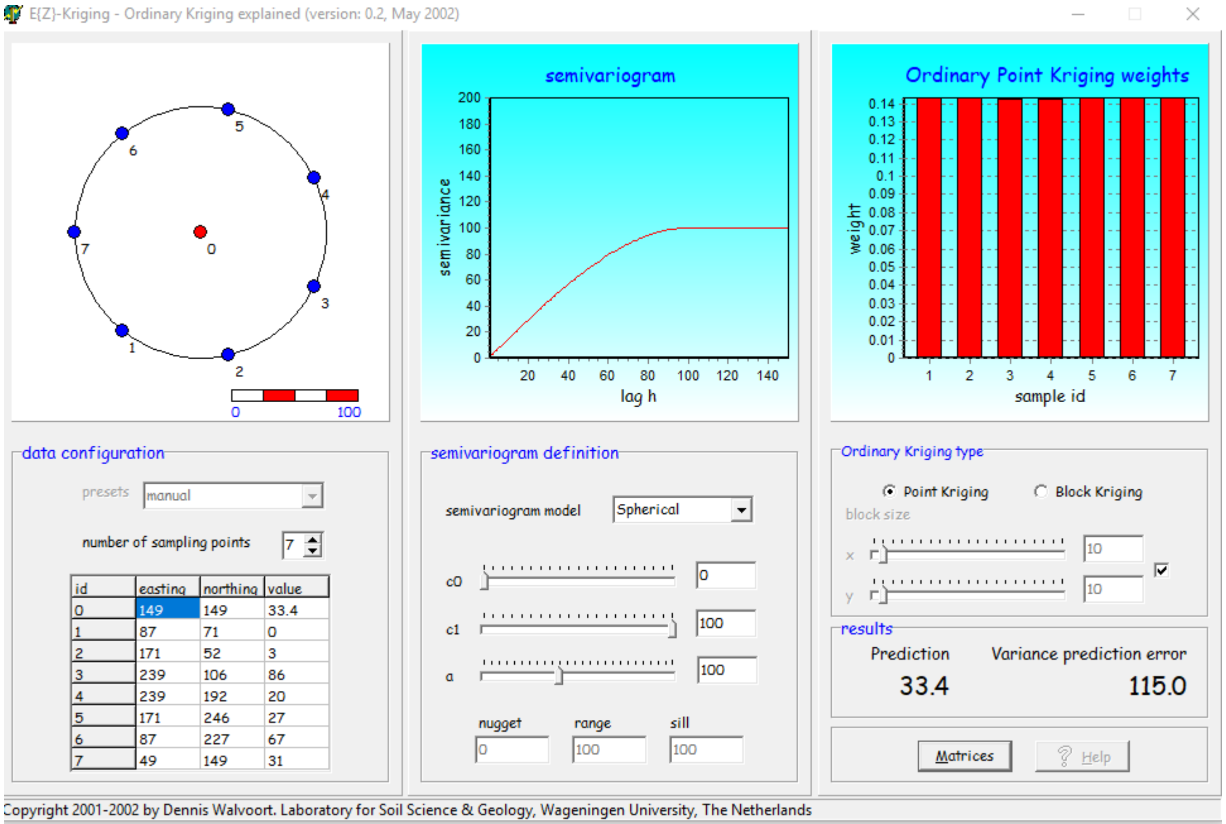
$$\sum_{i=1}^{n} \lambda_i = 1$$

Ordinary Kriging: example



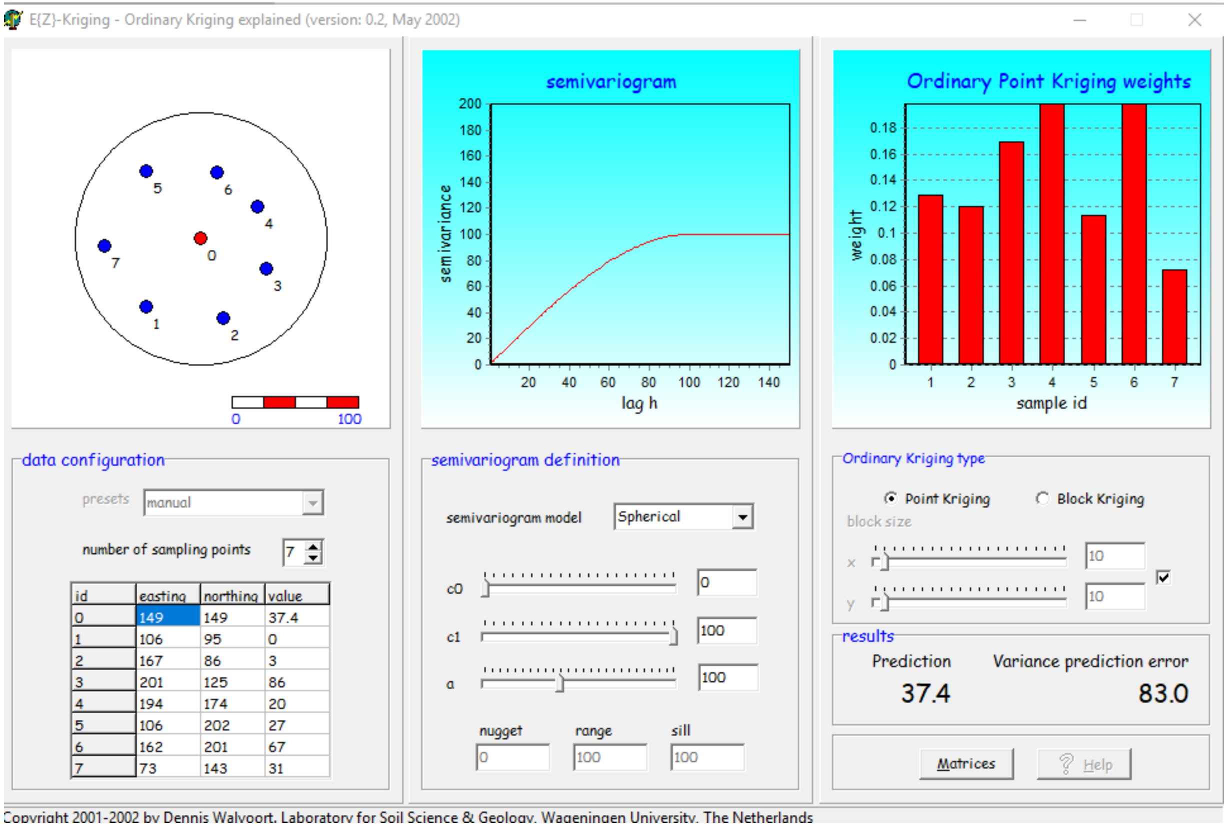
Estimation (left) and standard deviation (right) maps of pH by ordinary kriging. (Source: Heuvelink, 2017, ISRIC)

E{Z}-Kriging



https://wiki.52north.org/pub/AI_GEOSTATS/SWEZKriging/E_Z_Kriging.pdf

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

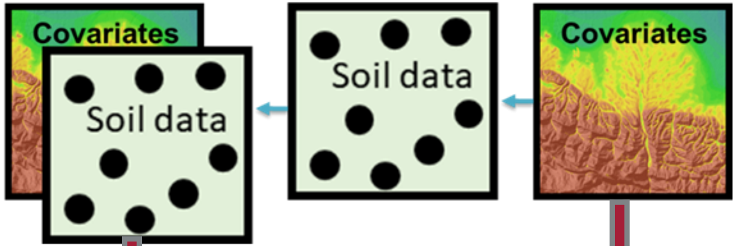
- Regression + kriging of residuals: is a hybrid spatial interpolation technique
- Linear regression is used to make a relationship between soil data and covariates
- Kriging is used to model the residuals

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

Soil data Model Covariates spatially dependent residuals

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

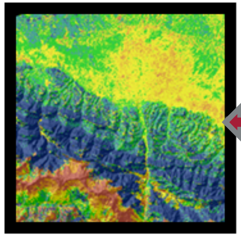
apply the model to all unobserved locations

Kriging of residuals

Map 2

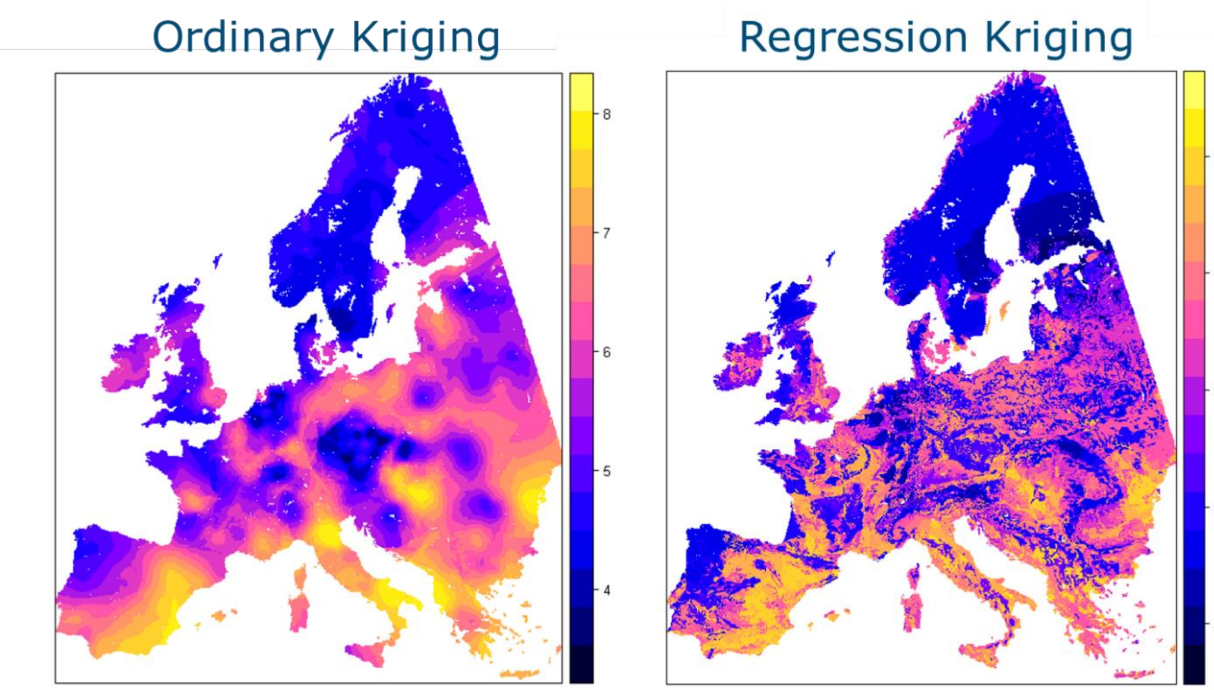
Map 1

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)

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

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.

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

1

```
#Preparing geodatabase
# import raster layers of covarites (RS data + DEM)
covariates <- stack(list.files("./cov/", pattern="\\.tif$", full.names = TRUE))
names(covariates)
 "Blue"  "DEM"   "Green" "Nir"   "Red"   "SWIR1" "SWIR2"
```

2

```
# import the point soil data
soil <- read.csv ("soil.csv", header = TRUE)
str(soil)
'data.frame':    183 obs. of  3 variables:
 $ x : num  777853 779844 781270 780613 783887 ...
 $ y : num  3589067 3589470 3589750 3588612 3589697 ...
 $ EC: num  138.9 232 232 183 5.67 ...
```

3

```
# extract values
cov_soil = raster::extract(covariates,soil,method='simple')

# inspect the covariates + soil data
str(cov_soil)
'data.frame':    183 obs. of  8 variables:
 $ Blue : num  0.217 0.216 0.213 0.218 0.207 ...
 $ DEM  : num  1004 1003 1006 1008 1017 ...
 $ Green: num  0.252 0.253 0.247 0.254 0.232 ...
 $ Nir  : num  0.376 0.38 0.373 0.379 0.333 ...
 $ Red  : num  0.32 0.324 0.315 0.322 0.284 ...
 $ SWIR1: num  0.395 0.407 0.404 0.408 0.385 ...
 $ SWIR2: num  0.359 0.364 0.368 0.373 0.353 ...
 $ EC   : num  138.9 232 232 183 5.67 ...
```

4

```
# Linear Regression
linear_fit <- lm(EC ~ Blue+DEM+Green+Nir+Red+SWIR1+SWIR2,
                 data=cov_soil)
```

5

```
#preparing soil maps
map_linear <- raster::predict( covariates,linear_fit)
```

Useful Resources