Introduction to soil property mapping

Dr. V.L. Mulder[a] & Dr. B. Kempen[b]
a) Soil Geography and Landscape group, Wageningen University, The Netherlands
b) ISRIC World Soil Information, Wageningen, The Netherlands

Training “Digital Soil Organic Carbon Mapping: towards the development of national soil organic carbon stock maps”

Outline

1. Overview global soil maps
   - Soil classification maps
   - Soil property maps
2. Conventional soil mapping
3. Digital soil mapping
4. Model evaluation
   - Uncertainty
   - Validation

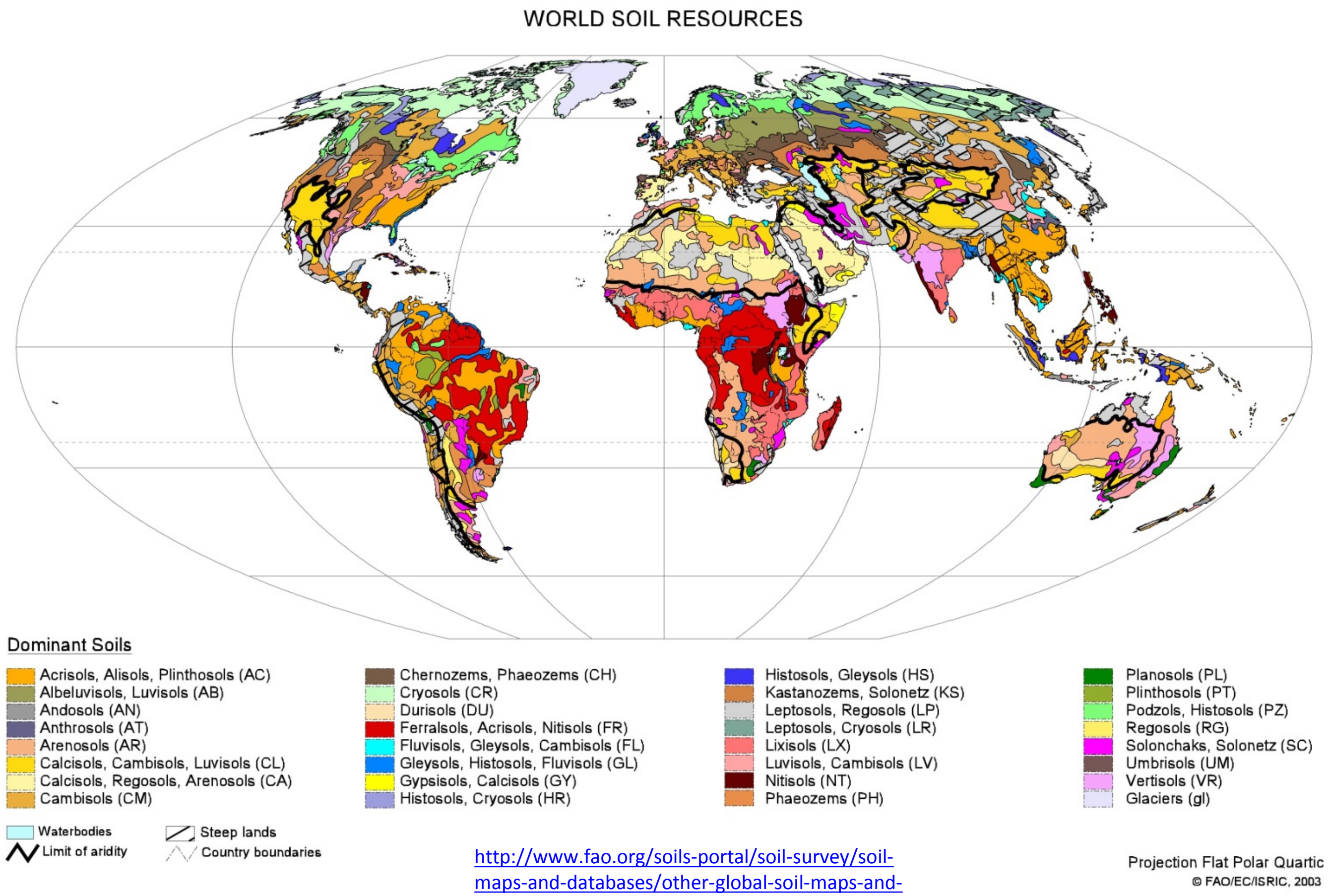
1. OVERVIEW SOIL MAPS

1. Overview soil maps
Definitions & objectives

- Soil property maps
  - Represent spatial information about soil properties for a certain depth interval or for soil horizons
- Traditional/Conventional soil maps
  - Polygon maps with mapping units
  - Mapping units are associated to representative soil profiles
- Digital Soil Mapping
  - Spatial continuous mapping and modelling of soil properties (and soil classes)
  - Statistical inference between sampled sites and environmental datasets
  - Spatial prediction of soil properties and quantification of the prediction error
- GSP implementation: 1km resolution, grid-based soil property maps
  - Whenever possible use Digital Soil Mapping

1. Overview soil maps

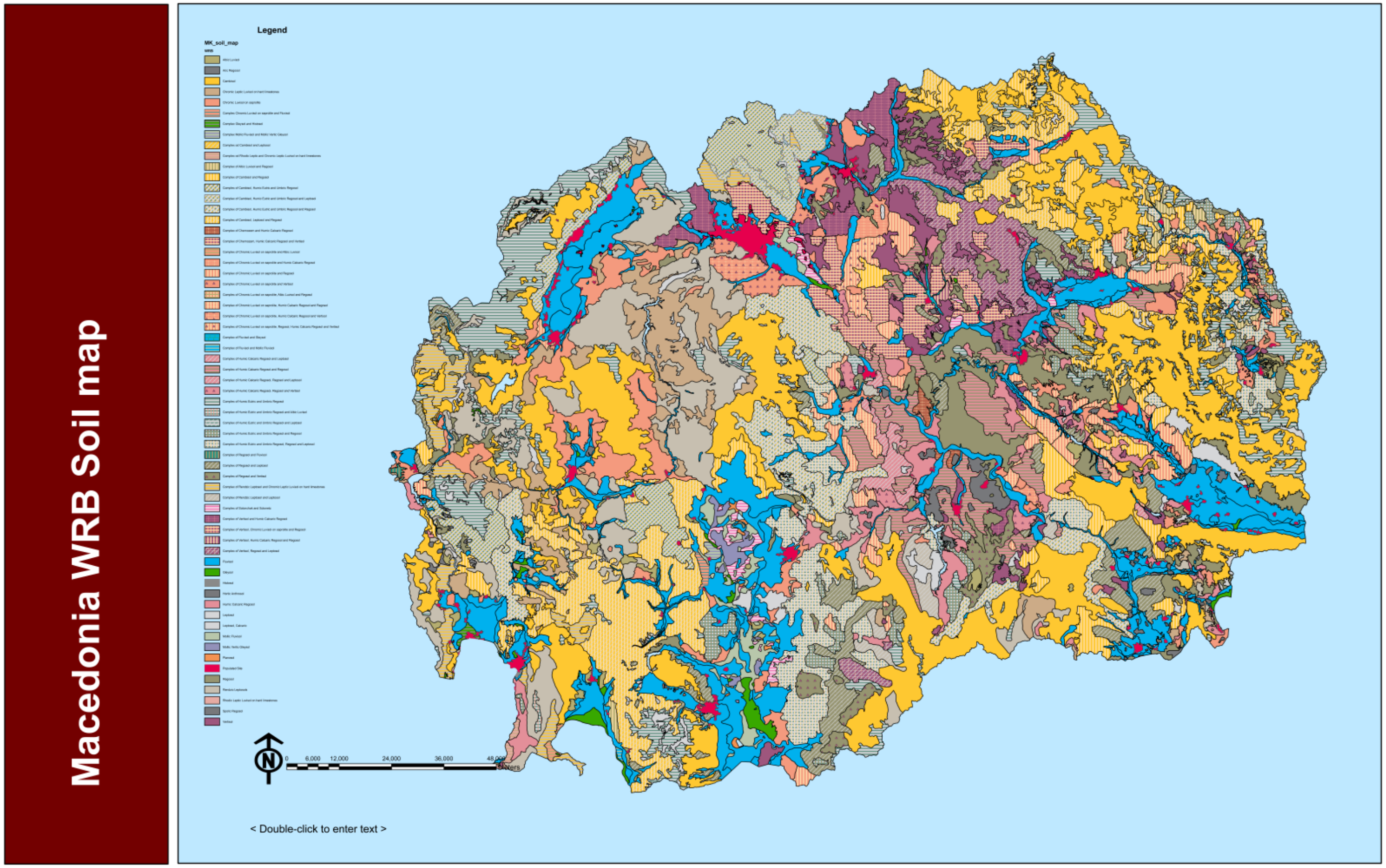
WRB Soil Map of the World



http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/other-global-soil-maps-and-databases/en/

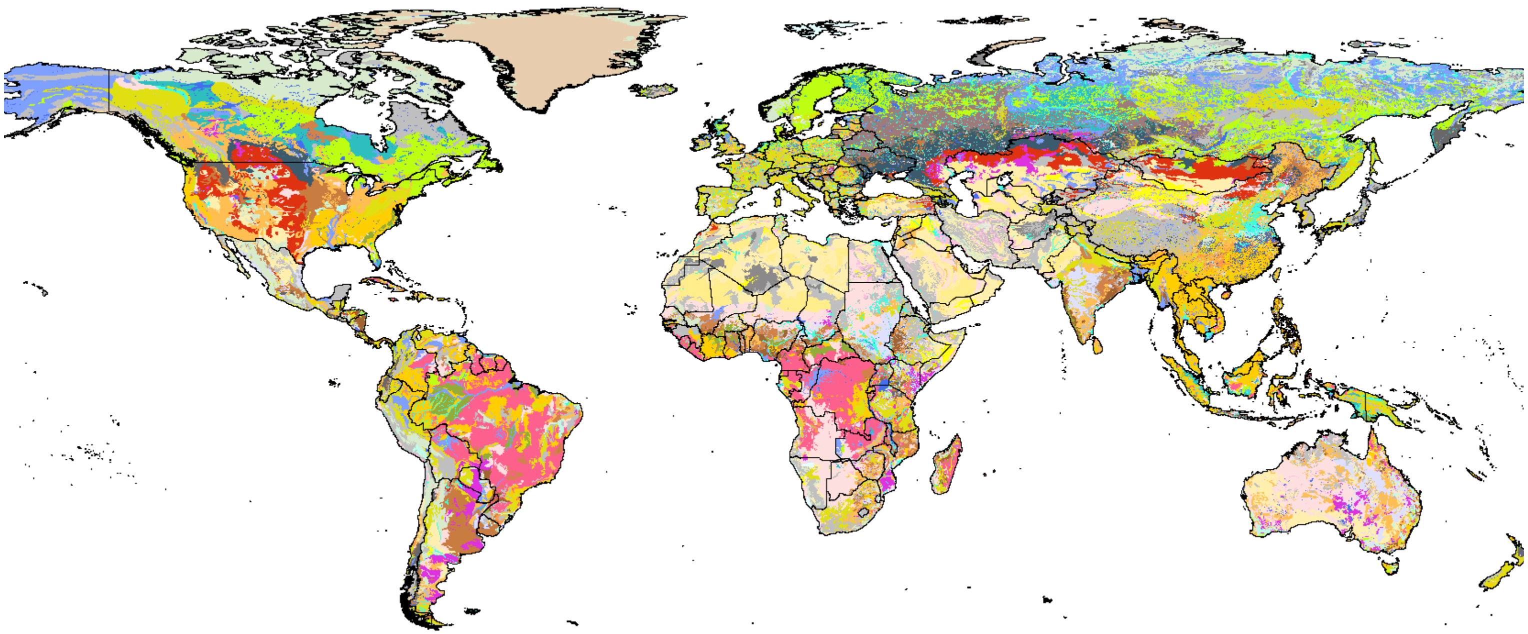
1. Overview soil maps

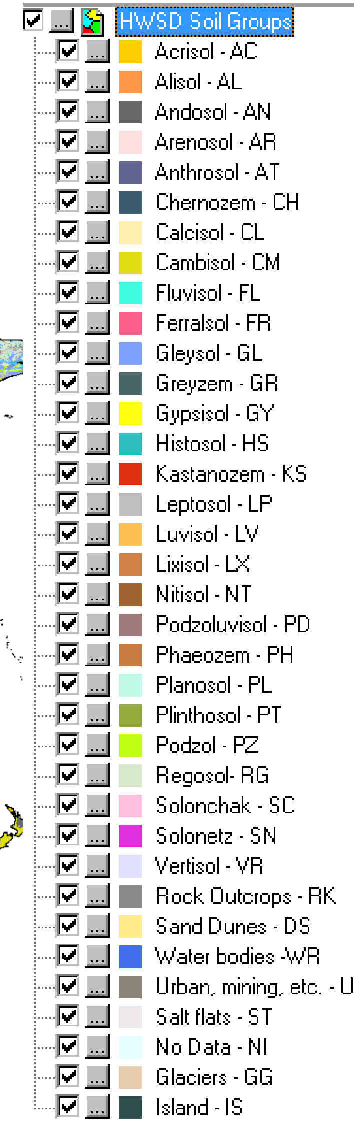
Macedonia WRB soil polygon map



1. Overview soil maps

Harmonized World Soil Database



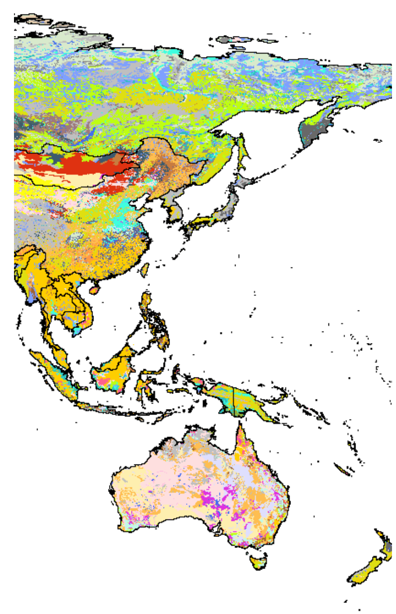


http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/

1. (

Ha abase

Coverage	DSMW			
Soil Mapping Unit	4640			
Dominant Soil Group	CM - Cambisols			
Sequence	1	2	3	4
Share in Soil Mapping Unit (%)	50	20	20	10
Database ID	43514	43516	43515	43517
Soil Unit Symbol (FAO 74)	Bd	L	A	G
Soil Unit Name (FAO74)	Dystric Cambisols	Luvisols	Acrisols	Gleysols
Topsoil Texture	Medium	Medium	Medium	Medium
Reference Soil Depth (cm)	100	100	100	100
Drainage class (0-0.5% slope)	Moderately Well	Moderately Well	Moderately Well	Poor
AWC (mm)	150	150	150	150
Gelic Properties	No	No	No	No
Vertic Properties	No	No	No	No
Petric Properties	No	No	No	No
Topsoil Sand Fraction (%)	41	48	49	38
Topsoil Silt Fraction (%)	40	39	30	27
Topsoil Clay Fraction (%)	22	20	22	24
Topsoil USDA Texture Classification	loam	loam	sandy clay loam	loam
Topsoil Reference Bulk Density (kg/dm3)	1.41	1.41	1.4	1.39
Topsoil Bulk Density (kg/dm3)	**1.3**	**1.45**	**1.4**	**1.33**
Topsoil Gravel Content (%)	**10**	**5**	**9**	**4**
Topsoil Organic Carbon (% weight)	**1.45**	**0.8**	**1.03**	**1.27**
Topsoil pH (H2O)	5.1	6.3	4.9	5.8
Topsoil CEC (clay) (cmol/kg)	35	30	36	15



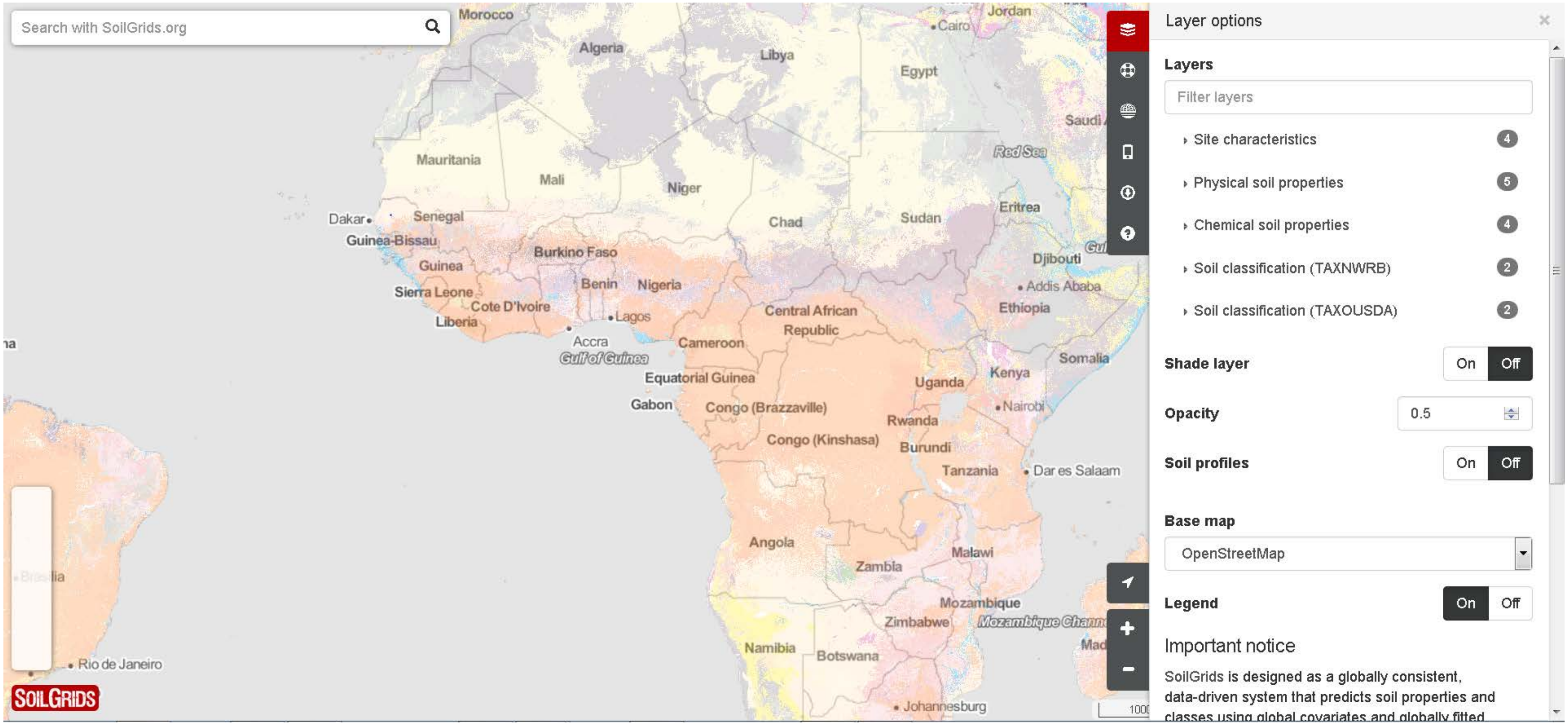
1. Overview soil maps
SoilGrids

http://www.soilgrids.org

Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12(2): e0169748. https://doi.org/10.1371/journal.pone.0169748

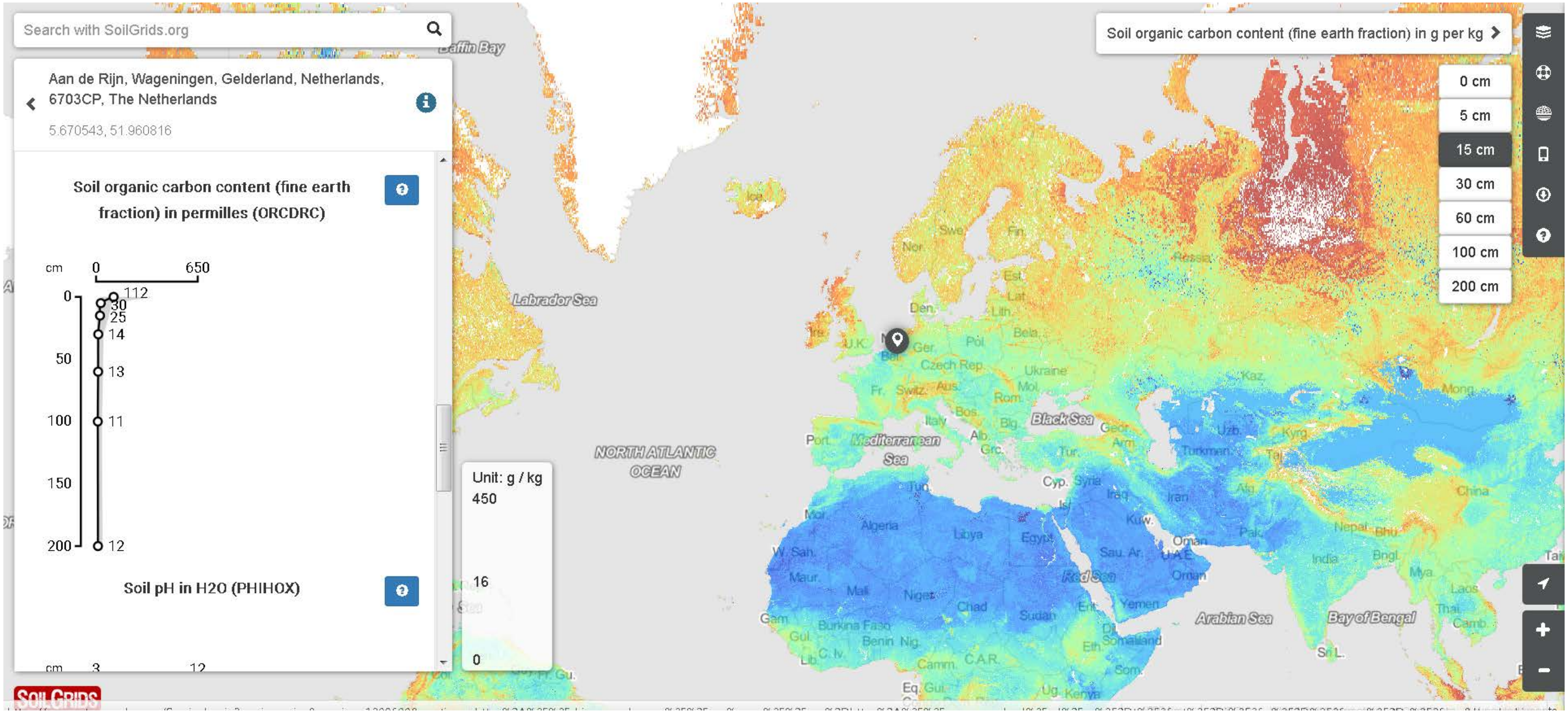
1. Overview soil maps

SoilGrids



1. Overview soil maps

SoilGrids: SOC at 15 cm depth



Food and Agriculture Organization of the United Nations
ISRIC World Soil Information

1. Overview soil maps

S-World: A Global Soil Map for Environmental Modelling

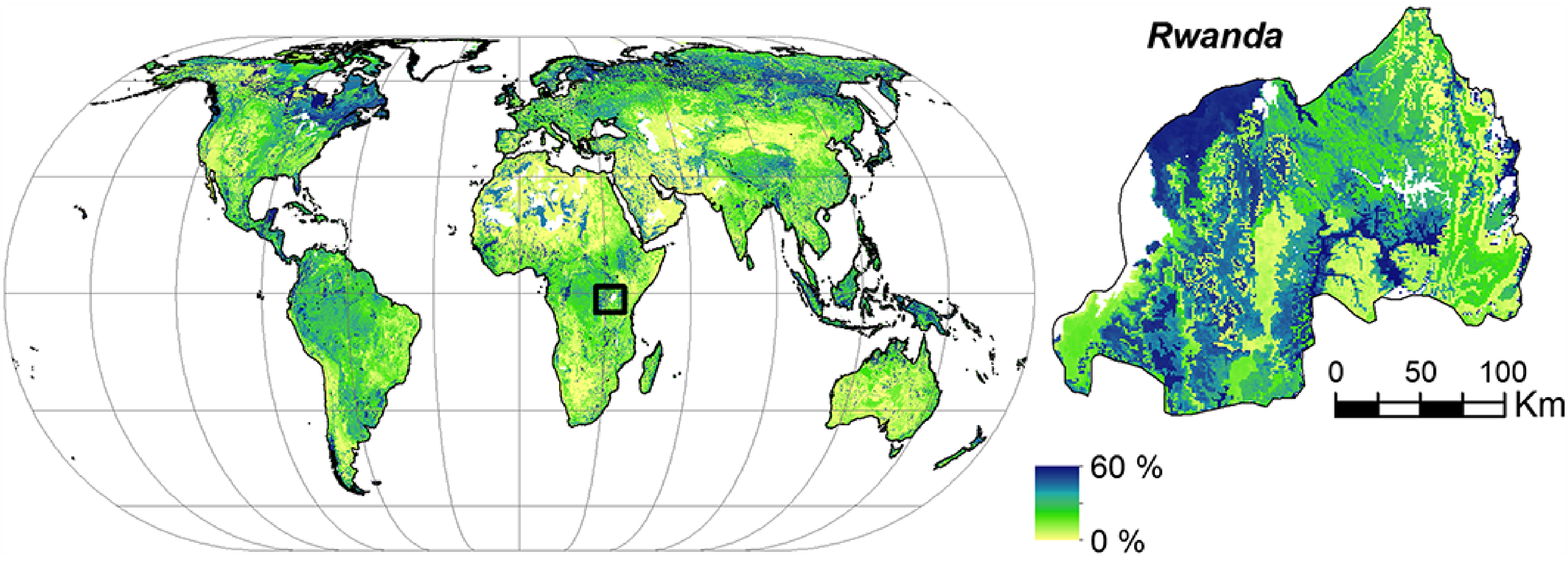


Figure 6: S-World results for modelled SOC (%) for the world and Rwanda (topsoil) (white areas represent areas with no data as there is no soil).

Land Degradation & Development
Volume 28, Issue 1, pages 22-33, 1 DEC 2016 DOI: 10.1002/ldr.2656
http://onlinelibrary.wiley.com/doi/10.1002/ldr.2656/full#ldr2656-fig-0005

2. CONVENTIONAL SOIL MAPS

2. Conventional soil maps
Theory

- Soil Survey
  - **Systematic study of the soil** of an area including classification and mapping of the properties and distribution of soil units
- Determine the pattern of the soil cover (delineated areas)
  - Divide pattern into relatively homogeneous units
    - Traditionally using aerial photographs combined with field observations
  - Map the distribution of these units (Soil Mapping Units)
    - Polygon map
    - Homogenous in soil and landscape variability, not by definition 1 value for the whole polygon (compound mapping unit)
- Characterize the SMUs
    - Soil classification, soil qualifiers, soil series, soil properties
    - Useful information for subsequent studies, e.g. land use potential and model expected response to changes in management
  - Soil Survey Manual and FAO guidelines:
    - ftp://ftp.fao.org/FI/CDrom/FAO_Training/FAO_Training/General/x6706e/Index.htm

2. Conventional soil maps
Potentials and limitations

- The soil classification, including soil qualifiers does provide a lot of useful information about the SMUs. However:
  - Fixed scale, not always useful for applications (too coarse)
  - Through aggregation in SMUs information is lost
  - Uncertainty typically not quantified
  - Polygon data model not easy to integrate with gridded biophysical data
- Whenever possible, DSM is recommended
  - More accurate spatial mapping of soil properties
  - Spatial quantification of the prediction error

2. Conventional soil maps
Mapping SOC stocks

- Matching the Soil Mapping Units with soil profile information
- Use other environmental gridded data
- Training: Conventional upscaling methods

3. DIGITAL SOIL MAPPING

3. Digital Soil Mapping

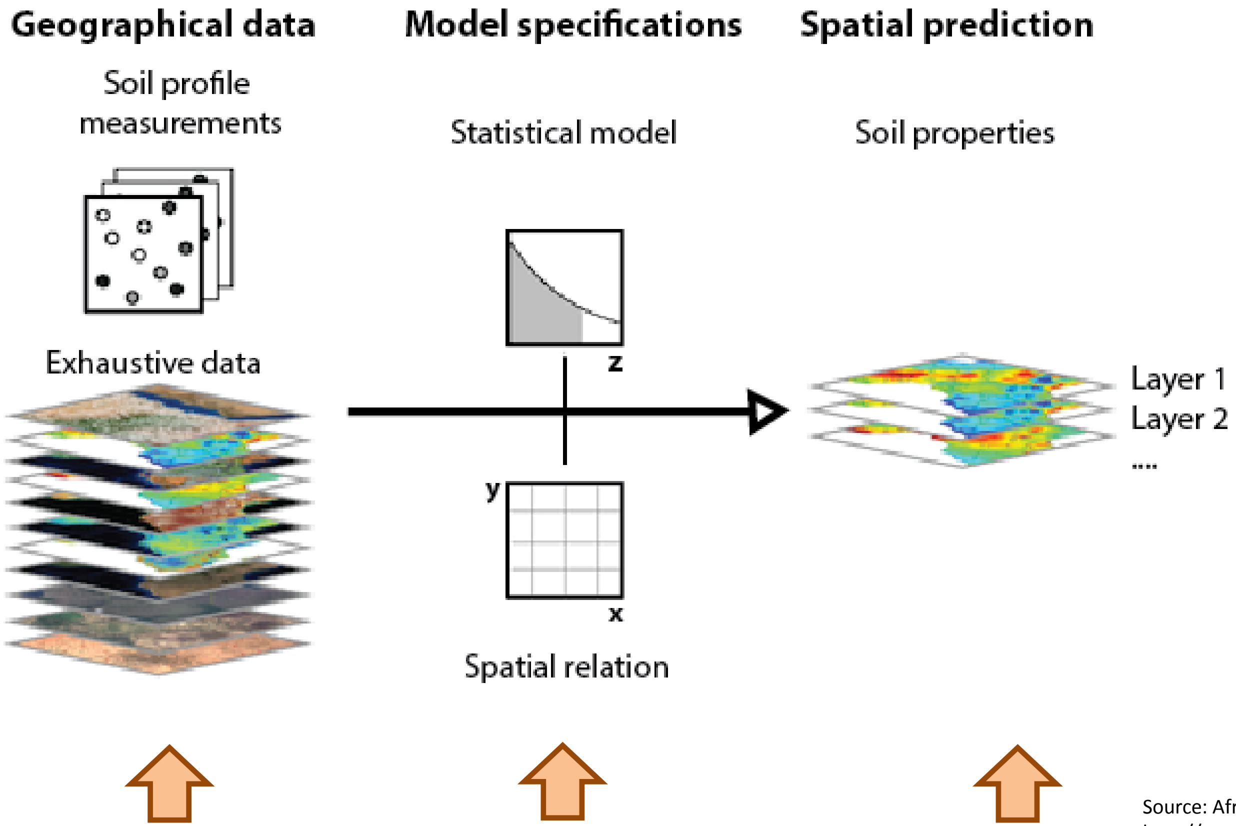
Pedometrics

- Pedometrics is a branch of soil science that applies **mathematical** and **statistical methods** for the study of the distribution and genesis of soils
- The goal of pedometrics is to achieve a better understanding of the soil as a phenomenon that varies over **different scales** in **space** and **time**
- **Predicting** the properties of the soil in space and time, with **sampling** and **monitoring** the soil and with modelling the soil's behaviour.
- Developing and applying **quantitative methods**
  - sampling designs and strategies
  - geostatistical methods for spatial prediction
  - linear modelling methods
  - novel mathematical and computational techniques (machine learning, wavelet transforms and fuzzy logic)

Source: http://www.pedometrics2017.org/what-is-pedometrics/

3. Digital Soil Mapping

Theory

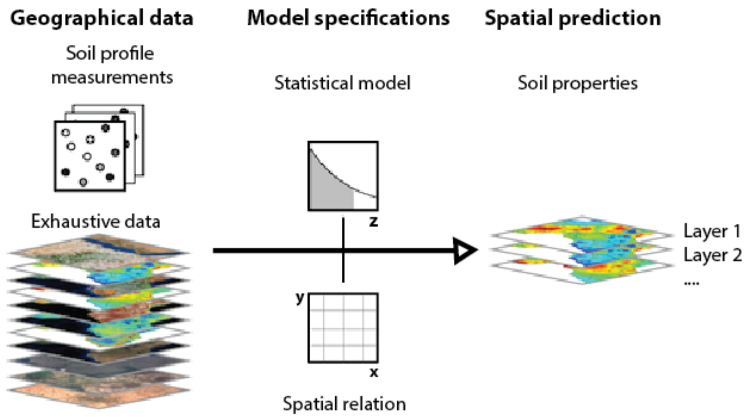


Source: Africa Soil Information Service, http://www.africasoils.net/data/digital-soil-mapping

3. Digital Soil Mapping
Theory

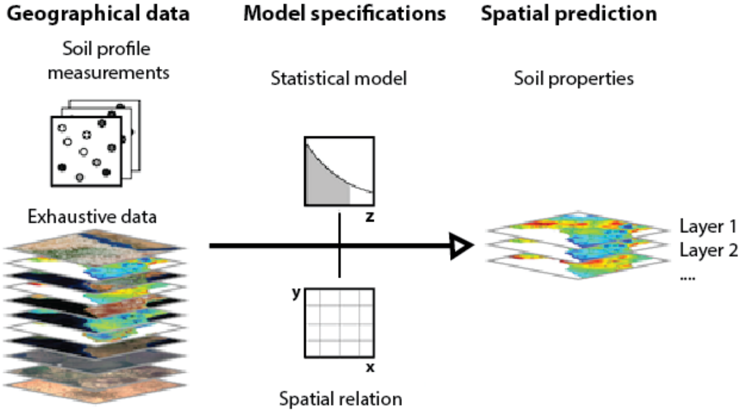
- Soil-landscape paradigm (Jenny, 1941)
- The soil forming factors: s=f(cl,o,r,p,t)
  - Cl = climate
  - O = organisms
  - R = relief
  - P = Parent Material
  - T = Time
- S=f(s,c,o,r,p,a,n) (McBratney et al., 2003)
  - with s being known soil properties
  - With n being the spatial or geographic position
- Environmental gridded dataset -> covariates

3. Digital Soil Mapping
Theory



1. Carry out a survey where the **target variable** is being sampled and measured.
2. Identify potential **covariate data** that are available with the appropriate resolution and extent.
3. Correlate the target variable to you covariate data and derive a **statistical model** (e.g. linear regression)
4. Apply the statistical model to your covariate data. This will give you a first estimate of the variability.
5. The statistical model will not always explain all the variability that we observe in our data. We therefore calculate **the prediction residuals** of our observations, i.e. the difference between the predicted values and the measured values.
6. Check whether the residuals are **spatially correlated** (variogram). If this is the case, the model over-predicts in certain areas and under-predicts in other parts. If the residuals are spatially correlated then we can interpolate the residuals to see where we over- and under-predict (**kriging of residuals**).
7. Finally we can 'correct' the estimate from Step 4 with the residuals.
8. **Model evaluation** and **map accuracy**

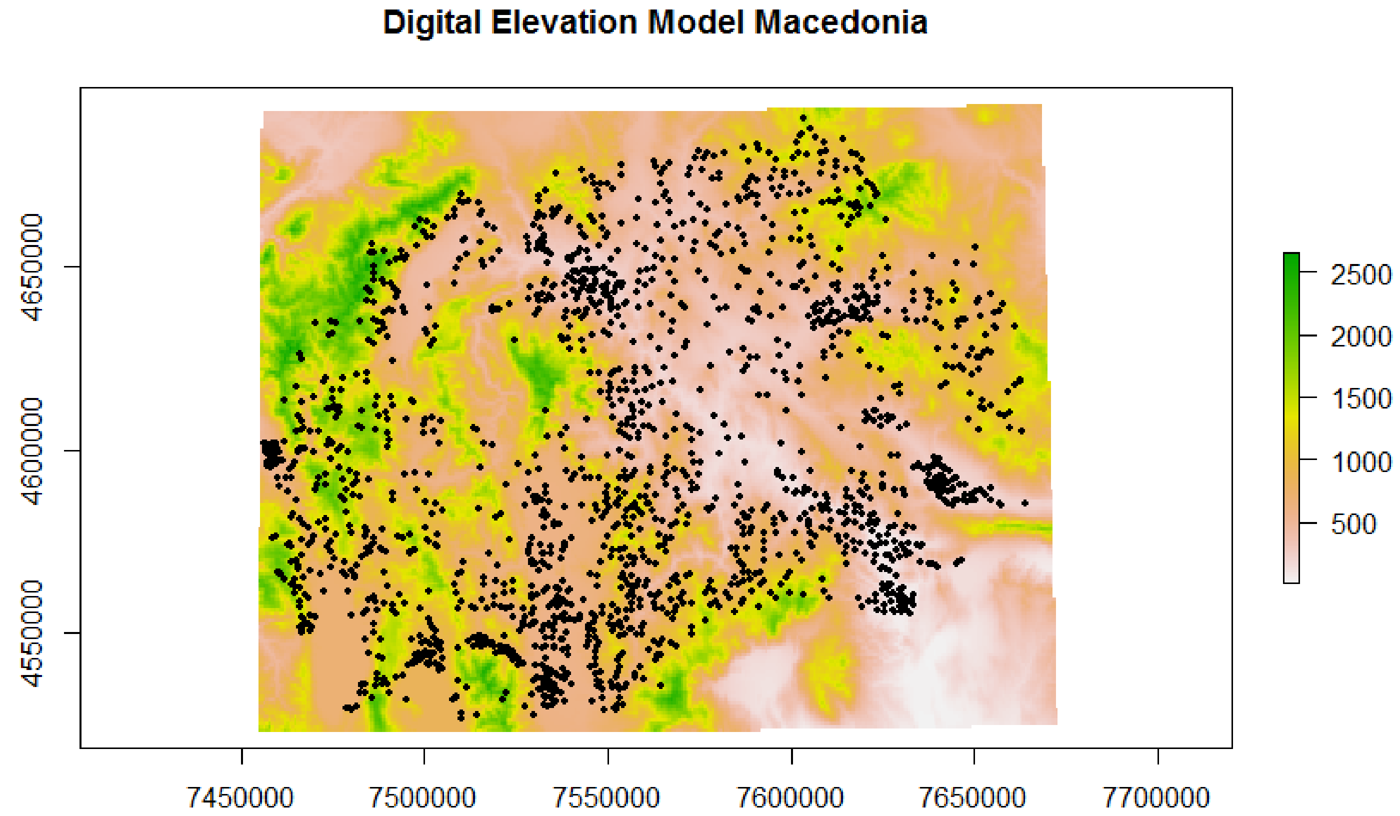
3. Digital Soil Mapping
Theory



1. Carry out a survey where the **target variable** is being sampled and measured.
2. Identify potential **covariate data** that are available with the appropriate resolution and extent.
3. Correlate the target variable to you covariate data and derive a **statistical model** (e.g. linear regression)
4. Apply the statistical model to your covariate data. This will give you a first estimate of the variability.
5. The statistical model will not always explain all the variability that we observe in our data. We therefore calculate **the prediction residuals** of our observations, i.e. the difference between the predicted values and the measured values.
6. Check whether the residuals are **spatially correlated** (variogram). If this is the case, the model over-predicts in certain areas and under-predicts in other parts. If the residuals are spatially correlated then we can interpolate the residuals to see where we over- and under-predict (**kriging of residuals**).
7. Finally we can correct the estimate from Step 4 with the residuals.
8. **Model evaluation** and **map accuracy**

3. Digital Soil Mapping

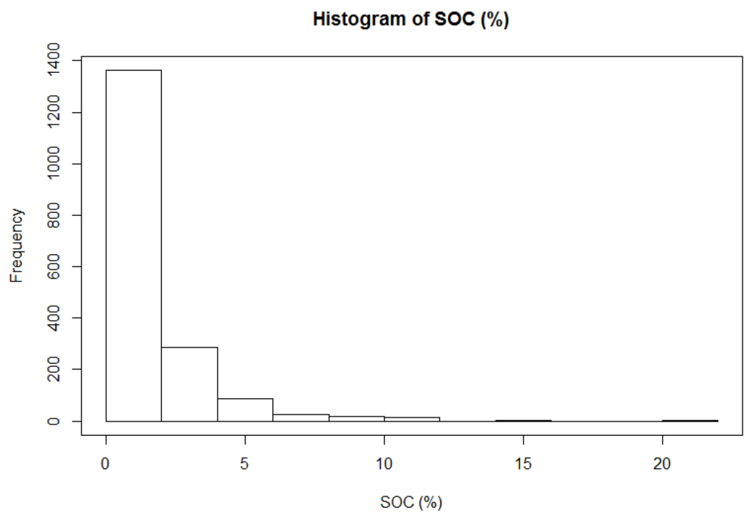
Example statistical inference: Data



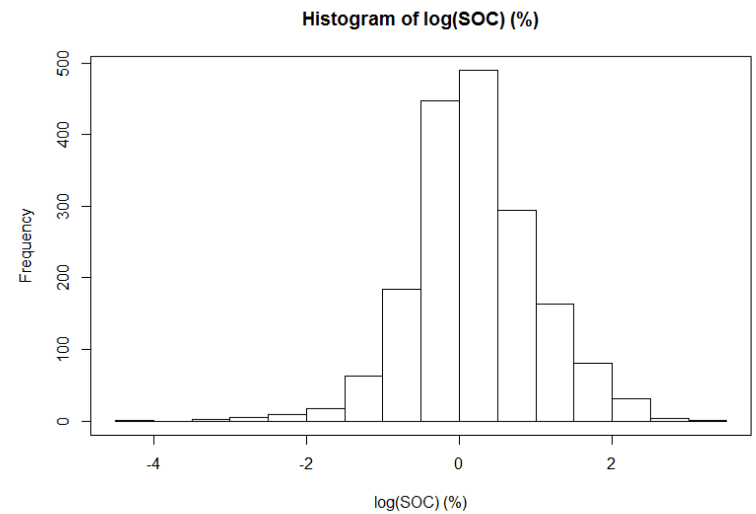
3. Digital Soil Mapping

Example statistical inference: Data

Original SOC (%)



Log-transformed SOC (%)



	Min.	1Q	Median	3Q	Max
SOC (%)	0.02	0.75	1.169	1.93	21.2

3. Digital Soil Mapping

Example statistical inference: Regression model

```
#### multiple linear regression model
lm_fit_log <- lm(log_SOC~ AWC+ Depth+ BD+ CEC+ Clay+ CoarseF+ DEM+Prec+ slope+ silt+ sand+Temp+TWI, data=pred_data)
summary(lm_fit_log)

# Stepwise Regression
step_lm_log <- stepAIC(lm_fit_log, direction="both")
step_lm_log$anova # display results
summary(step_lm_log)
```

Multiple Linear Regression

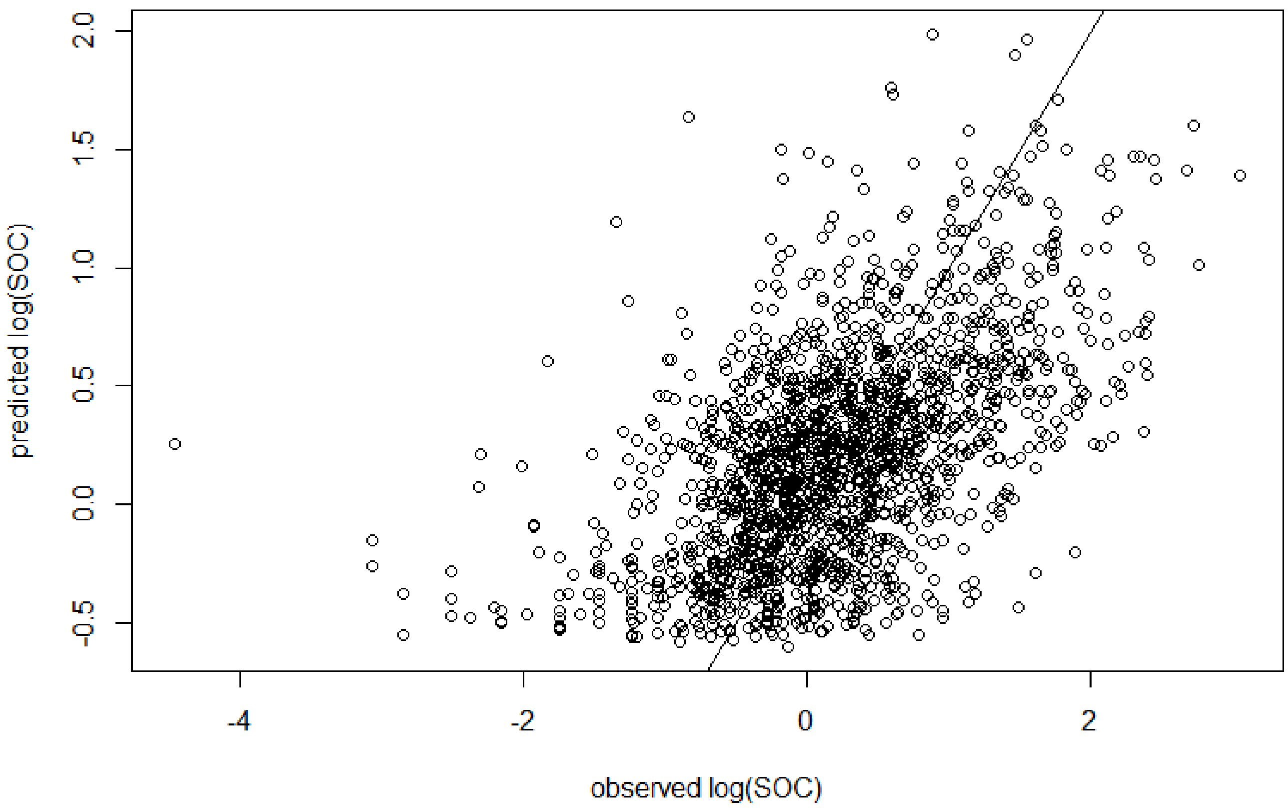
- R2 = 0.2993
- R2adj=0.2942
- Residual error = 0.6848
- 1780 degrees of freedom

Stepwise MLR

- R2 = 0.2986
- R2adj = 0.2951
- Residual error = 0.6844
- 1784 degrees of freedom
- Clay, Silt, AWC, CEC were excluded

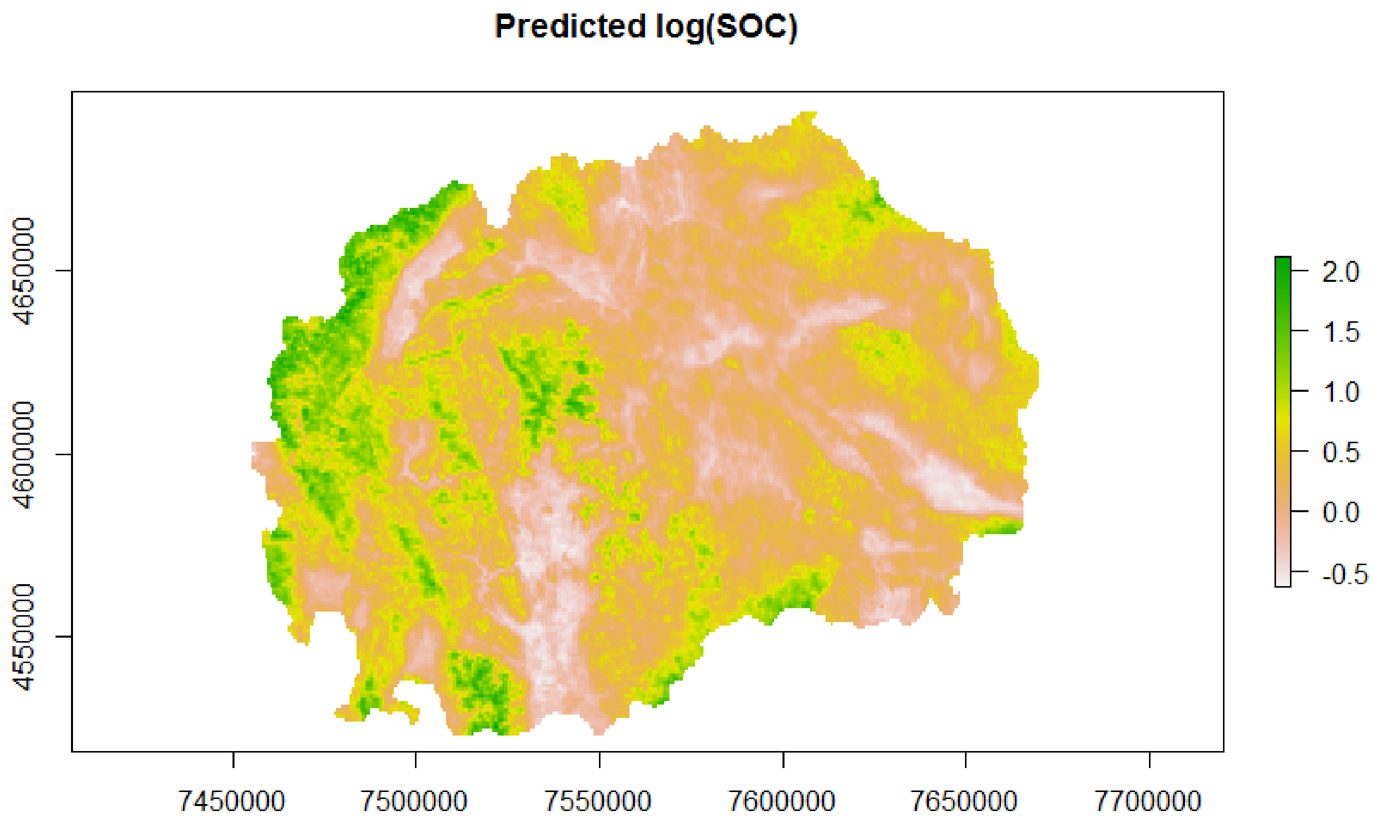
3. Digital Soil Mapping

Example statistical inference: Regression model

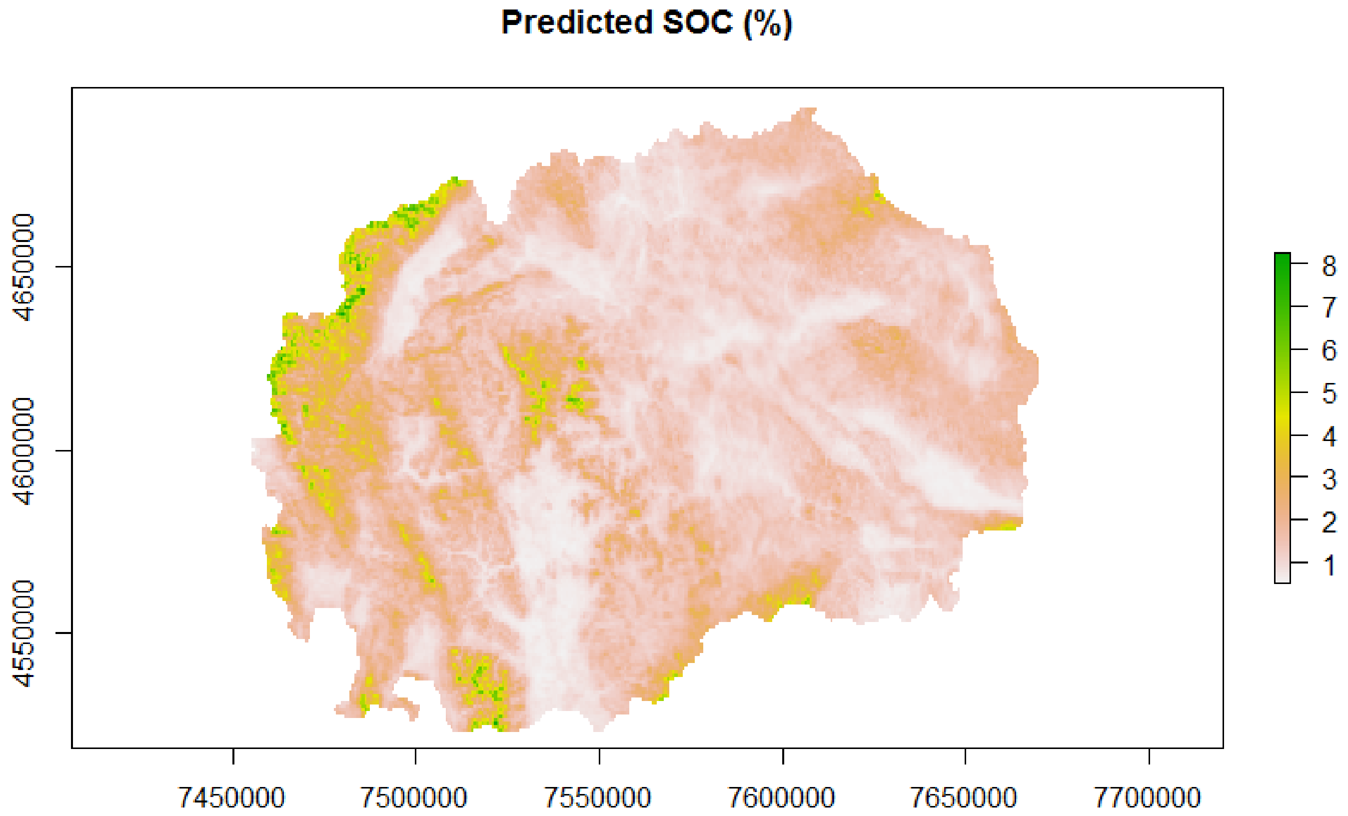


3. Digital Soil Mapping

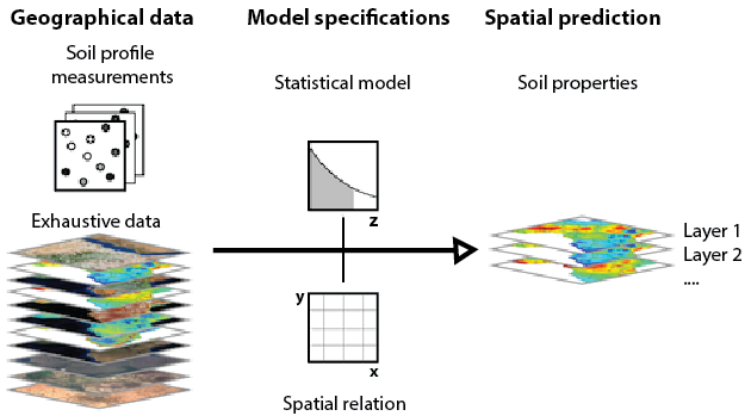
Spatial prediction of SOC



3. Digital Soil Mapping
Spatial prediction of SOC



3. Digital Soil Mapping
Theory

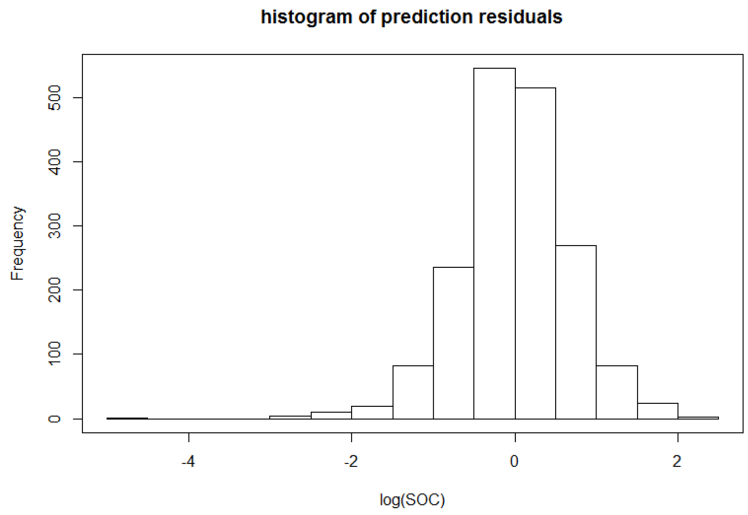


1. Carry out a survey where the **target variable** is being sampled and measured.
2. Identify potential **covariate data** that are available with the appropriate resolution and extent.
3. Correlate the target variable to you covariate data and derive a **statistical model** (e.g. linear regression)
4. Apply the statistical model to your covariate data. This will give you a first estimate of the variability.
5. The statistical model will not always explain all the variability that we observe in our data. We therefore calculate **the prediction residuals** of our observations, i.e. the difference between the predicted values and the measured values.
6. Double check whether the residuals are **spatially correlated** (variogram). If this is the case, the model over-predicts in certain areas and under-predicts in other parts. If the residuals are spatially correlated then we can interpolate the residuals to see where we over- and under-predict (**kriging of residuals**).
7. Finally we can correct the estimate from Step 4 with the residuals.
8. **Model evaluation** and **map accuracy**

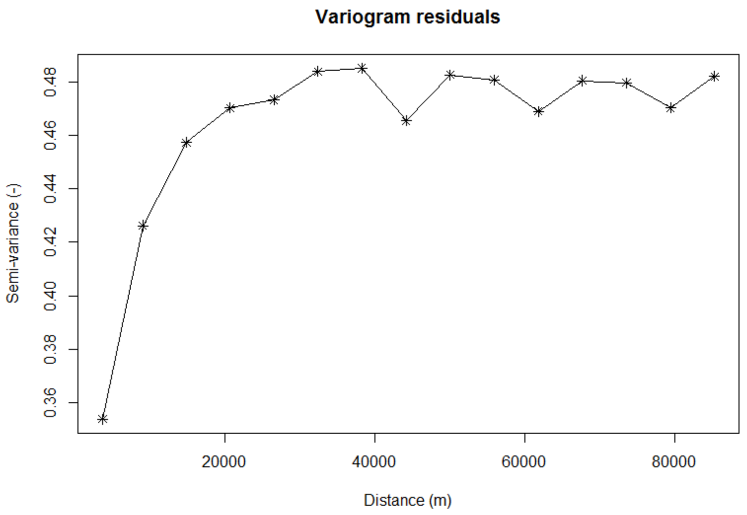
3. Digital Soil Mapping

Prediction residuals

Distribution of residuals



Spatial correlation in residuals



Kriging of residuals because the residuals are spatially autocorrelated
Geostatistical methods are needed to improve spatial quantification of soil properties!

3. Digital Soil Mapping

Geostatistics

- Training 'Basic geostatistics'
  - Quantifying spatial variation: semi-variogram
  - Ordinary Kriging
  - Quantification of prediction uncertainty
- Training Regression and Machine Learning Methods for DSM
  - Linear regression
  - Machine learning algorithms (random forest)
  - Fitting and interpreting a model in R

4. MODEL EVALUATION

4. Model evaluation

- Get **confidence** in a model
- Get to know how close the model is to **reality**
- Get to understand whether the **behaviour** of the model corresponds with reality
- Without model evaluation we would have no idea how **reliable** the model is

Slide courtesy: Gerard Heuvelink, ISRIC World Soil Information, Wageningen, The Netherlands

4. Model evaluation

- Model **evaluation** not the same as model **validation**
- **Model evaluation**: the process of analysing the validity of the model structure and appraising the performance of a model by comparing the model predictions with (independent) observations
- **Model validation (or verification)**: formal recognition that a model is legitimate, that it is an accurate representation of reality, or even that it represents the 'truth'
- **Beware**: no consensus on meaning of these terms (read today's literature!)

Slide courtesy: Gerard Heuvelink, ISRIC World Soil Information, Wageningen, The Netherlands

4. Model evaluation

Causes of deviation between model and reality

- **Human** error (e.g. a coding or typing error)
- Model is (partly on purpose!) a **simplification** of reality:
  - Less important processes ignored, lumped or represented in a simplified way (model **structural error**)
  - Errors in **model parameters** (e.g. because model is calibrated on a small dataset or a dataset from another area or time period)
  - Approximations in **solution methods** (e.g. numerical solution of differential equations)
- Model **inputs** poorly known (input error)

Slide courtesy: Gerard Heuvelink, ISRIC World Soil Information, Wageningen, The Netherlands

4. Model evaluation
Scientific evaluation

- Does model structure accord with **physical laws** such as conservation of mass, are cause-and-effect relationships properly represented, are model assumptions consistent with physical principles
- **Quality check** of computer implementation (standardized coding protocols, code checking, bug detection)
- Is model behaviour as expected (sensitivity analysis)?

Slide courtesy: Gerard Heuvelink, ISRIC World Soil Information, Wageningen, The Netherlands

4. Model evaluation

Statistical evaluation

- Training: Quantifying and assessing uncertainty
  - Spatial stochastic simulation
  - Uncertainty propagation
- Training: Validation of predicted soil property maps
  - Methods and measures
  - Sampling for validation and mapping

Training programme – GSP Global SOC

GLOBAL SOIL PARTNERSHIP

1. Introduction to soil property mapping
2. Data preparation
3. Basic Geostatistics
4. Conventional upscaling methods
5. Regression Kriging
6. Uncertainty
7. Validation

Questions?!

Next: lecture Data Preparation