Investigating soil data with **R**

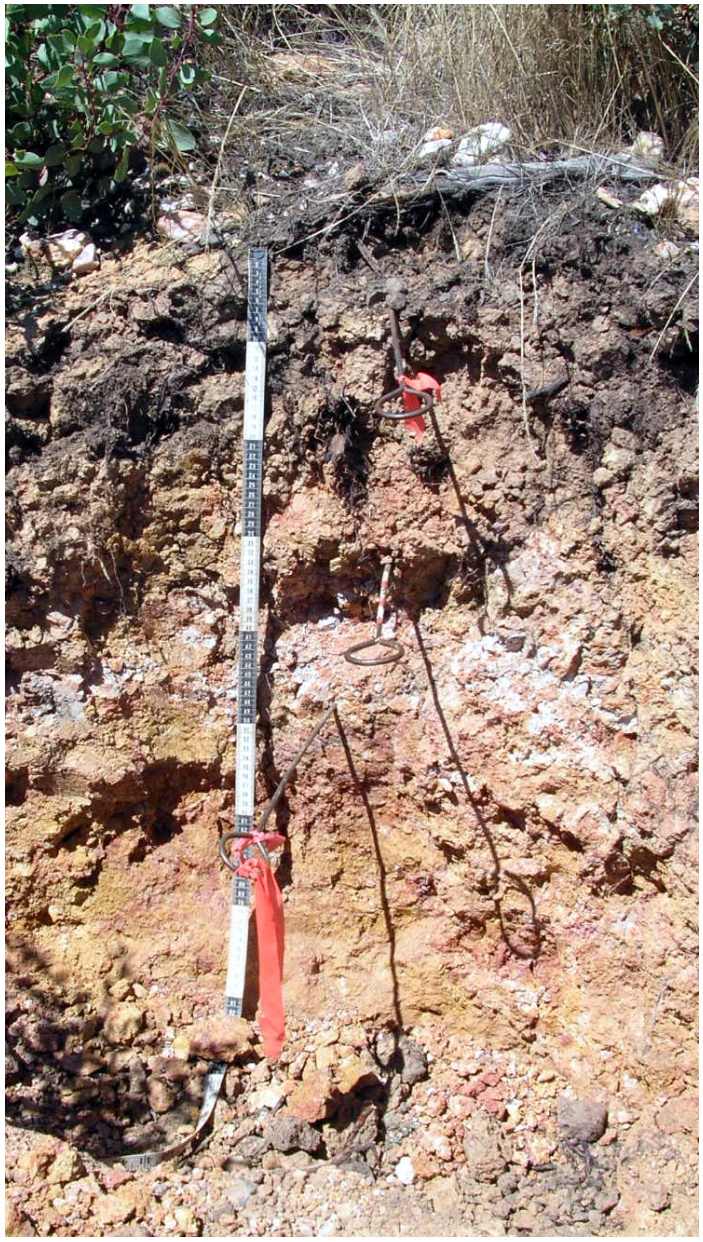
Pierre Roudier[1] and Dylan E. Beaudette[2]

[1] Landcare Research – Manaaki Whenua, New Zealand
[2] Natural Resources Conservation Service, USDA, USA

Acknowledgement

Talk Outline

1. The soil resource
   - Soil matters!
   - Soil science
   - Soil data and its analysis
2. The aqp package
   - Visualisation
   - Classification
   - Harmonisation
   - Analysis and modelling
3. Recent and future developments
   - Introduction of S4 classes
   - Design
4. Conclusions and further work



Talk Outline

1. The soil resource
   - Soil matters!
   - Soil science
   - Soil data and its analysis
2. The aqp package
   - Visualisation
   - Classification
   - Harmonisation
   - Analysis and modelling
3. Recent and future developments
   - Introduction of S4 classes
   - Design
4. Conclusions and further work

Soil matters!

Soil

- a very **thin** layer between parent rock and atmosphere,
- a very **complex** body – physical, chemical and biological interactions,
- a support for almost all terrestrial ecosystems – and thus **food production**

Soil matters!

Soil

is not dirt.

A threatened resource

- How to feed 7 billion+ people?
- Growing tensions on arable land
- Urbanisation
- Erosion
- Etc.

$\hookrightarrow$ Important to **provide soil information** to a wide range of decision makers.

A threatened resource

"Man has only a thin layer of soil between himself and starvation." –Bard of Cincinnati

Soils today

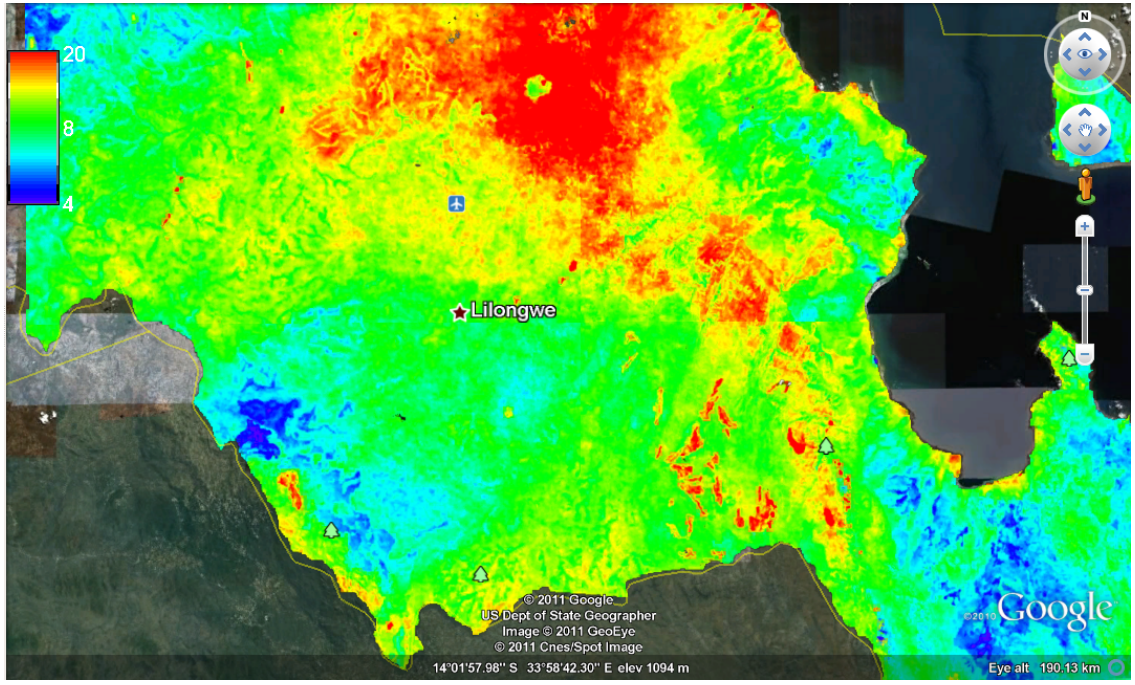
- Long been only regarded as a producer of crops
- But soils are back on the global agenda
- New challenges through global projects

Soils today

- Long been only regarded as a producer of crops
- But soils are back on the global agenda
- New challenges through global projects

GlobalSoilMap.net– a leading project

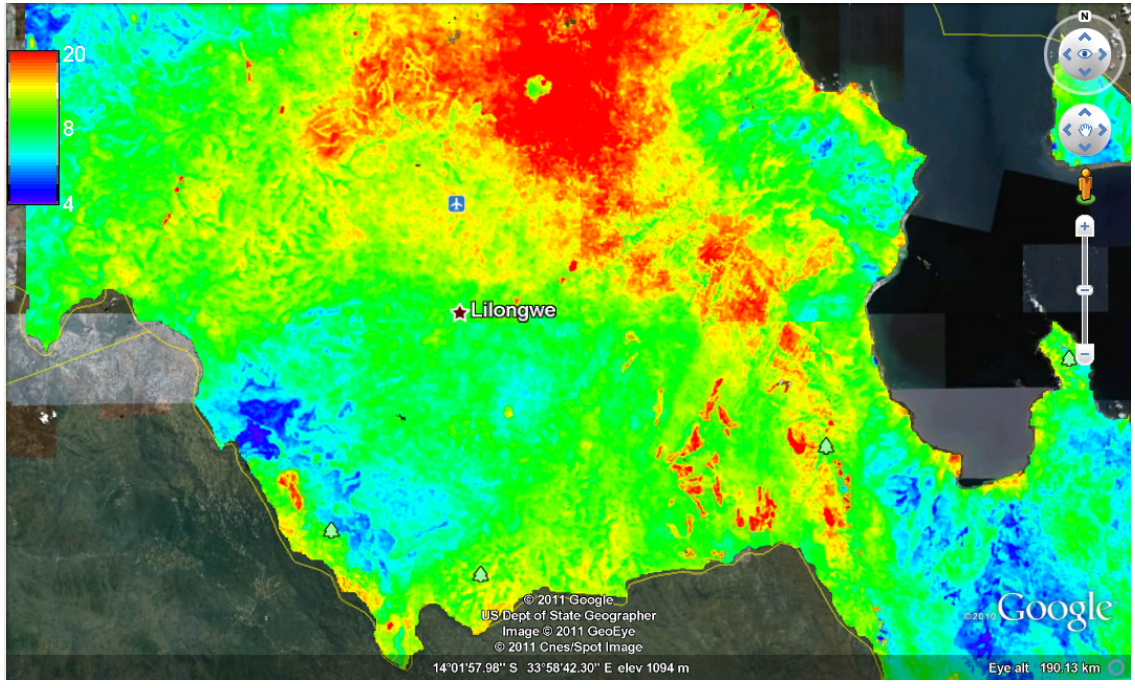
- International network of soil scientists
- Generate and provide a $\approx$ 100m grid of 9 key soils attributes globally
- R has been identified as a key platform for the various stages involved



Data courtesy of Tomislav Hengl.

GlobalSoilMap.net– a leading project

- International network of soil scientists
- Generate and provide a $\approx$ 100m grid of 9 key soils attributes globally
- R has been identified as a key platform for the various stages involved



Data courtesy of Tomislav Hengl.

$\hookrightarrow$ The way to do soil science is changing **significantly** (*"pedometrics"*)

The soil profile

The soil profile

Soil data:

- Highly multi-dimensional
- Point support
- Soft and hard data
- Importance of legacy data
- Most of the time associated with environmental covariates

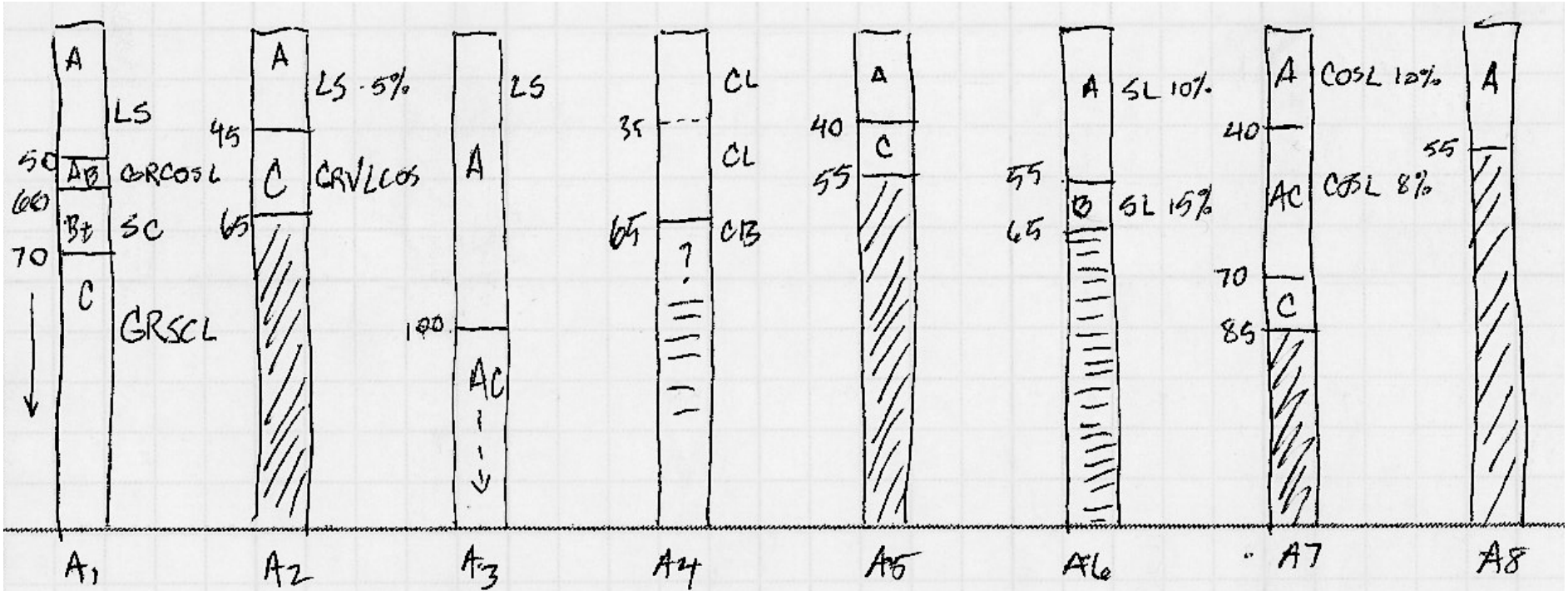
Talk Outline

1. The soil resource
   - Soil matters!
   - Soil science
   - Soil data and its analysis
2. The aqp package
   - Visualisation
   - Classification
   - Harmonisation
   - Analysis and modelling
3. Recent and future developments
   - Introduction of S4 classes
   - Design
4. Conclusions and further work

The aqp package

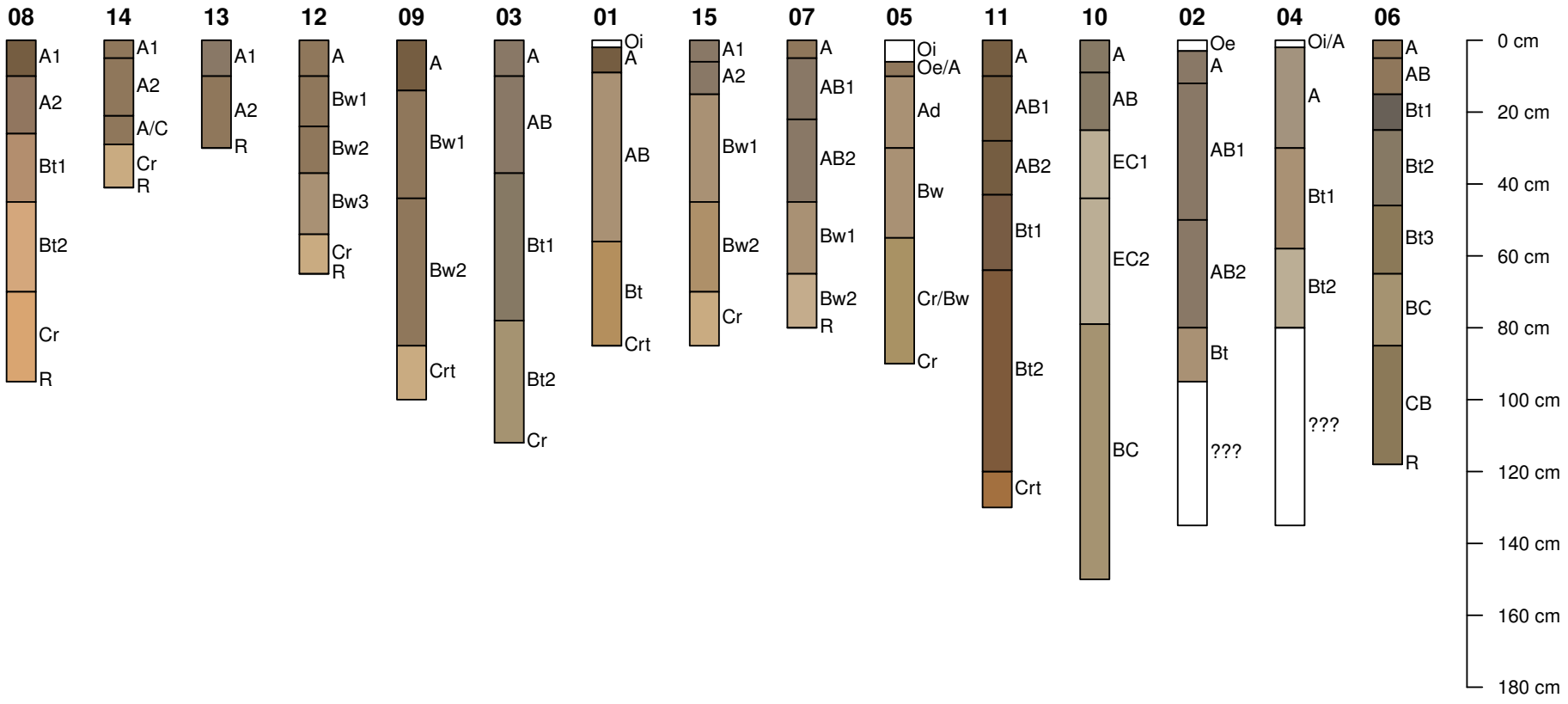
aqp – Algorithms for Quantitative Pedology

Soil profile sketches

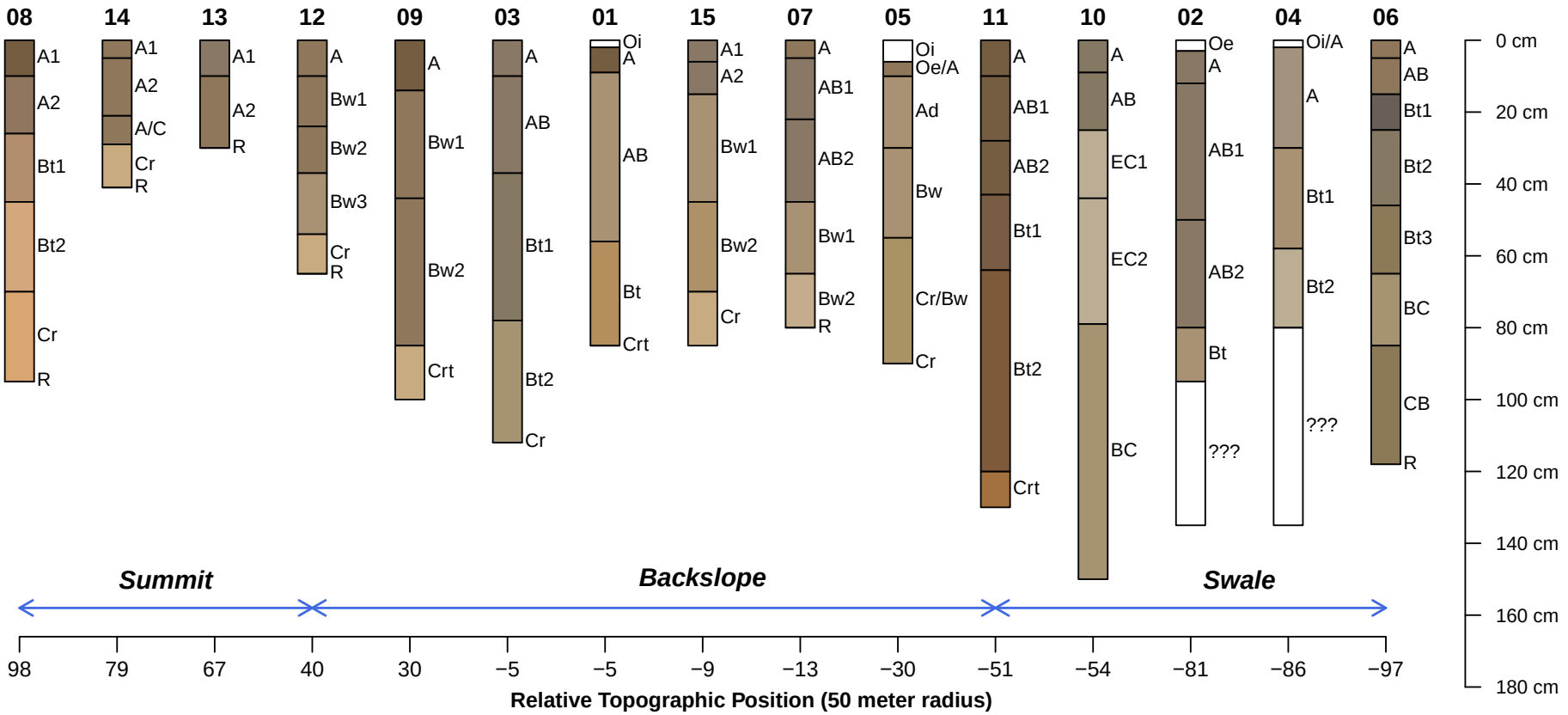


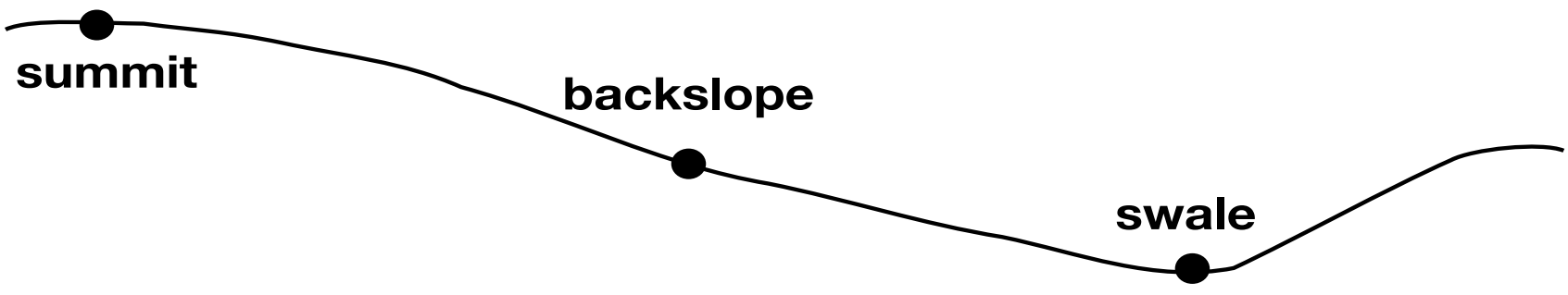
An example of profile sketches manually created from soil profile observations collected as part of the Pinnacles National Monument soil survey. Horizon designations, sequences, boundaries, and soil texture classes are usually sufficient for describing complex soil-landscape relationships.

(Digital) Soil profile sketches

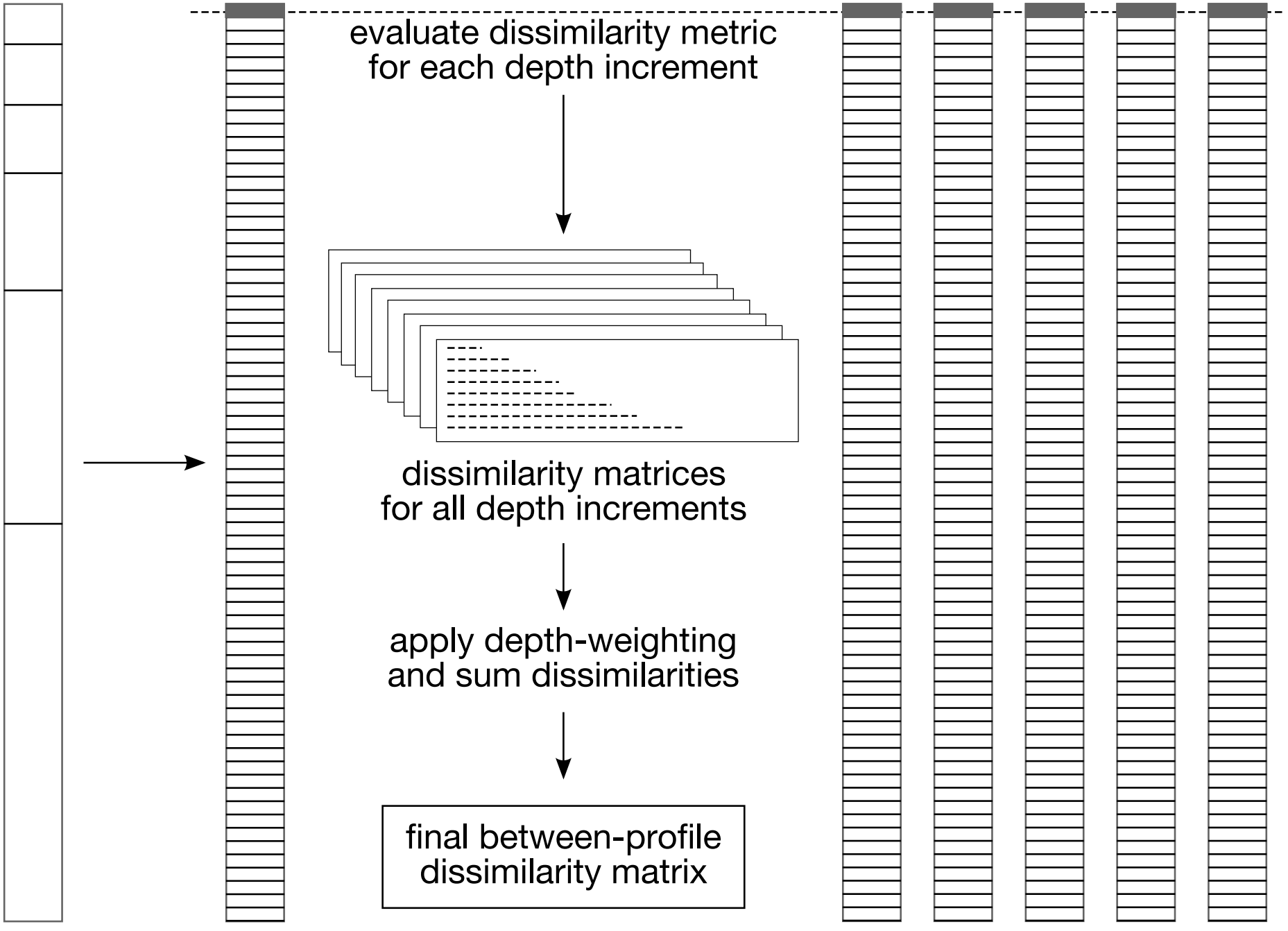


(Digital) Soil profile sketches

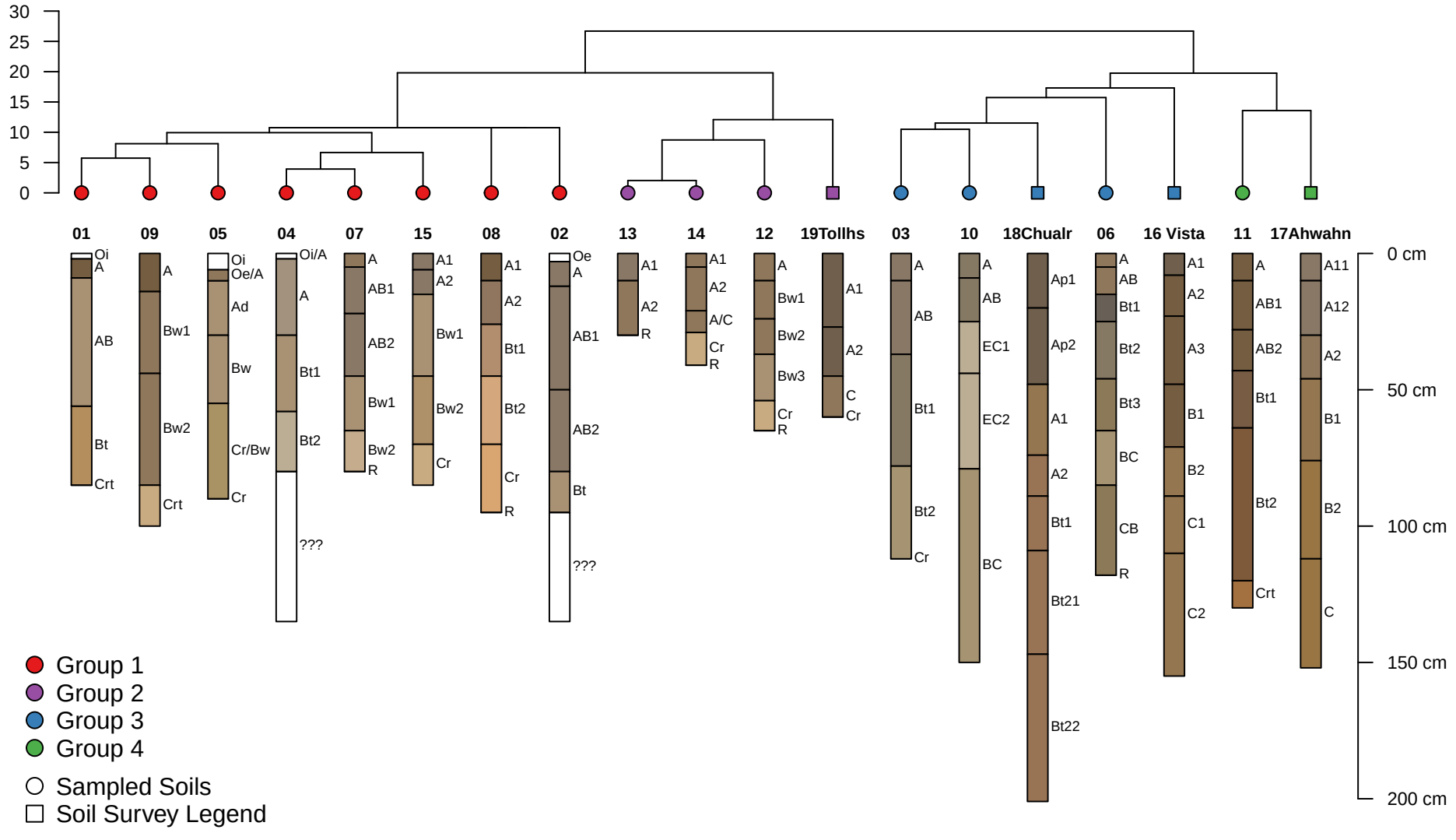




Numerical soil classification



Numerical soil classification

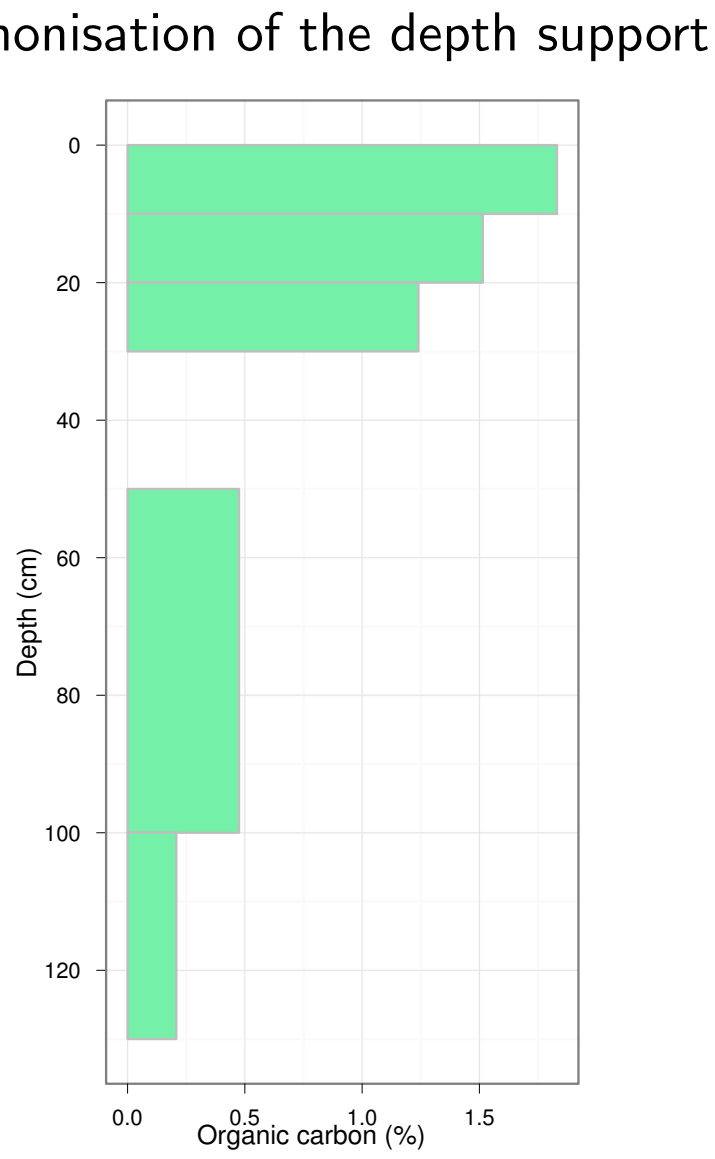


Soil data harmonisation

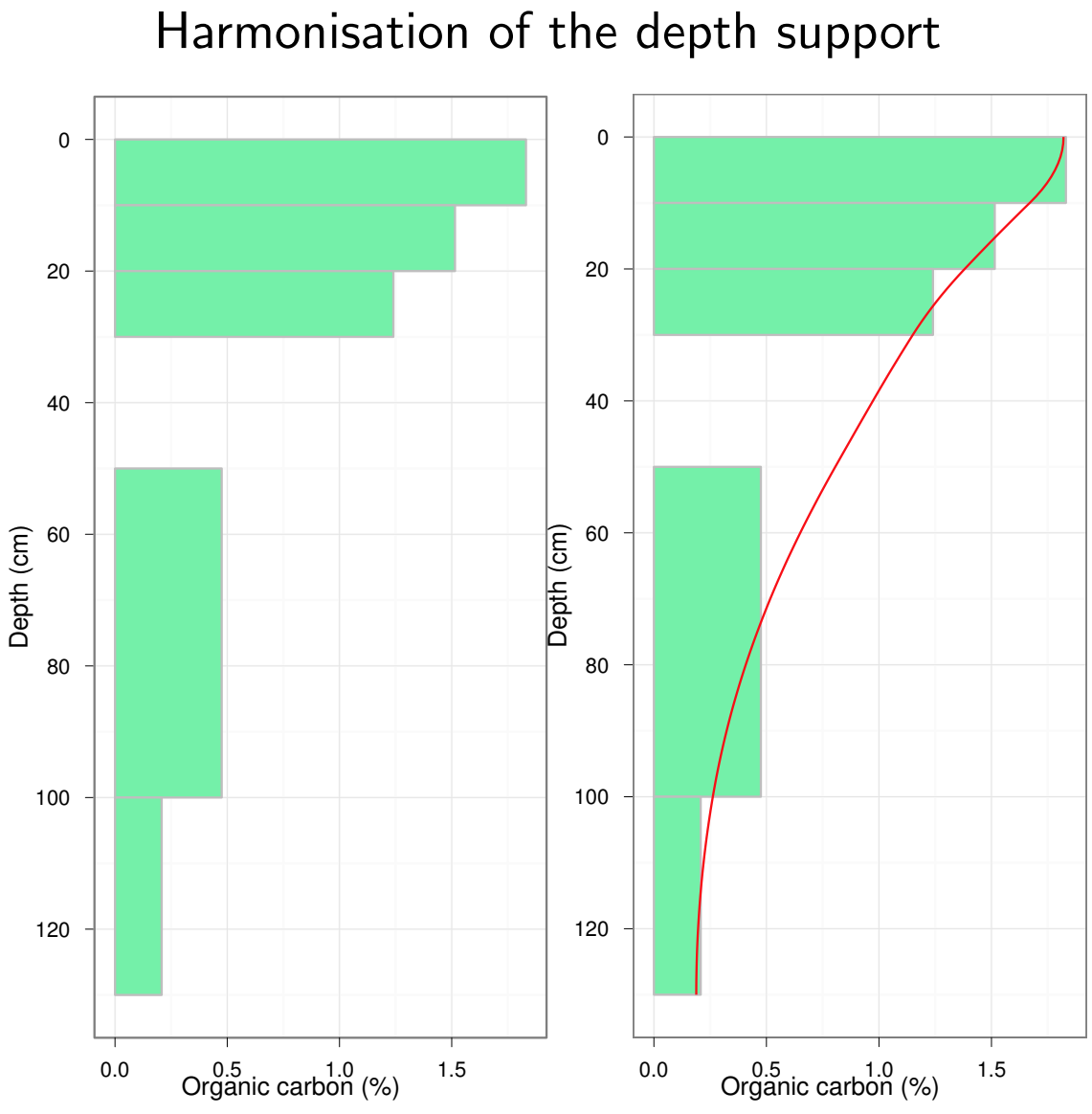
Harmonisation of heterogeneous data sets

- Legacy data
- Diversity of measurement methods
- Etc.

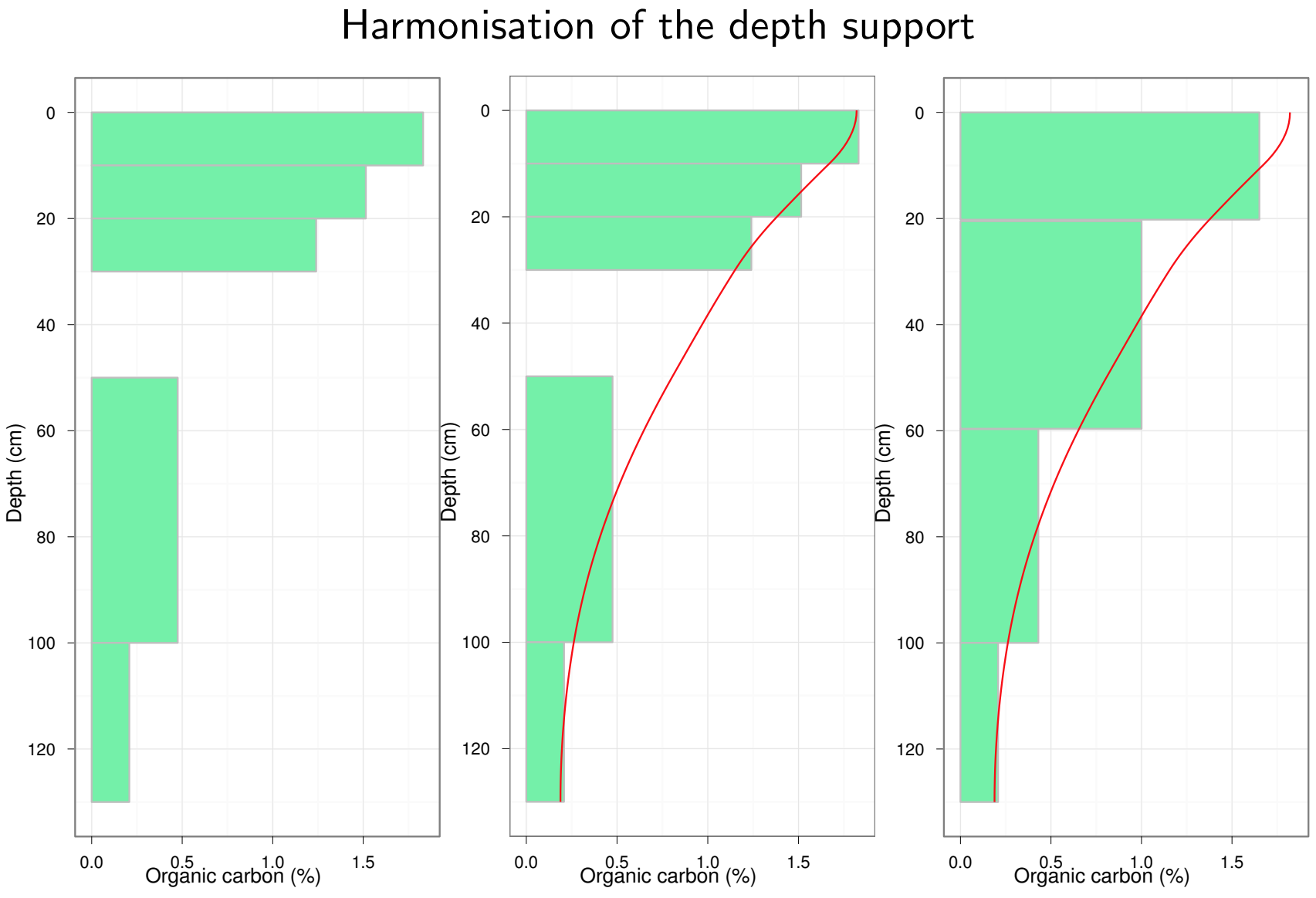
Soil data harmonisation



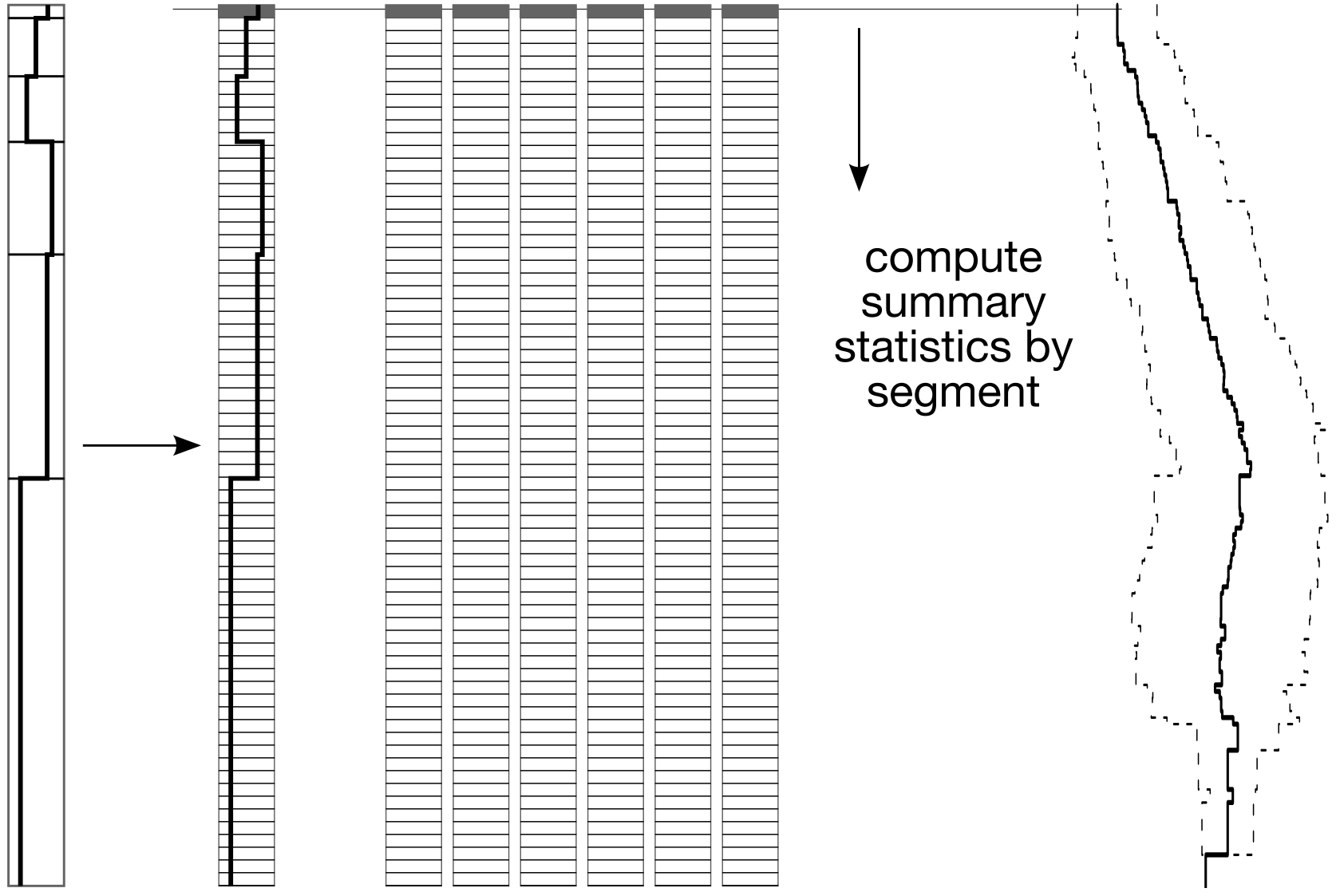
Soil data harmonisation



Soil data harmonisation



Aggregation of soil properties

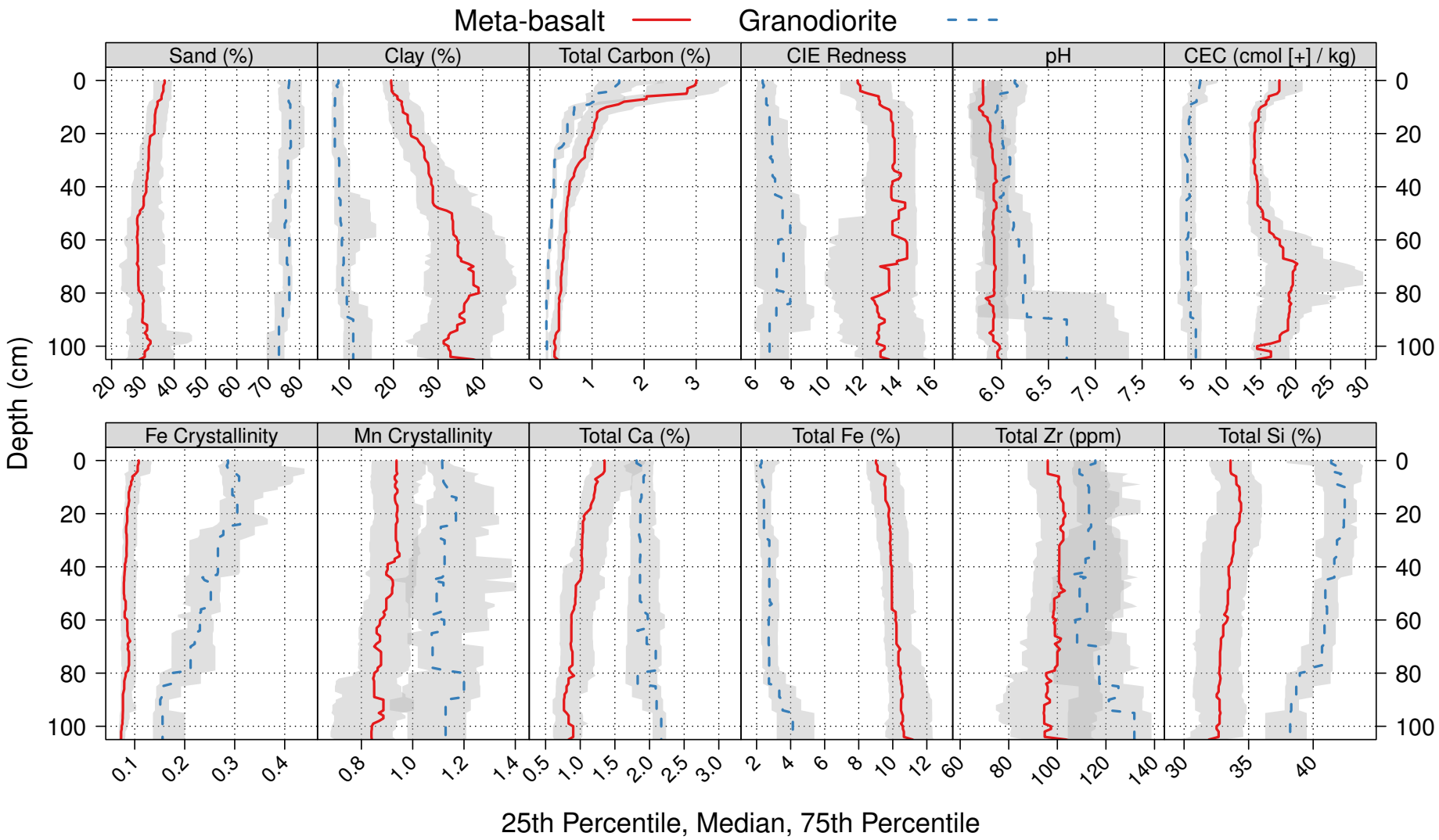


segment
each profile

collect profiles

representative depth function
and characteristic variability

Aggregation of soil properties



Pedotransfer functions

"*Pedotransfer functions*" are models predicting a soil attribute from other(s) soil attribute(s) and environmental covariates.

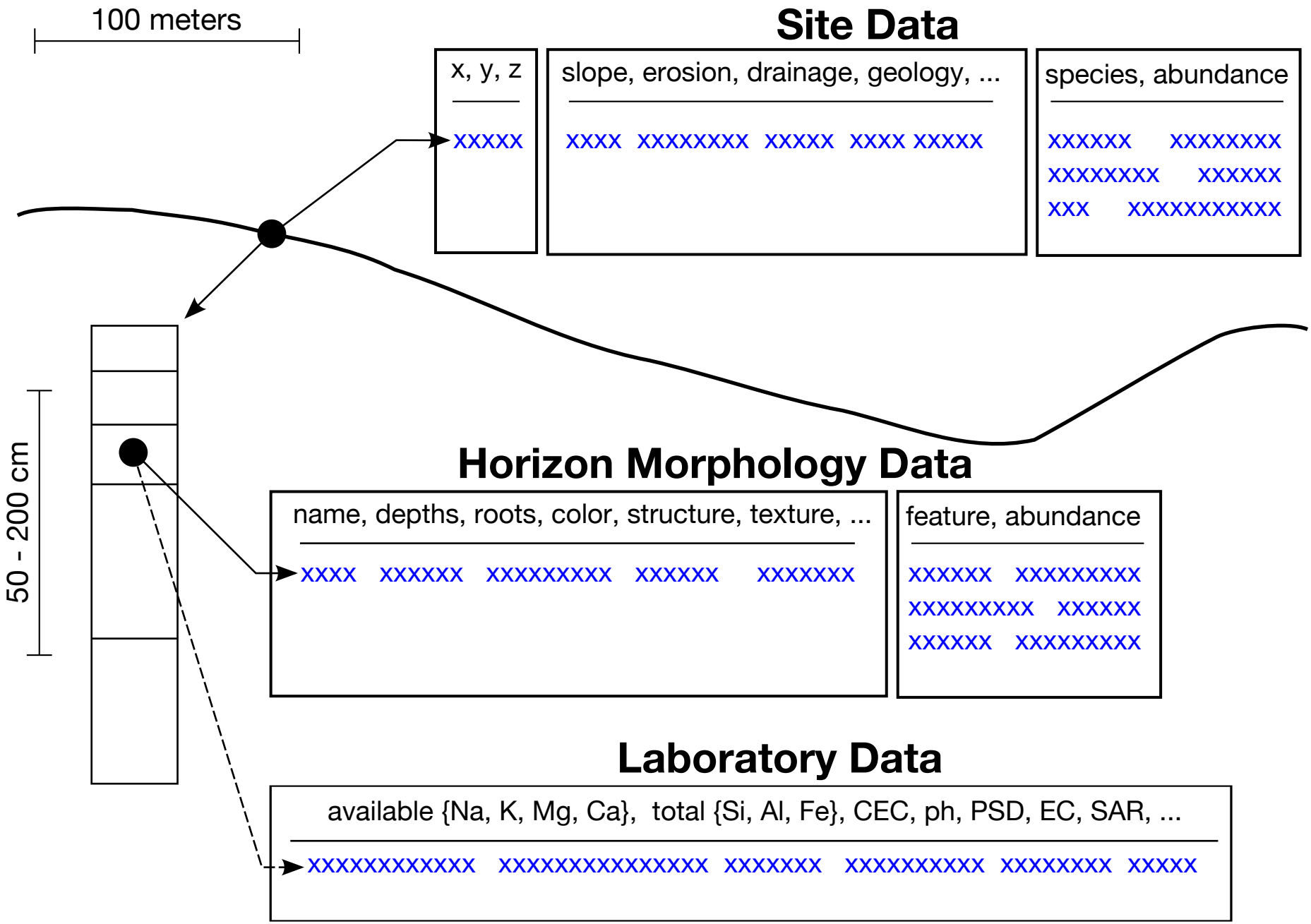
- From simple approach with multivariate linear models...
- ... to state-of-the-art machine learning algorithms (ANN, random forests, etc.)

$\hookrightarrow$ R is of course a natural platform for these tasks.

Talk Outline

1. The soil resource
   - Soil matters!
   - Soil science
   - Soil data and its analysis
2. The aqp package
   - Visualisation
   - Classification
   - Harmonisation
   - Analysis and modelling
3. Recent and future developments
   - Introduction of S4 classes
   - Design
4. Conclusions and further work

Structuring soil data

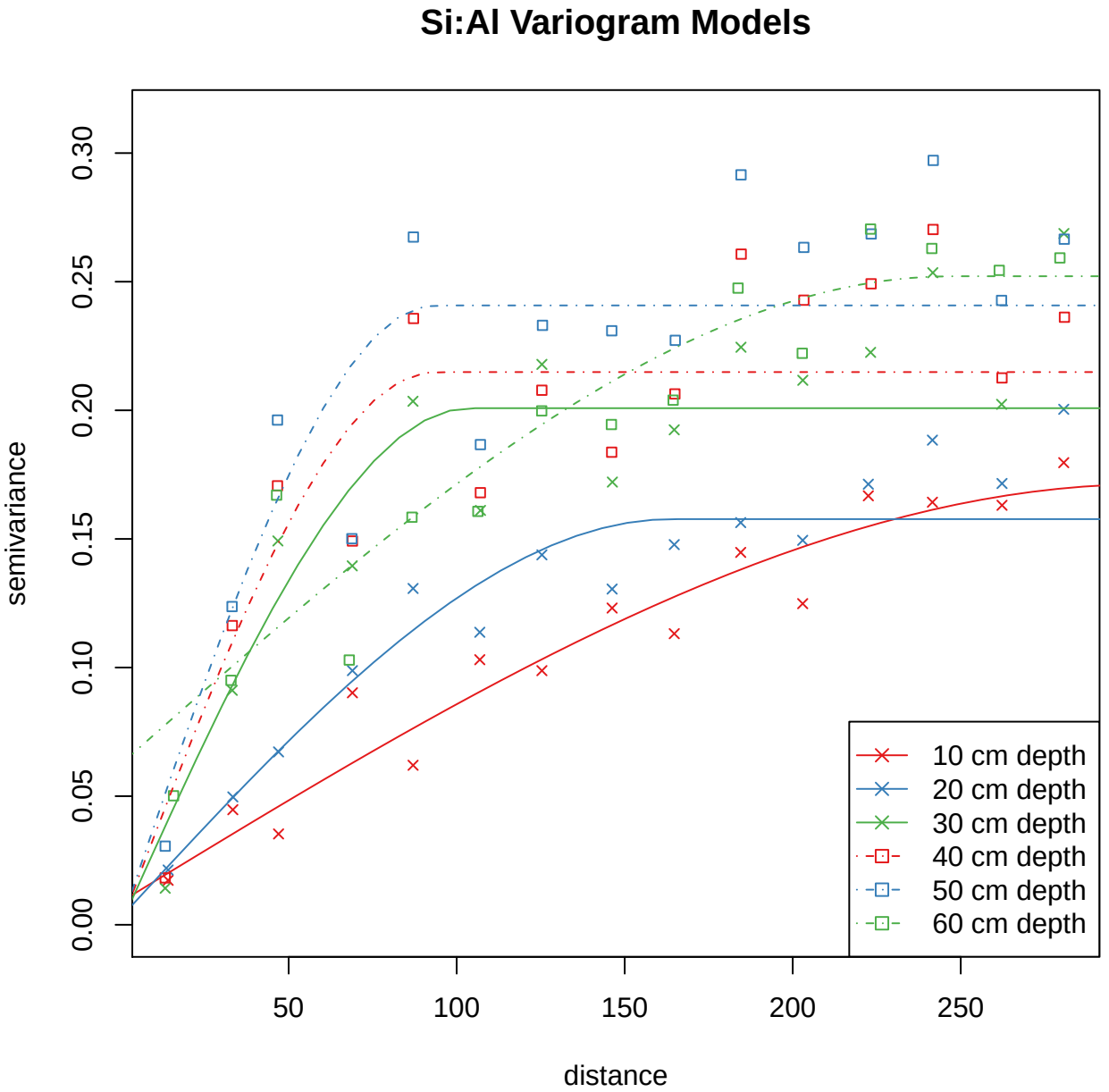


Spatial analysis

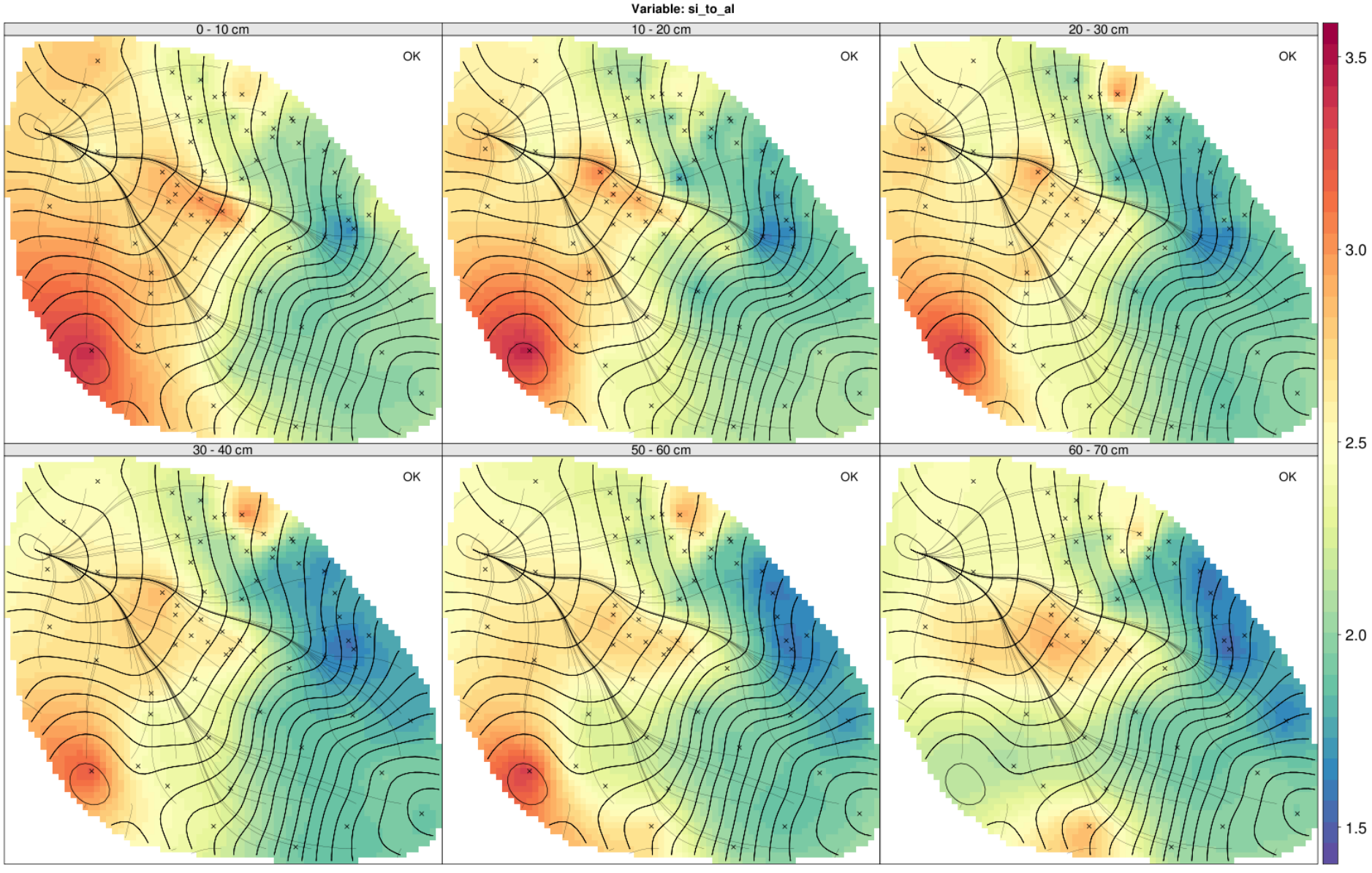
Spatial analyis is an example where we need bindings to other packages.

- Slice the soil collection on any depth interval
- Cast it as a `sp` object (`SpatialPointsDataFrame`)
- Apply `sp`/`raster` methods for spatial data analysis

Spatial analysis



Spatial analysis



S4 Classes and Methods for Soil Information

```
Formal class 'SoilProfileCollection' [package ".GlobalEnv"] with 4 slots
  ..@ id : Named chr "P001"
  .. ..- attr(*, "names")= chr "id"
  ..@ depths : int [1:6, 1:2] 0 2 14 49 57 89 2 14 49 57 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:6] "1" "2" "3" "4" ...
  .. .. ..$ : chr [1:2] "top" "bottom"
  ..@ units : chr "cm"
  ..@ horizons:'data.frame': 6 obs. of 4 variables:
  .. ..$ texture : Factor w/ 14 levels "C","CBVSCL","FSL",..: 8 8 8 8 9 NA
  .. ..$ stickiness : Factor w/ 4 levels "MS","SO","SS",..: NA NA NA NA NA NA
  .. ..$ plasticity : Factor w/ 4 levels "MP","PO","SP",..: NA NA NA NA NA NA
  .. ..$ field_ph : num [1:6] 7.9 7.7 8 7.9 7.4 NA
  ..@ site:'data.frame': 6 obs. of 2 variables:
  .. ..$ bound_topography: Factor w/ 3 levels "B","S","W": 2 2 2 2 2 NA
  .. ..$ elevation: int [1:6] 13 7 9 14 21 NA
```

S4 Classes and Methods for Soil Information

```
Formal class 'SoilProfileCollection' [package ".GlobalEnv"] with 4 slots
  ..@ id : Named chr "P001"
  .. ..- attr(*, "names")= chr "id"
  ..@ depths : int [1:6, 1:2] 0 2 14 49 57 89 2 14 49 57 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:6] "1" "2" "3" "4" ...
  .. .. ..$ : chr [1:2] "top" "bottom"
  ..@ units : chr "cm"
  ..@ horizons:'data.frame': 6 obs. of 4 variables:
  .. ..$ texture : Factor w/ 14 levels "C","CBVSCL","FSL",..: 8 8 8 8 9 NA
  .. ..$ stickiness : Factor w/ 4 levels "MS","SO","SS",..: NA NA NA NA NA NA
  .. ..$ plasticity : Factor w/ 4 levels "MP","PO","SP",..: NA NA NA NA NA NA
  .. ..$ field_ph : num [1:6] 7.9 7.7 8 7.9 7.4 NA
  ..@ site:'data.frame': 6 obs. of 2 variables:
  .. ..$ bound_topography: Factor w/ 3 levels "B","S","W": 2 2 2 2 2 NA
  .. ..$ elevation: int [1:6] 13 7 9 14 21 NA
  ..@ spatial:'SpatialPoints'
  ..@ metadata:'data.frame'
```

S4 Classes and Methods for Soil Information

```
# Initialization
depths(spc) <- id ~ top + bottom

# Adding site data
site(spc) <- ~ slope + aspect + curvature + x + y + z

# Adding spatial coordinates (sp)
coordinates(spc) <- ~ x + y
```

S4 Classes and Methods for Soil Information

```
# Initialization
depths(spc) <- id ~ top + bottom

# Adding site data
site(spc) <- ~ slope + aspect + curvature + x + y + z

# Adding spatial coordinates (sp)
coordinates(spc) <- ~ x + y

# Accessors
units(spc) # the units for depths
depths(spc) # get depth matrix
depthsnames(spc) # get names of depth columns
profile_id(spc) # get profile IDs
horizons(spc) # get horizon data as dataframe
```

S4 Classes and Methods for Soil Information

```
# Initialization
depths(spc) <- id ~ top + bottom

# Adding site data
site(spc) <- ~ slope + aspect + curvature + x + y + z

# Adding spatial coordinates (sp)
coordinates(spc) <- ~ x + y

# Accessors
units(spc) # the units for depths
depths(spc) # get depth matrix
depthsnames(spc) # get names of depth columns
profile_id(spc) # get profile IDs
horizons(spc) # get horizon data as dataframe

# Overloads
min(spc) # min depth within collection
max(spc) # max depth within collection
length(spc) # number of profiles

# Coercicion
as.data.frame(spc) # convert back to original dataframe

# dataframe-like interface
spc$property # read property
spc$property <- runif(length(spc)) # write property
```

Talk Outline

1. The soil resource
   - Soil matters!
   - Soil science
   - Soil data and its analysis
2. The aqp package
   - Visualisation
   - Classification
   - Harmonisation
   - Analysis and modelling
3. Recent and future developments
   - Introduction of S4 classes
   - Design
4. Conclusions and further work

Conclusions

- Soil science is changing, and **needs new tools**
- aqp aims at providing soil scientists with a R-based pedometrics **platform**

Conclusions

- Soil science is changing, and **needs new tools**
- aqp aims at providing soil scientists with a R-based pedometrics **platform**
- R is a great platform for soil science:
  - The ultimate Excel-killer!
  - Data and covariates in the same object
  - Provides advanced visualisations of soil data
  - Huge source of regression/classifion methods...
  - ... but also allows to test/extend/create
  - Spatial-literate environment
  - Supports big data (parallelisation backends)