



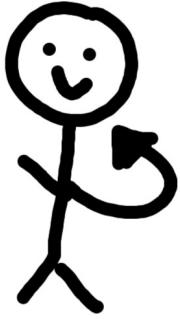
# Mastering Machine Learning for Spatial Prediction

Quick teaser and an application

OpenGeoHub Summer School  
20 August 2020

**Madlene Nussbaum**  
CC BY 4.0

# on myself ...



- **Geographer** with diploma and PhD in applied statistics
- Current position: **Bern University of Applied Sciences** (BFH ≠ UniBE), Section for Agronomy. Research group "soil protection and land use".
- **Focus:**  
Interdisciplinary research at the intersection of GIS, soil science and statistics. Predictive (spatial) modelling, data handling in field survey, visualisations etc.  
Moreover: foster knowledge transfer and **encourage FOSS!**
- Main language: **R**
- Further I fancy: QGis, Latex, SAGA, GRASS, OpenLayers, PostGIS, Inkscape, Git, Debian



# Teaser of today's programme ...

11:00 – 13:00

## **Mastering machine learning for spatial prediction I**

Basic concepts, an overview and introduction of widespread methods like shrinkage, generalized additive models, tree based methods, neural networks, support vector machines. Concepts like bootstrap, boosting and model averaging.

14:00 – 16:00

## **Mastering machine learning for spatial prediction II**

Covariate selection with importance, interpretation of covariate-response relationships with partial dependence plots and maps, uncertainty of model predictions by bootstrapping

→ ***both: first lecture then hands-on exercises***

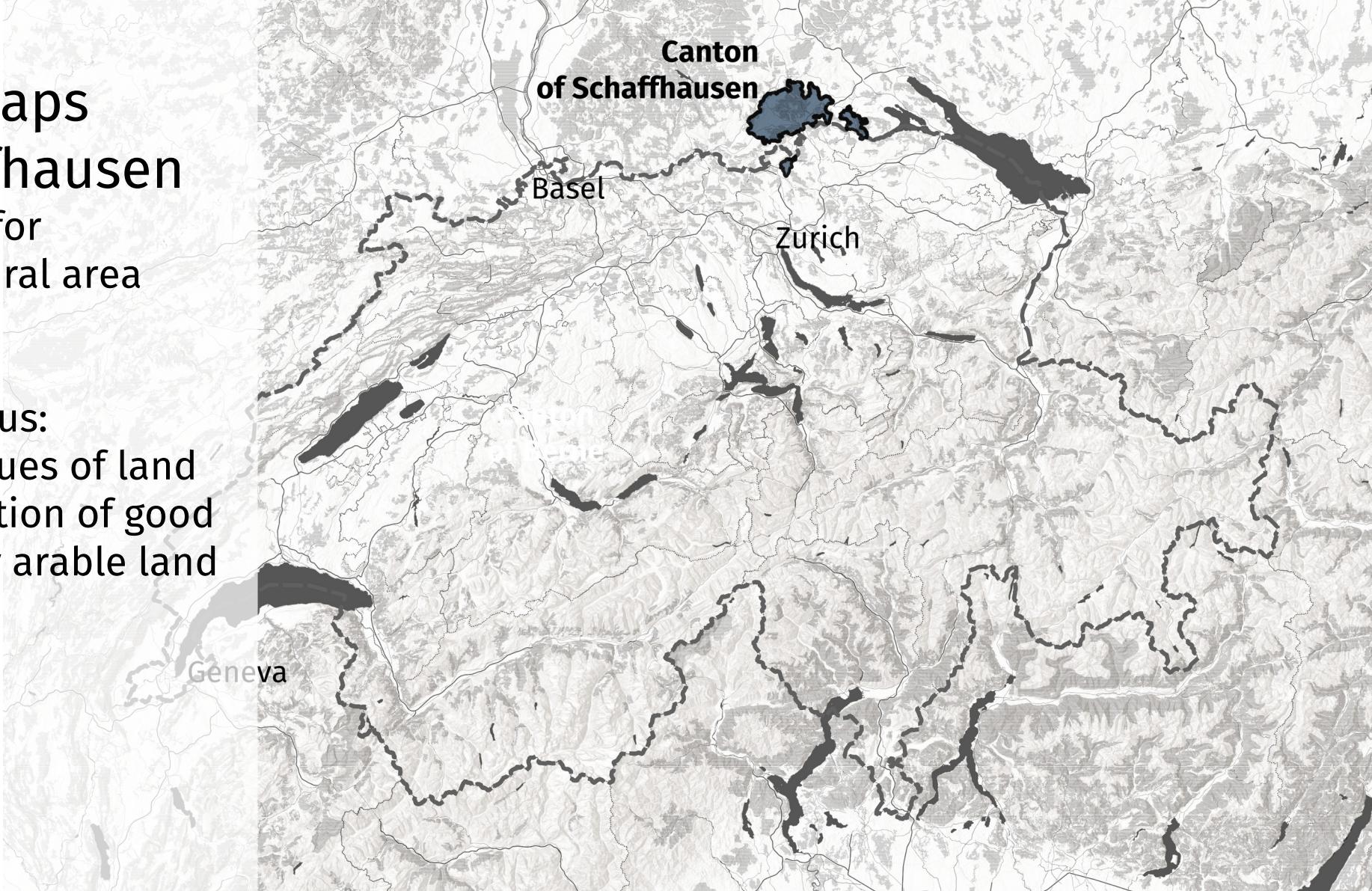
For you inspiration...  
a real world application from my current projects.

Exploit old soil maps to predict soil properties  
in unmapped areas

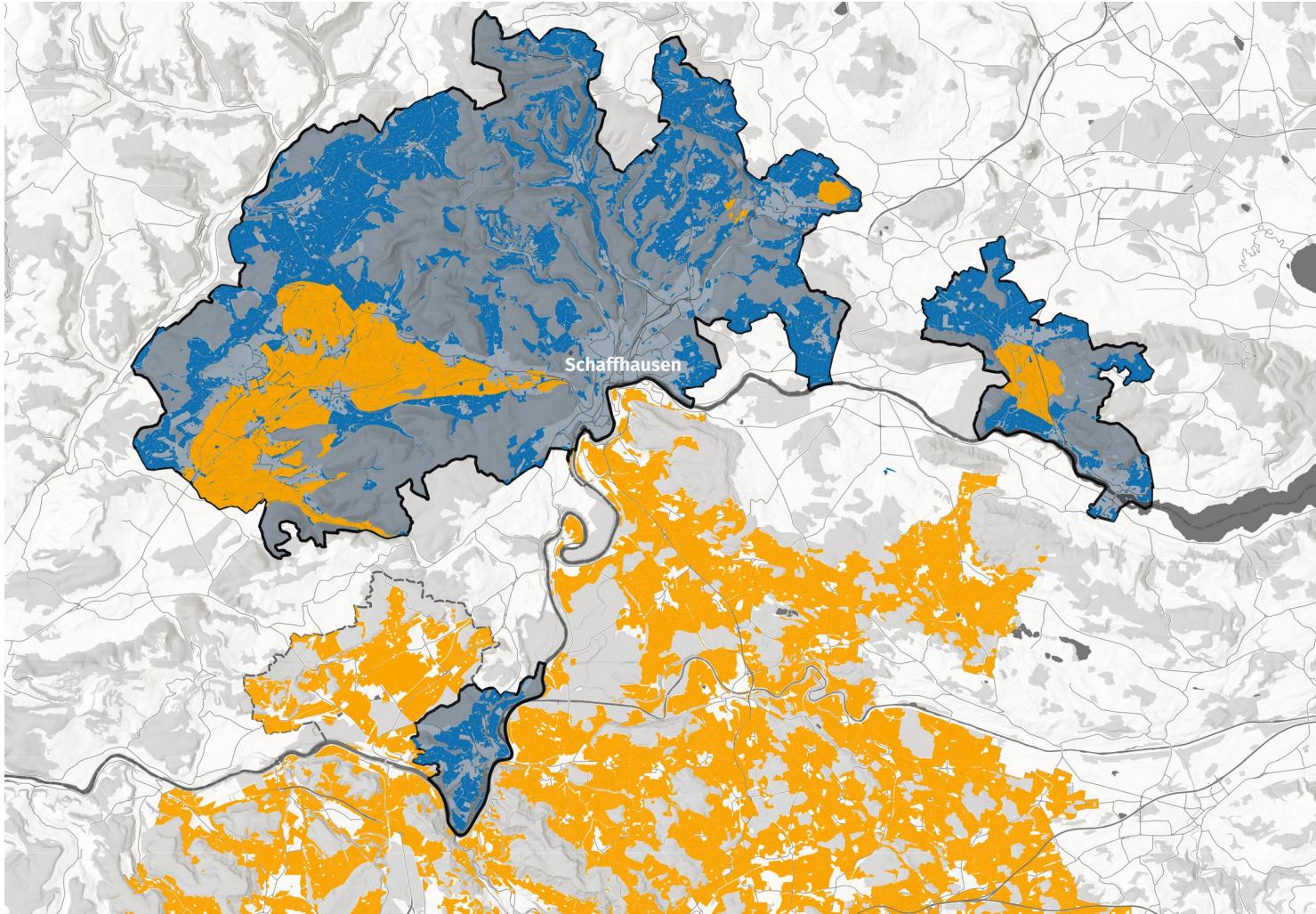
# Soil maps Schaffhausen needed for agricultural area needed.

Main focus:

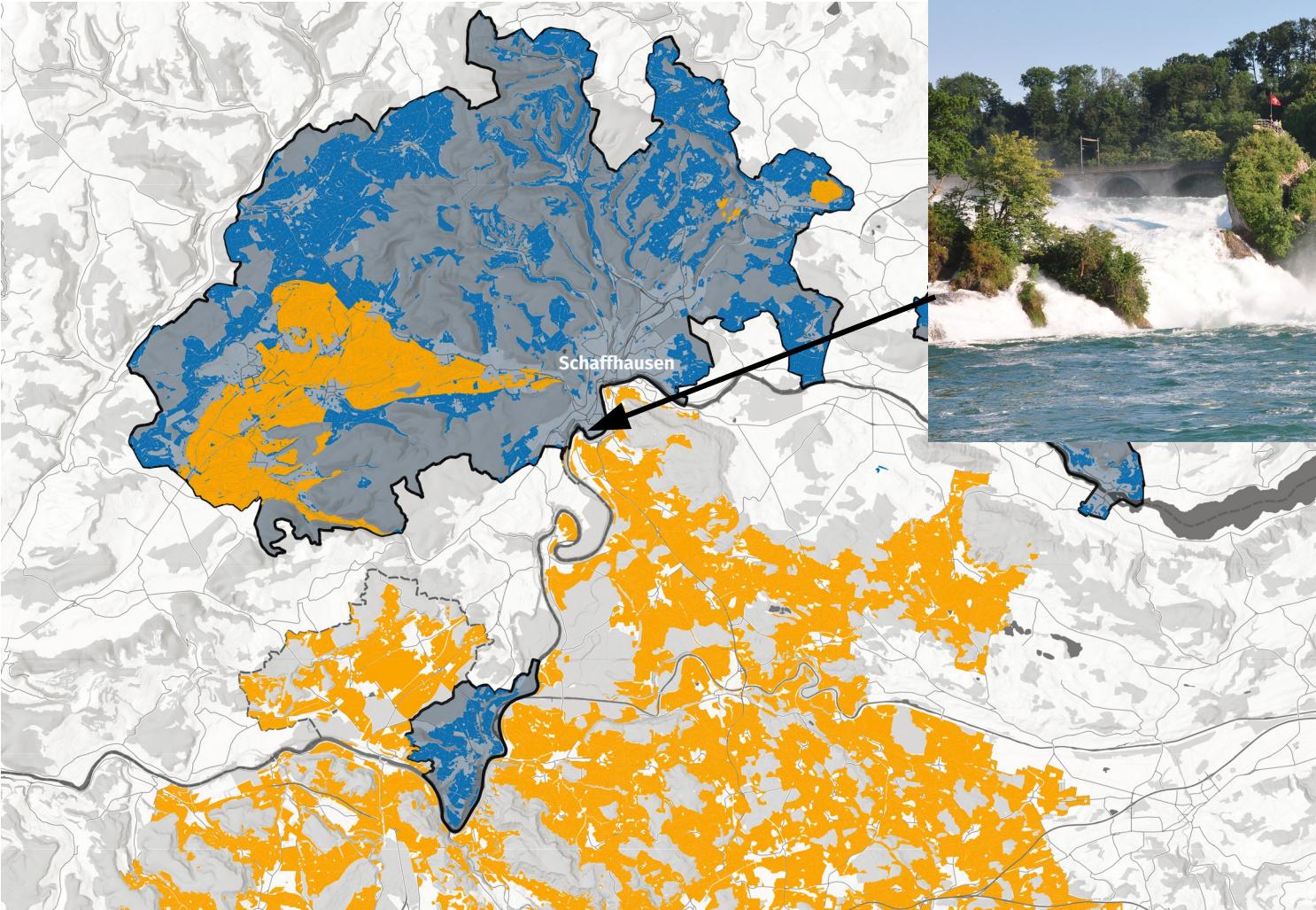
- tax values of land
- protection of good quality arable land



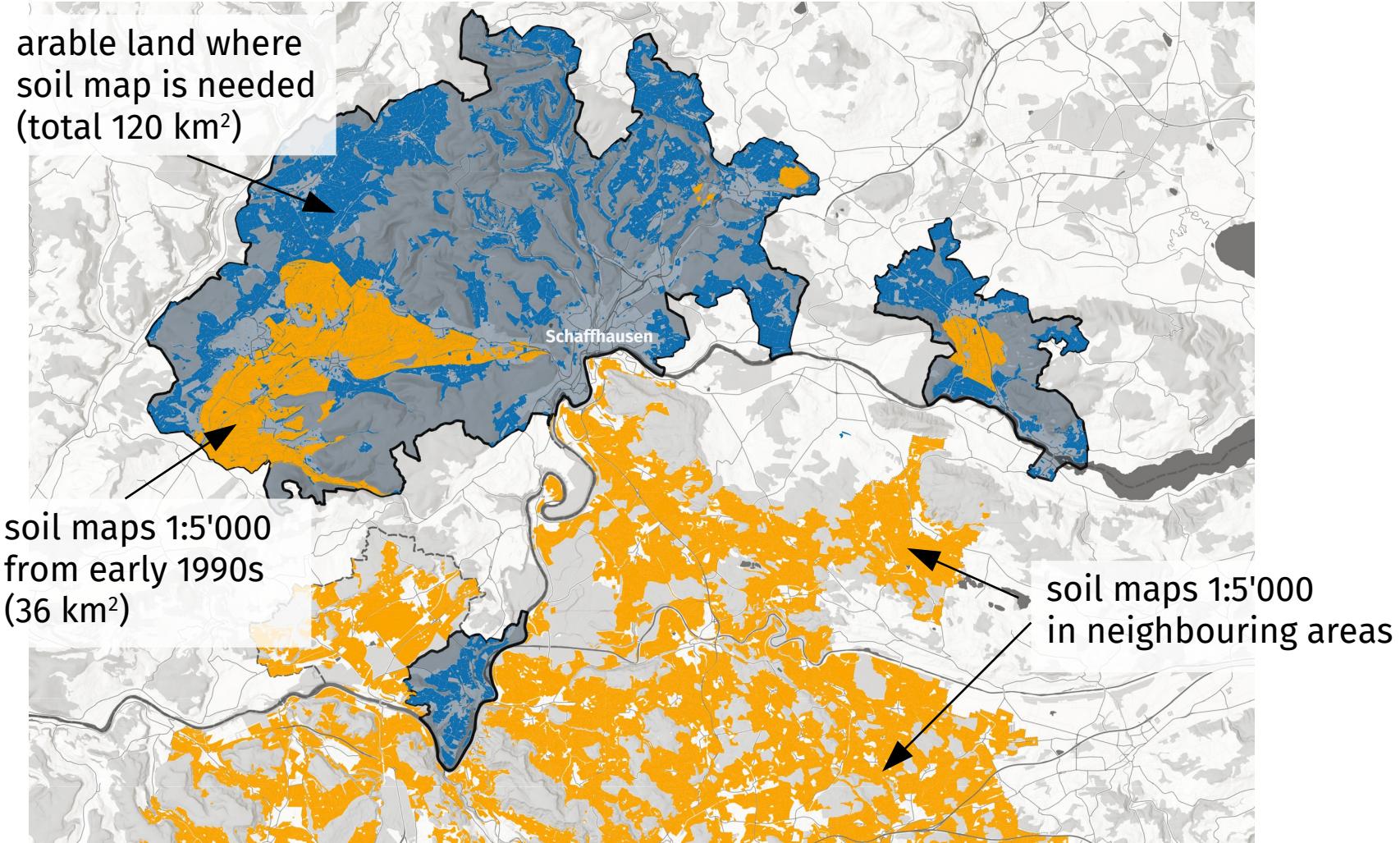
# Situation in Schaffhausen



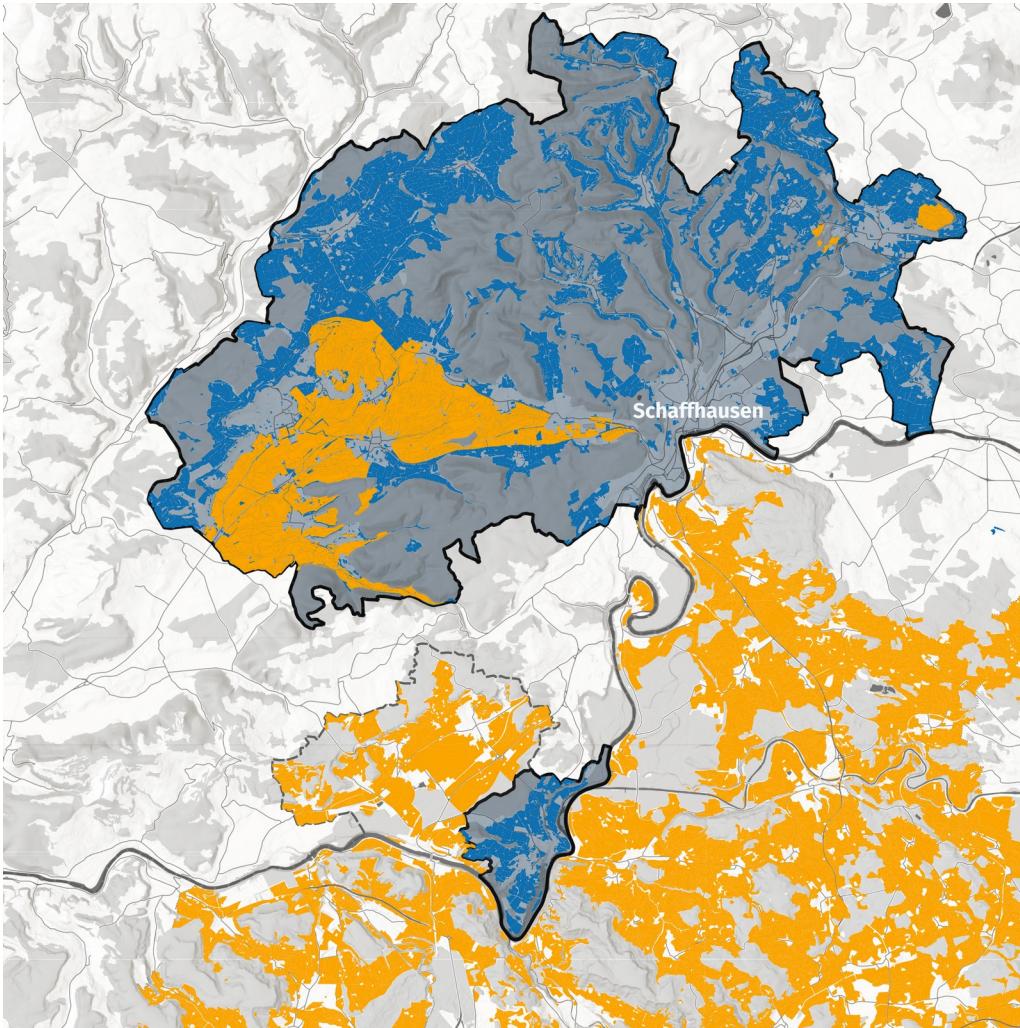
# Situation in Schaffhausen



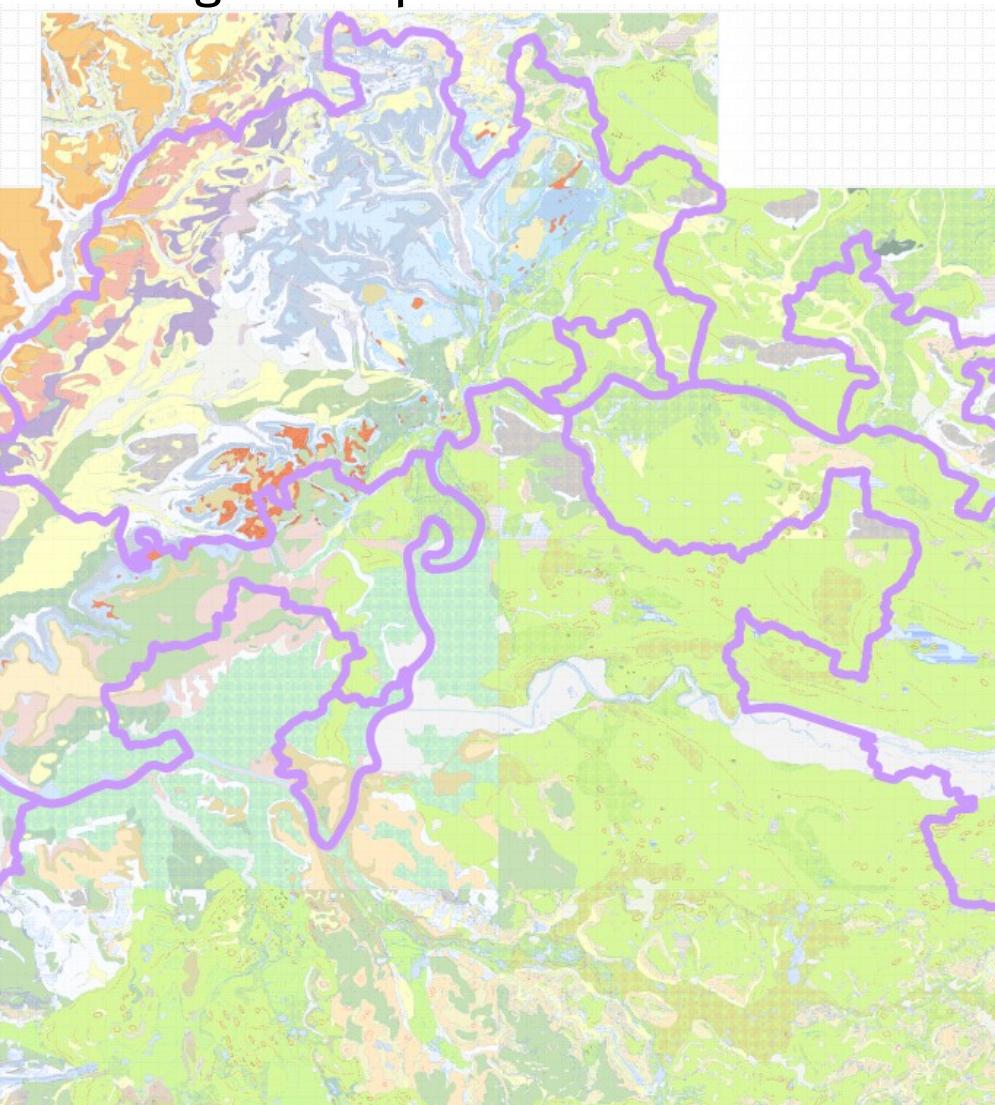
# Situation in Schaffhausen



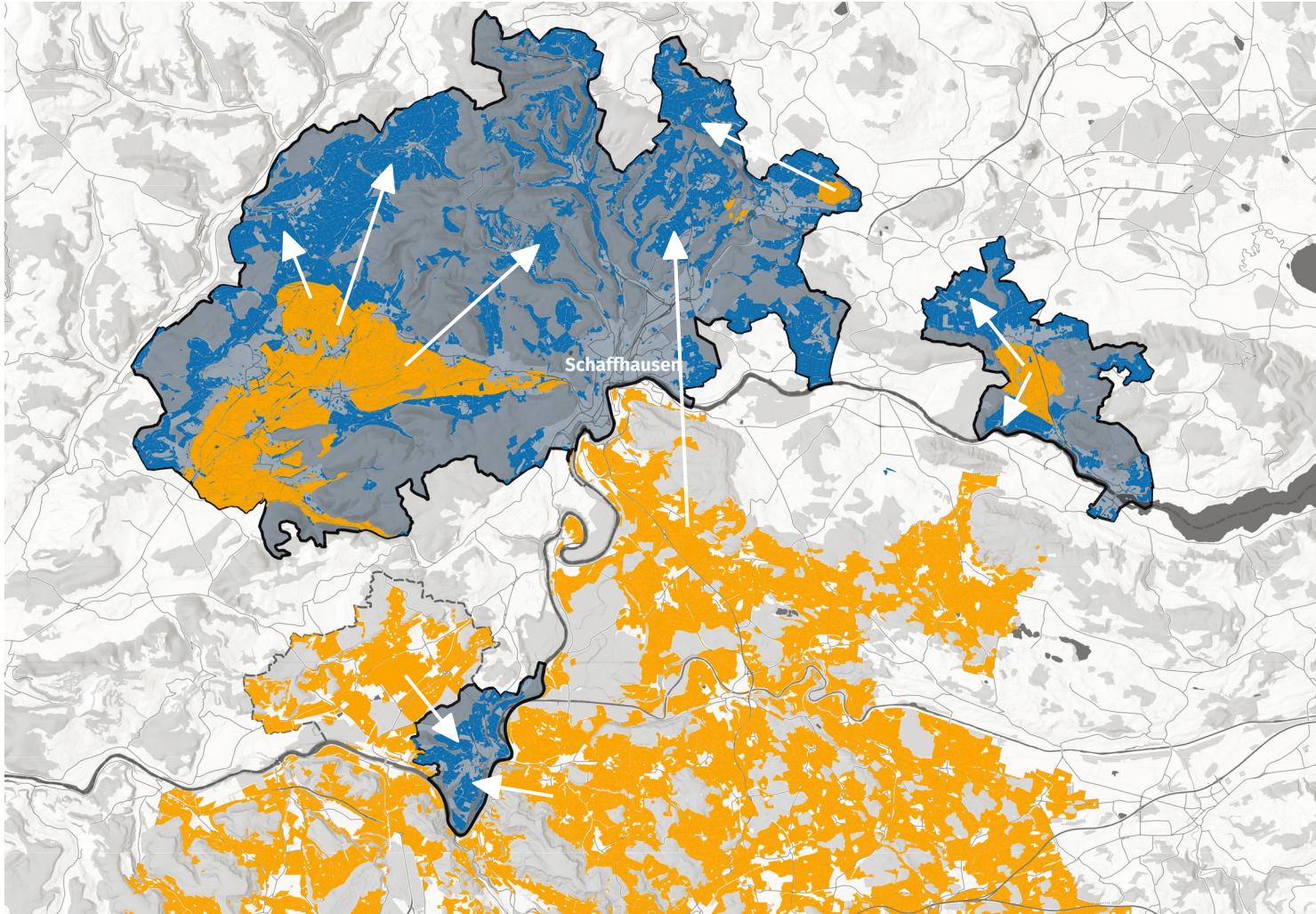
# Situation in Schaffhausen

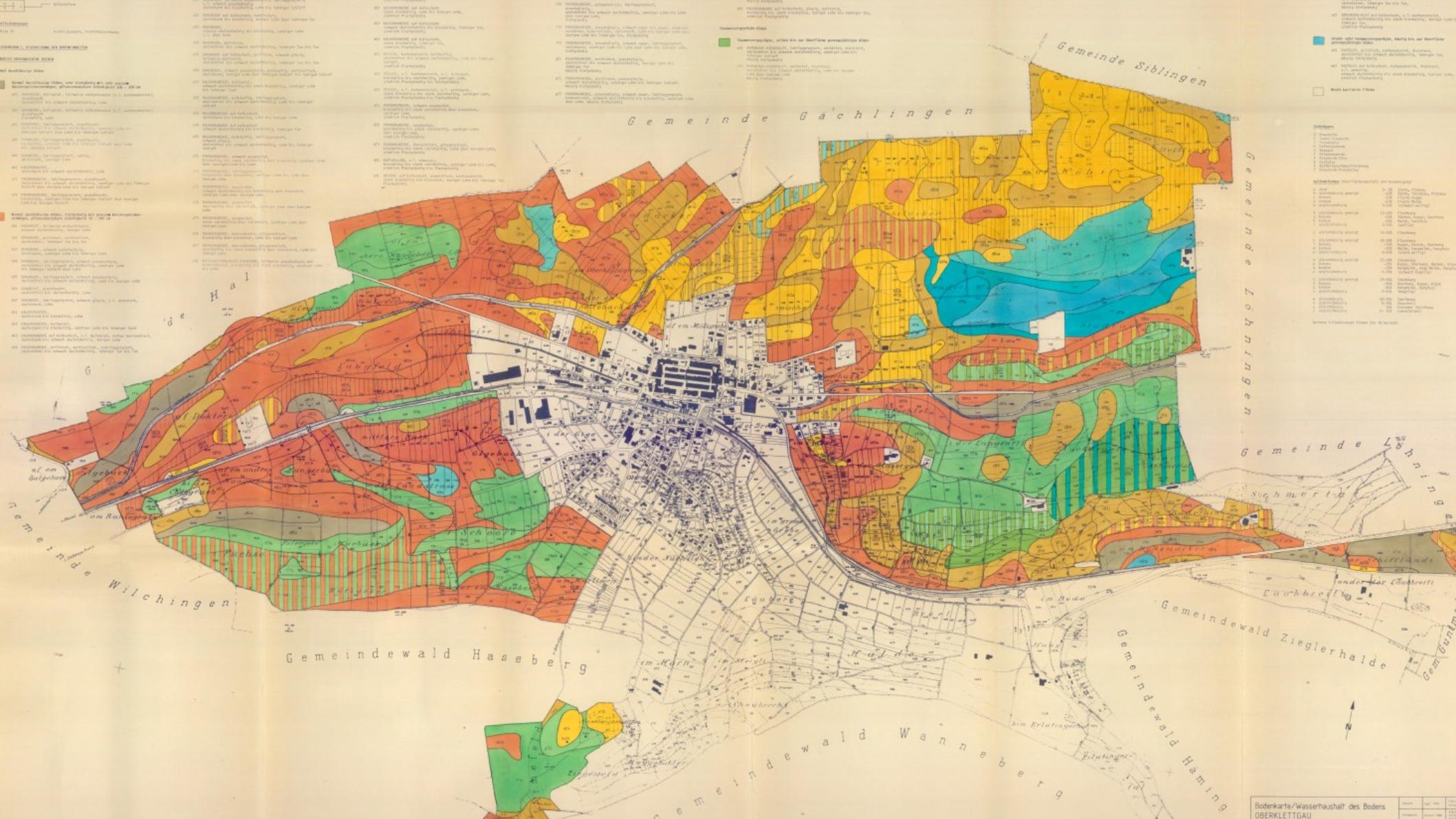


Geological Maps 1:25'000



# Extrapolating into non-mapped areas

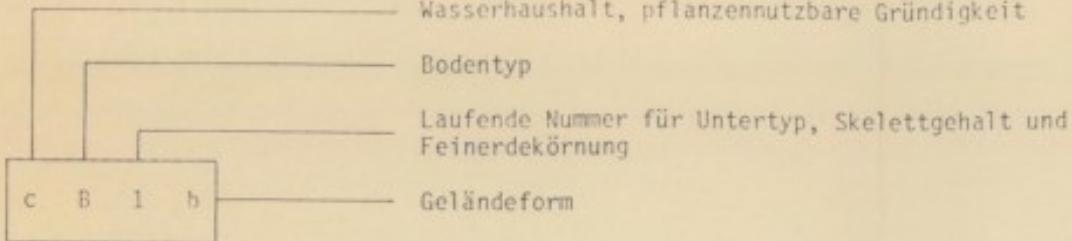




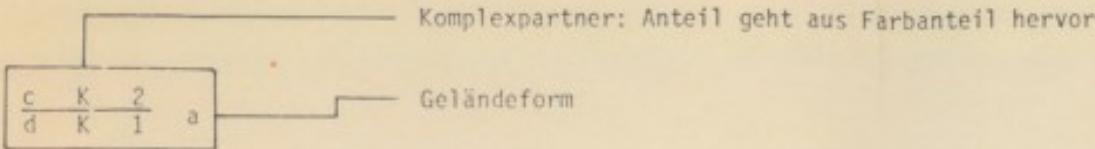
# Content of old soil map

## ERLÄUTERUNG DER SYMbole

### Reine Bodeneinheiten



### Zusammengesetzte Bodeneinheiten (Komplexe)



For each of these legend codes:  
further information in booklet (prose)

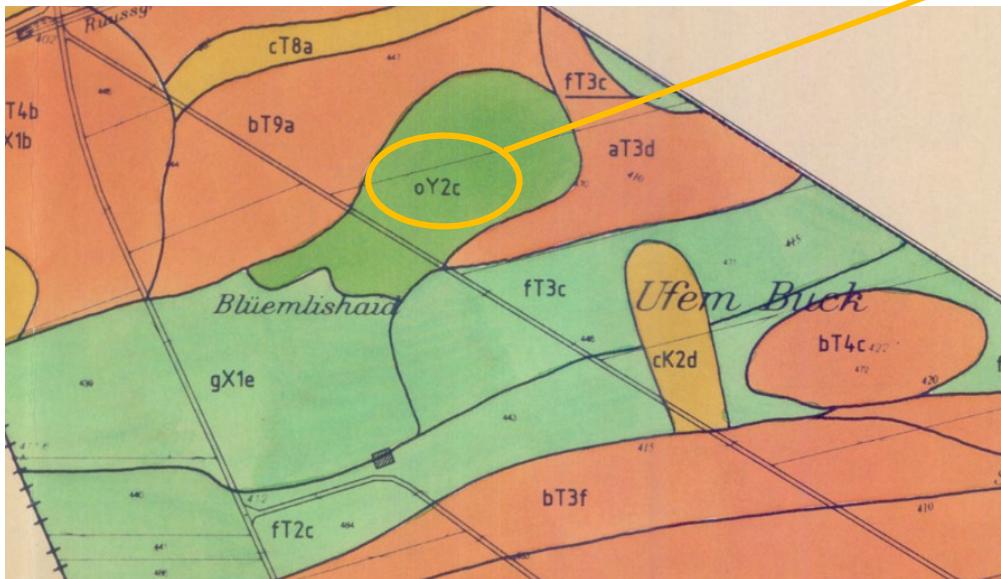
- aB3 BRAUNERDE, labilaggregiert, grundfeucht, skelettfrei bis schwach skeletthalig, sandiger Lehm bis lehmiger Schluff über Lehm bis lehmigem Schluff
- aB4 BRAUNERDE, labilaggregiert, grundfeucht, kieshaltig, sandiger Lehm bis lehmiger Schluff über Lehm bis lehmigem Schluff
- aB5 BRAUNERDE, labilaggregiert, sandig, skelettarm, sandiger Lehm
- aK1 KALKBRAUNERDE, skelettarm bis schwach skeletthalig, Lehm
- aT1 PARABRAUNERDE, labilaggregiert, grundfeucht, skelettfrei bis schwach skeletthalig, sandiger Lehm bis lehmiger Schluff über tonigem Lehm bis tonigem Schluff
- aT2 PARABRAUNERDE, labilaggregiert, grundfeucht, kieshaltig, sandiger Lehm bis lehmiger Schluff über tonigem Lehm bis tonigem Schluff
- Normal durchlässige Böden; tiefgründig mit grossem Wasserspeicher-vermögen, pflanzennutzbare Gründigkeit 70 - 100 cm**
- bB1 BRAUNERDE, teilweise entkarbonatet, schwach skeletthalig, toniger Lehm
- bB2 BRAUNERDE, pelitisch, vertisolisch, skelettfrei, lehmiger Ton bis Ton
- bB3 BRAUNERDE, schwach pseudogleyig, skelettarm, sandiger Lehm bis lehmiger Sand
- bB4 BRAUNERDE, labilaggregiert, schwach pseudogleyig, skelettfrei bis schwach skeletthalig, sandiger Lehm bis lehmiger Schluff über Lehm
- bB5 BRAUNERDE, labilaggregiert, schwach pseudogleyig, skelettfrei bis schwach skeletthalig, Lehm bis toniger Lehm
- bB6 BRAUNERDE, grundfeucht, skelettfrei bis skeletthalig, Lehm
- bB7 BRAUNERDE, labilaggregiert, schwach gleyig, z.T. drainiert, skelettarm, Lehm
- bK1 KALKBRAUNERDE, skelettarm bis kieshaltig, Lehm
- bK2 KALKBRAUNERDE, kolluvial, skelettarm bis kieshaltig, sandiger Lehm bis lehmiger Sand
- bK3 KALKBRAUNERDE auf Kalkschutt, z.T. mullreich, diffus horizontiert, skelettarm bis schwach skeletthalig, toniger Lehm
- bK4 KALKBRAUNERDE, pelitisch, vertisolisch, stabilaggregiert, skelettfrei bis schwach skeletthalig, lehmiger Ton bis Ton

# Legacy soil maps

scale 1:5'000

200 soil profiles in Swiss soil database

→ many augerings, only recorded as polygons



## Legend booklet

### STAUVASSERGEPRÄGTE BOEDEN

selten bis zur Oberfläche porengesättigte Böden

oY1 BRAUNERDE-PSEUDOGLEY, labilaggregiert, verhärt skelettfrei bis schwach skeletthaltig, sandige lehmiger Schluff, mässig tiefgründig

oY2 BRAUNERDE-PSEUDOGLEY, verhärtet, drainiert, skelettfrei bis schwach skeletthaltig, Lehm über tonigem Lehm, mässig tiefgründig

Stauwassergeprägte, häufig bis zur Oberfläche porengesättigte Böden

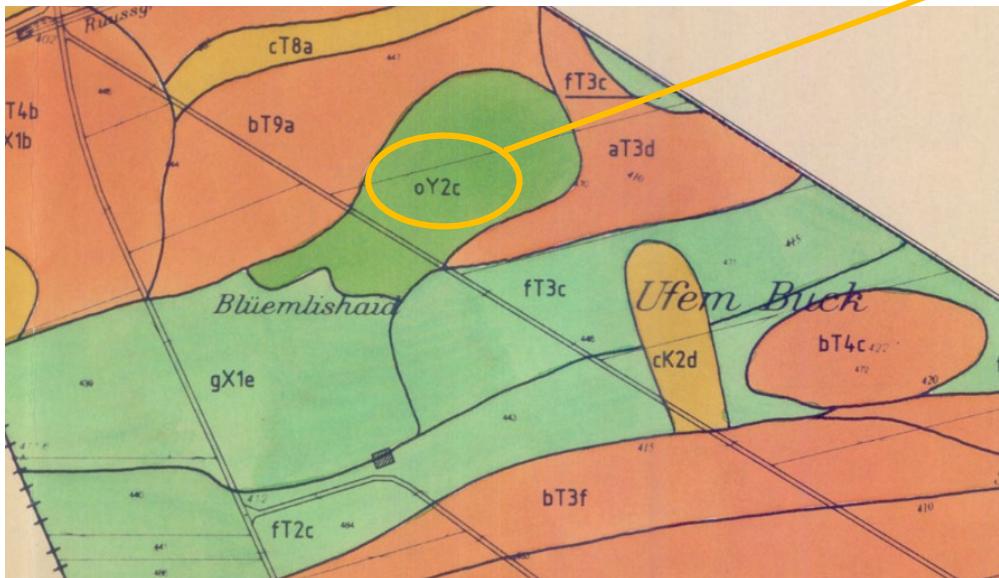
qI1 PSEUDOGLEY, labilaggregiert, verhärtet, drainiert, schwach skeletthaltig, Lehm bis lehmiger Schluff

# Legacy soil maps

scale 1:5'000

200 soil profiles in Swiss soil database

→ many augerings, only recorded as polygons



## Legend booklet

### STAUVASSERGEPRÄGTE BOEDEN

selten bis zur Oberfläche porengesättigte Böden

**oY1** BRAUNERDE-PSEUDOGLEY, labilaggregiert, verhärt skelettfrei bis schwach skeletthaltig, sandige lehmiger Schluff, mässig tiefgründig

**oY2** BRAUNERDE PSEUDOGLEY, verhärtet, drainiert, skelettfrei bis schwach skeletthaltig. Lehm bi Lehm über tonigem Lehm, mässig tiefgründig

Stauwassergeprägte, häufig bis zur Oberfläche porengesättigte Böden

**qI1** PSEUDOGLEY, labilaggregiert, verhärtet, drainiert, schwach skeletthaltig. Lehm bis lehmiger Schluff

## Tables with value ranges

Skelettklasse	Skelettgehalt Vol. %
skelettfrei	< 5
schwach skeletthaltig	5 - 10
kieshaltig*	10 - 20
steinhaltig	10 - 20
stark kieshaltig*	20 - 30
stark steinhaltig	20 - 30
kiesreich*	30 - 50

# Legacy soil maps

## Legend booklet

### STAUVASSERGEPRÄGTE BOEDEN

selten bis zur Oberfläche porengesättigte Böden

oY1 BRAUNERDE-PSEUDOGLEY, labilaggregiert, verhärt skelettfrei bis schwach skeletthaltig, sandige lehmiger Schluff, mässig tiefgründig

oY2 BRAUNERDE PSEUDOGLEY, verhärtet, drainiert, skelettfrei bis schwach skeletthaltig. Lehm bi Lehm über tonigem Lehm, mässig tiefgründig

Stauwassergeprägte, häufig bis zur Oberfläche porengesättigte Böden

qII PSEUDOGLEY, labilaggregiert, verhärtet, drainiert, schwach skeletthaltig. Lehm bis lehmiger Schluff

### Tables with value ranges

Skelettklasse	Skelettgehalt Vol. %
skelettfrei	< 5
schwach skeletthaltig	5 - 10
kieshaltig*	10 - 20
steinhaltig	10 - 20
stark kieshaltig*	20 - 30
stark steinhaltig	20 - 30
kiesreich*	30 - 50

# Legacy soil maps

## harvested soil properties

pH		carbonates		gravel %		topsoil clay %		subsoil clay %		topsoil silt %		subsoil silt %		soil depth	
from	to	topsoil	subsoil	from	to	from	to	from	to	from	to	from	to	from	to
.1	7	part.	yes	0	10	20	30	20	30	0	50	0	50	100	150
.1	7	part.	yes	10	20	20	30	20	30	0	50	0	50	100	150
.1	7	no	yes	0	10	10	20	10	30	0	100	0	100	100	150
.1	7	no	yes	10	20	10	20	10	30	0	100	0	100	100	150
.1	7	no	yes	0	5	5	20	10	20	0	50	0	50	100	150
.1	7	no	yes	0	5	40	50	40	60	0	50	0	50	100	150
.1	7	no	yes	10	20	40	60	40	60	0	50	0	50	100	150
7	8	yes	yes	0	10	20	30	20	30	0	50	0	50	100	150
7	8	yes	yes	5	20	50	60	50	60	40	50	0	50	100	150
.3	7	no	no	0	10	10	20	30	40	0	100	0	100	100	150
.3	7	no	no	10	20	10	20	30	40	0	100	0	100	100	150
.3	7	no	no	0	5	10	20	20	30	0	50	0	50	100	150
.1	7	no	yes	5	10	30	40	30	40	0	50	0	50	70	100
.1	7	no	yes	0	0	50	60	40	50	40	50	0	50	70	100
.1	7	no	yes	0	5	5	20	2	20	0	50	0	50	70	100
.1	7	no	no	0	10	0	20	20	30	0	100	0	50	70	100
.1	7	no	yes	0	10	20	40	20	40	0	50	0	50	70	100
.1	7	no	yes	0	20	20	30	20	30	0	50	0	50	70	100
.1	7	no	yes	5	10	20	30	20	30	0	50	0	50	70	100
.1	7	no	yes	0	5	10	20	10	25	0	50	0	50	70	100

## Legend booklet

### STAUWASSERGEPRÄGTE BOEDEN

selten bis zur Oberfläche porengesättigte Böden

oY1 BRAUNERDE-PSEUDOGLEY, labilaggregiert, verhärt skelettfrei bis schwach skeletthaltig, sandige lehmiger Schluff, mäßig tiefgrünig

oY2 BRAUNERDE PSEUDOGLEY, verhärtet, drainiert, skelettfrei bis schwach skeletthaltig. Lehm bi Lehm über tonigem Lehm, mäßig tiefgrünig

Stauwassergeprägte, häufig bis zur Oberfläche porengesättigte Böden

qII PSEUDOGLEY, labilaggregiert, verhärtet, drainiert, schwach skeletthaltig. Lehm bis lehmiger Schluff

## Tables with value ranges

Skelettklasse	Skelettgehalt Vol. %
skelettfrei	< 5
schwach skeletthaltig	5 - 10
Kieshaltig*	10 - 20
steinhaltig	10 - 20
stark kieshaltig*	20 - 30
stark steinhardtig	20 - 30
kiesreich*	30 - 50

# Legacy soil maps

## harvested soil properties

pH		carbonates		gravel %		topsoil clay %		subsoil clay %		topsoil silt %		subsoil silt %		soil depth	
from	to	topsoil	subsoil	from	to	from	to	from	to	from	to	from	to	from	to
.1	7	part.	yes	0	10	20	30	20	30	0	50	0	50	100	150
.1	7	part.	yes	10	20	20	30	20	30	0	50	0	50	100	150
.1	7	no	yes	0	10	10	20	10	30	0	100	0	100	100	150
.1	7	no	yes	10	20	10	20	10	30	0	100	0	100	100	150
.1	7	no	yes	0	5	5	20	10	20	0	50	0	50	100	150
.1	7	no	yes	0	5	40	50	40	60	0	50	0	50	100	150
.1	7	no	yes	10	20	40	60	40	60	0	50	0	50	100	150
7	8	yes	yes	0	10	20	30	20	30	0	50	0	50	100	150
7	8	yes	yes	5	20	50	60	50	60	40	50	0	50	100	150
.3	7	no	no	0	10	10	20	30	40	0	100	0	100	100	150
.3	7	no	no	10	20	10	20	30	40	0	100	0	100	100	150
.3	7	no	no	0	5	10	20	20	30	0	50	0	50	100	150
.1	7	no	yes	5	10	30	40	30	40	0	50	0	50	70	100
.1	7	no	yes	0	0	50	60	40	50	40	50	0	50	70	100
.1	7	no	yes	0	5	5	20	2	20	0	50	0	50	70	100
.1	7	no	no	0	10	0	20	20	30	0	100	0	50	70	100
.1	7	no	yes	0	10	20	40	20	40	0	50	0	50	70	100
.1	7	no	yes	0	20	20	30	20	30	0	50	0	50	70	100
.1	7	no	yes	5	10	20	30	20	30	0	50	0	50	70	100
.1	7	no	yes	0	5	10	20	10	25	0	50	0	50	70	100

## Legend booklet

### STAUWASSERGEPRÄGTE BOEDEN

selten bis zur Oberfläche porengesättigte Böden

oY1 BRAUNERDE-PSEUDOGLEY, labilaggregiert, verhärt skelettfrei bis schwach skeletthaltig, sandige lehmiger Schluff, mäßig tiefgrünig

oY2 BRAUNERDE-PSEUDOGLEY, verhärtet, drainiert, skelettfrei bis schwach skeletthaltig, Lehm über tonigem Lehm, mäßig tiefgrünig

Stauwassergeprägte, häufig bis zur Oberfläche porengesättigte Böden

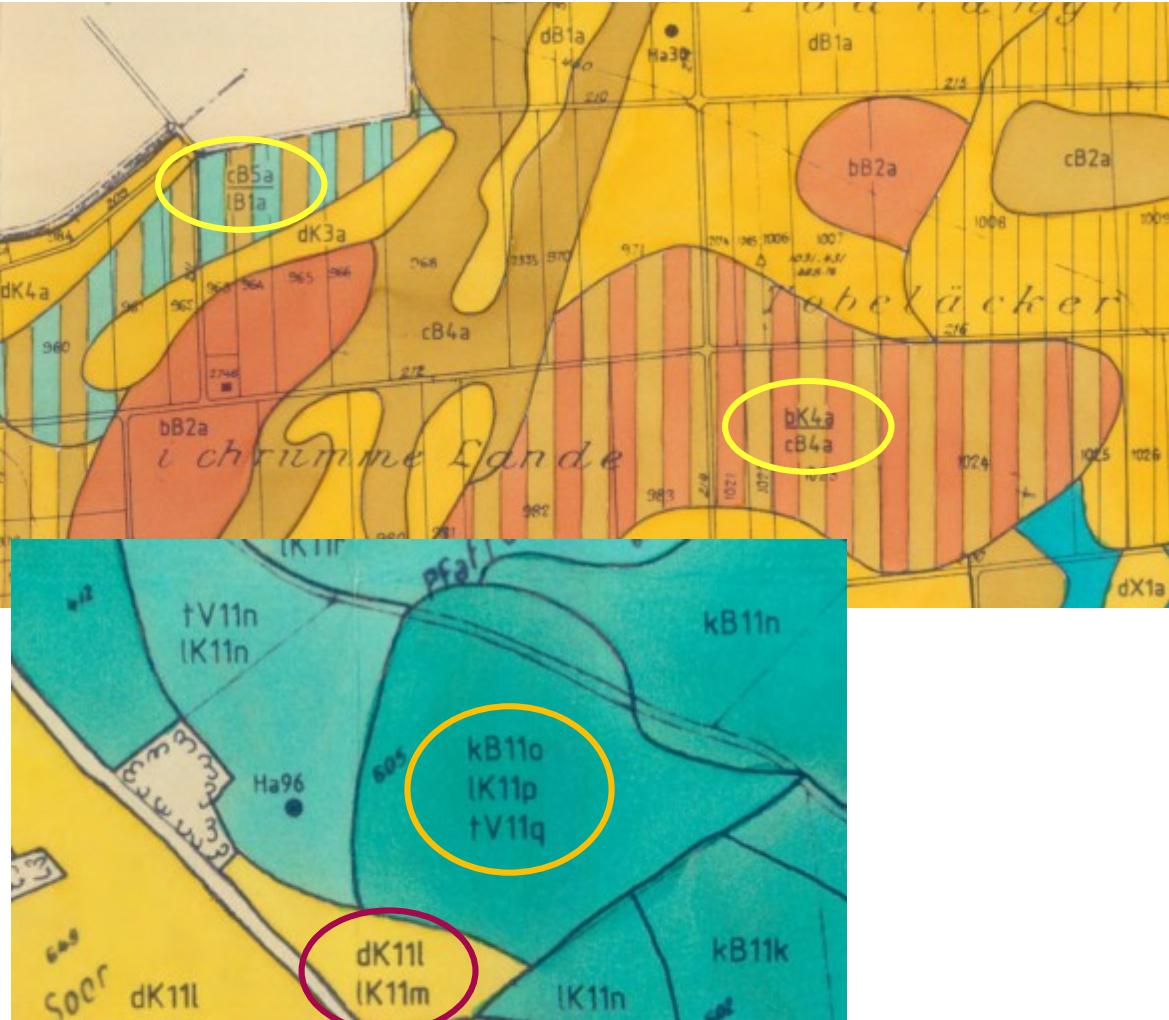
qII PSEUDOGLEY, labilaggregiert, verhärtet, drainiert, schwach skeletthaltig, Lehm bis lehmiger Schluff

## Tables with value ranges

Skelettklasse	Skelettgehalt Vol. %
skelettfrei	< 5
schwach skeletthaltig	5 - 10
Kieshaltig*	10 - 20
steinhaltig	10 - 20
stark kieshaltig*	20 - 30
stark steinhardtig	20 - 30
kiesreich*	30 - 50



# Complex soil map polygons



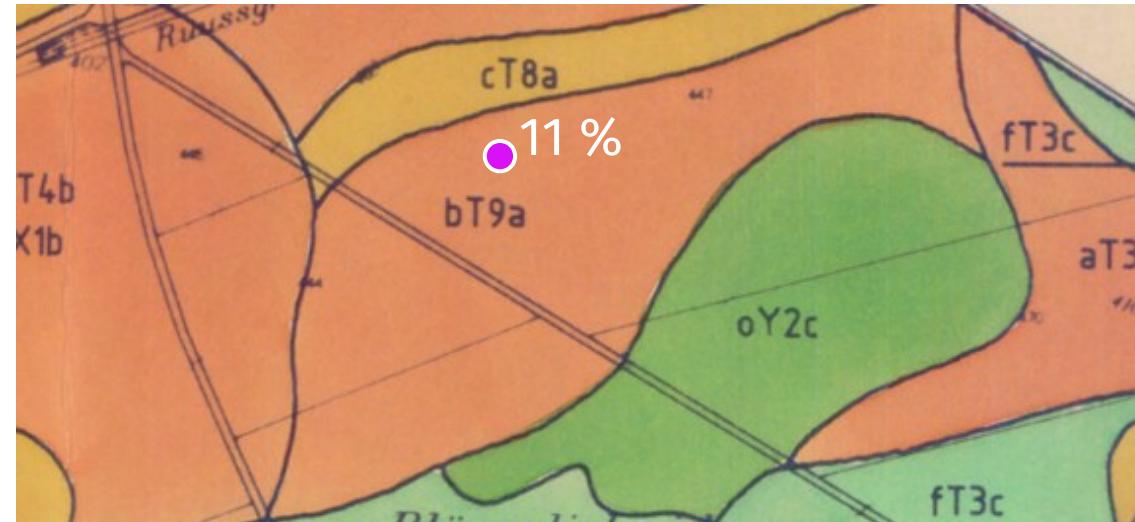
- small scale variability
- 2-3 units in one map polygon
- Ratio according to colouring  
→ could not be bothered to measure..
- Used default ratios from other soil map:  
2 units: 60 - 40 %  
3 units: 50 - 30 - 20 %

# Create “virtual samples” as real samples have not been recorded

Create response vector by

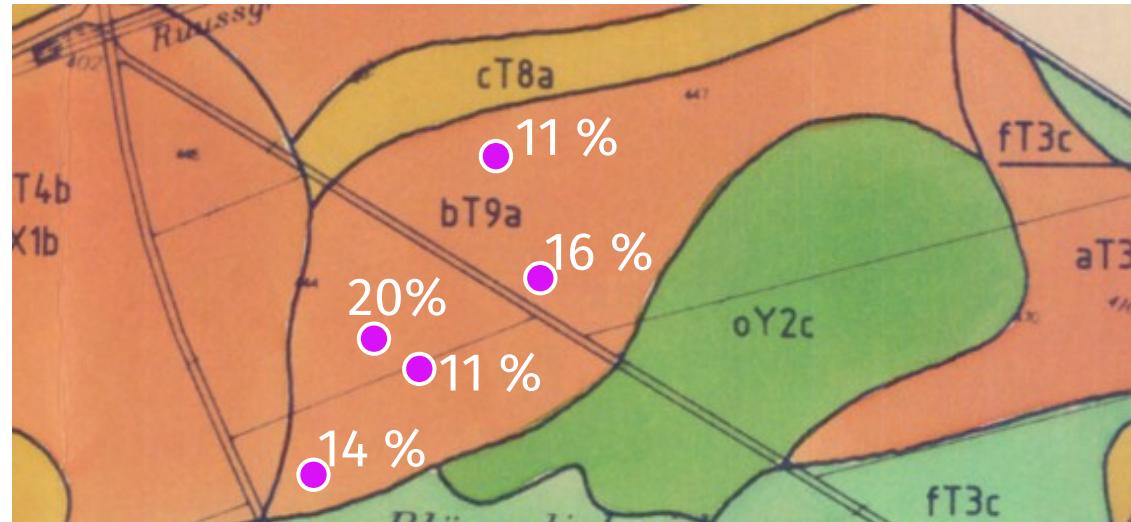
1. Select a point sample with 4 sampling strategies
  2. For complex polygons (>1 map units): sample unit code by weights
  3. Assign value out of range in legend (uniform distribution)
- Repeated 30 times for each sampling strategy.

Following DSMART-Algorithm  
(Odgers et al. 2014).



bericht	code	typ	topsoil clay %		subsoil clay %	
			from	to	from	to
Klettgau I	aB1	Braunerde	20	30	20	30
Klettgau I	aB2	Braunerde	20	30	20	30
Klettgau II	bT8	Parabraunerde	30	40	40	50
Klettgau II	bT9	Parabraunerde	10	20	20	30
Klettgau I	bX1	Aufschüttung	10	30	10	30

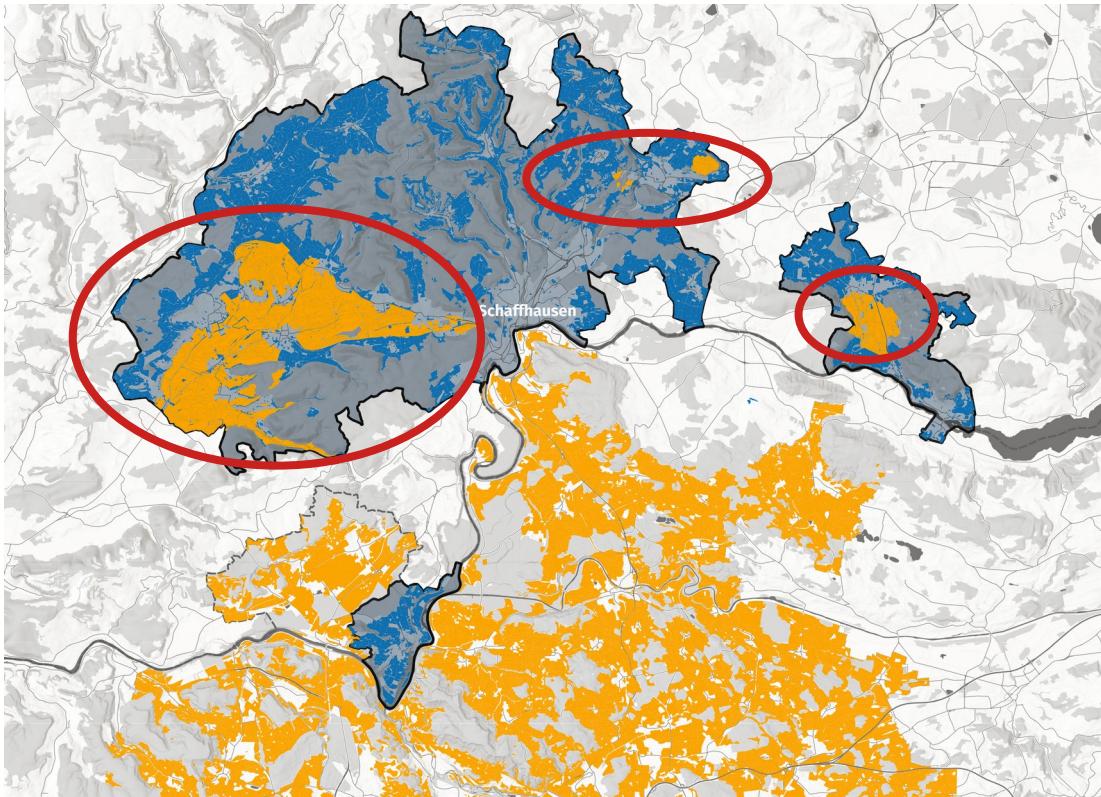
# Create “virtual samples” as real samples have not been recorded



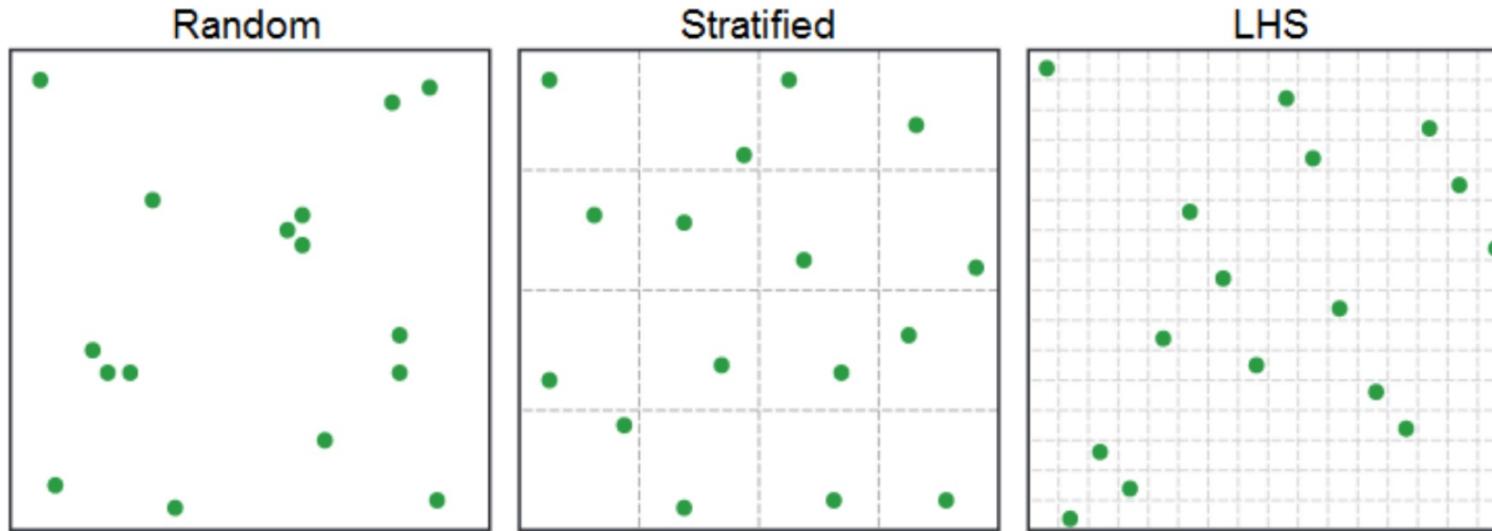
bericht	code	typ	topsoil clay %		subsoil clay %	
			from	to	from	to
Klettgau I	aB1	Braunerde	20	30	20	30
Klettgau I	aB2	Braunerde	20	30	20	30
Klettgau II	bT8	Parabraunerde	30	40	40	50
Klettgau II	bT9	Parabraunerde	10	20	20	30
Klettgau I	bX1	Aufschüttung	10	30	10	30

# Different sampling designs

- Simple random sample
  - soil maps Zurich and Schaffhausen
  - only soil maps Schaffhausen (red circles)
- Stratified random sample
  - Strata: geological units
  - Weights: area per unit in agricultural land to be mapped
- Conditional Latin Hypercube Sampling
- number of virtual samples:  
1 000 / 3 000 / 6 000 / 10 000 / 35 000



# Different sampling designs



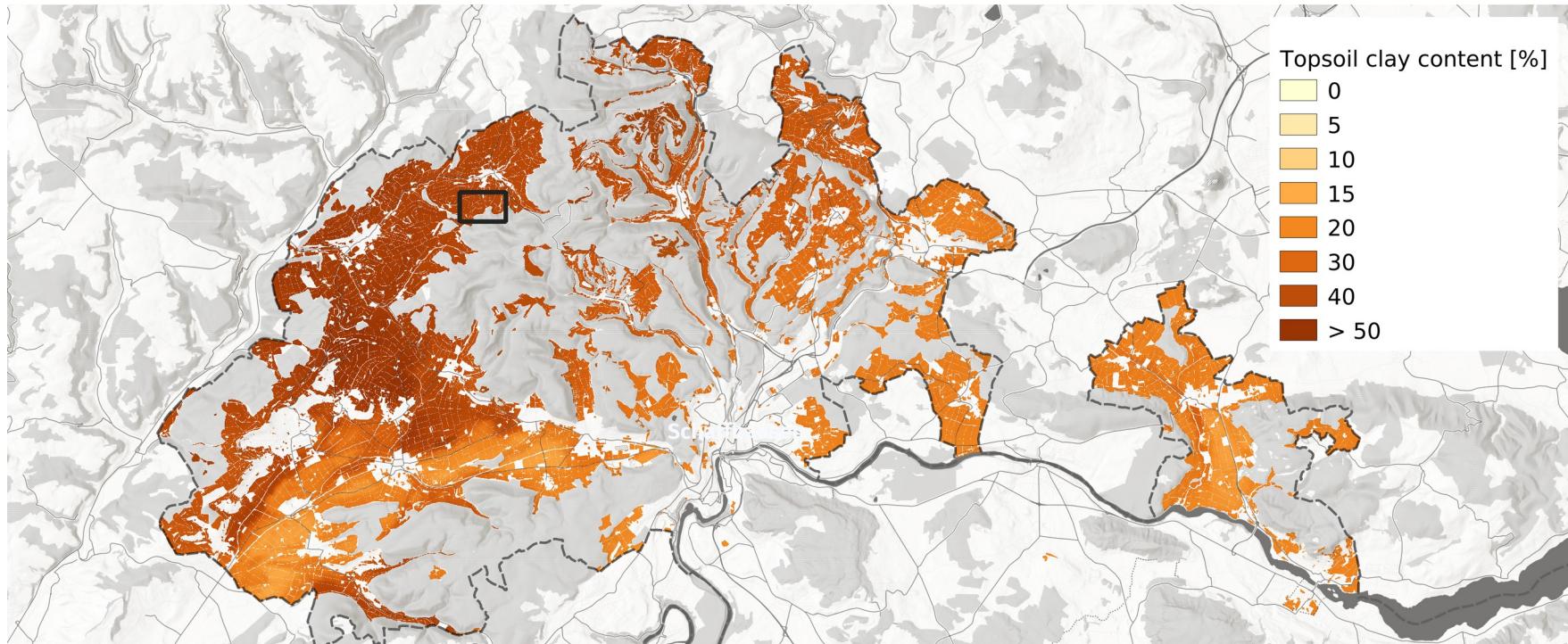
## LHS: conditioned latin hypercube sampling

- Number of strata = number of samples to be taken
- Hypercube: stratification on many covariates possible
- Conditioned: only combinations that actually exist are sampled
- Every combination only once → like a Sudoku

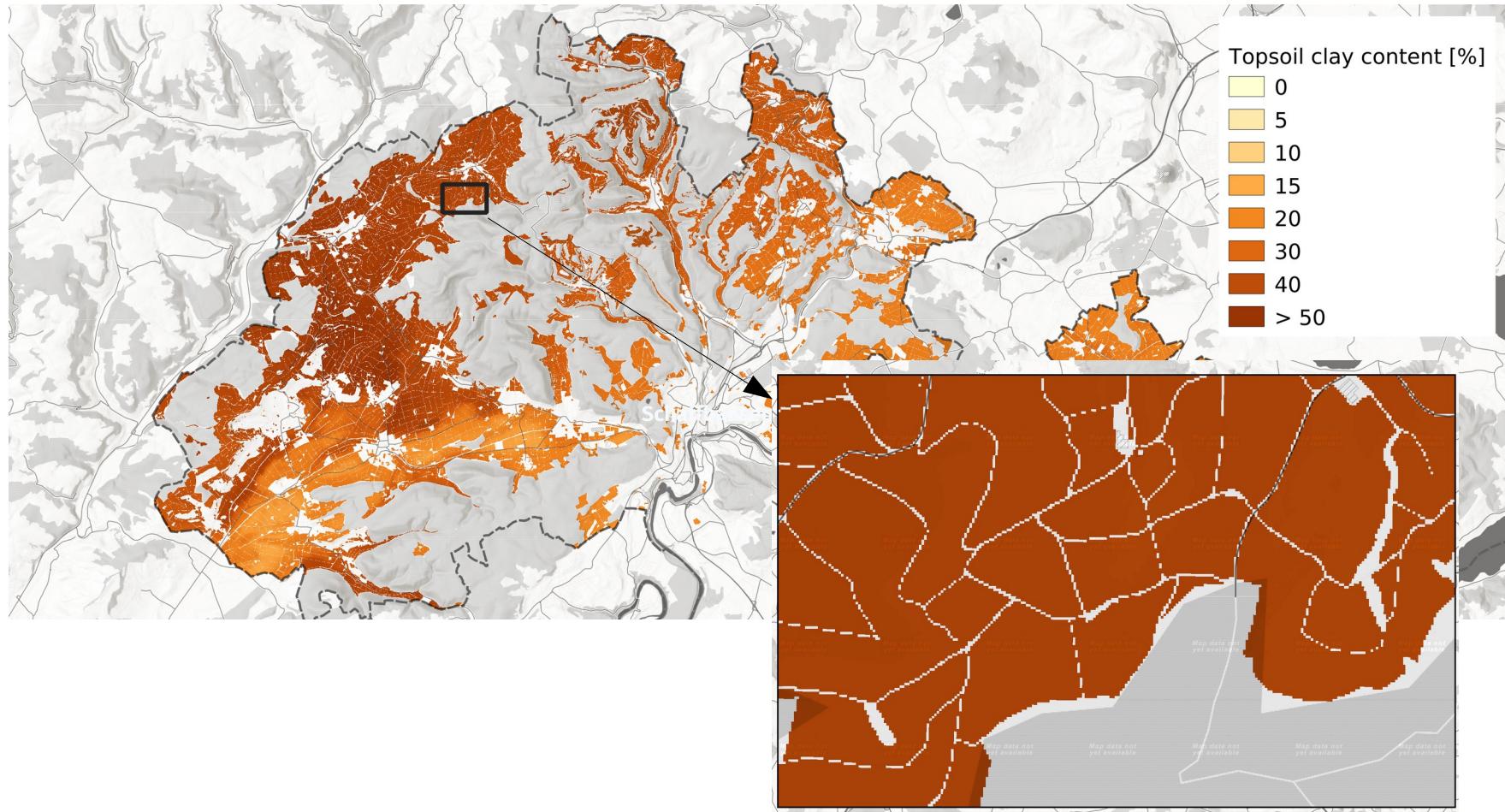
# Modelling

- 30 repetitions of sampling procedure for each sampling strategy
- Separate models for each soil property and each repetition
- Model:  
Random Forest (package ranger)  
simple fit without tuning (first go).
- Covariates:  
Geological units, simple and complex terrain attributes,  
vertical and horizontal distances to rivers

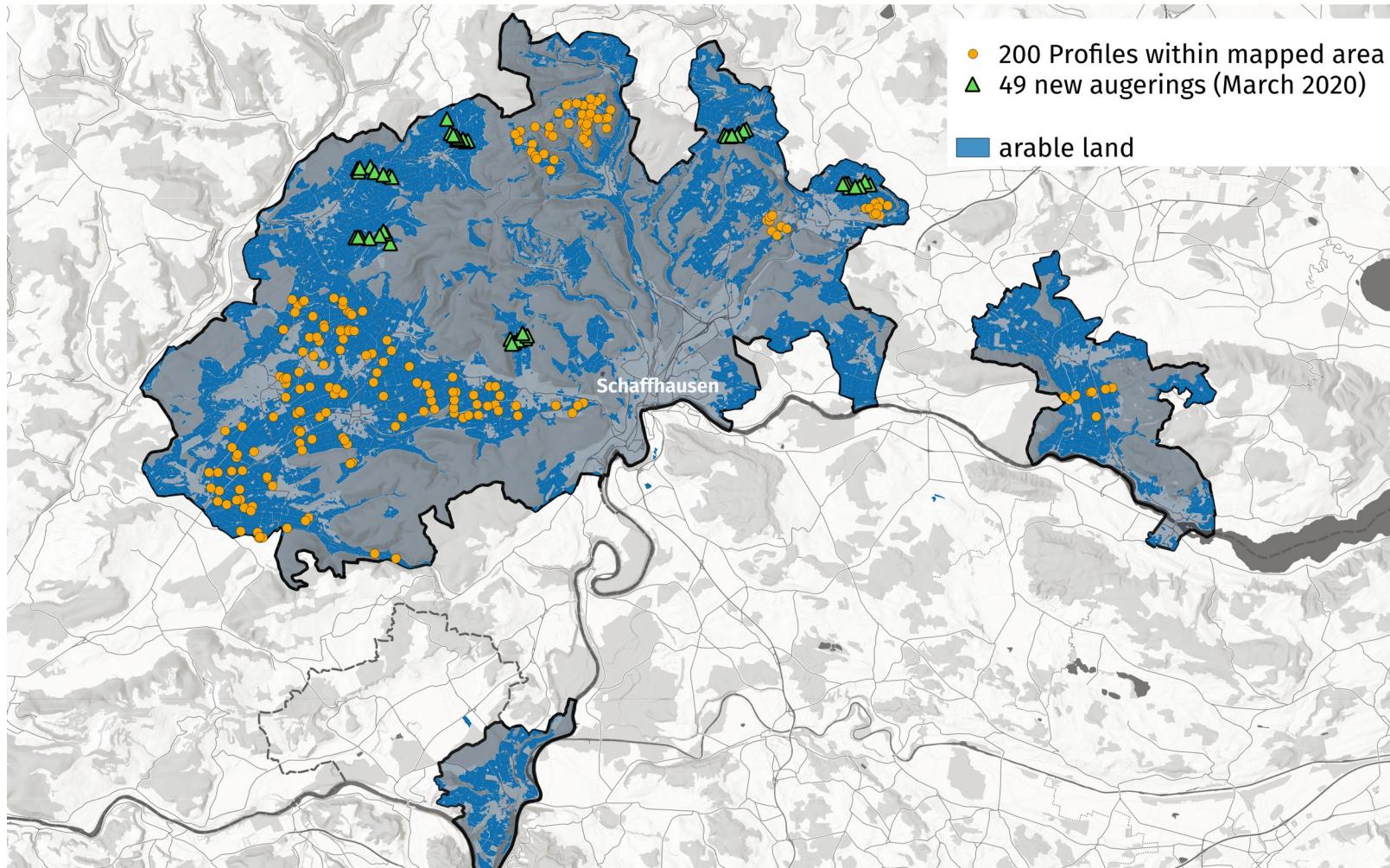
# Resulting map



# Resulting map

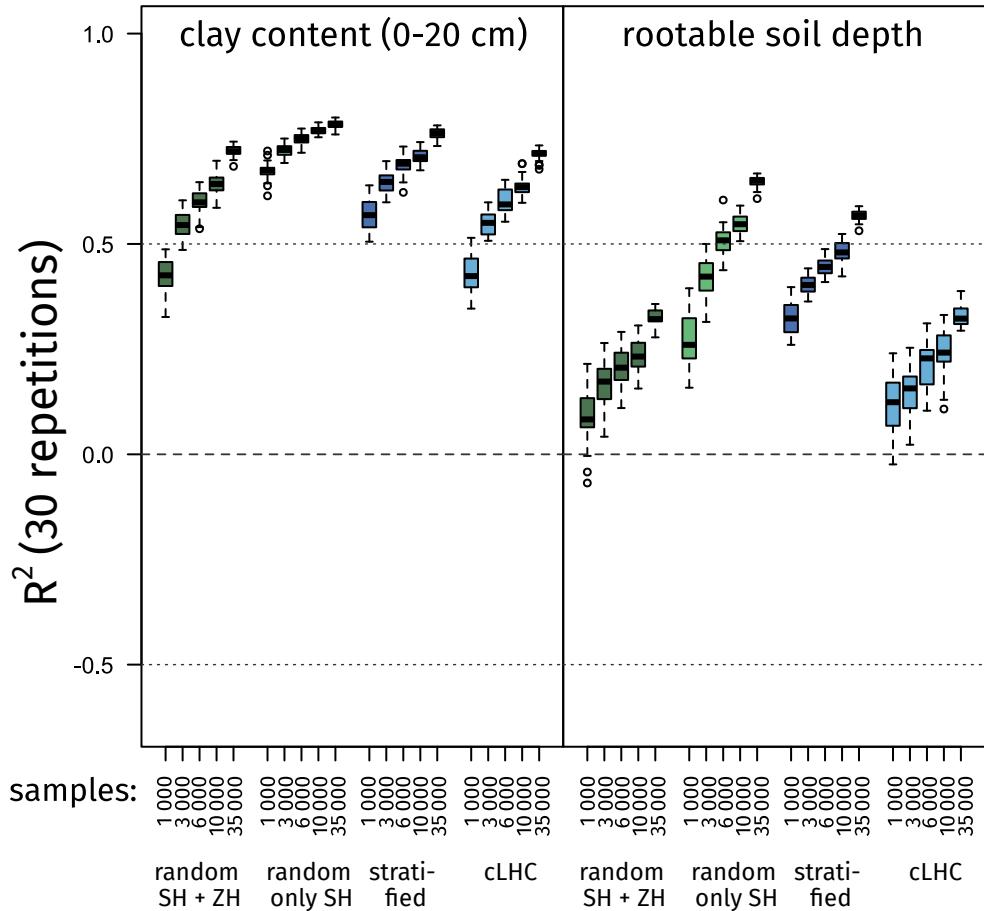


# Results – Validation Samples



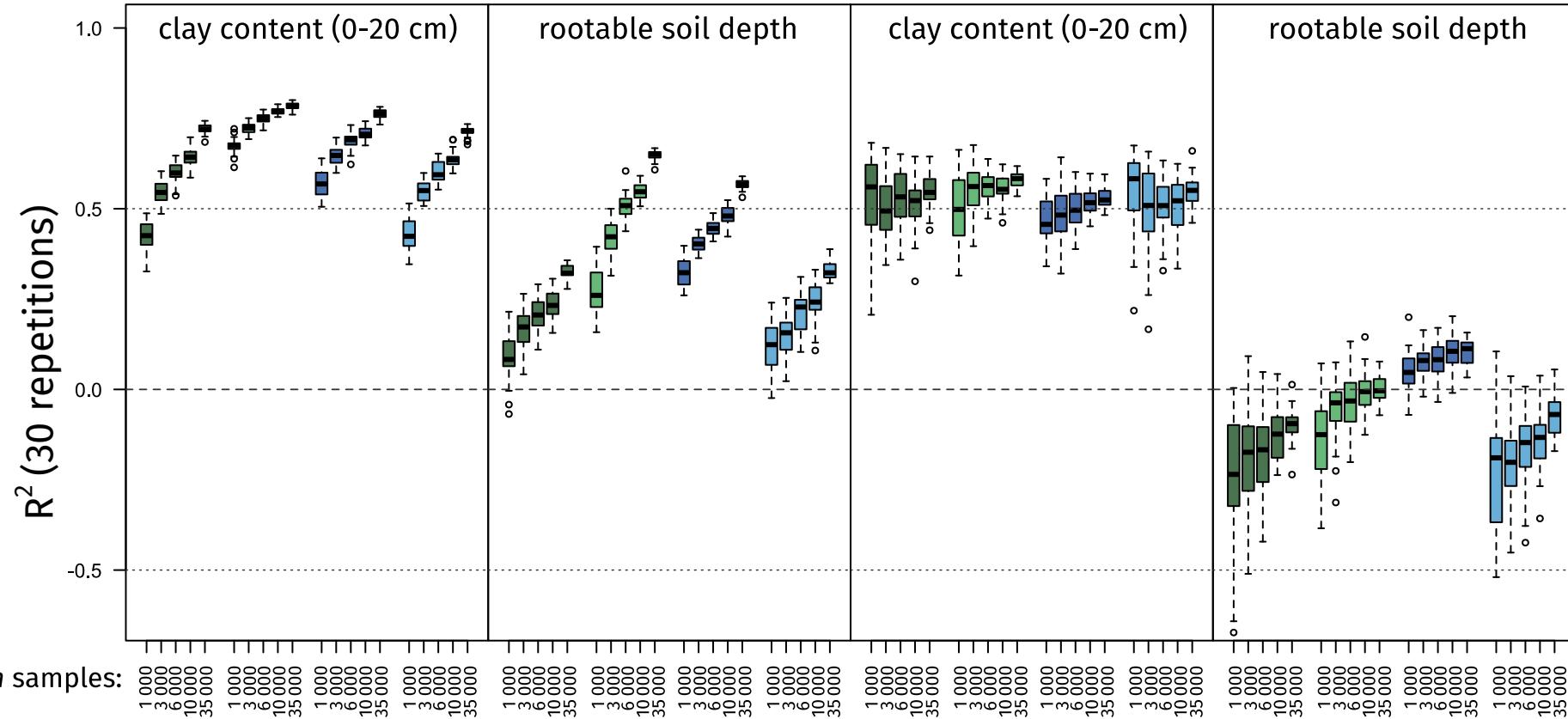
# Validation – $R^2$

200 soil profiles within legacy soil maps



# Validation – R<sup>2</sup>

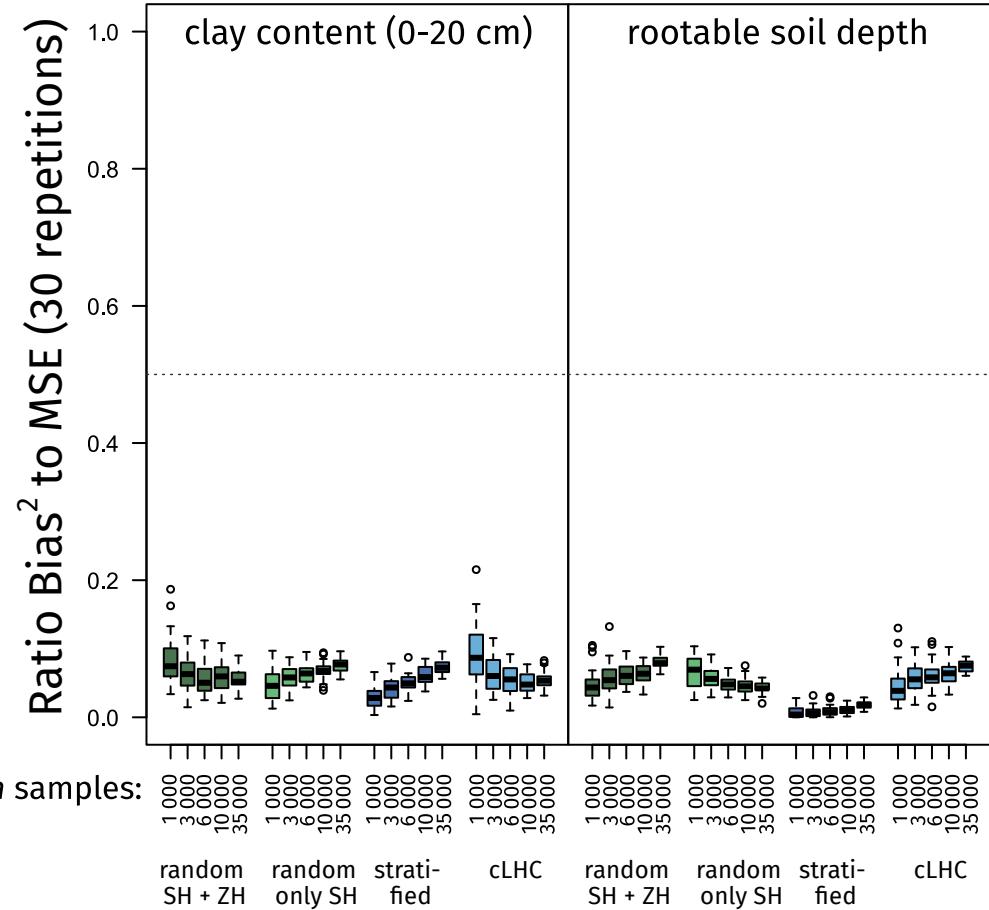
## 200 soil profiles within legacy soil maps



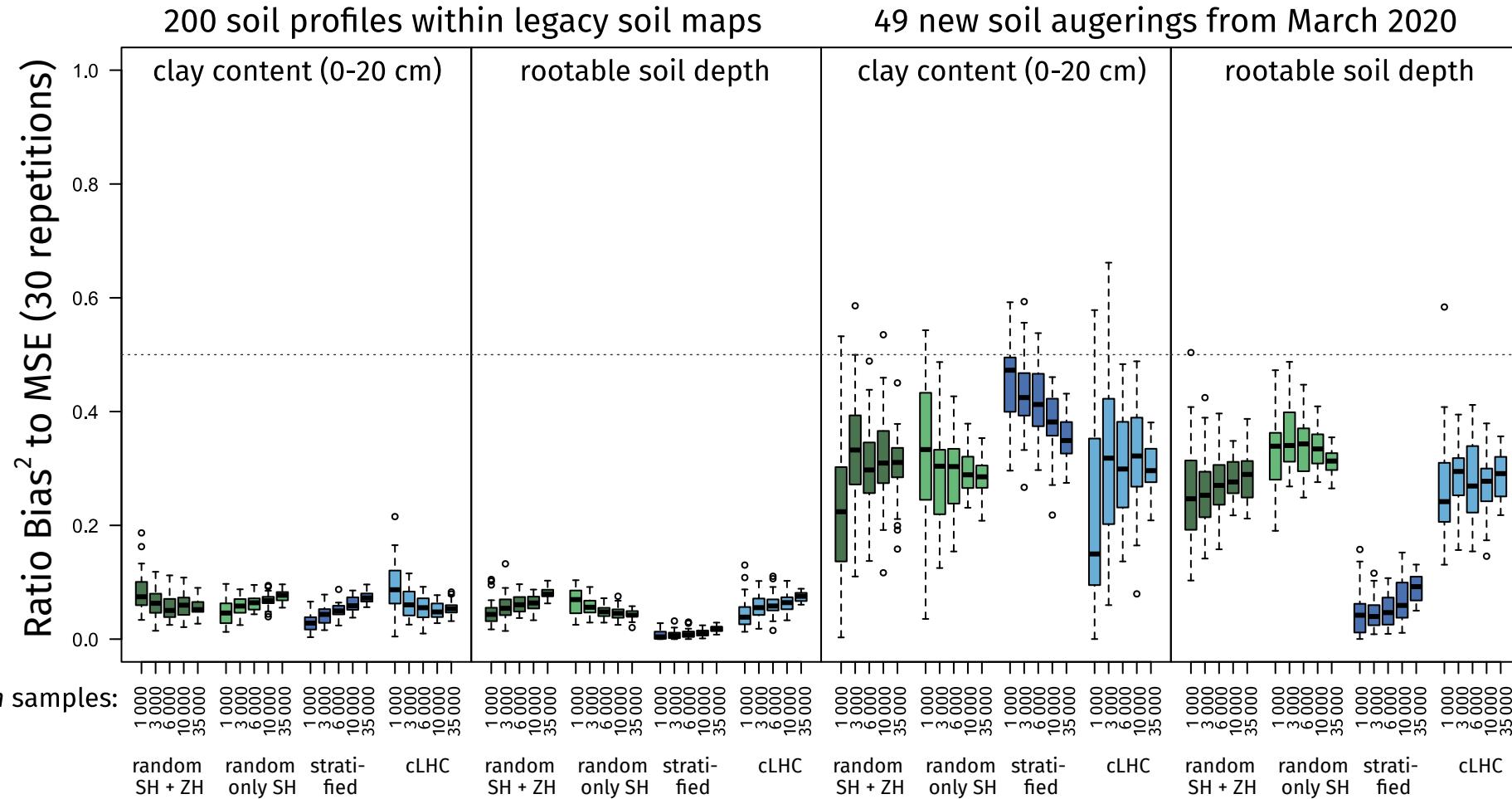
## 49 new soil augerings from March 2020

# Validation – Bias

200 soil profiles within legacy soil maps



# Validation – Bias



# Conclusion and Outlook

## Models based on legacy soil maps

### Extrapolation into new areas

- possible, but with care!
- large dependence on target soil attribute
- no decision on optimal sampling design

### Further work:

- Validation 300 new soil augerings and 10 soil profile pits until end of 2020
- Integration of interpreted crop map and property tax values (1950s, based on soil sampling)

# Many thanks to ...

Stéphane Burgos – Stefan Oechslin – Liv Kellermann – Christian Sprecher – Adrian Hochreutener

## Funding: Canton of Schaffhausen

