Modelling of Soil Air Permeability – Towards a Spatial Continuous Map

Eric Petermann & Peter Bossew

German Federal Office for Radiation Protection, Berlin
Section „Radon and NORM“

GARRM Workshop – Prague, 18 September 2018

Bundesamt für Strahlenschutz

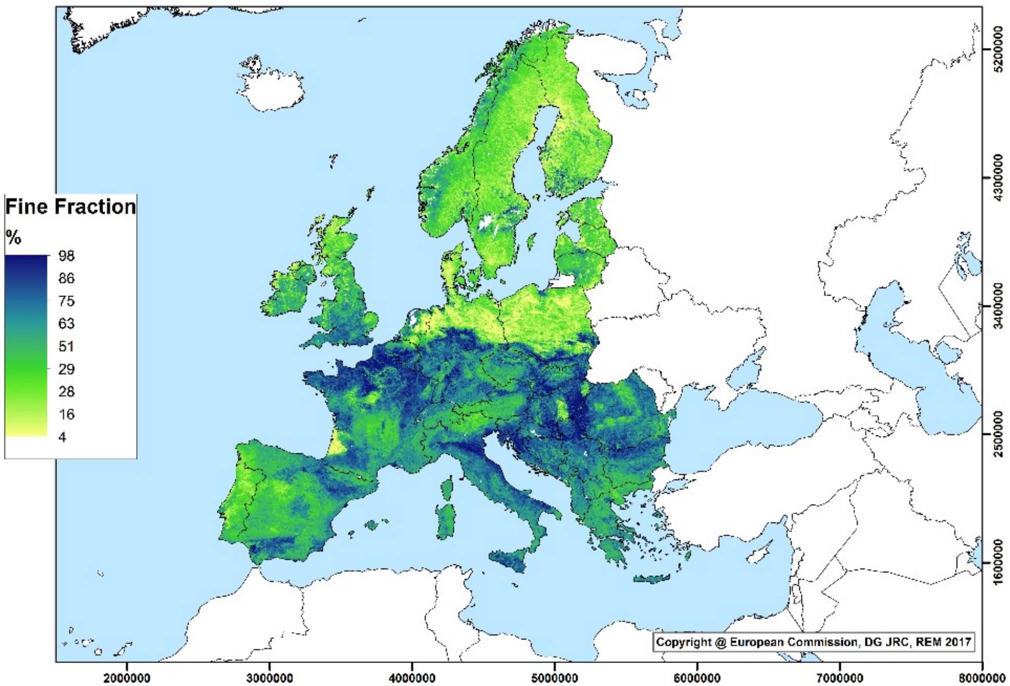
Agenda

1. Rationale
2. Soil air permeability driving forces
3. Data
4. Physical model
5. Machine Learning
6. Conclusion

Motivation

- Soil air permeability is a key driver of Rn entry into houses
- No permeability map available, only proxies
- Challenge: small-scale spatial and temporal variability
  → high uncertainty
  → geostatistical analysis is difficult

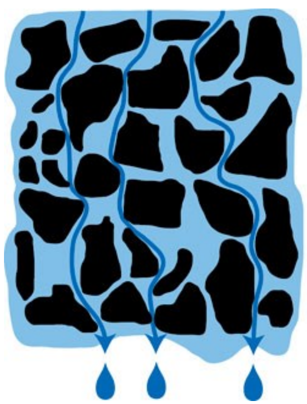
European map of permeability, April 2017

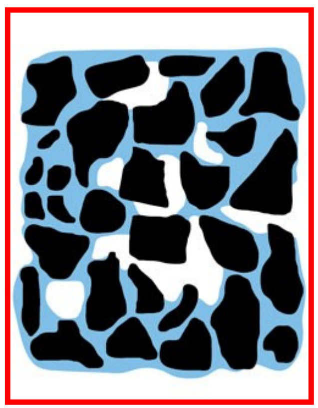


Cinelli et al. 2018

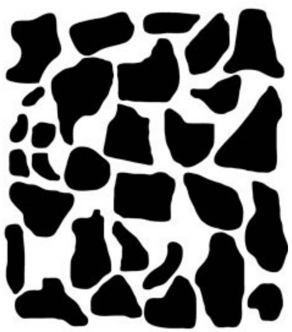
Soil Air Permeability: Processes and driving forces

Water saturated





Air saturated



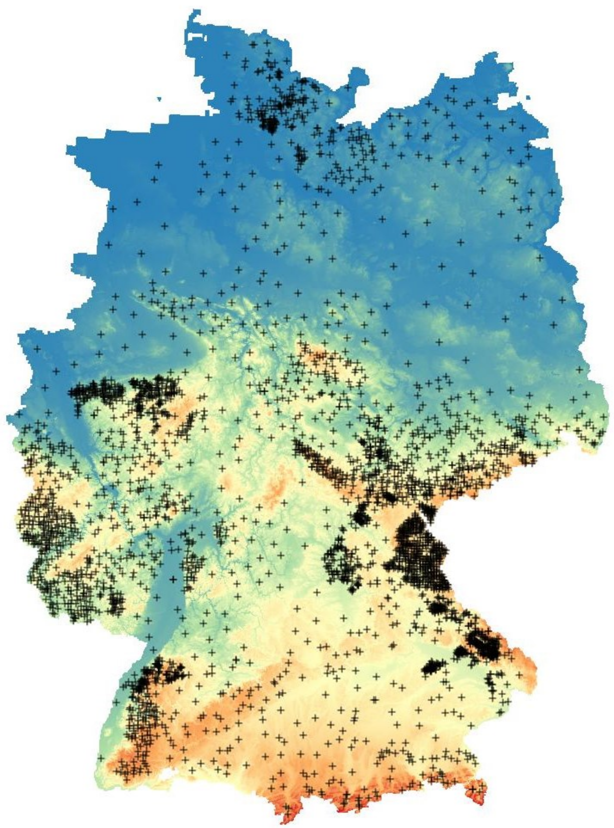
- Connection of pore spaces

→ Flow paths

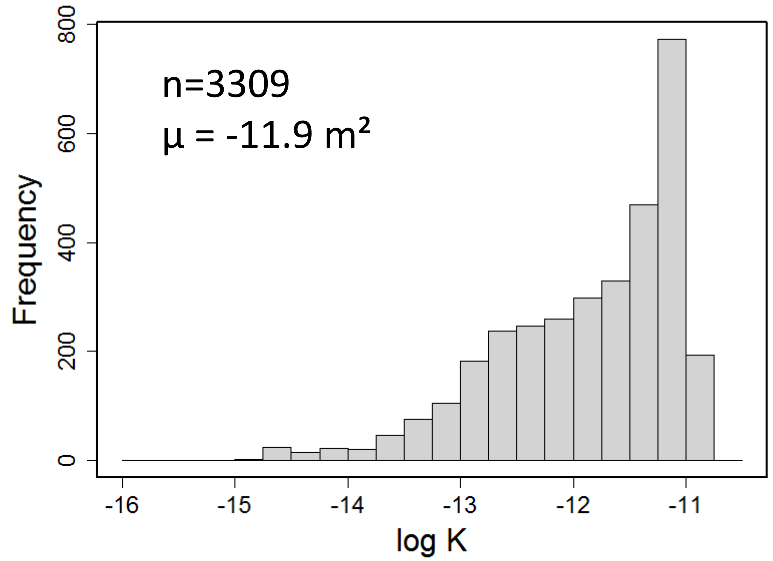
- "Sealing" for high soil moisture

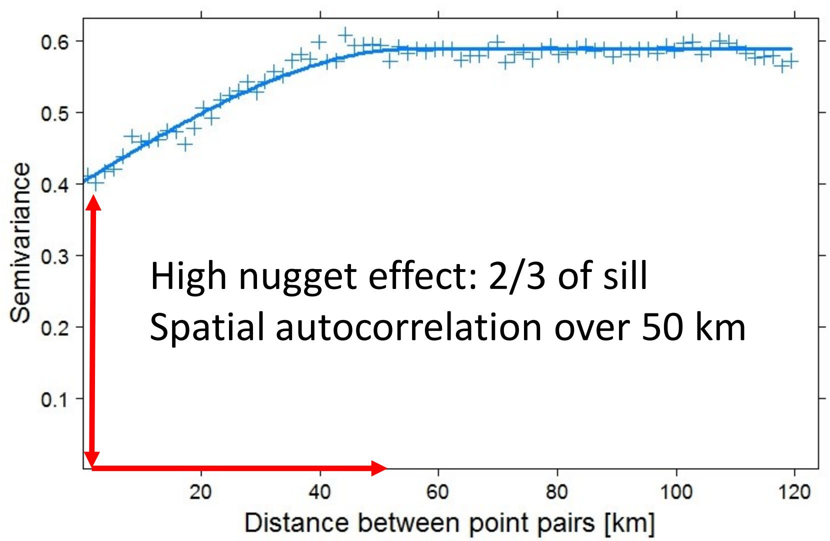
Soil Air Permeability: data in germany

Location of soil gas measurements



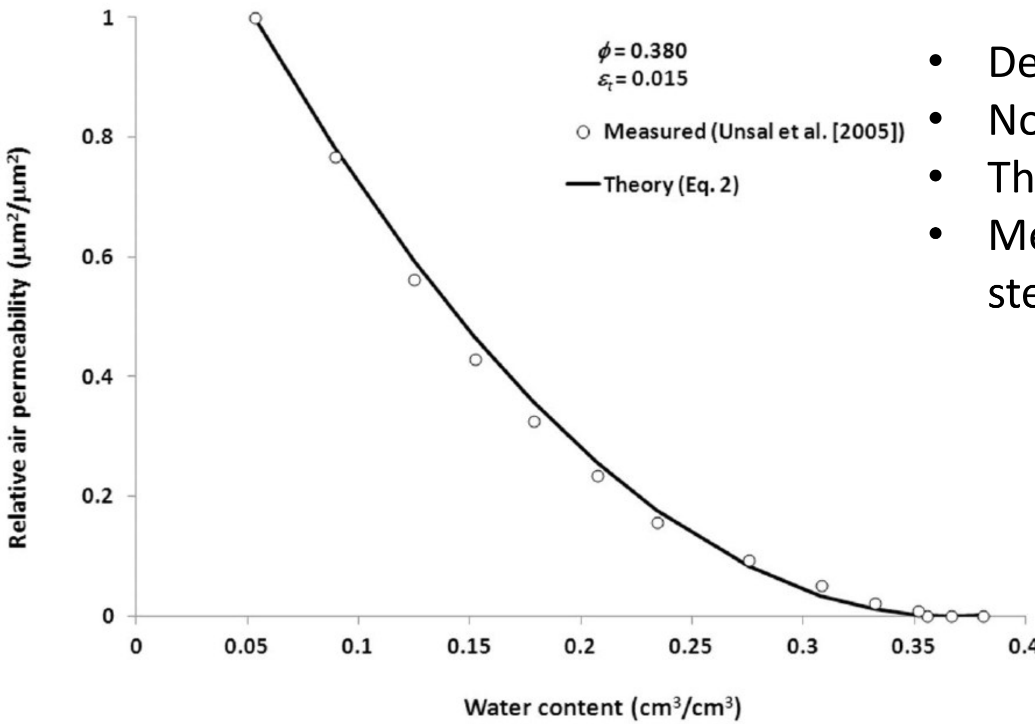
Underlying data: elevation model





Physical Model

Ghanbarian-Avijeh & Hunt (2012); Water Resources Research



- Dependence on water/air saturation
- Non-linearity
- Threshold values
- Measurements do not reflect a steady-state

$$k_a(\varepsilon) = k_a(\varepsilon = \phi)\left(\frac{\varepsilon - \varepsilon_t}{\phi - \varepsilon_t}\right)^2$$

Relative permeability

ε - air content

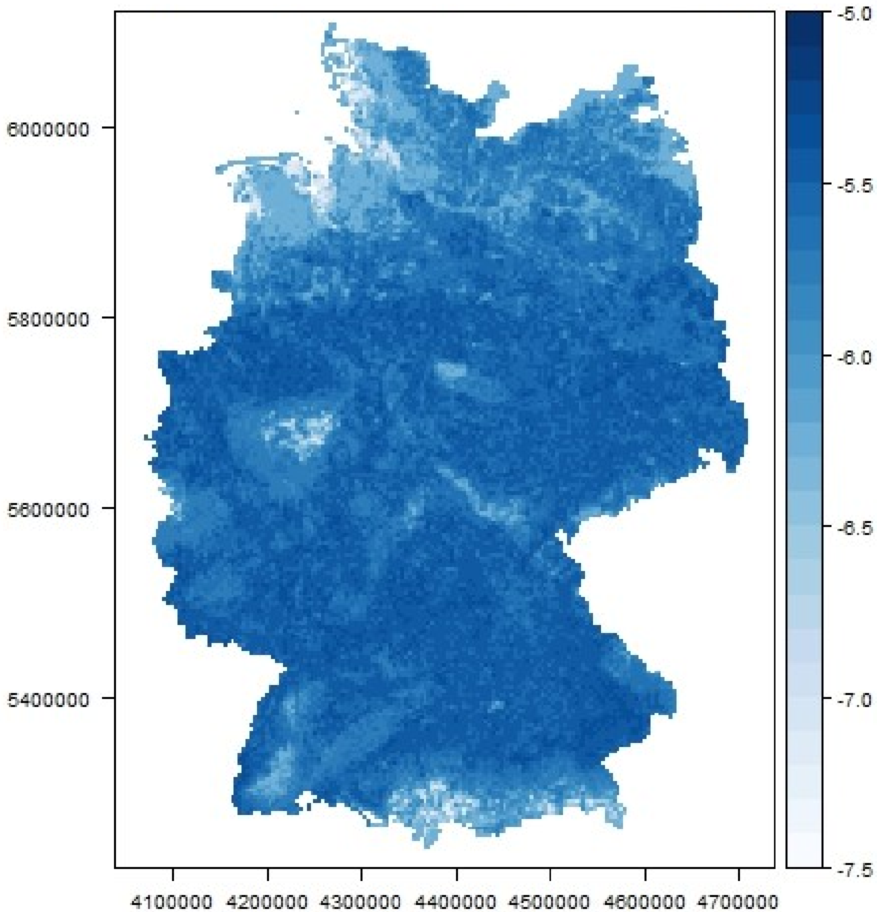
ε_t - minimum air content for percolation

ø - porosity

Maximum permeability

Physical Model – relation to hydr. conductivity

Saturated hydraulic conductivity 2m soil [m/s]

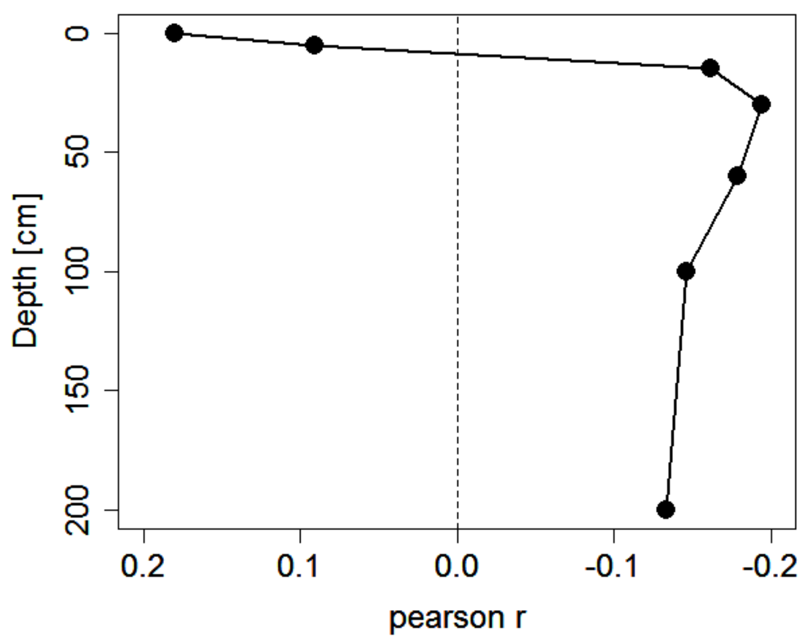


Data from: Toth et al. 2017, Hydrol. Process.

Sat. permeability ~ Sat. conductivity / (7.5 * 10^6)

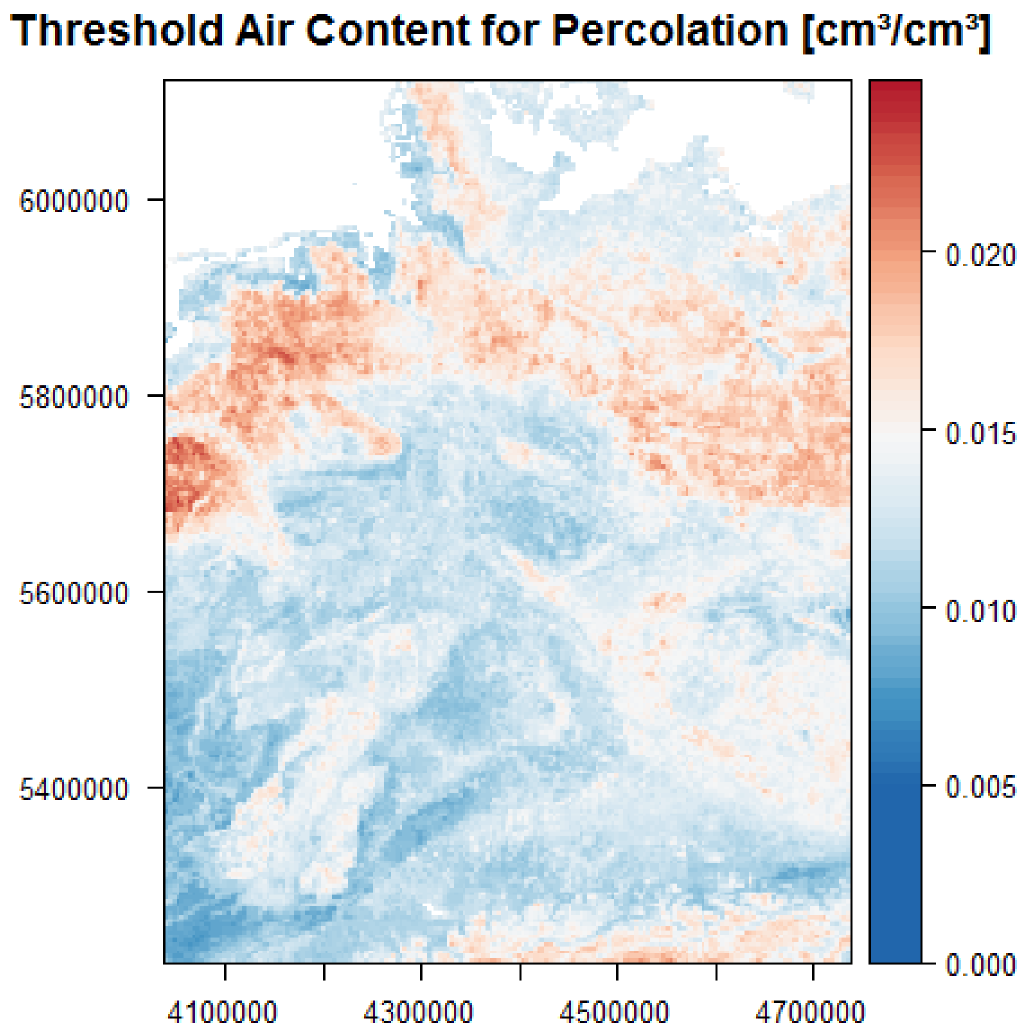
$$k_a(\varepsilon) = k_a(\varepsilon = \phi)\left(\frac{\varepsilon - \varepsilon_t}{\phi - \varepsilon_t}\right)^2$$

Observed K vs. hydraulic conducitivity



- Poor correlation
- Only positive for slices 1 +2
- Slices 1+2 (top soil) as matching points

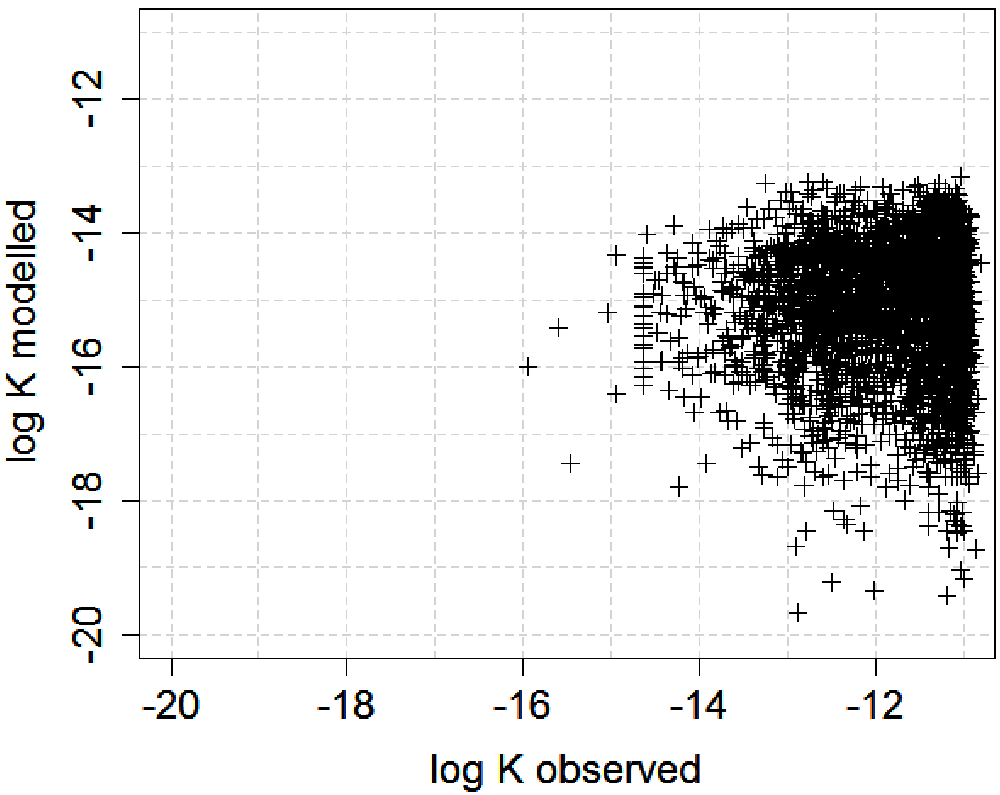
Physical Model – relation to hydr. conductivity

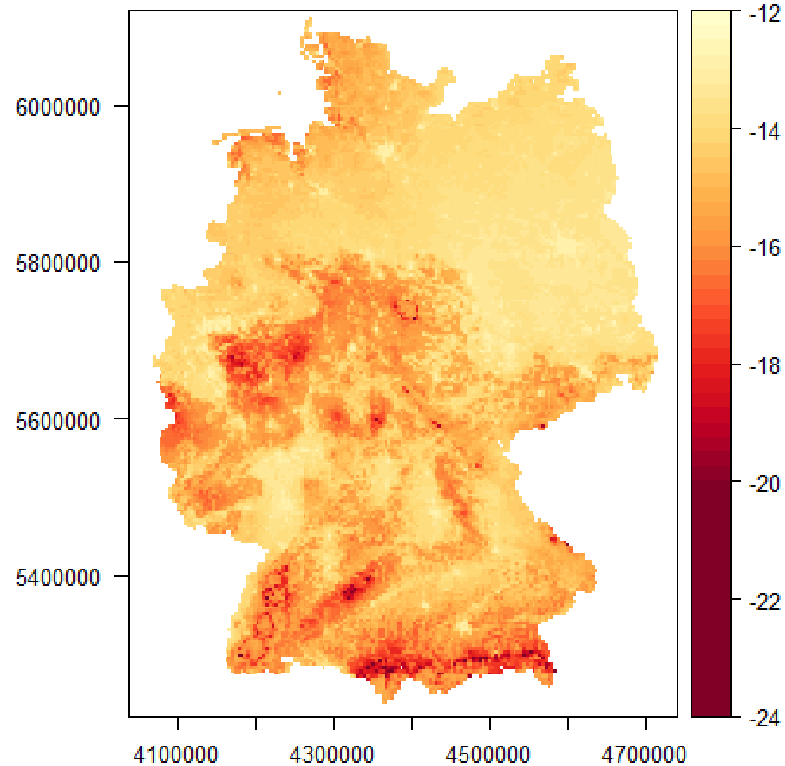


$$k_a(\varepsilon) = k_a(\varepsilon = \phi)\left(\frac{\varepsilon - \varepsilon_t}{\phi - \varepsilon_t}\right)^2$$

Based on data from Toth et al. (2017)*, Hydrol. Process., method adapted from Ghanabarian-Avijeh & Hunt (2013), WRR*

Physical Model: Results

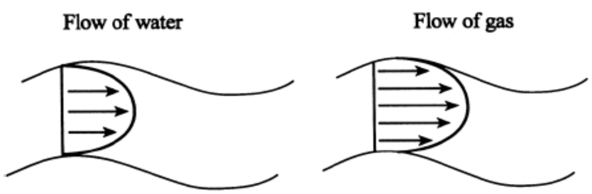




Assumption: $K_a \sim K_w$

Physical Model: Discussion

Klinkenberg effect



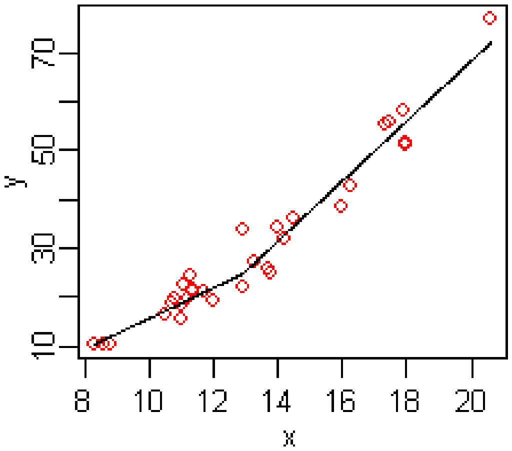
Model evaluation:

- Non-existent correlation
- Systematic underestimation

Possible reasons:

- Proportionality K_a ~ K_w ? K_w as matching point not suitable
  → non-linearity
  → at low saturation $K_a > K_w$
- Temporal variability of K_a due to soil moisture fluctuations
  -> soil moisture during sampling correction?
- Effect of macro-pores is not considered

Machine Learning:
Multivariate **A**daptive **R**egression **S**plines (MARS)



General:
- Well suited for high dimensional problems
- Allows for continuous and categorical input
- Creates pairs of linear basis functions
- Data-driven algorithm → estimates function

Model Building:

(1) Forward model building: adding basis functions aiming at reducing the training error (residual sum of squares - RSS)

→ including many model terms/usually over-fitted

(2) Backward pass: removes model terms aiming at improving generalizability

→ Tradeoff between model complexity and generalizability

→ Reducing “Generalized cross-validation” criterion f(RSS, model terms)

Machine Learning: Predictors

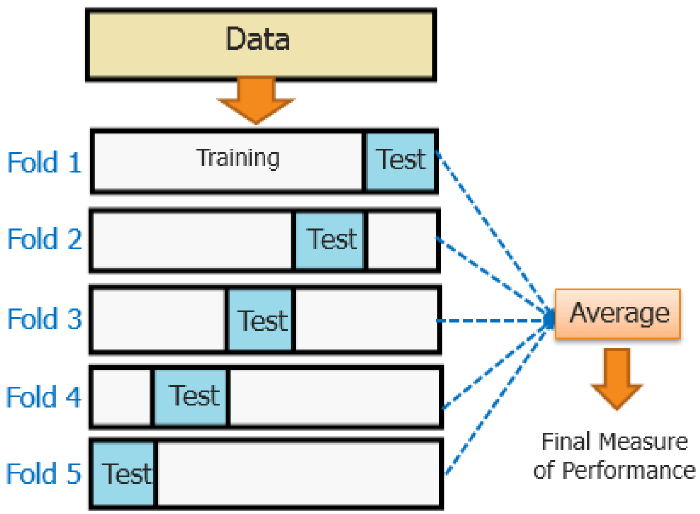
- Geology and soil attributes – BGR Fed. Inst. for Geosc. and Natural Resources
  - Categorical data (mostly 1:1.000.000)
  - e.g.: Petrography, stratigraphy, generation, C_{org} content, soil types, hydrogeological units
- Soil hydraulic properties - JRC *Toth et al. 2017;* *rol. Proc.*
  - Resolution 1 km / 250 m
  - 7 depths (0, 5, 15, 30, 60, 100, 200 cm)
  - e.g. saturated hydr. conductivity, saturated water parameters
- Topsoil physical properties LUCAS – JRC *Ballabio et a*
  - 500 m resolution
  - Grain size distribution, available water capacity, bulk density
- European Atlas of Natural Radiation – JRC *Cinelli et al. 2018, J. Environ. Rad.*
  - 10 km resolution
  - U, Th, K_20
- Soil moisture data – UFZ *Zink et al. 2017, Hydrol. Earth Sys. Sci.*
  - Modelling of daily soil moisture in Germany from 1951-2010
  - 4 km resolution
  - Cell-specific percentiles (dry 10%il, average 50%il, wet 90%il)

Total: 62 predictors
(so far)

Machine Learning: Model Building

- R-package *Earth*
- Best single predictors
- All numerical predictors
- Best numerical + Geology

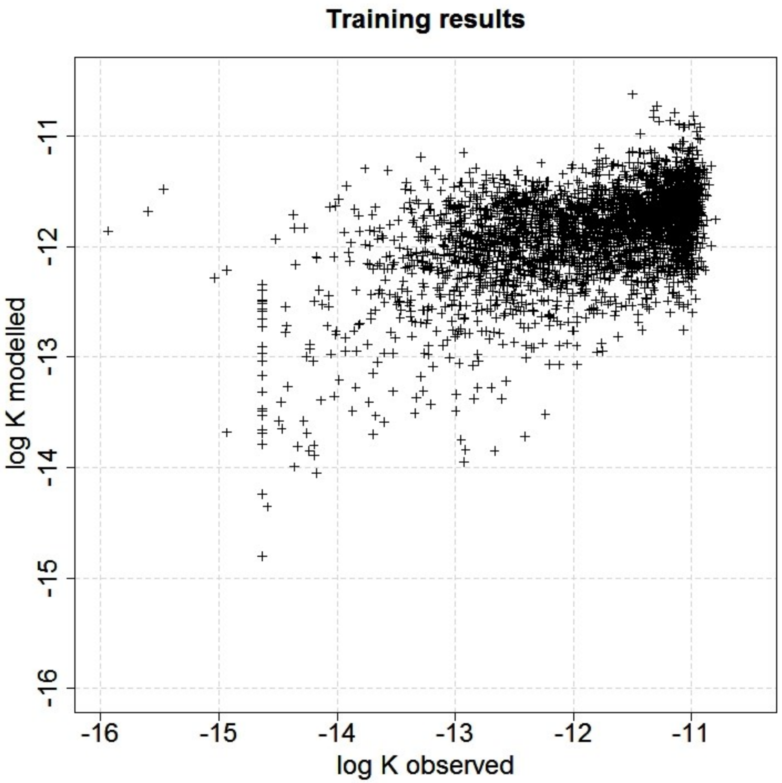
Repeated 5-fold cross-validation

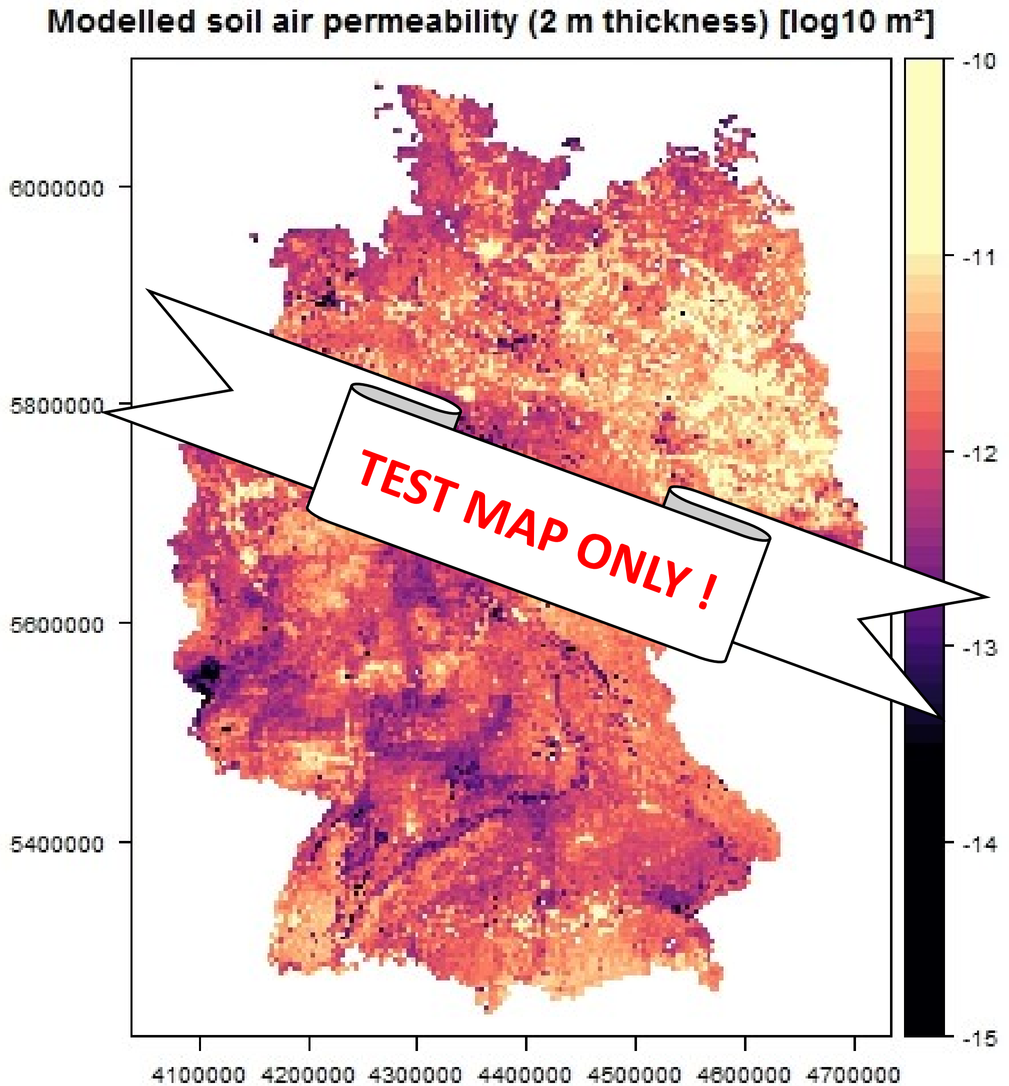


https://blog.contactsunny.com/data-science/different-types-of-validations-in-machine-learning-cross-validation

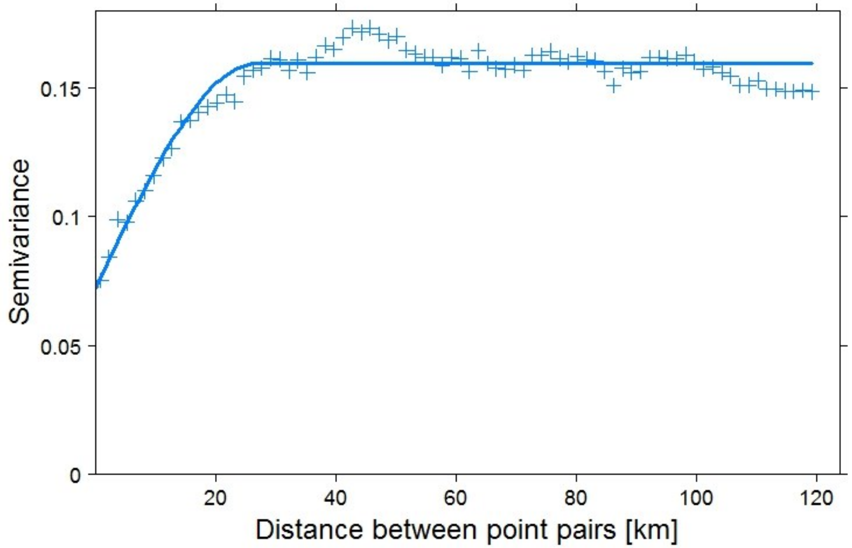
Predictor	Training R^2	Cross-validated R^2
Petrography	0.145	0.088
Geology	0.156	0.084
Stratigraphy	0.140	0.084
Hydrogeol. region – sub-section	0.154	0.084
Major soil landscape	0.105	0.083
Uranium content	0.066	0.059
Hydrogeol. region – section	0.078	0.056
Genesis	0.068	0.042
Soil Type	0.072	0.040
Hydr. conductivity 15 cm	0.051	0.040
Sat. water content 10 cm	0.046	0.039

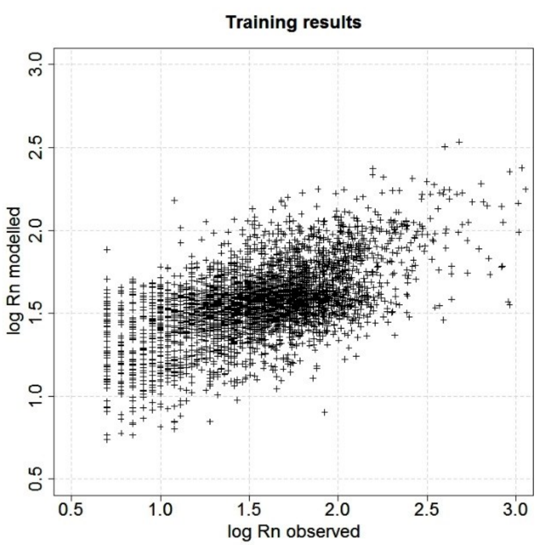
Machine Learning: K map v0.1

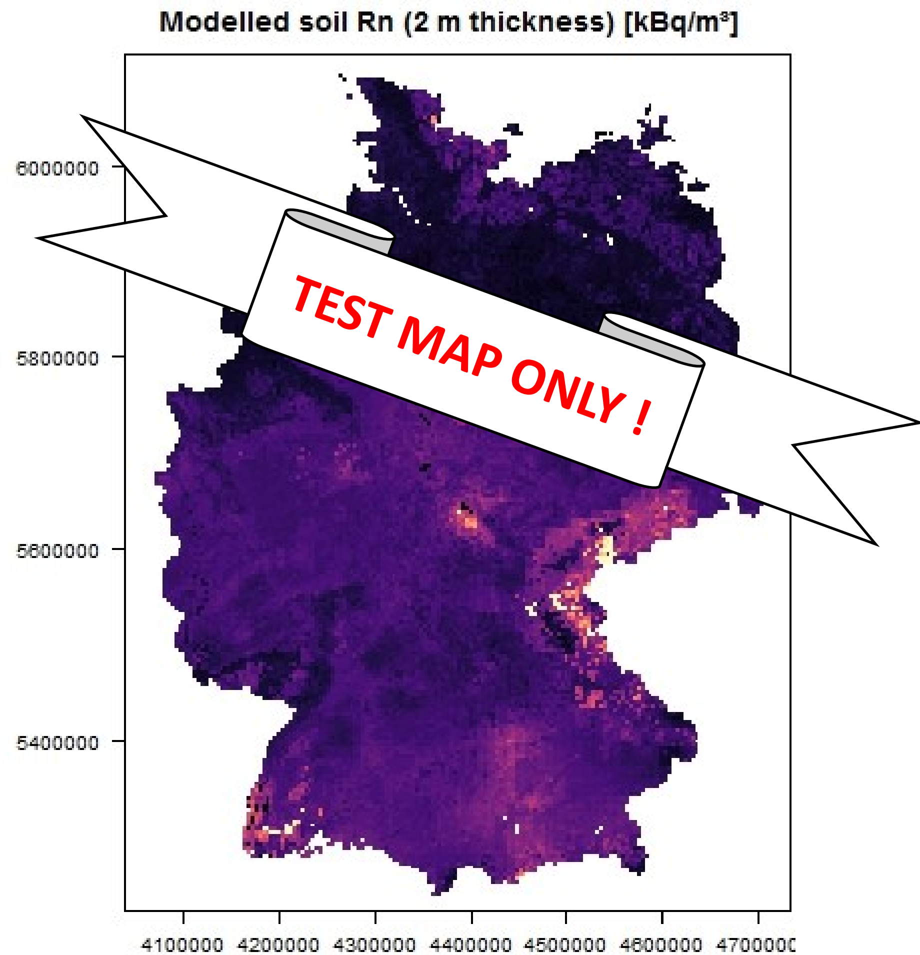




Machine Learning: Exemplary test maps Rn in soil v0.1







Machine Learning: Next steps

- Additional predictors:
  - More soil hydraulic properties (at higher resolution?)
  - Digital Elevation Model (DEM) derivatives (slope, curvature etc.)
  - Geological information at higher resolution (1:250 000)
  - Faults
- Possibly correction for effect of soil moisture “anomaly” during sampling
- Systematic procedure for model development predictor selection
- Find “best” model
- Estimating prediction uncertainty
- Applying other algorithms:
  - Classification and Regression Trees (CART)
  - Random Forest (RF)
  - Artificial Neural Networks (ANN)
  - …

Conclusion/Outlook

- Alternative to geostatistics
- Machine learning performs better than physical model
- Focus on machine learning approach
- Results relevant for mapping of GRP / defining radon priority areas
- Spatial resolution ≤4 km
- Comparison to geostatistics: better predictions? Less uncertainty?

Questions?!