Research Proposal

Master's Thesis

Using Machine Learning methods for Geogenic Radon potential Mapping in Hessen

By: Augustine M Gbondo
TropHEE

Sublogo

Contents

Sublogo

About Me

- Nationality – Sierra Leonean
- TropHEE Student at TU Darmstadt
- One and half years experience with Python
  -Data Analysis
  -Data Science
- GIS/ML and Hydroinformatics
- Darmstadt/Giessen

Sublogo

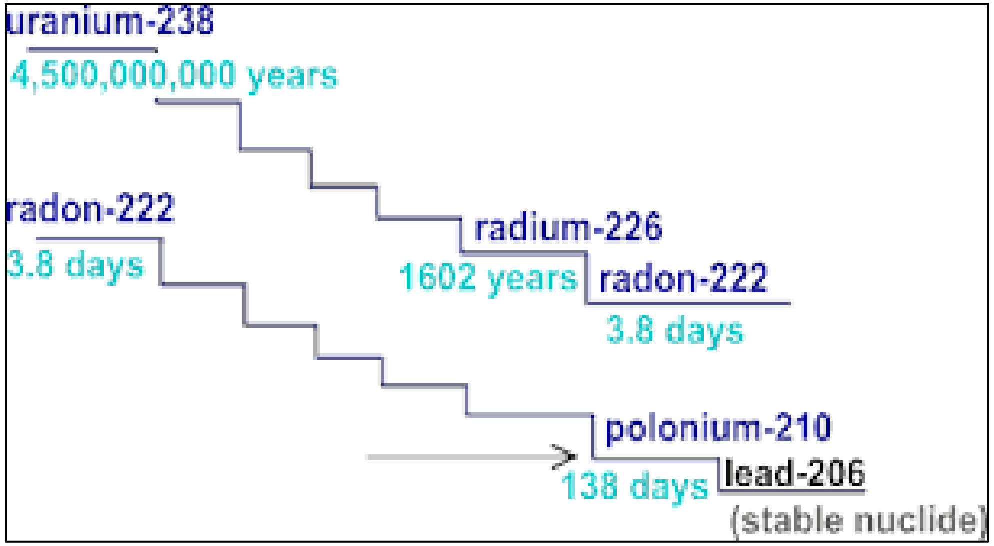
Introduction

TECHNISCHE UNIVERSITÄT DARMSTADT

- 3 naturally occurring (Rn) isotopes.
  -219 Rn (actinon), 220Rn (thoron), and 222Rn, (Radon),
  - Produced by radio-decay of 226Radium
  - Uranium is removed through weathering, but Radium remains insitu (Appleton, 2007)

- Originated from granitic and Uraniferous metamorphic rocks
  - Sandstones, limestones, phosphatic and organic shales are also host rocks. (Appleton, 2007)
  - Often found in planar discontinuities

- Transport by diffusion transport in gas/water carrier fluids



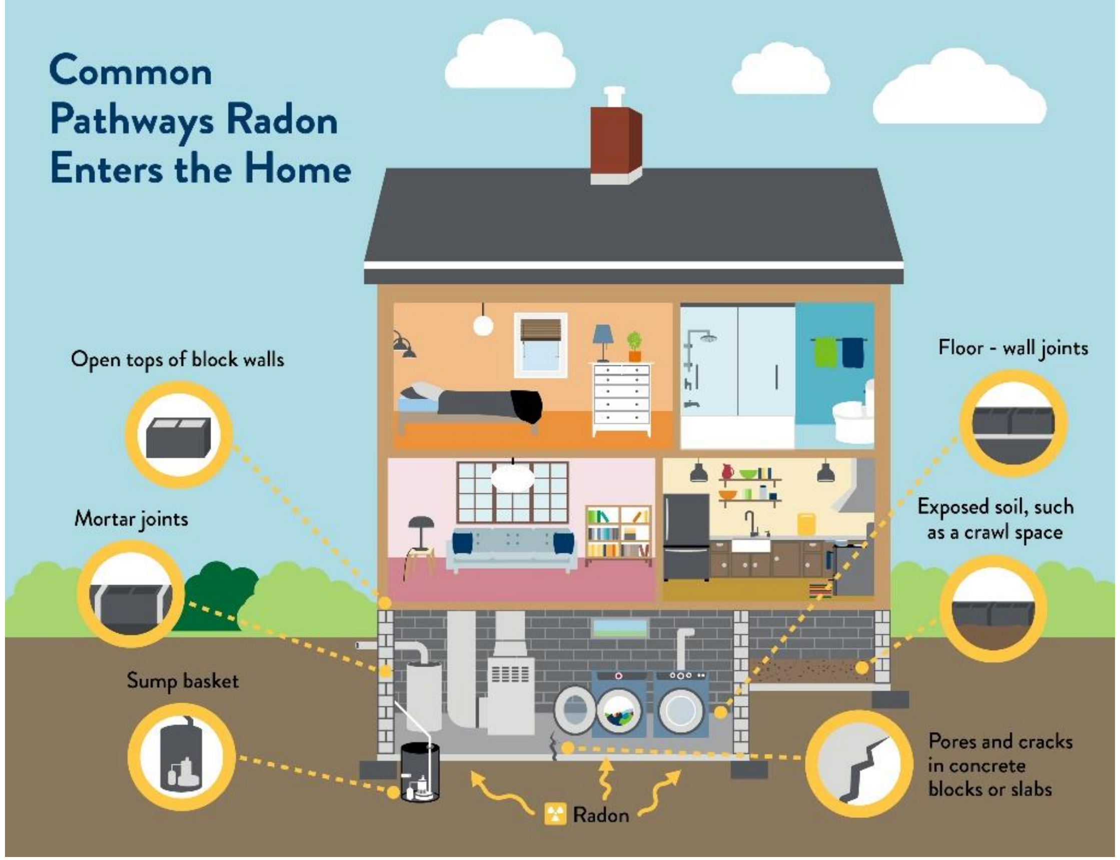
- Usually measured as Becquerel per m3/litre (Bq/l) or (Bq/m3)

Where 1 picocurie [pCi] = 0.037 becquerel [Bq]

Sublogo

Radon and Health

- 222Rn - air pollutant, carcinogenic and detrimental on human health (e.g.WHO (2009))
- Generally, people receive the highest exposure to Rn indoors, i.e. in homes and at workplaces.
- In Germany, the number of lung cancer deaths p/a by residential Rn was estimated with ~1900 (Menzler et al., 2008)
- Geogenic radon; natural e.g in rocks. geology, soil properties, and hydrology,
- Exceedance limit in Germany is 300 Bq/m3 (bfs.de)



https://www.health.state.mn.us

Sublogo

Summary of previous work

TECHNISCHE UNIVERSITÄT DARMSTADT

Defining a regional Randon Index

- Geogenic Rn hazard index (GRHI) in 2010 for harmonization of regional maps in Europe (EURAMET & MetroRADON project)
- Difficulty in Harmonizing maps since different predictors were used
  - Geogenic approach: weighted mean
  - Optimal approach; best predictor
- GRP = Soil Radon conc. X soil perm.
- GRP strong predictor of IRC

- ***Mapping of GRP in Germany***
  - Geostatistical method producing a 10x10Km Map (Bossew, 2015)

ML Applications in Randon.

Rest of Europe

- Ensemble regression trees for predicting indoor Rn concentration in Switzerland (Kropat et al., 2015),
- neural networks for Mean IRC in the Czech Republic (Timkova et al., 2017)
- Time-series analysis (Torkar et al. (2010), Janik et al. (2018)).

Germany

- Mapping GRP (Peterman et. Al, 2020)
- Mapping Indoor GRP by Logistic Regression (Peterman & Bossew, 2021)

Sublogo

Previous studies

Bossew (2015)

(*GRP and Randon Prone Areas (RPA) for effective resource allocation*)

- Actual (IRC) is a factor of GRP and anthropogenic factors

Geostatistical method

- Statistic related to IRC exceeds a threshold, mean or spatial mean
  -Mapping the geogenic factor
- Geological normalization of the GRP data, interpolated and plotted in Surfer
- RPA map based on areas >100 Bq/m2
- *Outlook*: Clustering technique of Geology, merging or sub-divisions

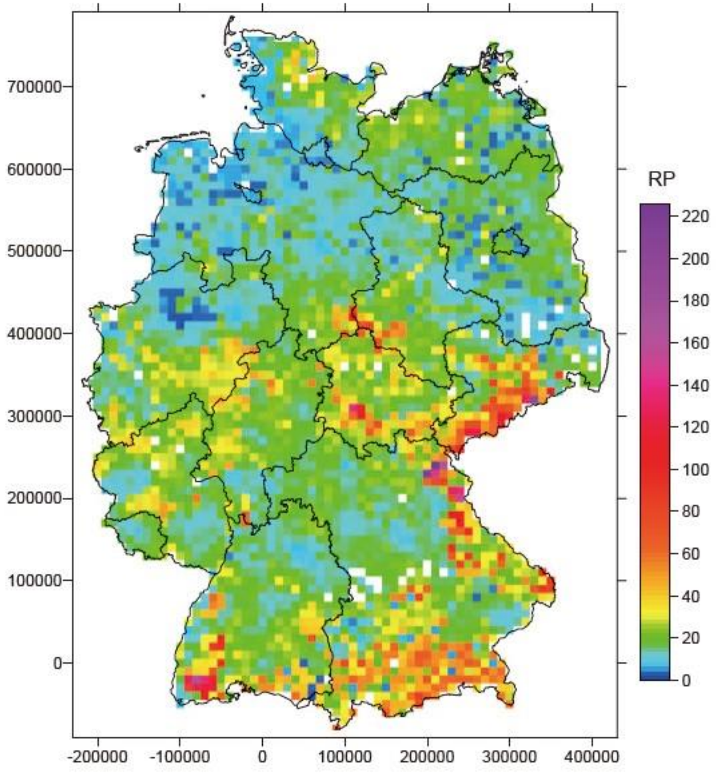
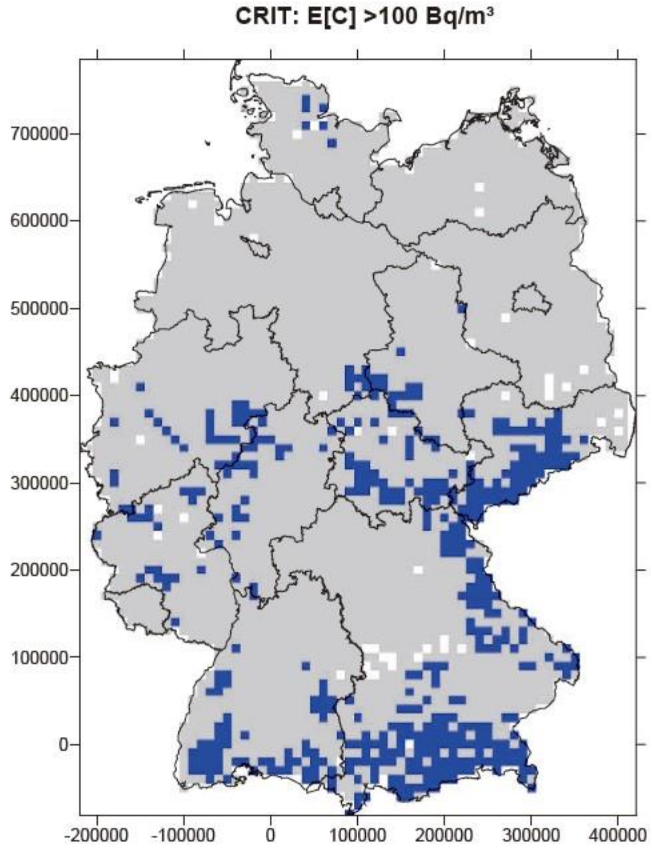


Fig. 1. Estimated expectation of the radon potential over Germany, based on 100 realizations of sequential Gaussian simulation. Coordinates: projected (Lambert azimuthal equal area), units: m. Cell size: 10 km×10 km.



(Bossew, 2015)

Sublogo

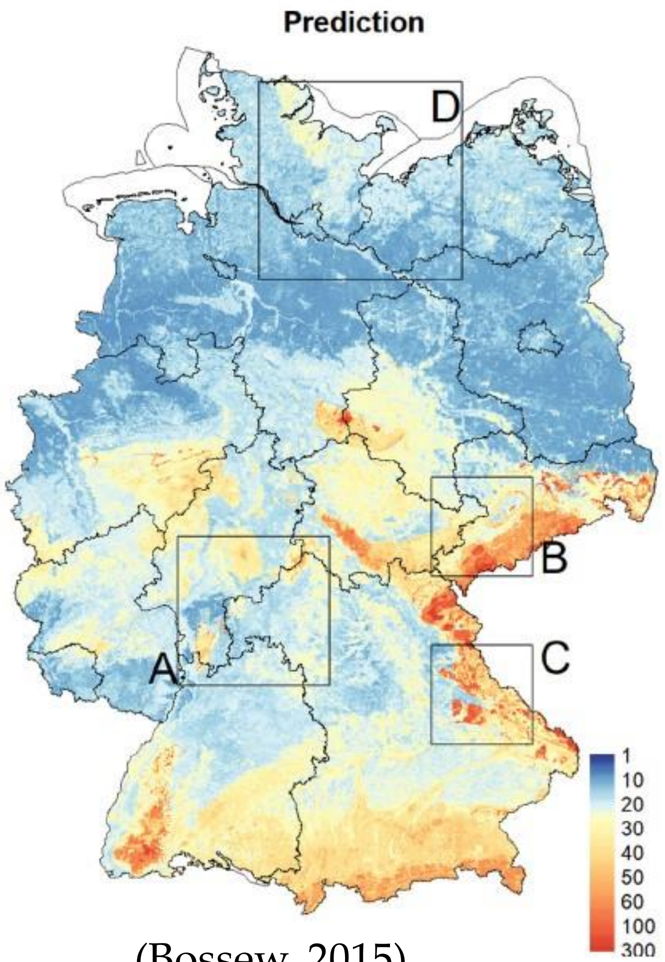
Previous studies

Peterman et. al (2020)

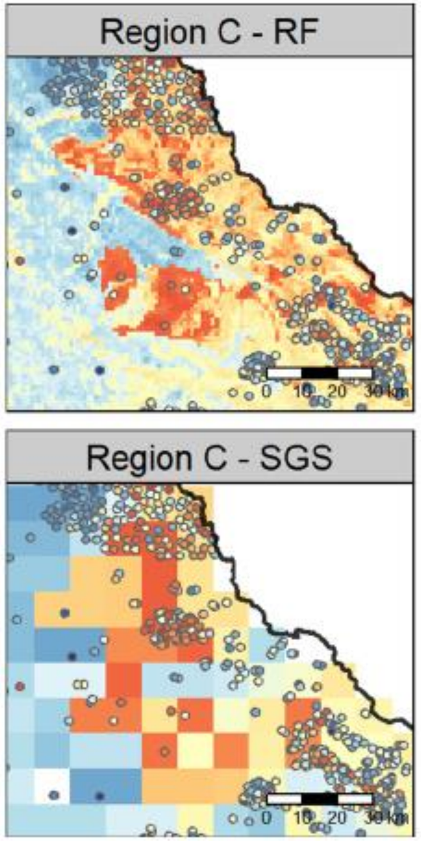
- 4400 measuring points. (1990–2003)
- A Machine learning approach
- MARS (Multivariate adaptive regression splines), SVM and Random Forest (best model)

- *Input parameters*
- Geology, soil(physical-chemical & hydr. Properties, Uranium content and Climate
- *Result*: Geology, winter temperature, Hydraulic conductivity are important predictors

- More field data and better geological map and implement deep learning techinques



(Bossew, 2015)



RF model shows more details

Sublogo

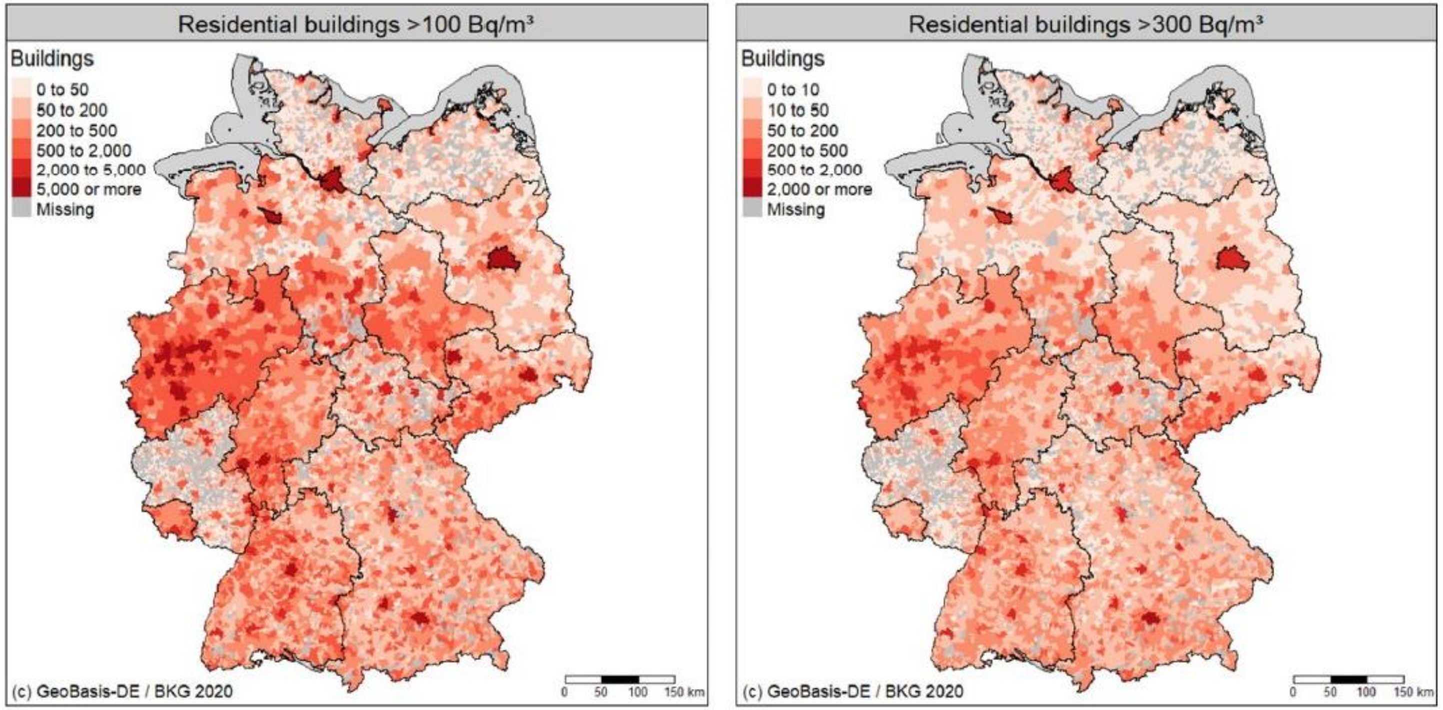
Previous studies

- **Peterman & Bossew (2021)**
  (Mapping Indoor Rn)

- Geological map 1:250 k instead of 1:1m
- Data on residential buildings, 1:250,000 and normalized for house basement and average hazard (exceedance probability)

- *Modelled* IRC as a function of GRP via a Logistic Regression model
- P(E) that IRC exceeds 100 or 300 Bq/m3,

- *Results* showed 350k (1.9%) of the 17.8mi residential buildings with over 300 Bq/m3



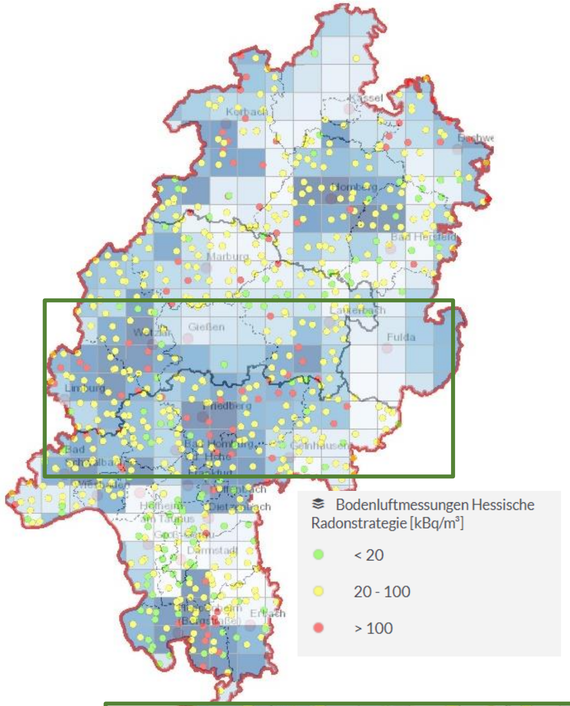
Peterman & Bossew (2021)

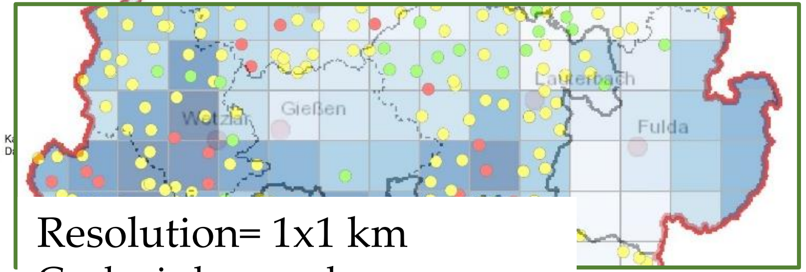
Model (left) is more accurate

Sublogo

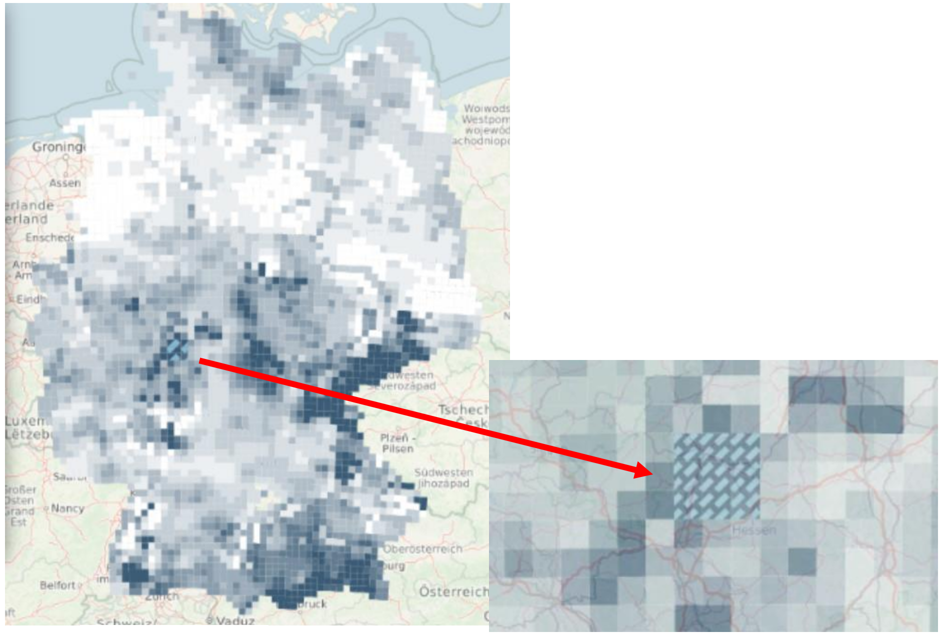
Research Objectives

- GRP Mapping of Hessen with a finer spatial resolution
- Best ML model for predicting GRP in Hessen
- Effect of a lineament Map and high-resolution Geology Map on predicting GRP in Hessen





Resolution= 1x1 km
Geologie hessen.de



Bfs.de

Marburg, Lollar, Allendorf..
Showed anomalous readings

Sublogo

Data and Methods

TECHNISCHE UNIVERSITÄT DARMSTADT

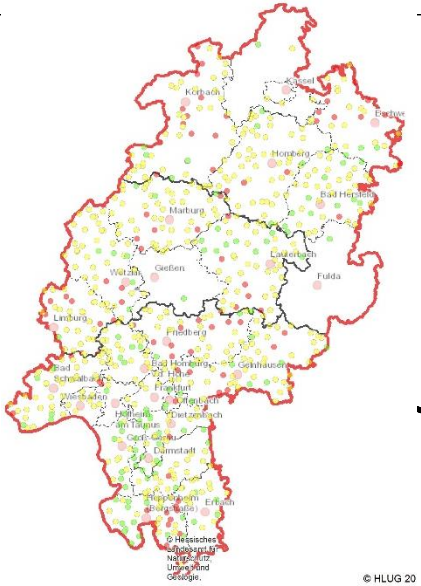
- Field Data
- (GRP data)
  normalized based
  on seasons

- Predictor Data
- Lineament Map
- 1:25,000 Geological Map

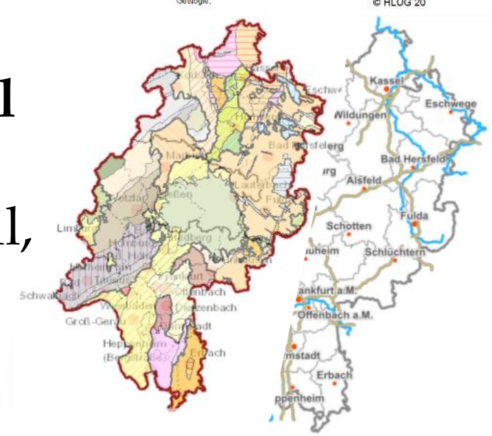
- Geological class based on the geological map of Germany, scale 1:1,000,000 (BGR, 1993). Data was re-classified based on the classification used previously for the GRP map of Germany (Bossew, 2015) and further simplified into 30 classes. Classification was mainly done by geological criteria (stratigraphy, petrography and genesis). Further, classes with similar statistical properties were merged to reduce the number of classes for computational reasons and to allow a minimum number of observations in each class. For details see Table 2 (Appendix).
- Soil hydraulic properties in 1000 m resolution (Tóth et al., 2017):
  - o saturated hydraulic conductivity
  - o saturated water content
  - o field capacity
  - o wilting point
  - o parameter α of the hydraulic conductivity curve
- Soil physical properties in 500 m resolution (Ballabio et al., 2016):
  - o clay content
  - o silt content
  - o sand content
  - o coarse fraction
  - o available water capacity
  - o bulk density
- Soil chemical properties in 500 m resolution (Ballabio et al., 2019):
  - o pH in H_2O
  - o cation exchange capacity
  - o carbon:nitrogen ratio
  - o concentration of calcium carbonate
  - o concentration of nitrogen
  - o concentration of phosphorous
  - o concentration of potassium
- Soil uranium concentration in 10 km resolution (Cinelli et al., 2019)
- SAGA wetness index derived from the digital elevation model of Germany (resolution 25 m) (BKG, 2018)
- Climate data in 1000 m resolution (DWD, 2018a, 2018b, 2018c):
  - o Temperature: annual and seasonal means 1981–2010 (DWD, 2018a)
  - o Precipitation: annual and seasonal means 1981–2010 (DWD, 2018b)
  - o Soil moisture: annual and seasonal means 1991–2010 (DWD, 2018c)

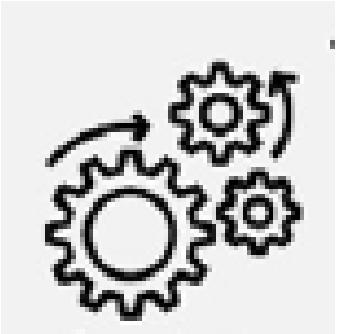
Data and Methods

- **Field data**
  - Updated
  - Correction factor

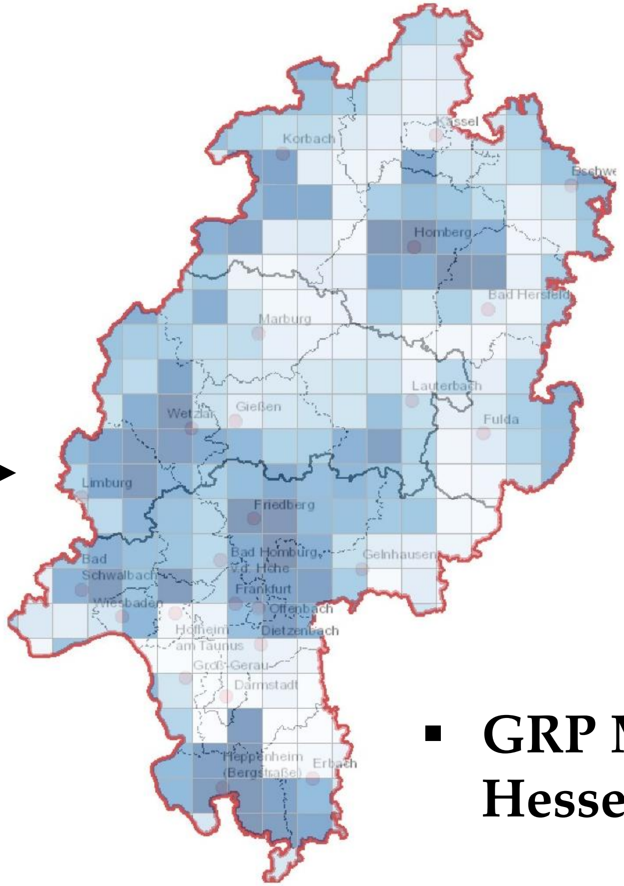


- **Environmental co-variables**
  - Geology, soil, climate etc.





ML Model



- **GRP Map of Hessen**

Sublogo

Data Processing

- **Measured/Field data**
  - Exploratory Data Analysis and cleaning
  - Handling missing data
  - Detecting and filtering unwanted outliers
  - Handling of missing data

- Tools : Xcell, python

- **Predictor Data**
  - Geological map, soil maps, climate
  - Geo-reference or digitize where necessary
  - Resampling where necessary
  - Feature extraction

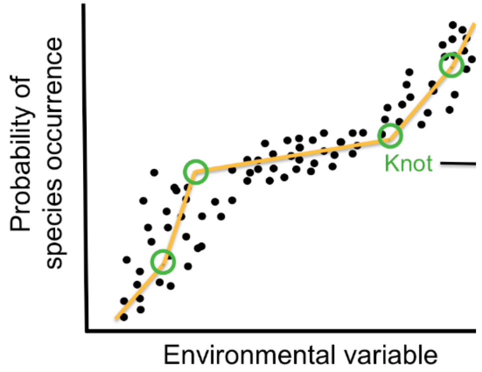
- Tools : ArcMap, python

Sublogo

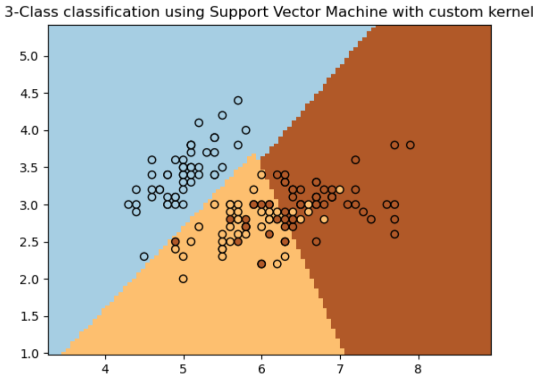
Model selection and Performance Metrics

TECHNISCHE UNIVERSITÄT DARMSTADT

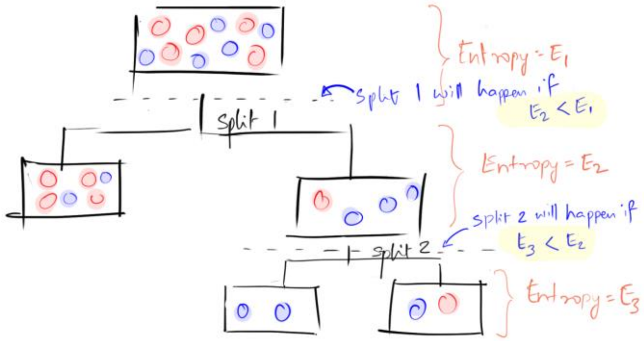
Models



- Multivariate adaptive regression splines (MARS)



- Support Vector Machines



- Random Forest

- These models showed promising result in spatial mapping of soil (Ballabio et al., 2016; Hengl et al., 2018; Liess et al., 2016), landslide hazard mapping (Bui et al., 2016) or mapping of atmospheric particulate matter (Choubin et al., 2020).

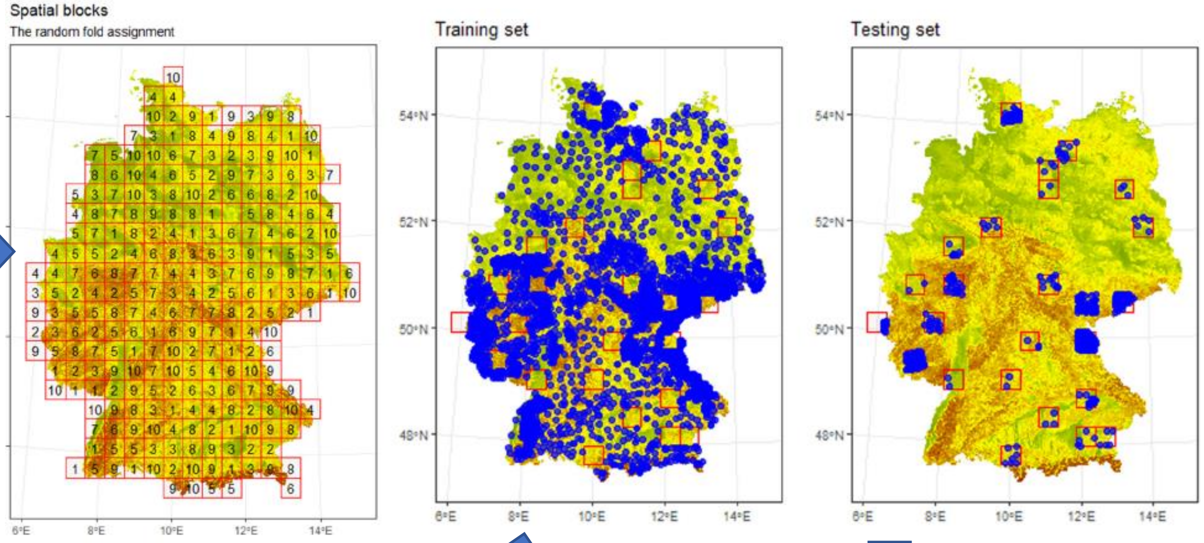
Sublogo

Modelling Strategy

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Predictor selection**
  - Removal of collinear and non-informative variables

- **Building a final model**
  -Cross validation



- **Hyperparameter tuning**
  - Maximizes the model's performance
  - Weight*(wij)* and biases *(bij)* assigned random values

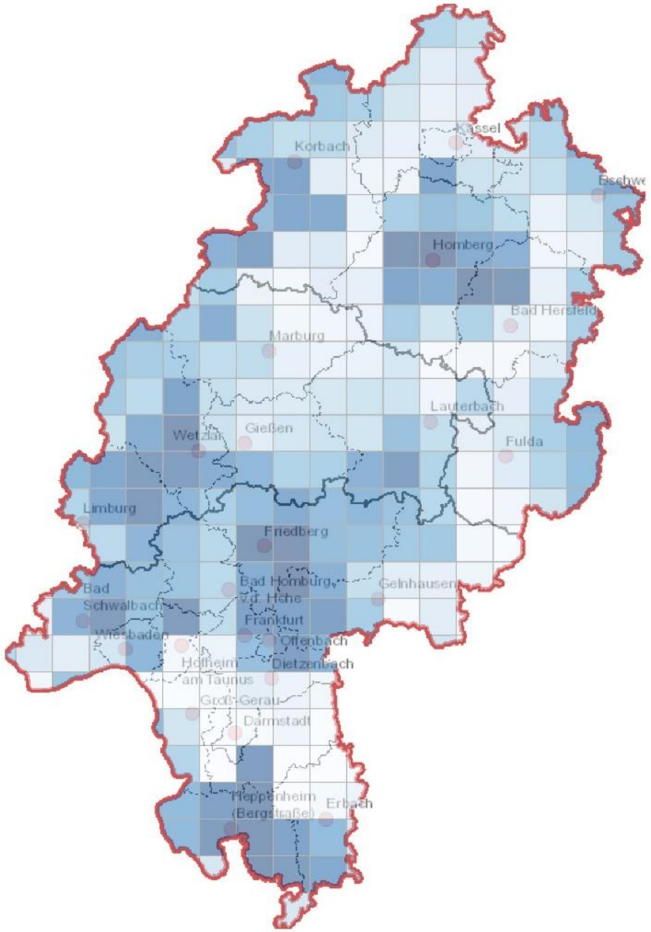
- **Performance assessment**
  - r2, RMSE, MAE, RRSE and RMSLE

- **Mapping** (i.e., applying the model for spatial prediction)

Targeted Outcome of the study

TECHNISCHE UNIVERSITÄT DARMSTADT

- A completed Masters Thesis
- GRP Map (seasonal – as modelled from the given data)
- Online interactive visualization of the Maps



Sublogo


Thank you