Federal Office for Radiation Protection

AGES
Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH

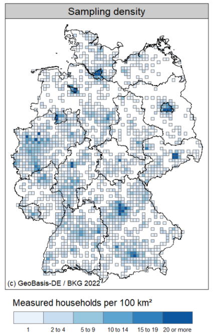
A new high-resolution residential radon map for Germany using a machine learning based probabilistic exposure model

Eric Petermann[1], Peter Bossew[1,2], Jochen Kemski[3], Valeria Gruber[4], Nils Suhr[1], Bernd Hoffmann[1]

1 – Federal Office for Radiation Protection (BfS), Berlin, Germany
2 – retired
3 – Sachverständigenbüro Dr. Kemski, Bonn, Germany
4 – Austrian Agency for Health and Food Safety (AGES), Linz, Austria

Workshop Geological Aspects of Radon Risk Mapping

19 Sep 2023, Prague



1. Motivation

- New indoor survey
- Representativeness
- Objectives

2. Model

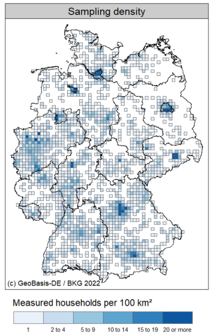
- Approach
- Predictors
- Quantile regression forest
- Population-weighting

3. Results

- Model interpretation
- Model performance
- Maps

4. Discussion

- Implications
- Limitations
- Conclusion



1. Motivation

- New indoor survey
- Representativeness
- Objectives

2. Model

- Approach
- Predictors
- Quantile regression forest
- Population-weighting

3. Results

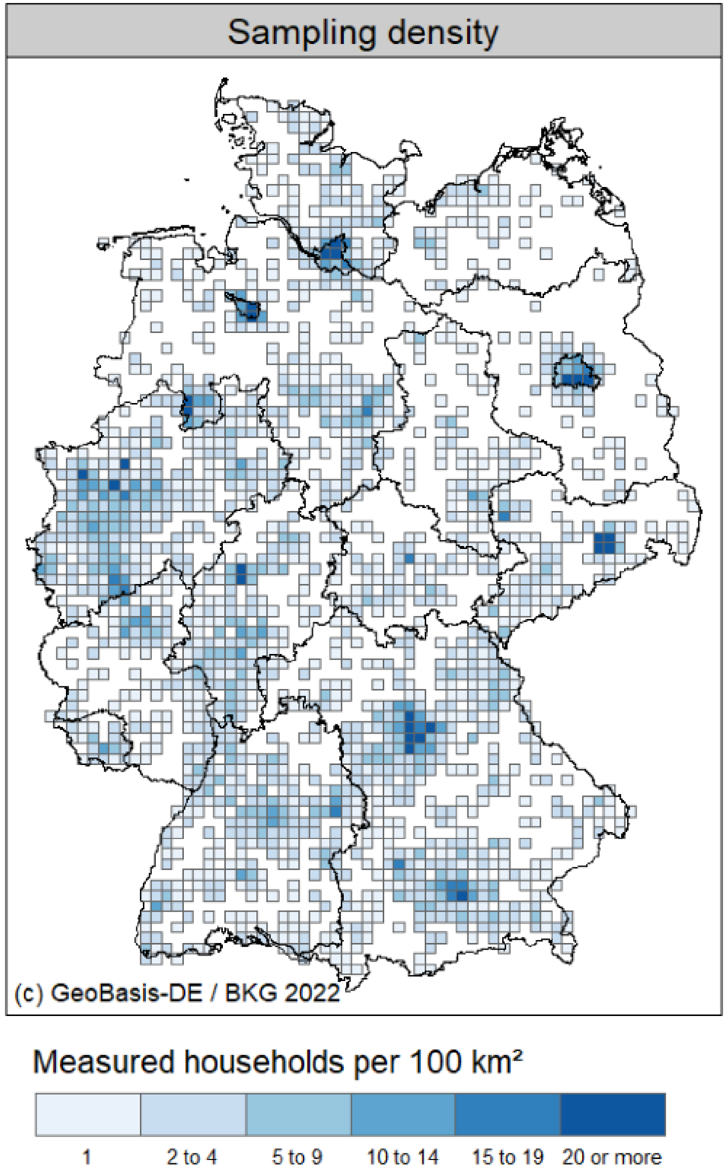
- Model interpretation
- Model performance
- Maps

4. Discussion

- Implications
- Limitations
- Conclusion

New indoor radon survey in Germany

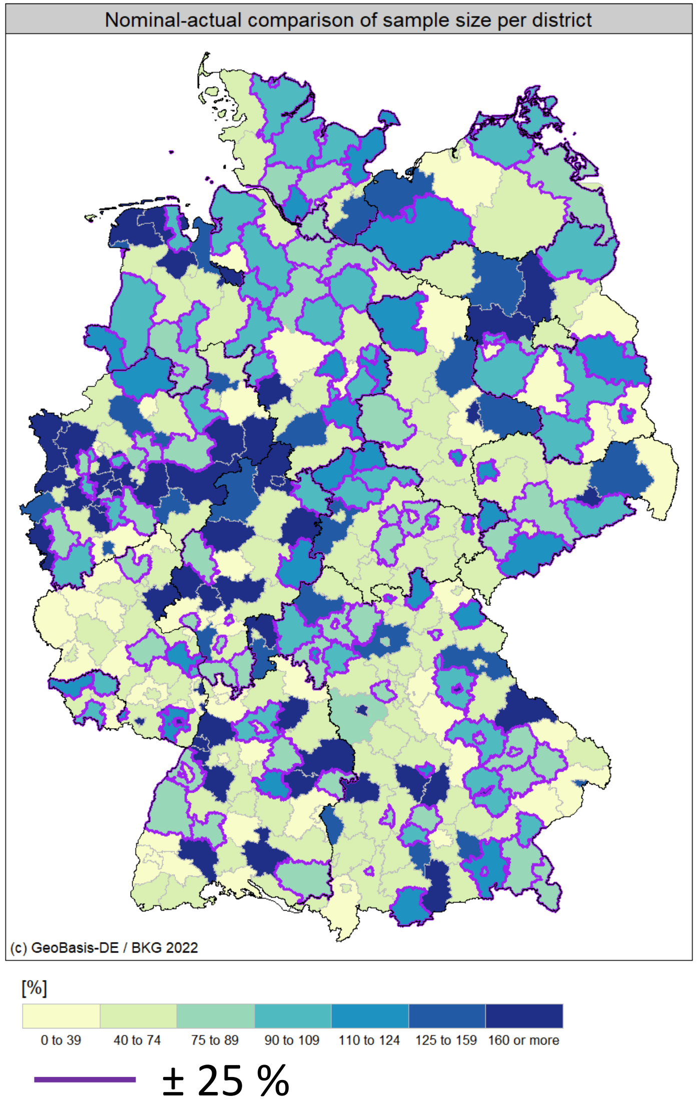
- Time period: 2019 – 2021
- In total 7500 households, 2 detectors each in occupied rooms
- Annual measurement using SSNTD detectors
- Sampling density proportional to population density on the district level

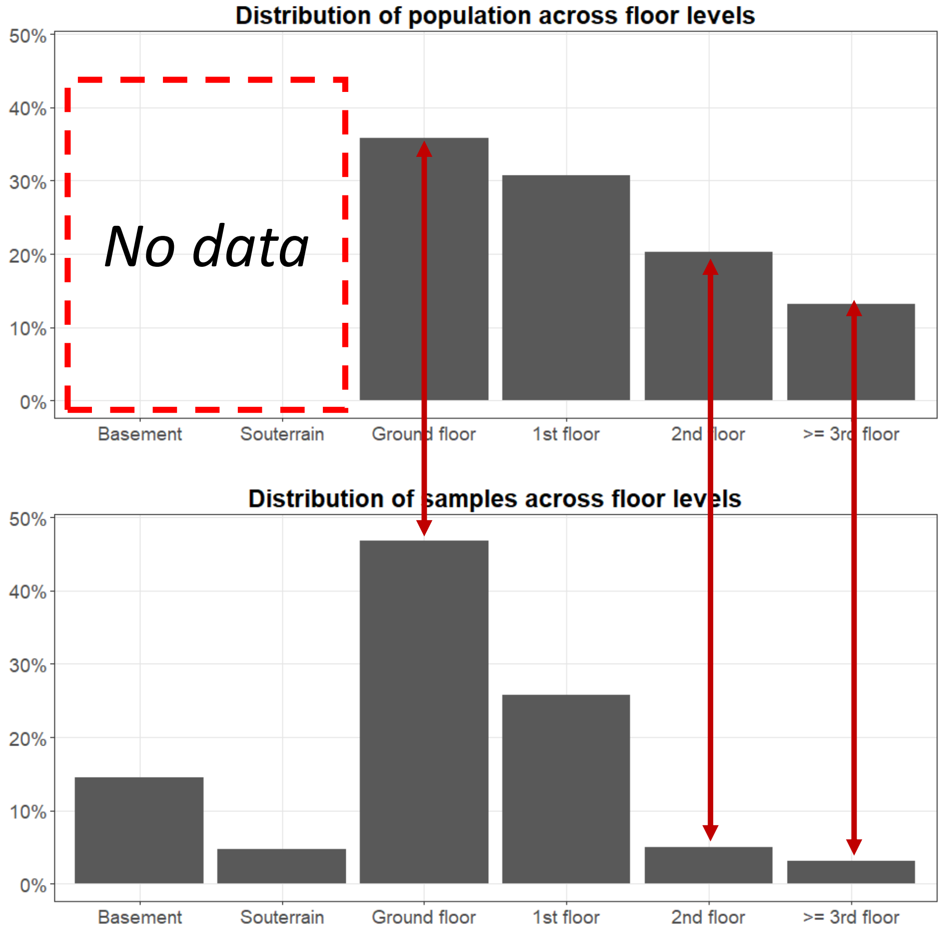


Representativeness?

- Overall spatial coverage satisfying, although deviations from target sample size in some districts
- Too many samples on groundfloor, too less on higher floor levels

→ **Overestimation expected**





Objective

1) compensate for a potential sampling bias by a model-based approach under consideration of distribution of radon-relevant factors of the entire population
2) reflect the indoor radon distribution of the population by propagation of prediction uncertainty into variability by using a probabilistic approach (Monte Carlo sampling) -> exceedance probability, P95 etc.
3) allow estimation at the municipality level by using spatially highly resolved auxiliary data (7500 dwellings measured, but ~10,900 municipalities)

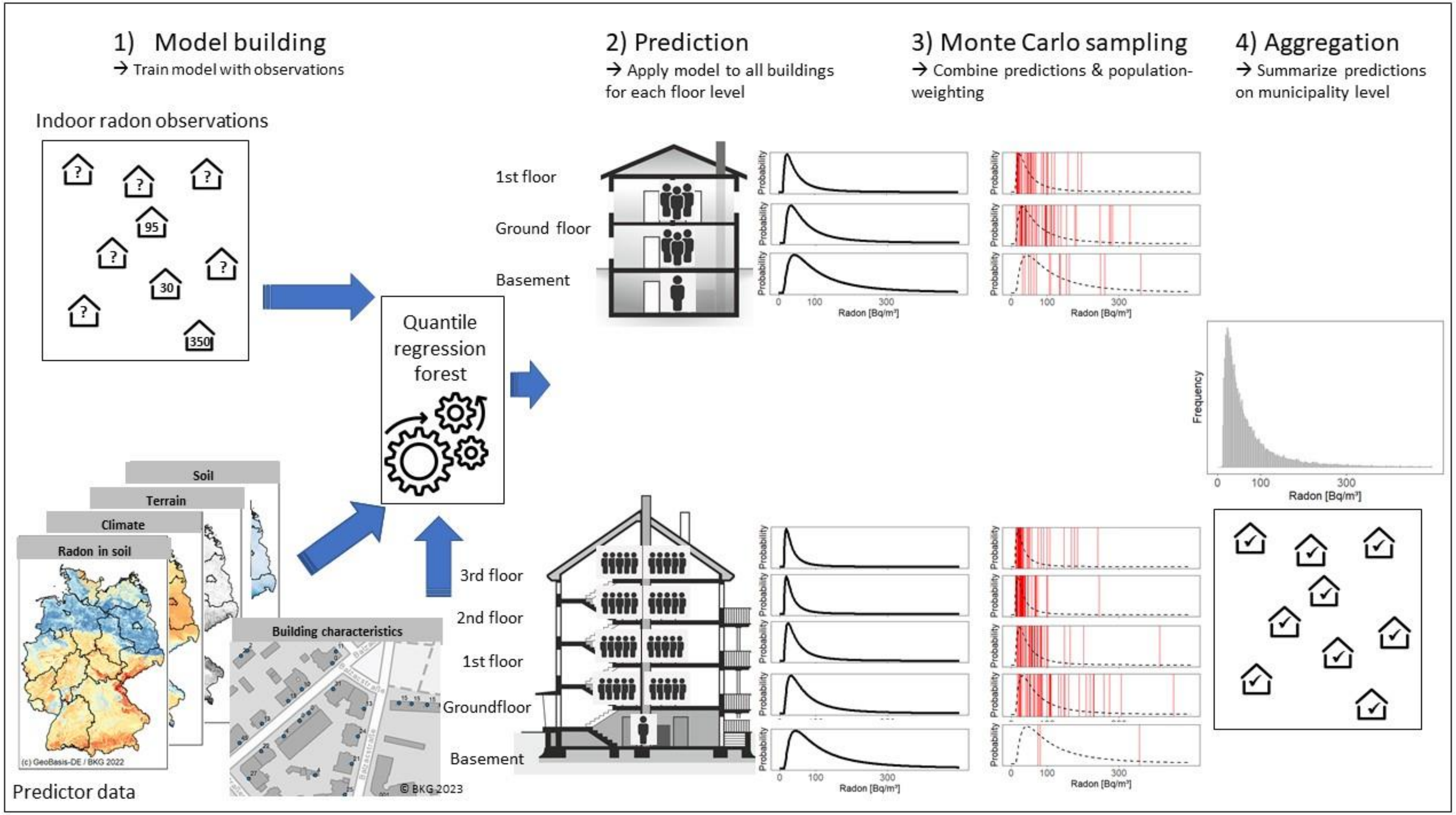
1. Motivation
- New indoor survey
- Representativeness
- Objectives

2. Model
- Approach
- Predictors
- Quantile regression forest
- Population-weighting

3. Results
- Model interpretation
- Model performance
- Maps

4. Discussion
- Implications
- Limitations
- Conclusion

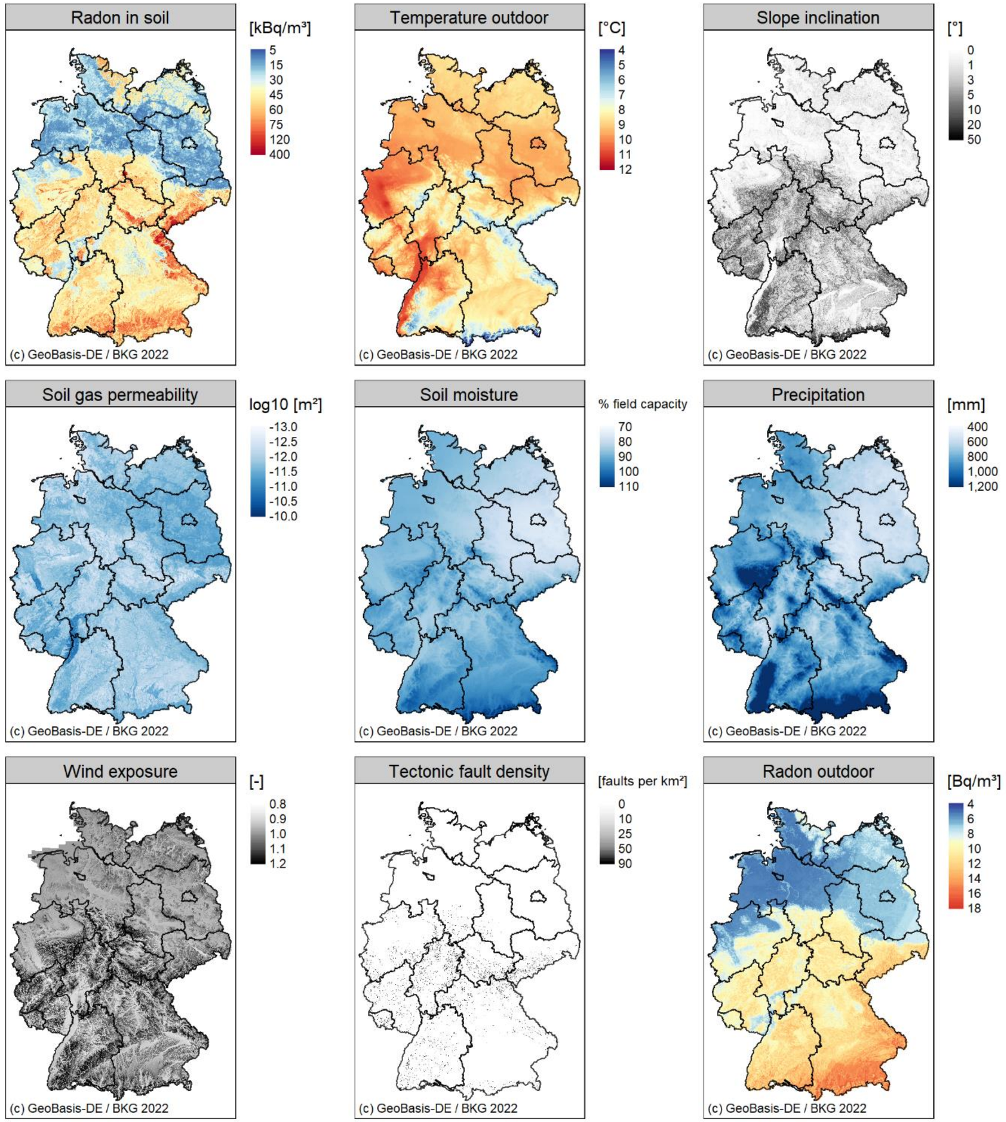
Modelling approach



Modelling approach

Predictors

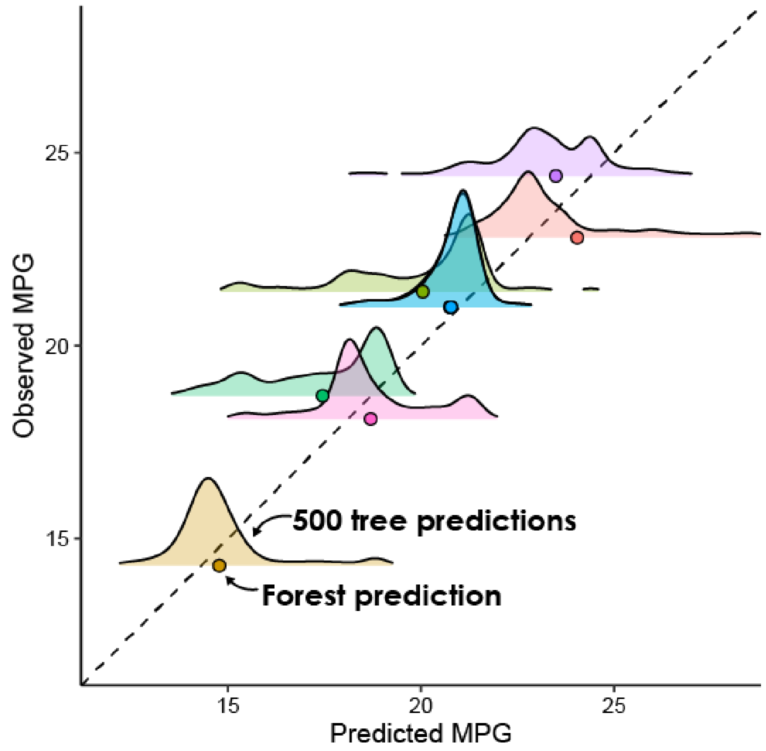
- Environmental predictors
  - Soil
  - Climate
  - Terrain
- Building characteristics
  - Floor level
  - Age
  - Type
  - Number of inhabitants



Predictor data

Model: Quantile regression forest

- Variety of a random forest model that allows preditions of full conditional distribution, not only conditional mean
- Can be easily implemented without much additional computational effort
- Estimation of conditional quantiles for each case -> each floor level of every residential building (n=21 million) in Germany

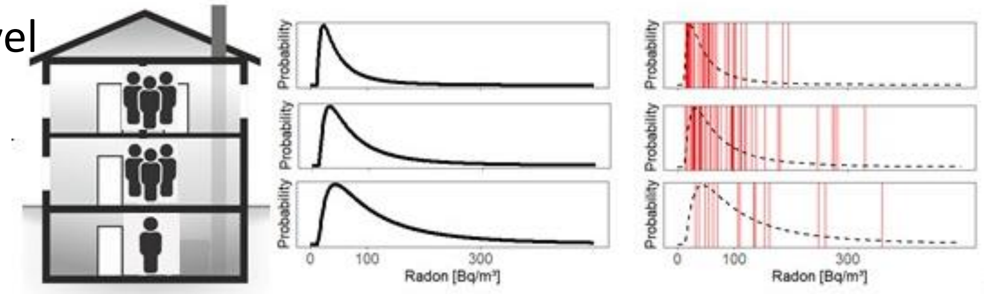


https://www.ryan-alcantara.com/projects/p89_random_forest_trees/

Probabilistic sampling

- High prediction uncertainty due to missing information, predictor data missing local phenomena etc.
- Consideration of predictive uncertainty required to give a realistic estimate of the distribution (e.g., probability to exceed 300 Bq/m³; 90%ile) because of smoothing tendency of regression models
- Goal: moving from quantification of prediction uncertainty to estimation of variability
- Implementation: random sampling from estimated floor level distribution, sample size proportional to number of inhabitants

Prediction uncertainty for individual case



Probabilistic sampling with sample size proportional to population size

1. Motivation
- New indoor survey
- Representativeness
- Objectives

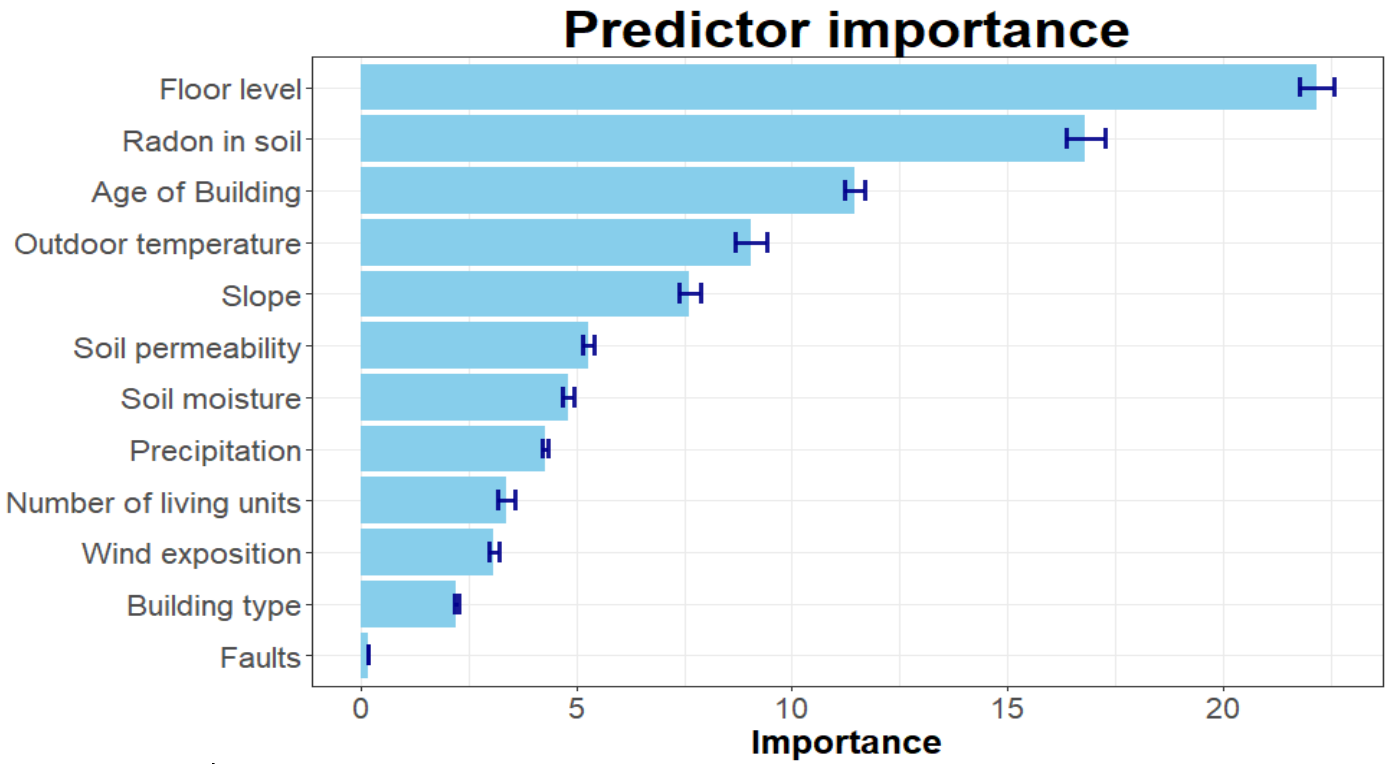
2. Model
- Approach
- Predictors
- Quantile regression forest
- Population-weighting

3. Results
- Model interpretation
- Model performance
- Maps

4. Discussion
- Implications
- Limitations
- Conclusion

Model interpretation – predictor importance

- 12 informative predictors
- Floor level and radon in soil most important
- By groups:
  - Building characteristics: 44 %
  - Soil: 30 %
  - Climate: 15 %
  - Terrain: 12 %

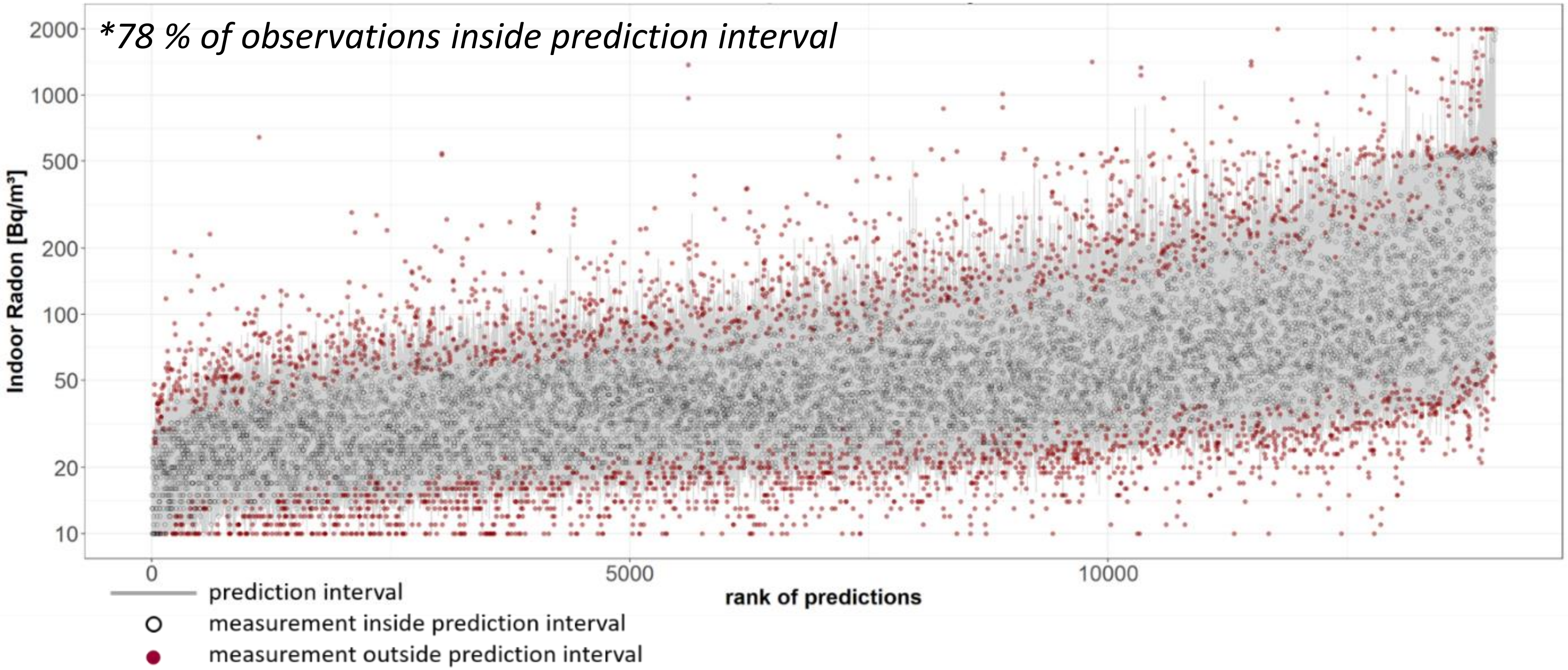


Predictor importance

Model performance

- 5 x 10-fold spatial cross-validation
- Observations outside prediction intervals ± homogeneously distributed across entire value range and prediction intervals
- Prediction uncertainty is large, but can be accurately characterized

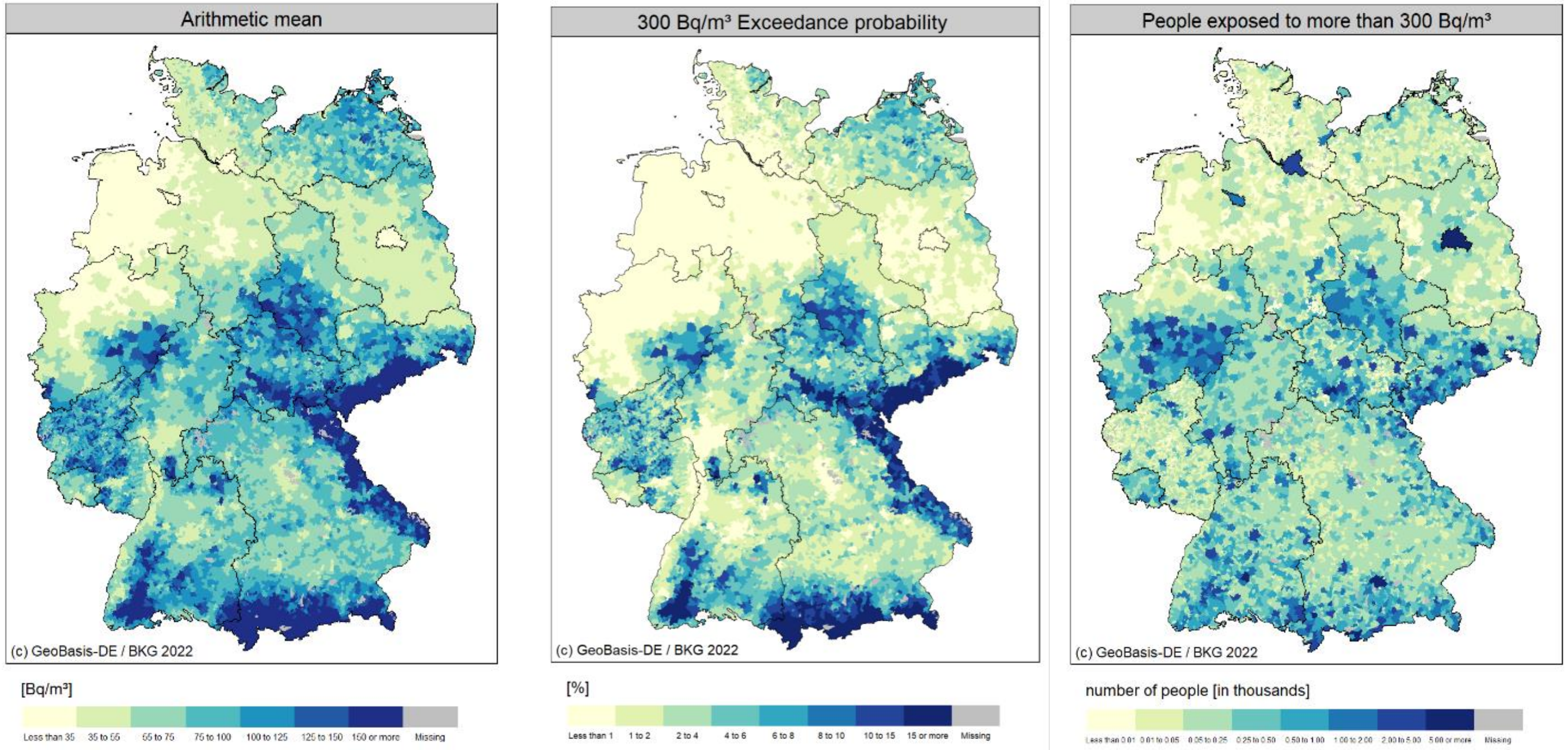
80 % prediction interval (10 to 90 %ile)



Quantile performance

Maps

- Predictions for each building and floor level + population-weighting
- Aggregation by municipality
- Spatial patterns follow radon in soil map
- Big cities with lower concentration (floor level effect)



Indoor radon maps

	Arith. Mean	Geom. Mean	Exceedance probability 300 Bq/m³	P95
National	63 Bq/m³ (SD: 147 Bq/m³)	41 Bq/m³ (GSD 2.27)	2.2 % (1.9 million)	180 Bq/m³
Municipality	24 Bq/m³ to 450 Bq/m³	20 Bq/m³ to 280 Bq/m³	<0.1 % to 48.0 %	54 Bq/m³ to 1500 Bq/m³

1. Motivation
- New indoor survey
- Representativeness
- Objectives

2. Model
- Approach
- Predictors
- Quantile regression forest
- Population-weighting

3. Results
- Model interpretation
- Model performance
- Maps

4. Discussion
- Implications
- Limitations
- Conclusion

Discussion – implications

- IRC estimates (AM: 63 Bq/m³; GM: 41 Bq/m³) higher than in previous study (Menzler et al. 2006; AM: 49 Bq/m³; GM: 37 Bq/m³)
  - ➢ Consideration of basement occupation
  - ➢ Temporal changes cannot be derived
- Estimate of 1.9 million people above 300 Bq/m³ consistent with 350.000 buildings above 300 Bq/m³ estimated by Petermann & Bossew 2021
- 210 municipalities are radon priority area in Germany (2 % of area; 1 % of population); but for ~ 900 municipalities population exposure was found > 10 % exceeding 300 Bq/m³
  - ➢ although, public exposure not being directly comparable to criteria for delineation of RPA, need for optimization of radiation protection
- Results useful for epidemiological studies such as estimation of lung cancer fatalities due to radon or indoor radon estimation in case-control studies with lack of radon data

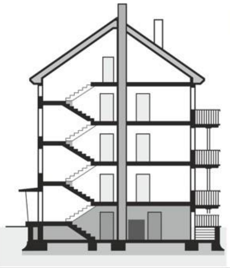
Discussion – uncertainty and limitations

- Uncertainty on basement prevalence and occupation -> parameterized by educated guess dependent on building type 30 % and 5 % of upper floor occupation
- Model performance evaluated at the global (i.e., national) scale; local deviations possible if predictor information incomplete or local specifics
- Period 2019 – 2021; not necessarily long-term mean (climate, living habits)

Example 1: single-family house, 2 floors, 2 inhabitants

Factor	Inh./floor	n
1	0.87	9
1	0.87	9
0.3	0.26	3

Example 2: appartment house, 4 floors, 25 inhabitants

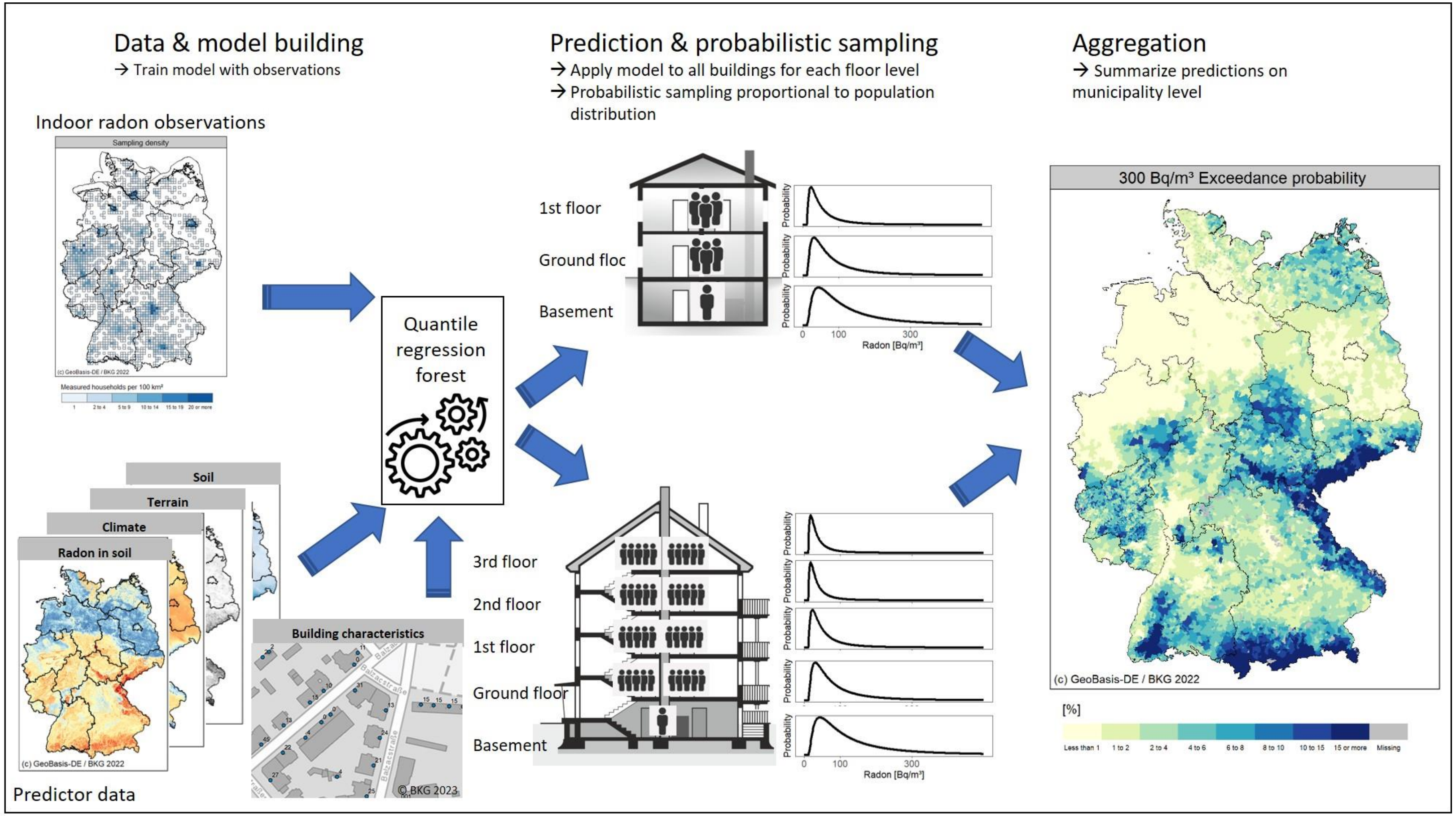


Factor	Inh./floor	n
1	4.95	50
1	4.95	50
1	4.95	50
1	4.95	50
0.05	0.25	3

Parameterization of basement occupation

Conclusion

- new estimate of indoor radon distribution for Germany was produced:
  - AM: 63 Bq/m³
  - GM: 41 Bq/m³
  - P95: 180 Bq/m³
  - 2.2 % > 300 Bq/m³ (1.9 million)
- Propagation of uncertainty of individual predictions into variability at the aggregated scale



Thank you!

Eric Petermann
Federal Office for Radiation Protection (BfS), Berlin, Germany
epetermann@bfs.de