RadoNorm
Managing risks from radon and NORM

Statistical methods for the identification of buildings and areas with high radon levels – a review

Sebastian Baumann, Sara Antignani, Joan Frédéric Rey, Valeria Gruber, Christian di Carlo, Joëlle Goyette Pernot, Francesco Bochicchio

WP 2.4 Institutions: ISS, IRSN, HES-SO, AGES, SURO

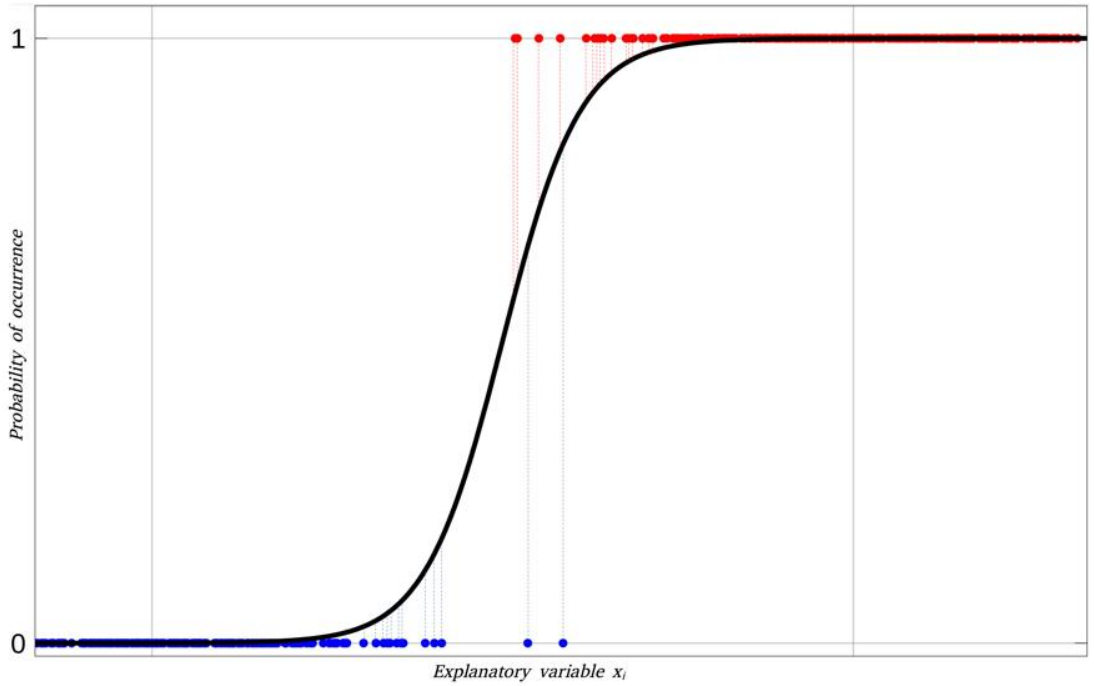
20.09.2023, GARRM, Prague

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Statistical methods to identify building and areas with high radon levels

Content:

- Overview
- Literature review
- Outlook



This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

RadoNorm
Managing risks from radon and NORM

Statistical methods to identify building and areas with high radon levels

Why?

- Estimate Radon Hazard
- Establish regulations and legislation
- Focus of resources
- Areas where the radon concentration exceeds the national reference level
- Remediation and Preventive measures

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Foundation

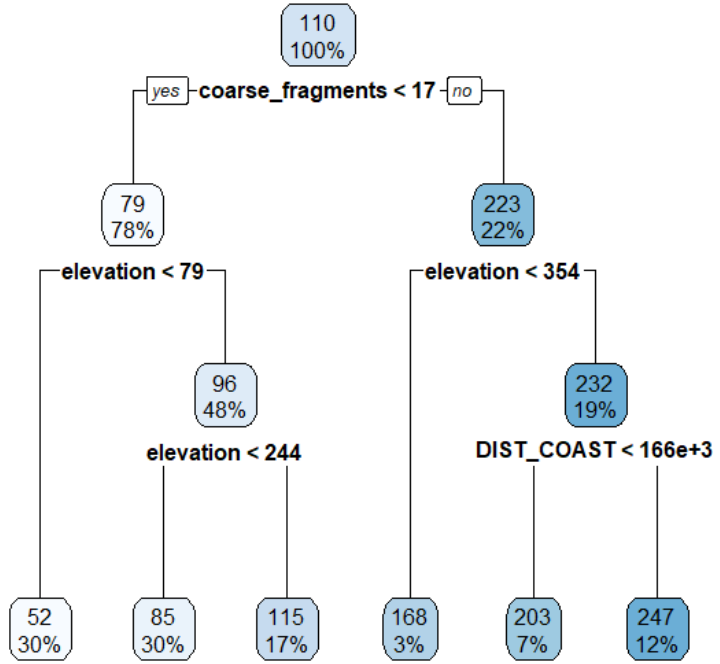
- How to identify buildings and areas with high radon levels?
  - Measurements
    - Radon in air, soil gas, uranium concentration, permeability …
  - Additional Information
    - Geology, building characteristics, climate, using habits …

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Statistical Models

- Target Variable ~ Model (Predictors)
  - Example :

IRC ~ Random Forest { Building Characteristics
Geogenic Factors
Climate ….



This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

RadoNorm
Managing risks from radon and NORM

Statistical Models

- Target Variable ~ Model (Predictors)
  - Example :

IRC ~ Random Forest { Building Characteristics; Geogenic Factors; Climate ….

Goals:

- Prediction:
  - Can we predict indoor radon concentrations with the gathered information accuratly?
- Inference:
  - Are there certain conditions under which certain radon levels are more likely?

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Literature Review

- Done seperatly for areas and buildings
- Key words (buildings):
  - in the title, the words "radon", or "(222)Rn", or "222Rn", or "Rn-222",
  - **AND** in the title, the word "building*", or "dwelling*", or "hous*", or "school*", or "workplace*", or "indoor*",
  - **AND** in the title or abstract, the word "factor*", or "characteristic*", or "feature*", or "parameter*",
  - **AND** in the title or abstract, the word "identif*", or "detect*", or "predict*", or "forecast*", or "affect*", or "influenc*", or "impact*",
  - **AND** in the title or abstract, the word "high*", or "elevated", or "exceed*", or "above", or "quantile*",
  - **AND** in the title or abstract, the word "approach*", or "method*".
- Databases: Web of Science, Scopus and PubMed

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Literature Review

- Done seperatly for areas and buildings
- Key words (areas):
  - in the title or abstract, the words "radon", or "(222)Rn", or "222Rn", or "Rn-222",
  - **AND** in the title or abstract, the words "radon prone area*" or "priority area*" or "high radon level*" or “elevated radon level*" or "high radon concentration*" or "high radon exposure*" or "high radon potential" or "high background" or "radon affected area*"
  - **AND** iii) in the title or abstract, the words "method*" or "approach*" or "mapping"
  - **AND NOT** in the title or abstract, the word "transform*".
- Databases: Scopus and PubMed

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

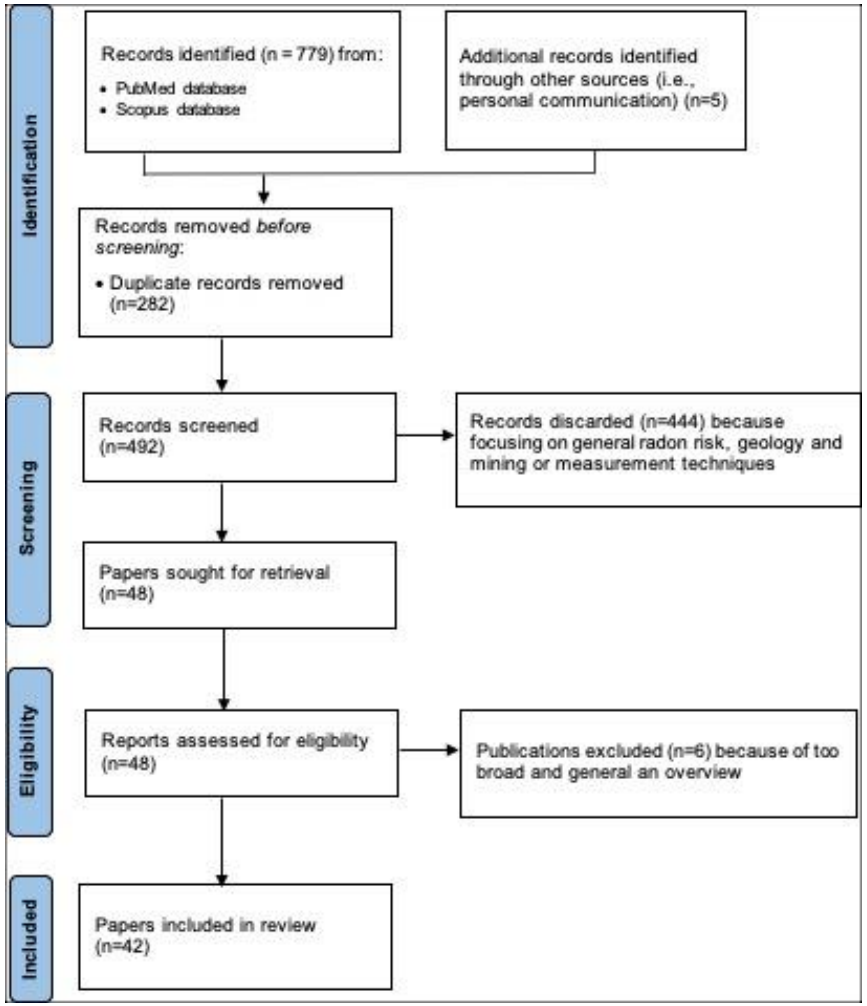
Literature Review

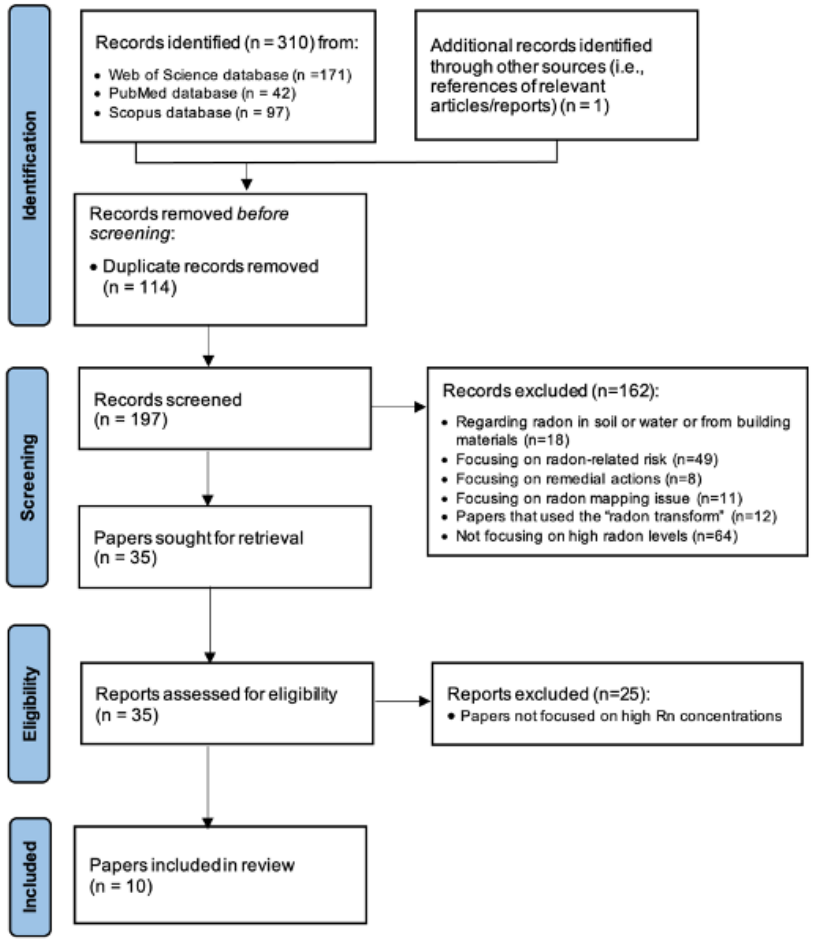
- Done seperatly for areas and buildings
- Number of publications selected (without duplicates):
  - 197 publications for buildings
  - 492 publications for areas

- Manual selection workflow:
  - Title
  - Abstract
  - Full Publications

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Literature review – selection workflow



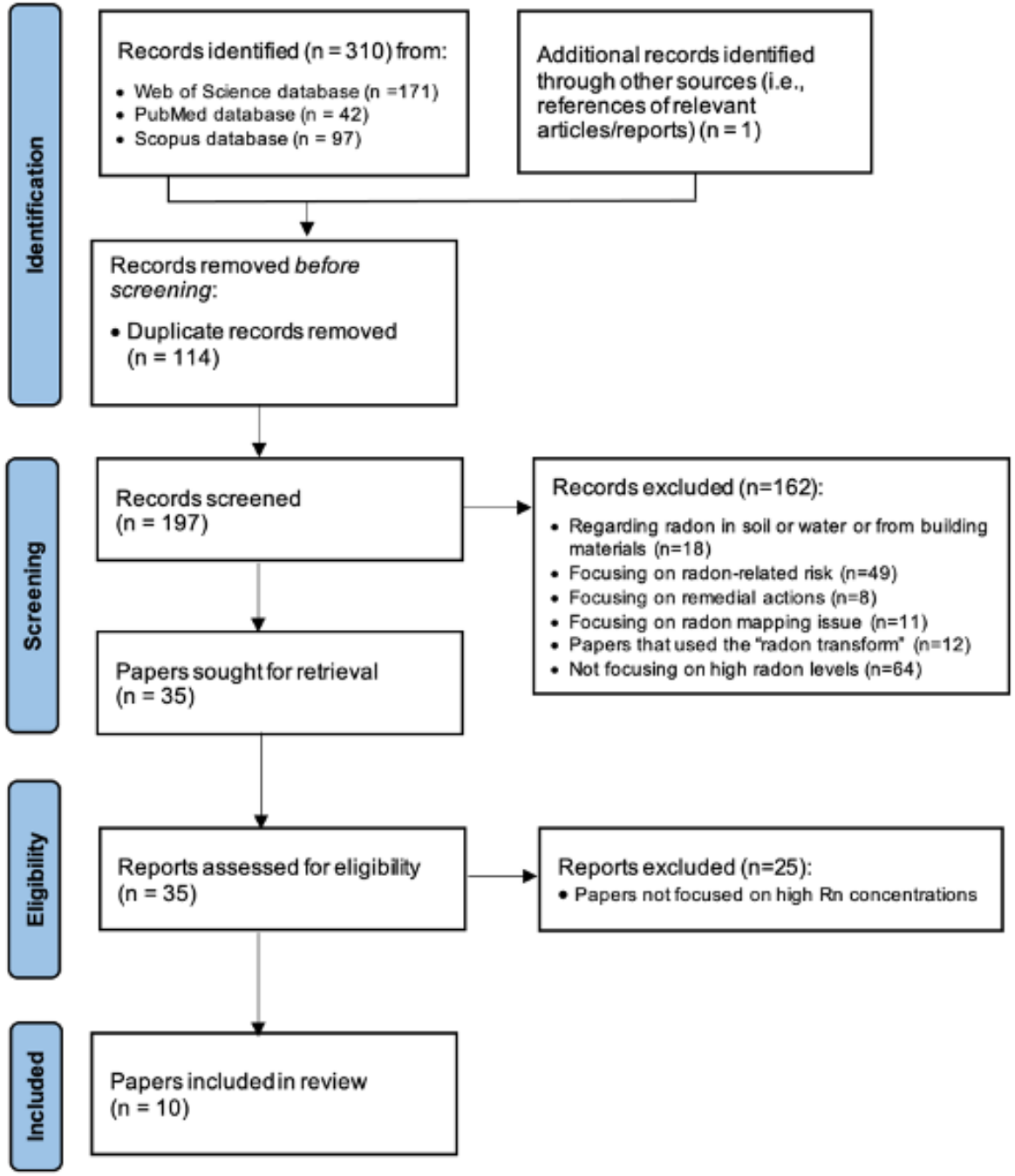


This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

RadoNorm
Managing risks from radon and NORM

Main reasons to reject:

- General radon risk
- Medical studies
- Measurement techniques
- Remediation
- Radon in Water
- Mining and Geology



This project has received funding from the Euratom research and t

Literature Review - selection workflow

- Number of publications selected after manual selection:
  - **10** (197) publications for buildings
  - **42** (492) publications for areas
- Analysis and characterisation of different statistical methods:
  - Statistical methods
  - Geostaistical methods
  - Machine learining methods

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Literature review findings

- Quantile and logistic regression

- Conditions of application:
  - Predifine quantile level or threshhold
  - Sound data set
- Main findings:
  - Well known predictors show impact (location, building age, floor level …)
  - Impact of predictors can change with selected quantile/threshhold

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

RadoNorm
Managing risks from radon and NORM

Literature review findings

- Geostatistical methods (Invrese distance weighting, Kriging, geographically weighted regression)

- Conditions of application:
  - Georeferenced and sound data
  - Predifine variogram, power parameters or spatial weights

- Main findings:
  - Used in many applications producing radon maps
  - Strongly depend on the spatial sample distribution of the data

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Literature review findings

- Machine Learning methods (Random Forest, Support Vector Machines, Artificial Neural Network)
- Conditions of application:
  - Sound data basis
  - Hyperparamater setting
  - Test and Training data split (consider spatial autocorrelation)
  - Performance evaluation
- Main findings:
  - Powerful models for prediction but not especially designed for prediction of high values
  - Modifications of the target value or the evaluation metric might be a valid strategy to focus on high radon levels.

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Literature Review - findings

- Usually statistical models are applied to predict the average radon concentration - good models can also predict high radon levels.
- Comparability of publications and their results is very limited, due to different …
  - Data basis
  - Data manipulation
  - Models
  - Performance evaluation

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Literature Review – findings/outook

- The overall workflow from data handling to performance evaluation might be more important than the actual method (machine learning perspective)
- Possible workflow adaptions to characterize high radon levels
  - a) Penalize samples with high values when training the model and for performance evaluation (weights)
  - Adopt performance evaluation (loss function) (e.g. quantile regression for machine learning methods)
  - b) Transform the regression task in a classification task, where the high values are a separate class
  - c) Duplicate samples with high values

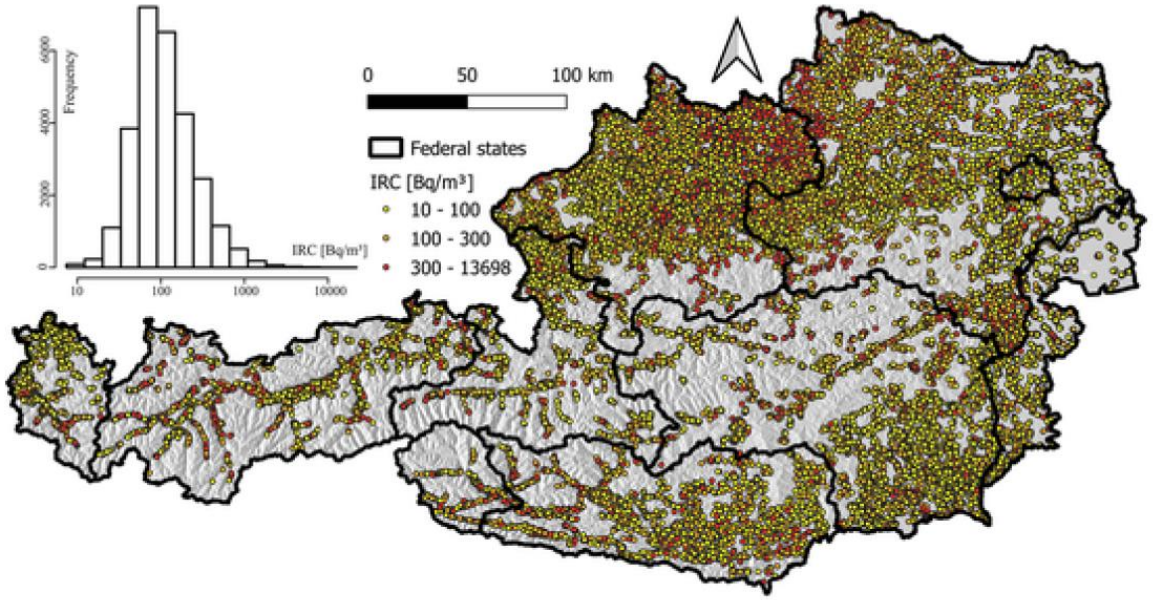
This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

WP 2.4.1 - outlook

- Evaluation and comparison of different methods on exhaustive IRC data sets:
  - Austria, France, Italy, Switzerland
  - ~ 200.000 measured buildings (IRC)
  - Additional Information as

    Building charcteristics, location
- To awnser:
  - Which methods and predictors to use

  for the identification of high radon levels



This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

WP 2.4.1 – additional research questions

- What are the predictors with the most impact regarding the identification of buildings and areas with high radon levels? – Does the impact change for the prediction of certain levels?
- If a model accuratly predicts high radon levels – is there a possible downside? (Sensitivity, Specificity)
- Trade of between weigthing the importance of high target values and the model depending on only a few samples?
- Is there a performance evaluation (loss function) that is better suited for the prediction of high values and can this be used for different methods (quantile regression)?

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

RadoNorm
Managing risks from radon and NORM

Statistical methods for the identification of buildings and areas with high radon levels – a review

Sebastian Baumann, Sara Antignani, Joan Frédéric Rey, Valeria Gruber, Christian di Carlo, Joëlle Goyette Pernot, Francesco Bochicchio

Thank you!

20.09.2023, GARRM, Prague

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.