RadoNorm
Managing risks from radon and NORM

# Tricks and tweaks for the global prediction of high radon levels and local pitfalls

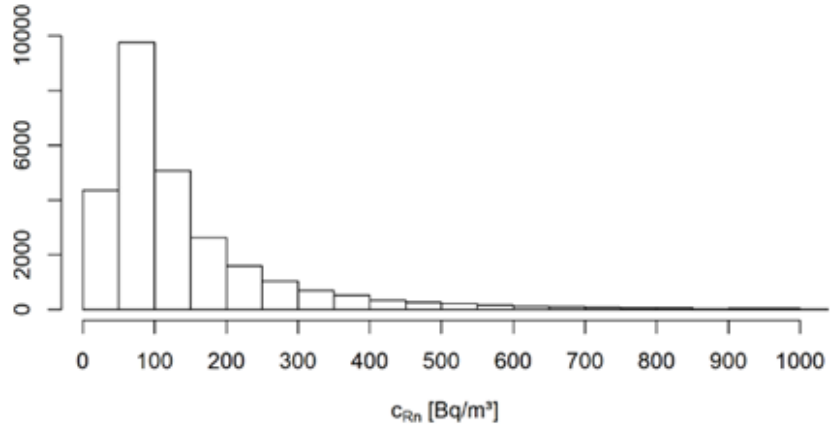
Sebastian Baumann, Sara Antignani, Joan Frédéric Rey,
Niccolo Loret, Mario Caprio, Valeria Gruber

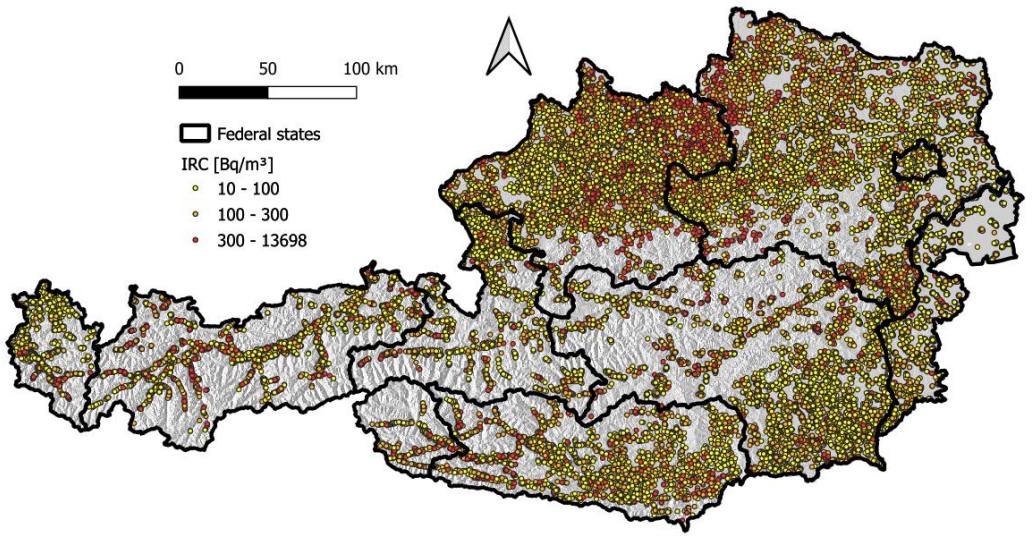
17.09.2025, GARRM, Prague

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

# Tricks and tweaks for the global prediction of high radon levels and local pitfalls

- Concepts
- Setup
- Performance
- Conclusion
- Suggestions




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

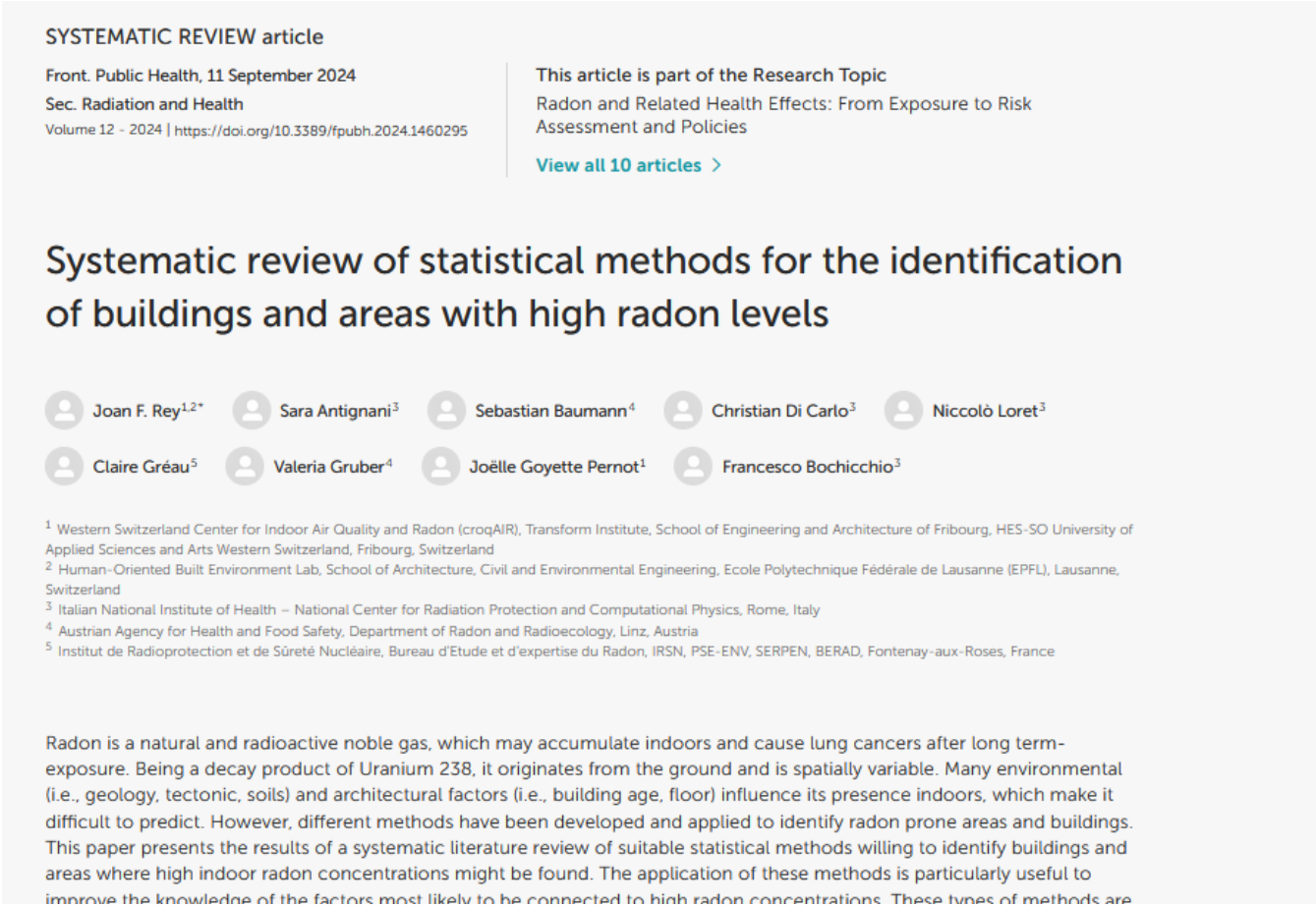
# Concept

- Radiation protection measures usually follow conservative approaches
- Consequences for the prediction of indoor radon concentrations?
- What are the aims of predicting IRC?
- Are high radon levels more relevant as low levels?
- Are there models and techinques better suitable to predict high IRC?

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

# Concept

- In Rey et. Al (2024) we suggested modifications of the target (IRC) in a predictive model to focus on the prediction of high radon levels.
- Today some of these model modifications and their performance are presented



SYSTEMATIC REVIEW article

Front. Public Health, 11 September 2024
Sec. Radiation and Health
Volume 12 - 2024 | https://doi.org/10.3389/fpubh.2024.1460295

This article is part of the Research Topic
Radon and Related Health Effects: From Exposure to Risk Assessment and Policies

View all 10 articles >

## Systematic review of statistical methods for the identification of buildings and areas with high radon levels

Joan F. Rey[1,2*] Sara Antignani[3] Sebastian Baumann[4] Christian Di Carlo[3] Niccolò Loret[3]
Claire Gréau[5] Valeria Gruber[4] Joëlle Goyette Pernot[1] Francesco Bochicchio[3]

[1] Western Switzerland Center for Indoor Air Quality and Radon (croqAIR), Transform Institute, School of Engineering and Architecture of Fribourg, HES-SO University of Applied Sciences and Arts Western Switzerland, Fribourg, Switzerland
[2] Human-Oriented Built Environment Lab, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[3] Italian National Institute of Health – National Center for Radiation Protection and Computational Physics, Rome, Italy
[4] Austrian Agency for Health and Food Safety, Department of Radon and Radioecology, Linz, Austria
[5] Institut de Radioprotection et de Sûreté Nucléaire, Bureau d'Etude et d'expertise du Radon, IRSN, PSE-ENV, SERPEN, BERAD, Fontenay-aux-Roses, France

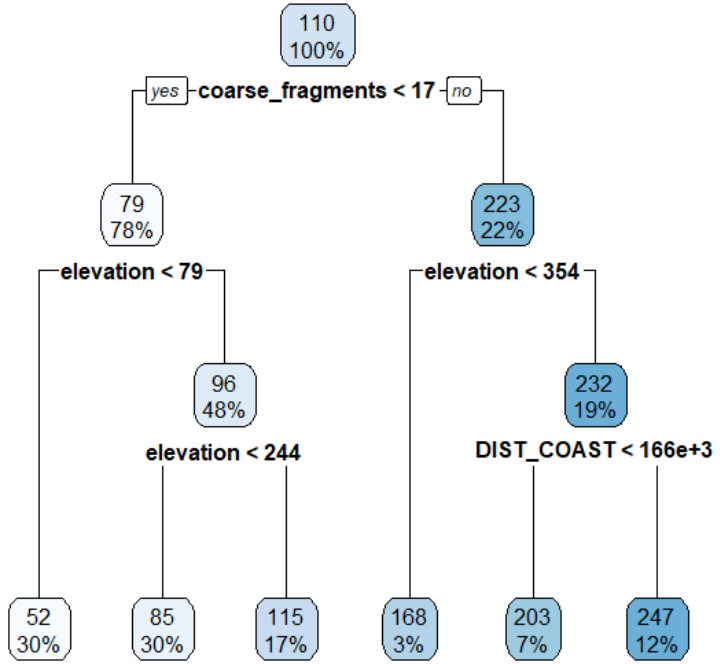
Radon is a natural and radioactive noble gas, which may accumulate indoors and cause lung cancers after long term-exposure. Being a decay product of Uranium 238, it originates from the ground and is spatially variable. Many environmental (i.e., geology, tectonic, soils) and architectural factors (i.e., building age, floor) influence its presence indoors, which make it difficult to predict. However, different methods have been developed and applied to identify radon prone areas and buildings. This paper presents the results of a systematic literature review of suitable statistical methods willing to identify buildings and areas where high indoor radon concentrations might be found. The application of these methods is particularly useful to improve the knowledge of the factors most likely to be connected to high radon concentrations. These types of methods are

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

# Setup

- Target Variable ~ Model (Predictors)

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




## Possible Goals:

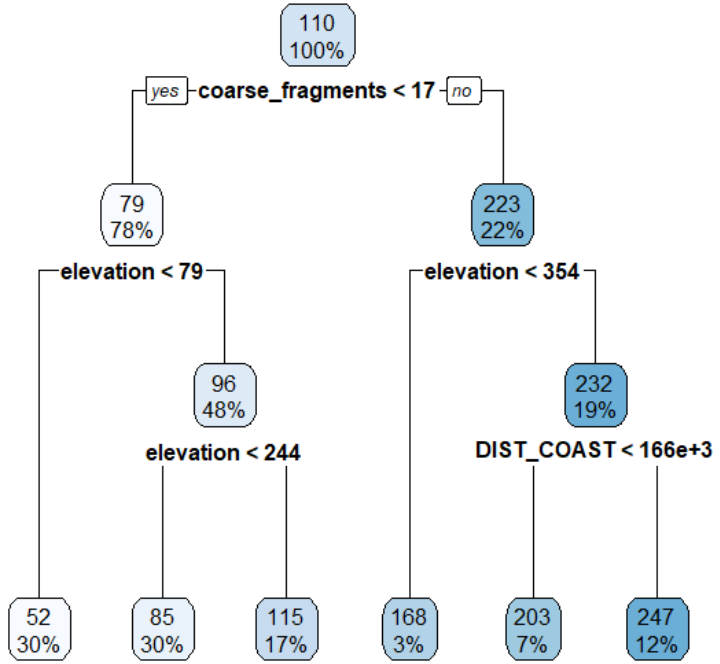
- Estimate the average IRC of a country
- Develop a radon map
- Estimate the amount of buildings with IRC greater 300 Bq/m³
- Evaluation of influencing factors

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

# Setup

- Target Variable ~ Model (Predictors)

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




## Model characteristics

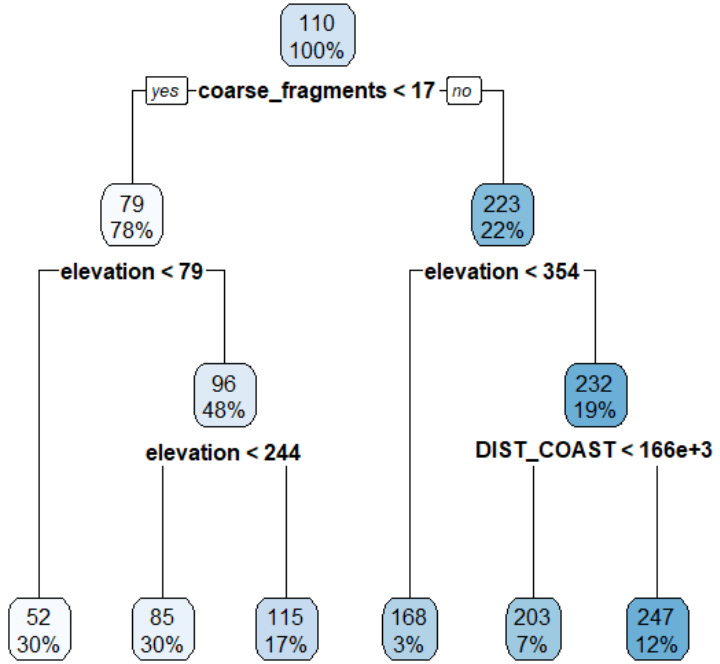
- ~ 50.000 indoor radon measurements
- Performance metrics on hold out random test set
- 20 Predictors from different categories: coordinates, floor level, geology, temperature, soil characteristics, building material …

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

# The model - Predicting high radon levels

- Target Variable ~ Model (Predictors)

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




## Modifications:

- No modification
- Log-transforming
- Quantile regression forest
- Classification instead of regression

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

# Performance – Example log-transformation

- No Modification of the target variable compared to log-transformation
- MAE = Mean absolute error of prediction and observation (the lower the better)

| | Non modification | Log-transformed |
|---|---|---|
| Overall MAE Bq/m³ | **90** | **79** |

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

# Performance – Example log-transformation

- No Modification of the target variable compared to log-transformation
- MAE = Mean absolute error of prediction and observation (the lower the better)

**Transforming the target value:**

- Usually, IRC follow a log-normal distribution and for prediction the IRC are often log-transformed

| | Non modification | Log-transformed |
|---|---|---|
| Overall MAE Bq/m³ | **90** | **79** |

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

# Performance – Example log-transformation

- No Modification of the target variable compared to log-transformation
- MAE = Mean absolute error of prediction and observation (the lower the better)

**Transforming the target value:**

- Usually, IRC follow a log-normal distribution and for prediction the IRC are often log-transformed
- Consequences: Typically, predictions on the log scale show overall better performance than non-transformed.

| | Non modification | Log-transformed |
|---|---|---|
| Overall MAE Bq/m³ | **90** | **79** |

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

# Performance – Example log-transformation

- No Modification of the target variable compared to log-transformation
- MAE = Mean absolute error of prediction and observation (the lower the better)

**Transforming the target value:**

- Usually, IRC follow a log-normal distribution and for prediction the IRC are often log-transformed
- Consequences: Typically, predictions on the log scale show overall better performance than non-transformed. But for higher concentrations (e.g. > 300 Bq/m³) the non-transformed show a higher performance.

| | Non modification | Log-transformed |
|---|---|---|
| Overall MAE Bq/m³ | **90** | **79** |
| MAE IRC > 300 Bq/m³ | **333** | **398** |

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

# Performance – Example log-transformation

- No Modification of the target variable compared to log-transformation
- MAE = Mean absolute error of prediction and observation (the lower the better)

**Transforming the target value:**

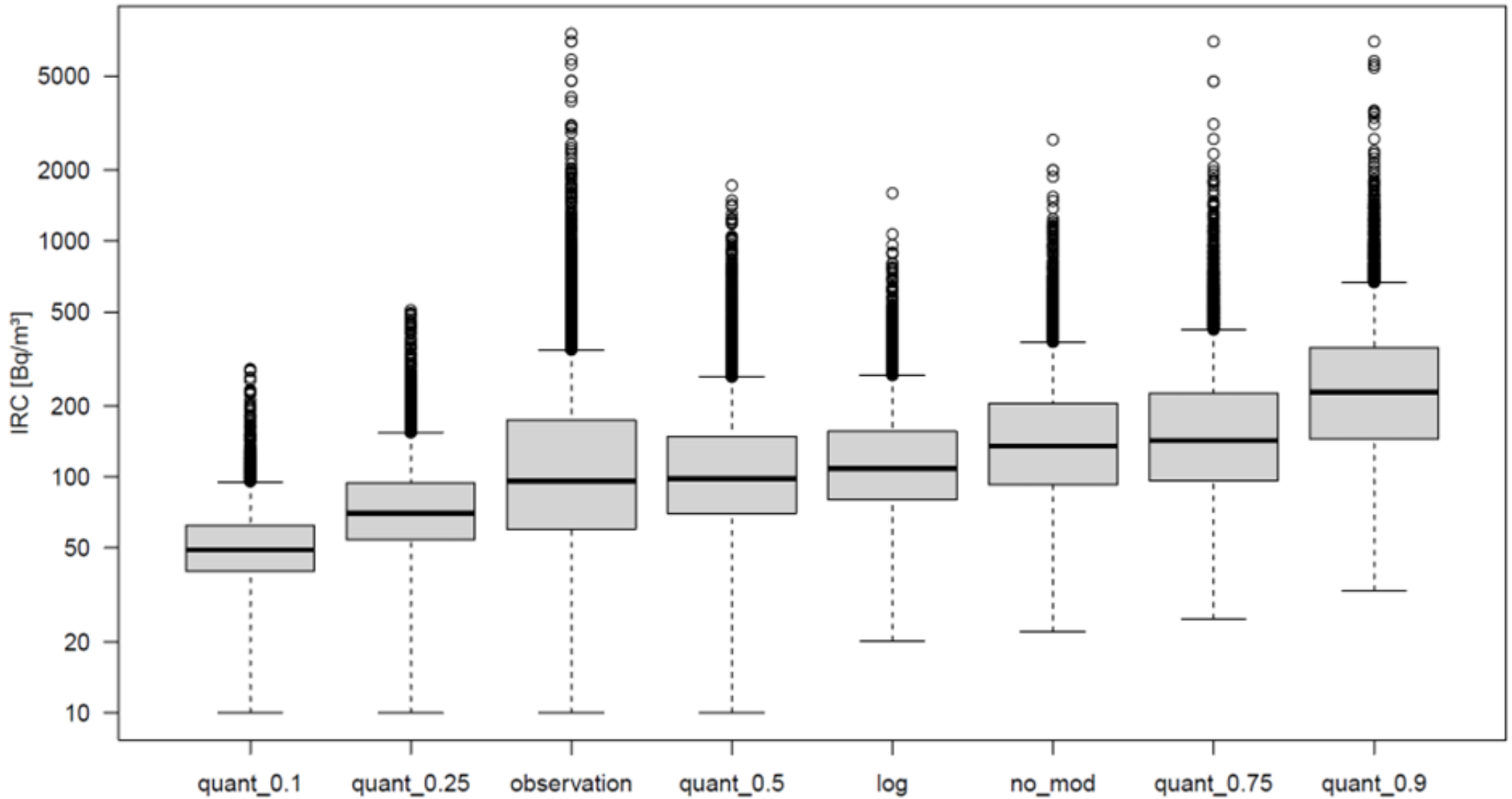
- Usually, IRC follow a log normal distribution and for prediction the IRC are often log-transformed
- Consequences: Typically, predictions on the log scale show overall better performance than non-transformed. But for higher concentrations (e.g. > 300 Bq/m³) the non-transformed show a higher performance.

| | Non modification | Log-transformed |
|---|---|---|
| Overall MAE Bq/m³ | 90 | 79 |
| MAE IRC >= 300 Bq/m³ | 333 | 398 |
| MAE IRC < 300 Bq/m³ | 58 | 37 |

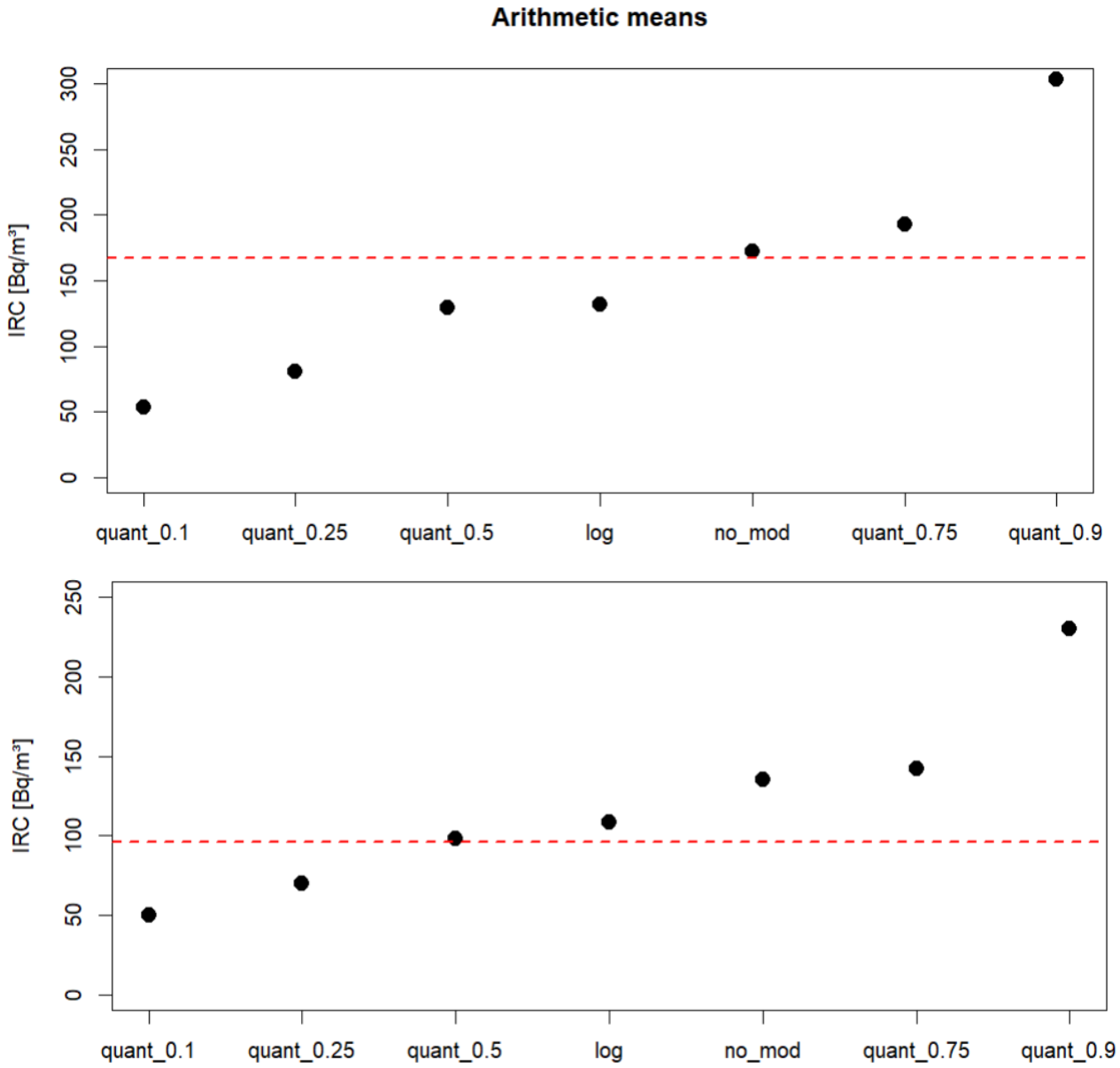
This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

RadoNorm
Managing risks from radon and NORM

# Performance - Distributions of predictions



# Performance – Aggregates

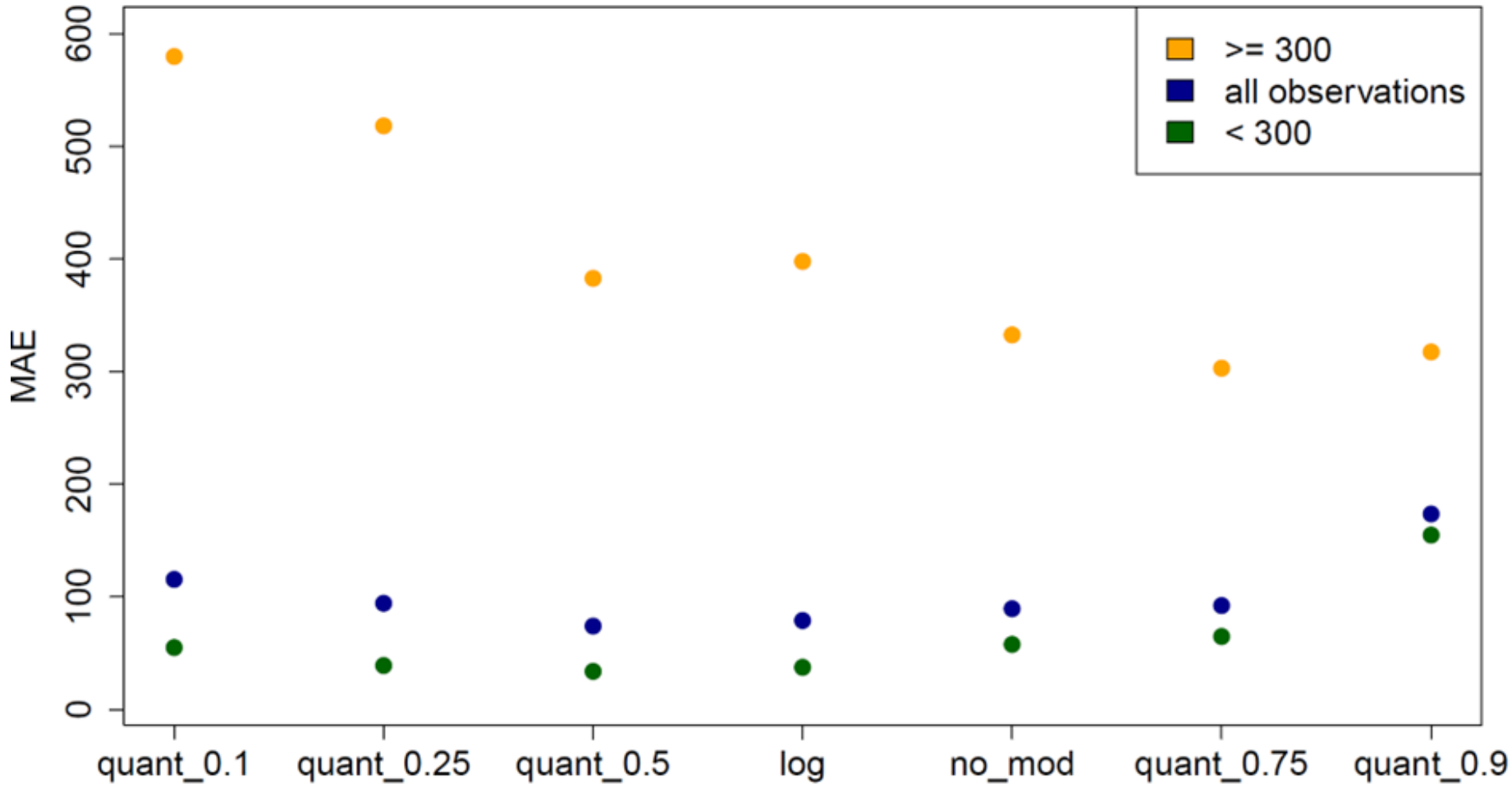



- Arithmetic means of the model predictions, compared to the observation
- Medians of the model predictions compared to the observation

ogramme 2019-2020 under grant agreement No 900009.

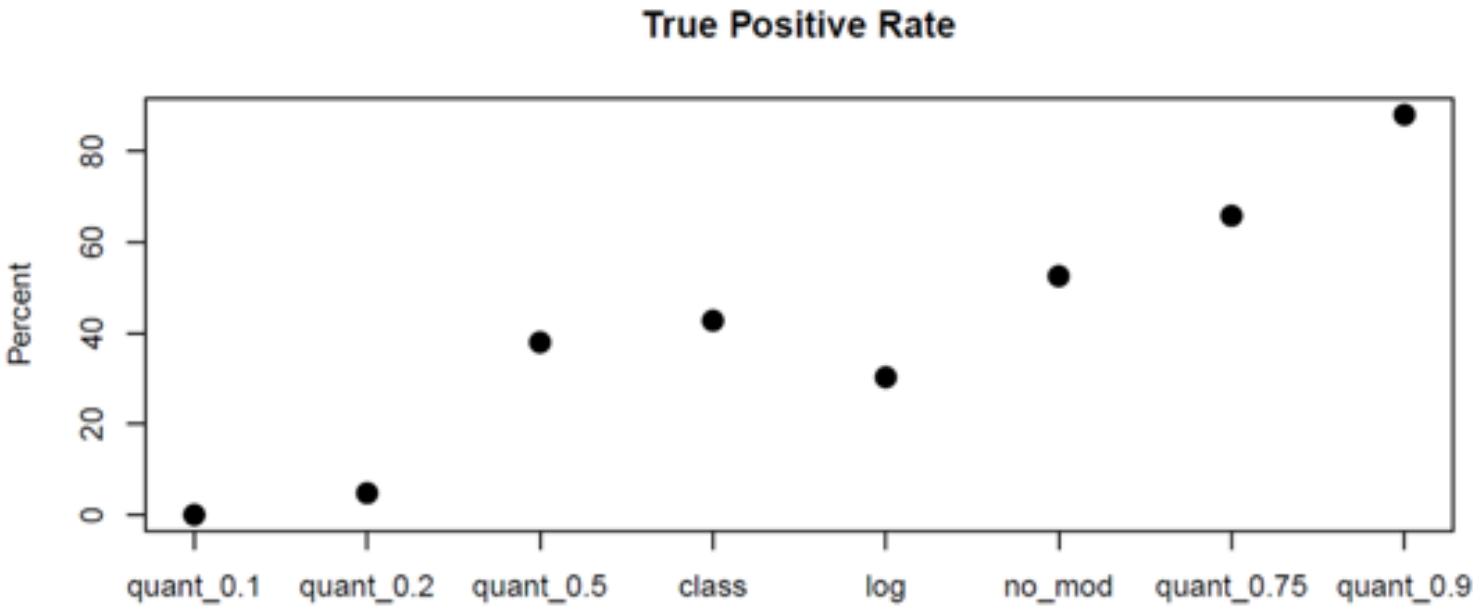
RadoNorm
Managing risks from radon and NORM

# Performance - MAE for different models



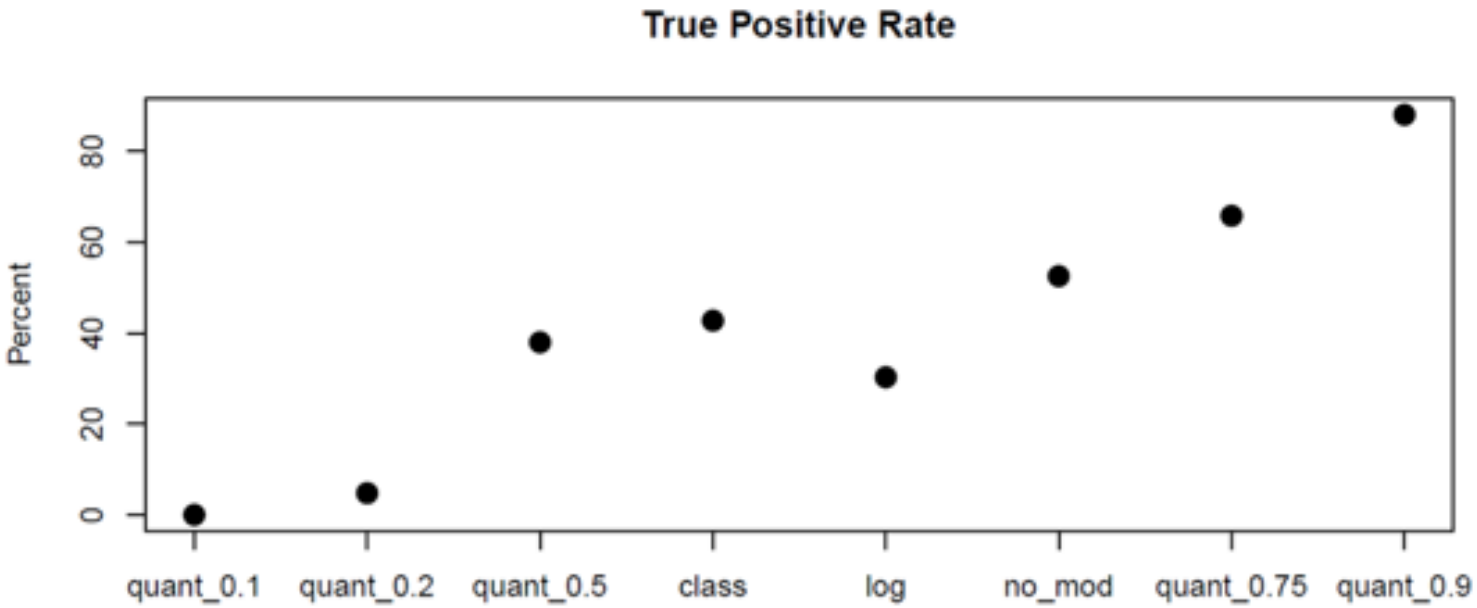

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

# Performance – TPR | TNR

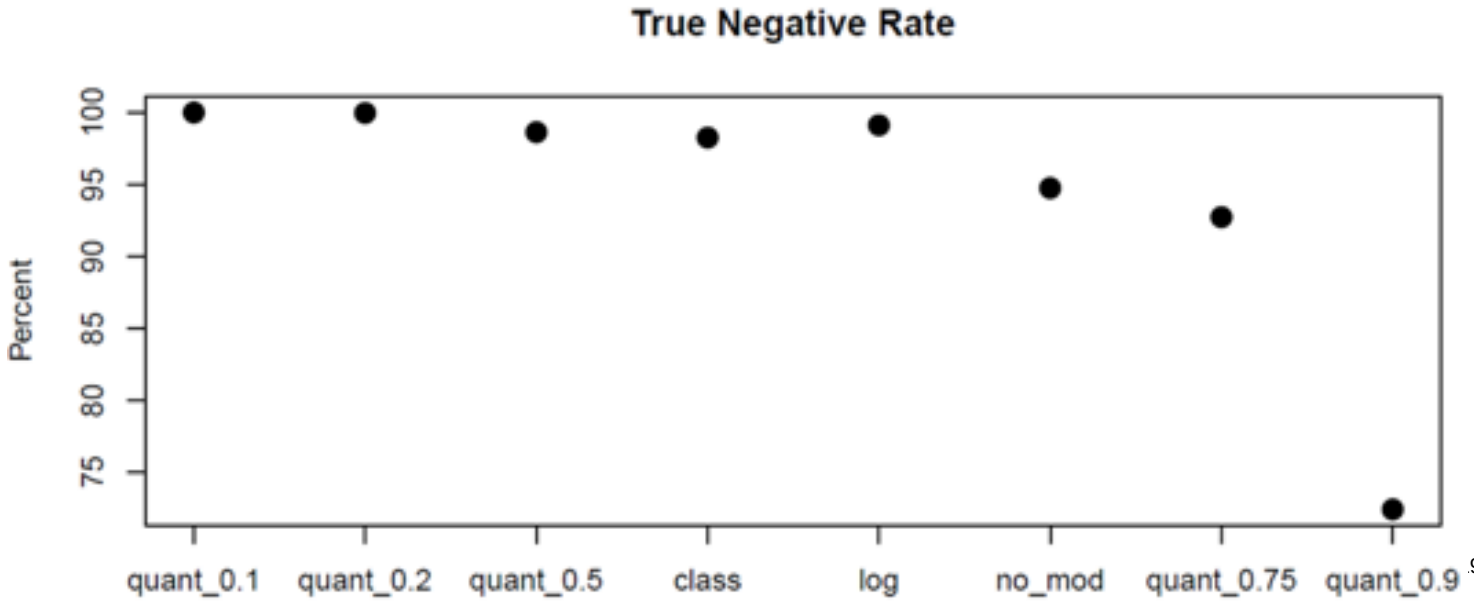



- TPR: how many observations >= 300 Bq/m³ were classified correctly [%]

# Performance – TPR | TNR




- TPR: how many observations >= 300 Bq/m³ were classified correctly [%]




- TNR: How many observations <300 Bq/m³ were classified correctly [%]

9-2020 under grant agreement No 900009.

# Conclusions

- Transforming the target (IRC) in a predictive model has consequences, especially if the target follwos a log-normal distribution.
- This is seldom dicussed when using predictive models
- There may be aims where the averall performace might be less relevant, than the performance for certain IRC – Populations (e.g. > 300 Bq/m³)
- What is the goal of my prediction?

| | Non modification | Log-transformed |
|---|---|---|
| Overall MAE Bq/m³ | 90 | 79 |
| MAE IRC >= 300 Bq/m³ | 333 | 398 |
| MAE IRC < 300 Bq/m³ | 58 | 37 |
| TPR % | 53 | 30 |
| TNR % | 94 | 99 |
| MEAN DIFF [Bq/m³] | + 6 | - 35 |
| MEDIAN DIFF [Bq/m³] | + 39 | + 12 |

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

# Suggestions

- There is no jack of all trades
- Use well working models (tree based) and focus on overall workflow:
  - DATA handling
  - Feature selection
  - Train / Test split (**Generalization!**)
  - Model evaluation
- Focus on the research question: What do I want to accomplish with my prediction? (Map, Mean, Distribution)

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

# Tricks and tweaks for the global prediction of high radon levels and local pitfalls

Sebastian Baumann, Sara Antignani, Joan Frédéric Rey,
Niccolo Loret, Mario Caprio, Valeria Gruber

**Thank you!**

17.09.2025, GARRM, Prague

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