Evaluating the Indoor Radon Concentrations in the Swedish Building Stock Using Statistical and Machine Learning

Wu, Pei-Yu; Johansson, Tim; Mangold, Mikael; Sandels, Claes; Mjörnell, Kristina

Published in:
13th Nordic Symposium on Building Physics (NSB-2023) 12/06/2023 - 14/06/2023 Aalborg, Denmark

DOI:
10.1088/1742-6596/2654/1/012086

2023

Document Version:
Publisher's PDF, also known as Version of record

Evaluating the indoor radon concentrations in the Swedish building stock using statistical and machine learning

Journal of Physics: Conference Series **2654** (2023) 012086 doi:10.1088/1742-6596/2654/1/012086

Evaluating the indoor radon concentrations in the Swedish building stock using statistical and machine learning

Pei-Yu Wu [1,2], Tim Johansson [1], Mikael Mangold [1], Claes Sandels [1], Kristina Mjörnell [1,2]
[1] RISE Research Institutes of Sweden, 412 58 Gothenburg, Sweden
[2] Department of Building and Environmental Technology, Faculty of Engineering, Lund University, 221 00 Lund, Sweden

Email: pei-yu.wu@ri.se

Abstract. Exposure to excessive indoor radon causes around 500 lung cancer deaths in Sweden annually. However, until 2020, indoor radon measurements were only conducted in around 16% of Swedish single-family houses and 17% of multifamily houses. It is estimated that approximately 16% of single-family houses exceed the indoor radon reference level of 200 Bq/m^3, and the corresponding situation in multifamily houses is unknown. Measuring indoor radon on an urban scale is complicated and costly. Statistical and machine learning, exploiting historical data for pattern identification, provides alternative approaches for assessing indoor radon risk in existing dwellings. By training MARS (Multivariate Adaptive Regression Splines) and Random Forest (RF) regression models with the data labels from the radon measurement records in the Swedish Energy Performance Certification registers, property registers, soil maps, and the radiometric grids, the correlations between response and predictive variables can be untangled. The interplay of the key features, including uranium and thorium concentrations, ventilation systems, construction year, basements, and the number of floors, and their impact magnitudes on indoor radon concentrations, are investigated in the study. The regression models tailored for different building classes were developed and evaluated. Despite the data complexity, the RF models can explain 28% of the variance in multifamily houses, 24% in all buildings, and 21% in single-family houses. To improve model fitting, more intricate supervised learning algorithms should be explored in the future. The study outcomes can contribute to prioritizing remediation measures for building stocks suspected of high indoor radon risk.

1. Introduction

Indoor radon has been regarded as the second leading cause of lung cancer worldwide; in particular, inhalation of radon gas and its progeny is fatal to cause damage to the human genome [1,2]. Radon-222 from uranium-238 and its radiative progenies attaching to dust particles and being inhaled into the lungs [3] is estimated to cause around 500 deaths annually in Sweden [4]. The exposure to high radon doses is especially severe in Nordic countries because of airtight energy conservation measures [2]. To monitor the health and safety risk of indoor radon and encourage affordable remediation measures, most countries adopt three radon concentration intervals: (i) 200 Bq/m^3 for residential and public buildings and as the highest acceptable level for new buildings, (ii) 400 Bq/m^3 for existing buildings, (iii) 1000 Bq/m^3 for obligatory decontamination. The baseline for indoor radon concentration was introduced in 1981 in Sweden and subsequently adjusted to 200 Bq/m^3. According to the Act of Environmental Goals, the radon levels in schools, preschools, and residential buildings should be lower than 200 Bq/m^3 [5].

The primary sources of indoor radon gas are ground, groundwater, and building materials, where 84-91% of indoor radon attributes to the ground [6]. Measuring radon gas and investigating radon sources in buildings are advised considering the atmospheric pressure synergies between ground radon

and radon from building materials. However, remediating indoor radon is challenging due to the complicated interplay between radon sources, geological conditions, and unique characteristics of individual buildings. The dynamic changes in radon gas concentrations depend on uranium levels in the ground, soil types, building typologies, construction materials, foundations, and ventilation types [3,4,7]. According to the latest report (2021:28) by the Swedish Radiation Safety Authority (SSM) [4], the surface uranium concentration and the soil types correlate to indoor radon levels positively. Olsthoorn et al. [13] verified the proposition by mapping the gamma radiation and household radon measurements with postal codes. Their findings further show that newly built houses have a relatively lower correlation to indoor radon than older buildings, which can be explained by the effectiveness of the ground radon protection techniques, such as radon vacuum or radon well, air diffuser or air cushion, and shield layer. Indoor radon levels are suspected to be affected by soil types, where buildings situated on clay are prone to a higher radon risk [4]. However, the uncertainties are high due to low sample numbers in the previous study, and more soil data on the national scale are needed to confirm the degree of correlation.

Other building-related factors can also affect indoor radon concentrations. Swedjermark [2] found that the construction year is closely associated with high radon levels in the Swedish building stock, with the highest average levels between the 1940s and mid-1970s. The upward trend reversed after 1980 when the production of radioactive concrete ceased, and the national reference level of indoor radon was introduced [8]. Likewise, the empirical results from the ELIB survey suggested that buildings built before 1980 with slab-on-ground foundations and basements tended to have higher radon levels [7]. Buildings built on a concrete slab appear more airtight than buildings with crawl space or load-bearing foundations [3]. On the other hand, ventilation systems facilitating air exchange with outdoor air can dilute indoor radon [9]. Natural or exhaust ventilation systems may create negative air pressure indoors and have an adverse impact on radon concentration by increasing the leakage of ground radon from the foundations or basements [7]. Yet balanced ventilation systems can lower radon concentration as indoor air is exchanged with outdoor air without negative pressure [10]. Single-family houses built before 1980 and multifamily houses built before 1940 in Sweden tend to be equipped with a natural ventilation system, making them more vulnerable to ground radon and ineffective in reducing radon from building materials. Therefore, the choice of ventilation systems as a remediation means for indoor radon depends on radon sources and building typologies and has to assess case by case [3]. Despite the complexity, several radon remediation measures are effective, including ground foundation sealing, enhancement of ventilation, and removal of radioactive building materials [3,9]. The empirical study has shown that reducing indoor radon concentration by 40%-85% is possible if remediation measures are implemented appropriately.

2. Scope of the paper

The study explores the possibility of predicting indoor radon levels by comparing statistical and machine learning techniques for residential and non-residential dwellings. The study adopts data analytics and regression to achieve the two objectives: (i) describe correlations between predictors and the response variable, and (ii) train and evaluate statistical and machine learning models for radon level prediction. Descriptive and predictive analyses for existing buildings were performed by coupling the indoor radon measurements, building registers, and the geological factors related to radioactive substances and soils. Exploiting data-driven approaches can overcome the contextual limitations of simulation case studies for specific buildings and increase model generalizability to the generic building stock. It could also enhance the knowledge of how and to what extent anthropogenic and geogenic factors interact and evaluate buildings prone to high radon risk.

3. Materials and methods

The radon dataset used in the study integrates data from the Swedish Energy Performance Certificates (EPCs), property registers, and geophysical aerial measurements of gamma radiation for uranium, potassium, thorium, and soil types. The collected raw datasets were merged into the radon dataset using Feature Manipulation Engine (FME), a spatial extract, transform, and load tool. Data pre-processing was performed, including data cleaning of duplicated values from aerial geophysical measurements of the radioactive substances, data aggregation of the average substance concentration for each building based on the building footprint maps, mapping the soil type for corresponding buildings in multiclass

classification, and removal of invalid radon measurements conducted less than two months or beyond the heating seasons. The extreme values outside the 95% confidence interval (CI), buildings containing radioactive concrete, and missing values of the yearly average radon levels were eliminated. The remaining dataset contains 156,072 properties built between 1930-2020 from 290 municipalities. This value distribution of the annual average radon concentration aligns with the national average from the previous BETSI and the SSM studies. Further, data were stratified based on the building class – single-family houses, multifamily houses, schools, and other buildings.

Subsequently, data analytics and visualization were carried out to understand the underlying data structure and correlations. The relationships between the radon concentrations and the other independent variables, including geological and geographical factors, building usage, and building parameters, were estimated. Using the interquartile range rule on measured values, such as annual average radon level, weighted average concentrations of uranium, thorium, and potassium, the outliers in each data subset were identified and removed from modeling. Similar criteria for eliminating extreme values were applied to the other four subsets. The data show the non-Gaussian distribution and non-linear patterns; thus, the statistical learning approach, i.e., Multivariate Adaptive Regression Splines (MARS), and the machine learning method, i.e., Random Forest (RF), were chosen for model comparison. Both models are non-parametric, generalized regression tree models that generate prediction results by hierarchically and successively pruning predictive variables and removing features until the optimal performance is reached in cross-validation [11]. However, the MARS algorithm features piecewise linear or cubic splines with hinge functions to improve data fitting [12]; while the random forest algorithm exploits the ensemble of decision trees trained with the bagging method. Derived variables such as area, stairwell, and apartments per floor were created. Then the data was partitioned into 70% for training and testing and 30% for validation. The ten best features were identified based on high F-scores for respective data groups for model training and comprehensive regression metrics, i.e., Mean-Absolute-Error (MAE), Mean-Squared-Error (MSE), Root-Mean-Squared-Error (RMSE), coefficient determination (R^2) was applied to evaluate models' performance. Lastly, residual plots illustrating the difference between the actual and the predicted value were created to ascertain the models' uncertainty.

4. Results

4.1 Data analytics and visualization

The clean radon dataset for the Swedish building stock with 123,000 observations consists of 49% single-family houses, 41% multifamily houses, 5% other buildings, and 5% school buildings. This distribution is representative given approximately a comparable proportion of each building class in Sweden - 93% residential dwellings, 7% non-residential dwellings - based on the data from Statistics Sweden in 2021 [13]. Around 23-28% of missing values were detected in variables including basements, number of floors, stairwells, and apartments. Figure 1 below shows the shares of buildings in radon levels grouped by key variables, including building classes, construction periods, ventilation types, and potassium, uranium, and thorium concentrations. The last row includes all observations as a baseline for the clustering comparison with other subgroups of the same category. The results show that around 12-13% of buildings exceed the highest accepted radon level of 200 Bq/m^3. Single-family houses are more likely to be exposed to high radon levels than other building classes. Yet, the exceptionally high radon levels (above 500 Bq/m^3) were measured more frequently in non-residential dwellings. Buildings built between the 40s, 50s, and 60s are measured with high radon concentrations, with proportions of 15%, 21%, and 19% of buildings above the reference level. Besides, 16% of buildings equipped with natural ventilation are exposed to radon above 200 Bq/m^3, whereas the percentage for buildings equipped with balanced ventilation is less than 8%. The radon pattern in buildings installed with exhaust ventilation aligns approximately with the baseline of the radon level in buildings regardless of ventilation type. Regarding the impact of the geological factors, the ground radioactive substances correlate to indoor radon positively, where the extreme concentrations account for the highest radon levels. Compared to potassium and thorium, the association between uranium and radon level is substantial, with evident increment across various levels. The baseline radon values in buildings correspond to the potassium concentrations of 2-3%, the uranium concentrations of 3-3.5 ppm, and the thorium concentrations of 6-10 ppm.

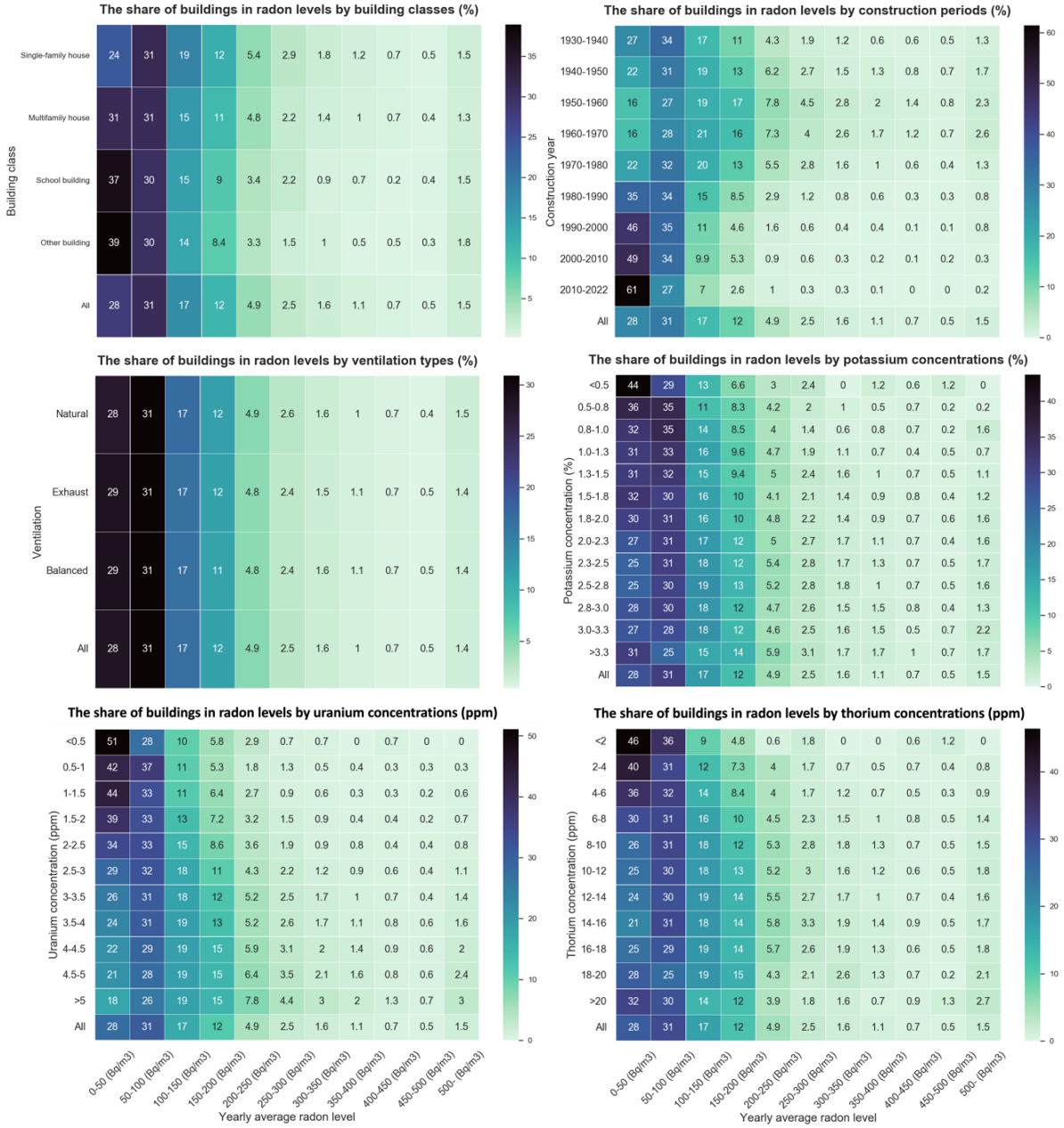

Figure 1. The shares of buildings across radon levels by building classes, construction periods, ventilation types, potassium, uranium, and thorium concentrations.

Further plotting the distribution of measured radon along the construction year and the radioactive substance concentrations by building class, basements, and soil types, as presented in Figure 2. The uneven distribution indicates that the relationship between response and predictive variables is highly complicated and non-linear. Thus, measured values were transformed into square root, and non-parametric algorithms were chosen for modeling.

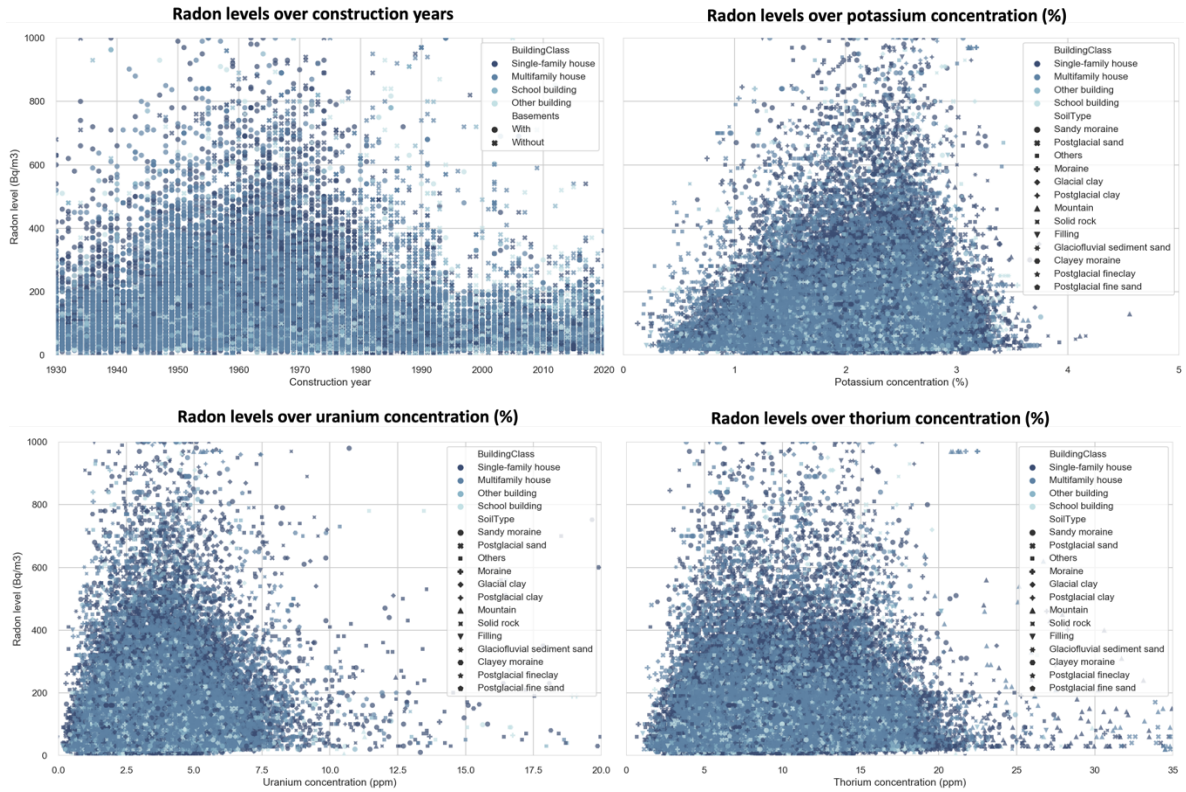

Figure 2. Measured radon distribution by construction year and radioactive substance concentrations.

4.2. Statistical and machine learning modeling

The confidence interval and variable correlation to radon of each building class subset were examined and presented in Table 1. The mean annual average radon concentration lies at 113 Bq/m^3. Among others, single-family houses have the highest mean radon level of 118 Bq/m^3, while school buildings have the least mean level of 97 Bq/m^3. The mean levels are similar between multifamily houses and other buildings, but the latter has a much higher standard deviation. The Pearson biserial correlation shows that uranium concentration is the most considerable factor to indoor radon, followed by natural ventilation, basements, and thorium concentration. On the other hand, construction year, balanced ventilation with heat exchange, and the number of floors have counter effects. The data on the number of stairwells and apartments are too few in schools and other buildings, and no coefficients were generated. Several soil types are found to be correlated to indoor radon, but their impact magnitude is not as profound as radioactive substances and building parameters. In general, the coefficients are significant and coherent across building types.

Table 1. Summary of the Pearson correlation between response and predictive variables.

Category	Predictive variable	Single-family house	Multifamily house	School building	Other building	All buildings
Radon mean confidence interval		118 ± 1	105 ± 1	97 ± 3	105 ± 20	111 ± 1
		Square root indoor radon level [Bq/m^3]				
Building parameter	Construction year	-0.28***	-0.22***	-0.14***	-0.16***	-0.23***
	Floor area [m^2]	0.03***	0.05***	0.10***	0.04**	-0.00
	Basements	0.22***	0.11***	0.13***	0.12***	0.16***
	Number of floors	-0.12***	-0.02***	0.11***	0.03*	-0.05***
	Number of stairwells	-0.02**	0.13***	N/A	N/A	0.06***
	Number of apartments	-0.03***	0.06***	N/A	N/A	0.02***
	Natural ventilation	0.18***	0.11***	0.04**	0.11***	0.17***
	Exhaust ventilation	-0.06***	0.13***	0.01	0.05***	0.04***
	Exhaust (heat pump)	-0.17***	-0.05***	-0.04**	-0.02	-0.09***
	Balanced ventilation	-0.02***	-0.00	0.03*	0.03*	-0.02***

	Balanced (heat exch.)	-0.12***	-0.21***	0.01	-0.08***	-0.20***
Radioactive substance	Sqrt. potassium [%]	0.09***	0.06***	0.05***	0.04***	0.07***
	Sqrt. uranium [ppm]	0.22***	0.23***	0.15***	0.17***	0.22***
	Sqrt. thorium [ppm]	0.15***	0.13***	0.10***	0.13***	0.13***
Soil type†	Sandy moraine	0.02***	0.06***	0.04**	0.05***	0.05***
	Others	-0.01**	-0.06***	-0.02	0.00	-0.04***
	Glacial clay	0.03***	0.03***	0.02	0.01	0.03***
	Moraine	-0.02***	-0.04***	0.12	0.02	-0.02***
	Postglacial clay	0.03***	0.05***	-0.06***	-0.00	0.03***
	Postglacial sand	-0.04***	-0.05***	-0.06***	-0.08***	-0.05***
	Filling	-0.02**	-0.03***	0.02	-0.02	-0.04***
	Glaciofluvial sed. sand	-0.02***	-0.03***	0.01	-0.04**	-0.02***
	Postglacial fine sand	-0.02***	-0.00	-0.03*	-0.04**	-0.01***
	Postglacial fine clay	0.01	0.00	-0.01	0.01	0.01**
	Clayey moraine	0.01	-0.01**	-0.03	-0.00	-0.00*
	Mountain	0.00	0.02***	0.03*	0.01	0.01***
Statistics	Count (N)	27,838	45,495	4,745	5,395	79,944

P-Value (The level of marginal significance within a hypothesis test): * p<.1, ** p<.05, ***p<.01

After numeral model training and testing iterations, the best-performed MARS and Random Forest regressions for each subset were identified using input features with the highest F-scores, described in Table 2. The fitted MARS models were described as the equation below (h stands for hinge). The finding shows that key features in both models are nearly identical and consistent with the coefficient's direction from the previous Pearson correlation. Next, the fitted models were applied to the validation subsets to evaluate their prediction performance with regression metrics. The results from Table 3 show that random forest models outperform MARS models in predicting radon levels in all buildings and residential buildings. However, MARS models have higher prediction power over schools and other buildings. The errors measured in the training subset are similar to those in the validation subset, indicating no overfitting or underfitting. The highest prediction performance is found in multifamily houses with an average R^2 of 0.275, followed by all buildings ($R^2 = 0.24$) and single-family houses ($R^2 = 0.21$). For school buildings, MAE, MSE, and RMSE are not specifically high compared to the others, yet their R^2 is much lower. Other buildings show the least predictable in both MARS and Random Forest regressions.

Table 2. Equations of MARS models and the selected features of the RF models are sorted by F-scores.

Subset	**Model**	**Equations and features**
Single-family house	MARS	Y = 77.90 - 0.04*Construction year - 3.14* h (Sqrt. uranium - 3.18) + 0.72* h (5 - NumFloors) + 0.55*Basements - 0.99*Balanced (heat exch.) - 1.13*Exhaust (heat pump) + 0.39* Natural ventilation + 2.84* h (Sqrt. uranium - 1.2)
	RF	Construction year, Basements, Sqrt. uranium, Natural ventilation, Exhaust (heat pump), Sqrt. thorium, AreaperFloor, NumFloors, Balanced (heat exch.), Sqrt. potassium
Multifamily house	MARS	Y = 57.68 - 20.14* h (Sqrt. uranium - 3.47) - 0.02*Construction year - 1.27* Balanced (heat exch.) + 4.36* h (Sqrt. potassium - 1.67) + 1.86* h (1.67 - Sqrt. potassium) + 0.61*Sandy moraine + 0.02* h (AreaperApartment – 137.31) - 0.02* h (AreaperApartment - 82) + 19.19* h (Sqrt. uranium - 3.07) - 2.59* h (3.07 - Sqrt. uranium) - 0.55*Exhaust (heat pump) + 1.32*Mountain
	RF	Construction year, Sqrt. uranium, Balanced (heat exch.), Exhaust ventilation, NumStairwells, Sqrt. thorium, Basements, Natural ventilation, AreaperFloor, Others
School building	MARS	Y = 63.20 - 0.01* h(155-AreaperStairwell) + 1.35* Sqrt. uranium - 0.03*Construction year - 1.86*Postglacial fine sand
	RF	Sqrt. uranium, Basements, NumStairwells, Construction year, NumFloors, AreaperStairwell, Area, Postglacial fine sand, AreaperFloor, Sqrt. thorium
Other building	MARS	Y = 61.34 – 1.71* h(2.77- Sqrt. uranium) – 0.03*Construction year - 0.73*Natural ventilation – 0.65*Balanced ventilation
	RF	Sqrt. uranium, Construction year, Sqrt. thorium, Basements, Natural ventilation, Balanced (heat exch.)

All buildings	MARS	Y = 51.2 - 0.02*Construction Year + 1.12* h (Sqrt. uranium - 1.95) - 2.23* h (1.95 - Sqrt. uranium) - 1.08*Balanced (heat exch.) - 0.15*NumFloors + 3.98* h (Sqrt. potassium - 1.66) + 1.03* h (1.66 - Sqrt. potassium) - 0.57*Exhaust ventilation + 0.28*Basements + 0.26*Sandy moraine
	RF	Construction year, Sqrt. uranium, Balanced (heat exch.), Natural ventilation, Basements, Sqrt. thorium, Exhaust (heat pump), Sqrt. potassium, NumStairwells, Sandy moraine

Table 3. Performance evaluation between the MARS and the Random Forest regression models.

		Training subset				Validation subset			
Subset	**Model**	**MAE**	**MSE**	**RMSE**	**R2**	**MAE**	**MSE**	**RMSE**	**R2**
Single-family	MARS	2.94	21.13	4.60	0.12	2.91	16.49	4.06	0.15
house	RF	2.20	7.55	2.75	0.21	2.22	7.65	2.77	0.21
Multifamily	MARS	2.81	13.90	3.73	0.14	2.88	14.77	3.84	0.12
house	RF	2.05	6.76	2.60	0.27	2.08	6.89	2.63	0.28
School building	MARS	2.77	14.36	3.79	0.08	2.81	17.24	4.15	0.08
	RF	2.25	7.59	2.75	0.05	2.24	7.67	2.77	0.07
Other building	MARS	3.04	21.45	4.63	0.03	3.09	66.50	8.15	0.02
	RF	2.80	7.87	2.81	0.01	2.20	7.50	2.74	0.02
All buildings	MARS	2.39	8.39	2.90	0.13	2.39	8.40	2.90	0.13
	RF	2.16	7.31	2.70	0.24	2.17	7.38	2.72	0.24

Figure 3 below displays the residual plots to quantify the models' uncertainty in terms of standardized residuals between the actual and the predicted values for all buildings. From the residuals and fitted plots, it is recognized that the residuals of the RF models are much lower than the MARS model, which lies within the boundary of ± 10. Yet, the clear trend of asymmetric distribution shows room for model improvement. On the contrary, the residuals in the MARS model have more even distribution along the X-axis, but the lower boundary is large, and outliers are detected. The normal Q-Q plots in the MARS model imply that the training and testing regressions align closely with potential data skew, while the RF model does not experience the same problem but needs to improve model fit.

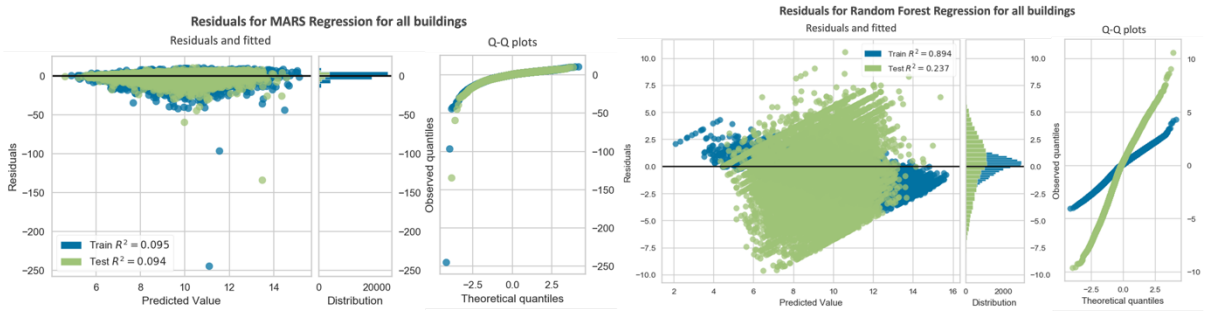

Figure 3. Residual plots for MARS regression (left) and Random Forest regression (right).

5. Discussions

The results from the data analytics and visualization aligned with the findings in the literature [4,9,14]. The yearly average radon in single-family houses lies at 118 Bq/m^{3,} with 13% of buildings exceeding 200 Bq/m^3, slightly lower than the previous estimation (128 Bq/m^3 and 16%). No detailed radon statistics specific to multifamily houses, schools, and other buildings were found; thus, the study compiles the pilot results for future cross-validation. Uranium concentration, construction year, ventilation types, basements, and the number of floors are recognized as the most pronounced indicators for radon risk screening, which are in good agreement with previous studies [8,15,16]. Despite the improved understanding of feature importance and intertwined effects, creating regression models in predicting indoor radon levels for the heterogenous building stock is extremely challenging. The effort in estimating the radon levels for individual buildings was explored using statistical and machine learning. Compared to other predictive radon mapping studies using kernel estimation [15] and RF [16] that explain 28-33% variance, the RF models in the study, with 28% explanation, perform slightly less well in predicting radon for individual buildings. On the other hand, the developed model is less complex

than the comprehensive model developed by Kropat et al. [16], making it more adaptable to other countries or regions. A higher dimensional dataset that includes geographical and atmospheric factors and more intricate models like deep neural networks should be considered in future research for indoor radon prediction.

6. Conclusions

The study assesses the national average indoor radon level affected by potential contributing factors and delineates the fraction of buildings exceeding the reference level. Uranium and thorium concentrations, construction year, ventilation types, basements, and the number of floors, are identified as closely associated with indoor radon in every building class. Random Forest regression models perform better than MARS models in predicting indoor radon in multifamily houses with the 28% variance explanation, while 24% for all buildings and 21% for single-family houses.

Acknowledgments

The study is funded by the EU BuiltHub project (grant agreement ID: 957026) and the scholarship from Maj och Hilding Brosenius Forskningsstiftelse, 2021.

References

[1] Stanley F K T, Irvine J L, Jacques W R, Salgia S R, Innes D G, Winquist B D, Torr D, Brenner D R and Goodarzi A A 2019 Radon exposure is rising steadily within the modern North American residential environment, and is increasingly uniform across seasons *Sci. Rep.* **9**

[2] Swedjemark G A 2002 Residential Radon Case 4 in the Swedish ICRP-project, SWIP

[3] Clavensjö B and Åkerblom G 2020 *Radonboken. Befintliga byggnader* (Stockholm, Sweden: Svensk byggtjänst)

[4] Rönnqvist T 2021 *Analysis of Radon Levels in Swedish Dwellings and Workplaces*

[5] Swedish National Board of Housing Building and Planning 2010 *Technical status in Swedish buildings - results from the BETSI project (Teknisk status i den svenska bebyggelsen - resultat från projektet BETSI)*

[6] Swedish National Board of Housing Building and Planning 2010 *Radon in the indoor environment (Radon i inomhusmiljö)*

[7] Sedin D and Hjelte I 2004 The Radon Situation in Sweden 3–5

[8] Khan S M, Pearson D D, Rönnqvist T, Nielsen M E, Taron J M and Goodarzi A A 2021 Rising Canadian and falling Swedish radon gas exposure as a consequence of 20th to 21st century residential build practices *Sci. Rep.* **11** 1–15

[9] Akbari K, Mahmoudi J and Ghanbari M 2013 Influence of indoor air conditions on radon concentration in a detached house *J. Environ. Radioact.* **116** 166–73

[10] Akbari K and Oman R 2013 Impacts of heat recovery ventilators on energy savings and indoor radon in a Swedish detached house *WSEAS Trans. Environ. Dev.* **9** 24–34

[11] Nisbet R, Miner G and Yale K P 2018 *Handbook of statistical analysis and data mining applications* (Academic Press)

[12] Kartal Koc E and Bozdogan H 2015 Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function *Mach. Learn.* **101** 35–58

[13] Statistics Sweden 2021 Number of dwellings by region and type of building (including special housing). Year 2013 - 2021 *Stat. Sweden*

[14] Olsthoorn B, Rönnqvist T, Lau C, Rajasekaran S, Persson T, Månsson M and Balatsky A V. 2022 Indoor radon exposure and its correlation with the radiometric map of uranium in Sweden *Sci. Total Environ.* **811**

[15] Kropat G, Bochud F, Jaboyedoff M, Laedermann J P, Murith C, Palacios Gruson M and Baechler S 2015 Predictive analysis and mapping of indoor radon concentrations in a complex environment using kernel estimation: An application to Switzerland *Sci. Total Environ.* **505** 137–48

[16] Kropat G, Bochud F, Jaboyedoff M, Laedermann J P, Murith C, Palacios M and Baechler S 2015 Improved predictive mapping of indoor radon concentrations using ensemble regression trees based on automatic clustering of geological units *J. Environ. Radioact.* **147** 51–62