

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/316923559>

Logistic regression model for detecting radon prone areas in Ireland

Article in *The Science of The Total Environment* · December 2017

DOI: 10.1016/j.scitotenv.2017.05.071

CITATIONS

45

READS

635

5 authors, including:



Javier Elío

Høgskulen på Vestlandet

58 PUBLICATIONS 614 CITATIONS

[SEE PROFILE](#)



Quentin Crowley

Trinity College Dublin

169 PUBLICATIONS 4,222 CITATIONS

[SEE PROFILE](#)



Stephanie Long

Science Foundation Ireland

45 PUBLICATIONS 531 CITATIONS

[SEE PROFILE](#)

Logistic regression model for detecting radon prone areas in Ireland

Elío J.⁽¹⁾, Crowley Q.⁽¹⁾, Scanlon R.⁽²⁾, Hodgson J.⁽²⁾, Long S.⁽³⁾

(1) Trinity College Dublin, Ireland, (2) Geological Survey Ireland, (3) Environmental Protection Agency of Ireland

Corresponding author: Elío J.

Email: elioj@tcd.ie – javiereliomedina@gmail.com.

Postal address: Museum Building, Geology, School of Natural Sciences, Trinity College, Dublin 2, Ireland.

ABSTRACT

A new high spatial resolution radon risk map of Ireland has been developed, based on a combination of indoor radon measurements ($n=31,910$) and relevant geological information (i.e. bedrock geology, Quaternary geology, soil permeability and aquifer type). Logistic regression was used to predict the probability of having an indoor radon concentration above the national reference level of 200 Bq m^{-3} in Ireland. The four geological datasets evaluated were found to be statistically significant, and, based on combinations of these four variables, the predicted probabilities ranged from 0.57% to 75.5%. Results show that the Republic of Ireland may be divided in three main radon risk categories: High (HR), Medium (MR) and Low (LR). The probability of having an indoor radon concentration above 200 Bq m^{-3} in each area were found to be 19%, 8% and 3%; respectively.

In the Republic of Ireland, the population affected by radon concentrations above 200 Bq m^{-3} is estimated at ca. 460k (about 10% of the total population). Of these, 57% (265k), 35% (160k) and 8% (35k) are in High, Medium and Low Risk Areas, respectively. Our results provide a high spatial resolution utility which permit customised radon-awareness information to be targeted at specific geographic areas.

Keywords: Radon; Radon mapping; Indoor radon; Geogenic factors; Logistic regression

1. Introduction

Radon is a natural radioactive gas present in all soils (Cothorn and Smith, 1987) and represents the highest source of natural ionizing radiation to the general population (UNSCEAR, 2000a). Radon may accumulate in dwellings and may cause lung cancer when its decay products are

inhaled (Field, 2015; US-EPA, 2003; WHO, 2010). In this regard, radon is the second cause of lung cancer after smoking (McColl et al., 2015), and the number of lung cancer deaths linked to radon exposure may vary from 3% to 20% depending on the country (Kim et al., 2016). In the United States of America, for example, around 21,000 annual deaths from lung cancer (approximately 10% of the annual lung cancer deaths) are linked to radon exposure (Casey et al., 2015; US Cancer Statistic Working Group, 2016), while in Europe radon exposure causes around 18,000 lung cancer deaths annually (about 8% of the annual lung cancer deaths; Gray et al., 2009). Accurate prediction of radon prone areas aids regulators and local authorities to develop a proper strategy to minimise exposure to this carcinogen, resulting in an enhanced quality of life and improvement of long-term health for the general population.

Although radon is a health hazard, it can be mitigated if appropriate radon measures are implemented (e.g. Long et al., 2013; US EPA, 2001). Radon is heterogeneously distributed and certain geographic areas are more likely to have high levels of radon than others. High risk areas can be identified on radon risk maps, which are useful to target homeowners and the construction industry so that remediation work can be carried out. Such maps can also be used to identify the areas where preventive measures in new buildings should be applied (McColl et al., 2015). Therefore, radon mapping has profound economic and social implications (Gray et al., 2009), and a high-resolution, accurate and statistically robust map is crucial to inform government policy allowing development and implementation of the most cost-effective approach to reduce radon exposure in the general population.

The criteria selected to develop a radon risk map depends on national radiological protection strategies and the data available. Data on which such maps are based are conventionally indoor radon measurements (e.g. Fennell et al., 2002; Ferreira et al., 2016) or soil-gas radon measurements (e.g. Bossew, 2015). The use of geological information, and other factors, assist to improve the accuracy of these maps (e.g. Bossew, 2015b, 2014; Ferreira et al., 2016; Miles and Appleton, 2005; Pásztor et al., 2016). Such an approach is warranted, as geology is known to be the main factor controlling indoor radon concentrations (Appleton and Miles, 2010; Watson et al., 2017).

Consequently, radon mapping can be divided in two activities, one for developing Indoor Radon Maps and another for Geogenic Radon Maps (e.g. Bossew et al., 2013; Miksova and Barnet, 2002). The principal advantage of the Geogenic Radon Maps is that it measures “*what earth delivers in terms of radon*” (Bossew et al., 2013) and, therefore, they are independent of anthropogenic factors. Recent studies suggest that this affirmation is not totally correct and some other influences exist (Bossew, 2014). On the other hand, the main advantage of Indoor Radon Maps is that radon is directly measured at the exposure point (i.e. dwellings). Nevertheless,

uncertainties in the exact location of tested homes and extensive areas with little or no data may impede the generation of an accurate Indoor Radon Map. Furthermore, indoor concentrations depends on multiple factors, both natural and anthropogenic (Gunby et al., 1993; Tollefsen et al., 2014), and even some dwellings may also have been subjected to preventive/remediation measures (e.g Long et al., 2013) which may hinder the correct interpretation of data. Finally, an appropriate sample design is crucial for accurate assessments of the radiological risk (Burke and Murphy, 2011).

In this study, the use of logistic regression models has been evaluated, combining indoor radon measurements and geological information. The primary objective of this study is to estimate the probability of having an indoor radon concentration above the Irish national reference level of 200 Bq m⁻³ (Fennell et al., 2002). Previous work has been carried out in Ireland by multivariate linear regression (Hodgson et al., 2014), however, logistic regression models are a better statistical tool for studying the probability of occurrence of categorical variables (James et al., 2013). In this regard, for radon mapping purposes indoor radon can be evaluated as a binary variable taking the value 0 or 1 if it is below or above the reference level, respectively, while simultaneously reporting the probability of belonging to one of the given explanatory variable categories (in this study Bedrock Geology - BG, Quaternary Geology - QG, Subsoil Permeability - SP, and Aquifer Type - AT). Such models are frequently used in epidemiology, for example to analyse the relationship between radon and lung cancer (Kreienbrock, 2001; Wang et al., 2002), but also to analyse natural hazards (Dong et al., 2011; Ettinger et al., 2016; Notario del Pino and Ruiz-Gallardo, 2015; Wang et al., 2016) and more recently for radon mapping (Kropat et al., 2017).

Indoor radon measurements were previously used to develop the Irish National Radon Map in grids of 100 km² (www.epa.ie/radiation/radonmap). Each grid was classified according to the probability of a dwelling having an indoor radon concentration above the reference level of 200 Bq m⁻³ (i.e., <1%, between 1% and 5%, between 5% and 10%, between 10% and 20% and >20%; Fennell et al., 2002), and a “High Radon Risk Area” was defined as the area in which this probability is 10% or higher. The same indoor radon data and the same criteria were used in this research, however, geological information (i.e. BG, QG, SP, and AT) has been included to add geogenic factors and develop a higher resolution radon risk map of the Republic of Ireland.

2. Material and methods

2.1. Indoor Radon Measurements

The indoor radon (InRn) measurements used in this study were carried out by the Environmental Protection Agency of Ireland (EPA) as part of a national survey in Irish dwellings (Fennell et al., 2002). More than 60,000 dwellings have been sampled by the EPA and local authorities across the country. However, only 32,108 measurements have been accurately geo-referenced by the Geological Survey of Ireland (Hodgson et al., 2014) and the Ireland's Health Services (HSE-Health Intelligence Unit; HSE, 2013), and are available for this study.

Indoor radon concentrations were measured by passive alpha track detectors (CR-39), located in homes for a minimum of 3 months and seasonally adjusted to give an annual value (Burke et al., 2010), being the limit of detection 10 Bq m^{-3} (Rochford et al., 2010). Then, the average outdoor radon concentration (i.e. 5.6 Bq m^{-3}) (Gunning et al., 2014) was subtracted from the data (Hodgson et al., 2014).

For the purpose of this study, indoor radon measurements were combined with a simplified bedrock geology map (scale 1:1 million), a simplified Quaternary geology map, a subsoil permeability map and an aquifer type map (using ArcGIS software). After completion of this overlay treatment, the total number of geo-referenced indoor radon data available for the model are 31,910 (Figure 1; the small reduction in the total number of data, from the initial 32,108, may be caused by small differences in the area covered by the different maps, and errors in the geo-referenced datasets, both the indoor radon measurements and the map boundaries).

The total number of dwellings found to have an indoor radon concentration above the reference level of 200 Bq m^{-3} is 3,851 (12% of the dataset used; red stars in Figure 1). Indoor radon concentration ranged from $0.1 (< \text{LD})$ to $9,708 \text{ Bq m}^{-3}$, with a median of 53.4 Bq m^{-3} and a mean of 109.7 Bq m^{-3} . The geometric mean (GM) and the geometric standard deviation (GSD) are 56.81 and 2.97 Bq m^{-3} , respectively, thus the overall probability of having an indoor radon concentration above the reference level in the Republic of Ireland is 12.41% (assuming a lognormal model).

2.2. Bedrock Geology

The Bedrock Geology (BG) of Ireland (scale 1:1 million) was used for the model. The BG was further simplified by reducing the initial 30 different units to 12 units because some geological units do not have sufficient indoor radon measurements to be statistically significant in the model (Table 1 and Figure 2). The combination and removal of units was principally achieved by similar geological formation (e.g. Group 6: Late Ordovician-Silurian deep marine greywacke, mudstone (Longford-Down), and Silurian deep marine mudstone, greywacke & conglomerate; or Group 8: Tournaisian limestone, and Tournaisian sandstone, mudstone, limestone). In some cases,

however, the criteria was the size of the unit. In this regard, if a specific unit had a small area and few indoor radon measurements (Table 1) it was merged to the neighbouring unit (e.g. the Devonian volcanic rocks was merged with the ORS, sandstone, conglomerate & mudstone, Group 5; and Carboniferous volcanic rocks with Visean limestone & calcareous shale, Group 11). The only exception in this criteria was the “Triassic sandstone, mudstone, evaporite” formation (Group 9). Although the area is very small and the number of data in this unit was 6, the high indoor radon concentrations (ranged from 46 to 438 Bq m⁻³, and 3 of 6 higher than 200 Bq m⁻³) led to the decision to retain this geological unit as a separate entity in our model.

2.3. Quaternary Geology

Small polygons (less than 2 km²) and man-made categories (e.g. urban, industrial or airfield) in the original Quaternary Geology (QG) map of Ireland (Figure 3a) were merged to neighbouring polygons. Then the different Quaternary sediments were grouped by similar compositional characteristics in 10 levels (Table 2 and Figure 3b).

2.4. Subsoil permeability

The Subsoil permeability (SP) Map was obtained from the National Groundwater Recharge Map and the Quaternary Map of Ireland (GSI). In the National Groundwater Recharge Map (Figure 4a), the Geological Survey of Ireland (GSI) used three classifications for subsoil permeability; “High”, “Moderate” and “Low”. These subsoil permeability classifications were based on direct measurements, observation of drainage patterns and vegetation, particle size analysis as a permeability predictor, and subsoil descriptions using the British Standard (BS)5930 (BSI, 1999) as a proxy for particle size analysis (Hunter Williams et al., 2013). Where subsoil was less than 3m thick (Depth to Bedrock, DTB, <3m), the GSI did not classify subsoil permeability as it can display spatial variability due to vegetation rooting, cracking and the influence of the underlying bedrock (Lee et al., 2008; Masterson et al., 2008). In these areas, the type of soil (i.e. Quaternary Geology Map) has been taken into account and assigned a permeability value based on the criteria described in Appleton et al., 2015 (Table 3 and Figure 4b). The summary statistics of the 31,910 indoor radon by subsoil permeability (i.e. High, Moderate and Low) are shown in Table 4).

2.5. Aquifer type

Information on aquifer type (AT) was obtained from the Groundwater Aquifer Map of Ireland (GSI; Figure 5a). The aquifer types were grouped into five categories (Table 5) and Figure 5b; GSI, 2005): a) karstified aquifers (Rkc, Rkd, Rk, Lk); b) productive fissured (Rf, Lm); c) poor-

moderate productive fissured which is moderately productive only in local zones (LI); d) unproductive aquifers (PI, Pu); and e) intergranular aquifers (Rg, Lg).

2.6. Logistic regression

The method of logistic regression is commonly used to analyse the probability of occurrence of binary variables; this is achieved by fitting the log odds and explanatory variables to a linear model (James et al., 2013).

$$\log\left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Where $Y = (0, 1)$ is the binary variable, $X = (X_1, \dots, X_n)$ are “n” explanatory variables, and $\beta = (\beta_0, \dots, \beta_n)$ are the regression coefficients to be estimated based on the data. In this work the “glm” function of the R package “stats” was used (R Core Team, 2016). The probability of occurrence is predicted by the logistic function as follows:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$$

For radon mapping, the binary variable (Y) is the indoor radon concentration (1 if it is higher than the Reference Level and 0 if not), modelling the probability of having an indoor radon concentration above the reference level (i.e. 200 Bq m⁻³) given the explanatory variables (e.g. n = 4; BG, QG, SP, and AT).

The explanatory variables used in the regression model were selected based on the Akaike Information Criterion or AIC, choosing the model which minimises the value of AIC ($AIC = -2 \log(\text{lik}) + 2p$; where (lik) is the verisimilitude of the model and p the number of parameters of the model; Akaike, 1974). Furthermore, likelihood ratio tests ($LRT = 2(\log \text{lik}(H_1) - \log \text{lik}(H_0))$) were also used to confirm the model that best fits the data. The p-value was calculated by a chi-squared with degree of freedom equal to the difference of the degree of freedom of the tested models ($df = df_1 - df_0$), and the significant level was 0.01.

2.7. Radon Risk Map

The four geological maps (i.e. BG, QG, SP and AT) were intersected, giving a total of 960 different combinations of the 4 geo-datasets in the Republic of Ireland. Each of these 960 combinations were merged in polygons and the probability of having an indoor radon concentration above the

reference level was predicted by the logistic regression model. Therefore, all polygons with the same combination of the four variable (i.e. BG x QG x SP x AT) have the same probability. Finally, a probabilistic radon map was developed based on the predictions, by dividing the country in five probability bands ($P < 1\%$; $1\% \leq P < 5\%$; $5\% \leq P < 10\%$; $10\% \leq P < 20\%$; and $P \geq 20\%$).

A binomial analysis was carried out to evaluate if the probabilities of the radon risk grouping developed based on the logistic regression (i.e. VL: $<1\%$; L: 1-5%, M: 5 – 10 %; H: 10 – 20% and VH: $\geq 20\%$) are in agreement with the probabilities of having an indoor radon concentration above the reference level, knowing the number of cases ($\ln Rn > 200 \text{ Bq m}^{-3}$) and fails ($\ln Rn \leq 200 \text{ Bq m}^{-3}$) in each area. The probability and the 95% confidence interval of having an indoor radon concentration above the 200 Bq m^{-3} reference level in each area were estimated by a binomial distribution (Exact Binomial Test, function `binom.test{stats}` of R software; R Core Team, 2016), assuming that the sampled dwellings were selected randomly.

Since the Irish building regulation only requires prevention measures in High Risk Areas (HRA; $\text{Prob}[\ln Rn > 200 \text{ Bq m}^{-3}] \geq 10\%$) a confusion matrix (Dong et al., 2011) was performed to evaluate the radon risk maps, estimating the percentage of dwellings correctly classified, and the errors, when the areas are classified as HRA or No HRA.

The estimation of the population that may be affected by high radon concentration was carried out based on the last census of Ireland, which counts the number of inhabitants by grids of 1km^2 (CSO, 2013), and averaging the estimated probability (p_i) in these grids (weighted arithmetic mean, $P = \frac{1}{A_T} \sum_{i=1}^n A_i \cdot p_i$ and $A_T = \sum_{i=1}^n A_i$. Where P and A_T are the average probability and the total area of a specific grid, respectively; and p_i and A_i the probability and the area of the different subdivisions present in the grid, respectively).

3. Results and discussion

3.1. Logistic model

The AIC Criterion values, the log likelihoods and the degree of freedom of the possible models from combinations of the four explanatory variables are included in Table 6. All models are significantly different, and fit better, than the empty model (NULL). However, as expected, the model with minimum AIC is the model with four explanatory variables (i.e. BG, QG, SP, AT). The likelihood ratio test confirms that the model with four variables adjusts better to the data, and is therefore the model used here to predict the probability of having an indoor radon concentration above 200 Bq m^{-3} .

The adjusted model is shown in Table 7, the intercept levels (i.e.: BG, Group 1; QG: Glacio-fluvial; SP: Low; and AT: Intergranular) have been defined as the ones with the smallest GSD (Table 1, Table 2, Table 4, Table 5). The log-odd ratio ($\log(\text{OR})$; $\text{OR} = p/(1-p)$; $p = P[Y=1|X]$) of the intercept is -4.699 (Table 7), thus the probability of having an indoor radon concentration above the reference level is 0.90% if the BG is Group 1; the QG is Glacio-fluvial; the SP is Low; and the AT is Intergranular.

When one category of these variables changes, the log-odds ratio is affected by the estimated regression coefficients accordingly. In this regard, the probability increases with respect to the intercept if the regression coefficient is positive, or decreases if the regression coefficient is negative. For example, changing the intercept variables for the following; BG: Group 5, QG: Sandstone and Shale till, SP: Moderate, and AT: Poor-Moderate Productive, the resultant log-odds ratio is -2.478 ($\log(\text{OR}) = -4.699 + 0.730 + 0.813 + 0.713 - 0.034$), and the probability 7.74% ($p = \text{OR}/(1+\text{OR})$; Table 7 and Table 9).

Although the overall influence of the four variables (i.e. BG, QG, SP, and AT) is complex and the probability of having an indoor radon concentration above the reference level should be estimated as a cumulative effect, individual inspection of each variable may help to understand the weighting of factors affecting radon accumulation in dwellings.

From Table 7 we can see that the regression coefficients in subsoil permeability are positive (0.713 and 1.057 for Moderate and Low, respectively). This is a statistical indication that the probability of having high indoor radon concentrations increases when the subsoil permeability also increases. Logically, subsoil permeability is related to effective mobility of radon in soil. If the permeability is high the soil-gas can be transported easily into the house by the stack effect (UNSCEAR, 2000b), while lower values impede this transport, and therefore the probability of having an indoor radon above the reference level is reduced.

The regression coefficients for AT indicate that the level with the highest probability of having an indoor radon concentration above the reference level of 200 Bq m^{-3} is the Karstified Aquifers ($\beta_4 = 0.743$; Table 7). The category with the next highest probability is the Intergranular Aquifers (intercept); followed by Poor-Moderate Productive Fissured ($\beta_4 = -0.034$), Productive Fissured ($\beta_4 = -0.386$) and finally Unproductive Aquifers ($\beta_4 = -0.387$). These results corroborate with the notion that the radiological risk for a given area increases when geological structures allow subterranean mobility of radon (i.e. karstified areas). Such enhanced mobility of radon and transfer to buildings may occur when radon dissolved in groundwater degasses near dwellings, or when radon gas is transported from deep sources, enhanced by a higher permeability. Such

scenarios agree with our model which ranks areas with karstified aquifers as the highest risk, compared to intergranular and moderate productive aquifers and finally the poor aquifers. These results confirm the importance of karstified and fracture areas in enhancing mobility of radon.

As radon in groundwater is considered to be predominantly sourced from uranium in rock, and not the dissolved uranium content in water (Yang et al., 2014), longer residence time of groundwater will favour the decay of this dissolved radon, reducing the influence of groundwater as a transport mechanism for radon. Such an effect explains the cause of high radon risk in karstified areas where groundwater may migrate long distances in a relatively short duration of time, so the dissolved radon does not significantly decay in the transport process. Furthermore, it also explains why areas underlain by intergranular aquifers have a greater radon risk compared to poor-moderate productive and productive fissured, and why poor aquifers (unproductive) have the lowest radon risk in our model.

The influence of BG and QG is, however, more complex and difficult to interpret. Several factors related to both explanatory variables could influence indoor radon concentration, possibly due to uranium and radium concentrations in the matrix, but also due to physical properties that influence the radon emanation process and its migration in the soil (e.g. grain size, mineralization, porosity, permeability and water saturation) (Cothorn and Smith, 1987).

In this study, the BG Groups which have the highest regression coefficient are Group 9 (Triassic sandstone, mudstone, evaporite) and Group 7 (granite and intrusive rocks; Table 7), probably related to an enrichment in uranium concentrations. Such a correlation between certain BG and an elevated radon risk is not unexpected; for example Scheib et al. (2013) reported uranium anomalies in the south of England related to “Triassic and Jurassic (Lias Group) mudstone, siltstone, limestone and sandstone sedimentary assemblages”, and a high percentage of homes exceeding the action level of 200 Bq m⁻³ overlying Triassic and Jurassic sedimentary rocks (i.e. up to 17%). Furthermore, some granites and intrusive rocks are well known as relatively high uranium rocks and sources of radon, and as such are generally termed radon prone areas (e.g. Adepelumi et al., 2005; Yang et al., 2014).

Following the Triassic-Jurassic sedimentary successions, granite and intrusive rocks, the regression coefficients seem to be reduced by relatively similar geological characteristics, however, further investigations are necessary to corroborate this hypothesis and possible causes. Thus, with respect to BG, the next groups defined in our regression model are Groups 12 and 10, with log-odds of 1.421 and 1.352 (Table 7), respectively (Devonian and Viséan sandstones, mudstones), Groups 2, 11 and 6 (log-odds: 1.123, 0.981 and 0.971; Ordovician slate, sandstone, greywacke, conglomerate; impure limestone; and Ordovician Silurian greywackes, mudstones,

slates, conglomerate) and the Groups 8 and 5 (log-odds: 0.738 and 0.730; Tournaisian sandstones, mudstones, limestones and ORS, sandstone, conglomerate & mudstone). The lowest influence corresponds to Group 4 (log-odd: 0.450; principally Neoproterozoic metasedimentary rocks – Dalradian), Group 3 (log-odd: 0.350; mainly Namurian shale, sandstone, siltstone & coal) and Group 1 (intercept; Cambrian greywacke, slate, quartzite).

The QG demonstrating the highest probabilities (Table 7) are the Shale tills ($\beta_2 = 1.858$) and Granite tills ($\beta_2 = 1.789$), probably due to an enrichment in uranium and radon (e.g. Appleton et al., 2008; Casey et al., 2015; Cothorn and Smith, 1987; Yang et al., 2014). Following these, the next highest probabilities are encountered when bedrock is at surface ($\beta_2 = 1.419$) and Sandstone tills ($\beta_2 = 1.314$). The lowest probabilities correspond to Marine deposits ($\beta_2 = -0.075$), Glacio-fluvial (“intercept”), Peat ($\beta_2 = 0.531$), Limestone tills ($\beta_2 = 0.772$), Sandstone and Shale tills ($\beta_2 = 0.813$), and Metamorphic tills ($\beta_2 = 1.054$). Interpretation of the cause of such a reduction in probability is difficult without further investigation, since it can be produced by multiple factors, such as different U or Ra concentration, soil permeability, texture, water moisture, or other physio-chemical characteristics.

Although it is instructive to examine probabilities for individual explanatory variables, it is important to note that the radon risk in a specific area is best estimated with a combination of the four variables, since use of a single variable may generate misinterpretations. For example, the probability in bedrock Group 3 (mainly Namurian shale, sandstone, siltstone coal formation), has one of the lowest values if only BG is considered (Table 1 and Table 7), however when other explanatory variables are taken into account the predicted probabilities (Appendix 1) ranged from 0.81% (0.48% - 1.37%) in conjunction with QG: marine deposits, SP: low and AT: unproductive aquifers to 24.43% (19.61% - 29.99%) in conjunction with bedrock in surface (QG), high subsoil permeability (SP) and karstified aquifers (AT). While the bedrock Group 7 (mainly Siluro-Devonian granitic rocks and appinite) is one of the highest probability values when solely examining the bedrock geology category, it ranged from 2.41% (1.50% - 3.85%) in conjunction with Quaternary marine deposits, low subsoil permeability and unproductive aquifers to 50.07% (45.20% - 54.94%) in conjunction with Quaternary granite till, moderate subsoil permeability and karstified aquifers. Therefore, if only BG is considered, it is possible that some Namurian areas would be considered as low risk, when they may be high radon risk areas. Conversely, it would be incorrect to assume that all areas underlain by granitic rock are high risk areas in that some of these areas may actually be classified as low risk.

3.2. Indoor Radon Map

The probability of having an indoor radon concentration above 200 Bq m⁻² predicted in the 960 different polygons after the intersection of the four explanatory variables (appendix 1) range from a maximum of 75.57% (i.e. BG: Group 9; QG: Sandstone and shale till; SP: Moderate; and AT: Karst) to a minimum of 0.57% (i.e. BG: Group 1; QG: Marine deposits; SP: Low; and AT: Unproductive (Figure 6; Table 8). From the 960 polygons, 20 polygons represent up to 50% of the total area of the country (Table 9), and 116 polygons (with areas >80 km²) more than 90%.

The probability of an indoor radon concentration greater than the reference level (i.e. 200 Bq m⁻³) in 960 different polygons from the intersection of four variables (i.e. BG, QG, SP, AT) are plotted in Figure 7a. The overall spatial distributions of probabilities on the actual map are generally consistent with the EPA Indoor Radon Map of Ireland which was developed based solely on indoor radon measurements (Figure 7b; Fennell et al., 2002), but with a higher spatial resolution. The predicted values also are in agreement with estimations in England (Scheib et al., 2013) and Northern Ireland (Appleton et al., 2015); these are countries with some similarities in geology, building standards and climate conditions.

For example, the probability in the Group 3 (mainly Namurian shale, sandstone, siltstone coal formation) ranged from 0.81% to 24.43% (appendix 1), while Scheib et al. (2013) estimated that for England the average probability in the Namurian sandstone was around 1% with maximum values up to 30.9%. On the other hand Appleton et al. (2015) have measured percentages in Namurian sandstones ranging from <0.1% to 6% in Northern Ireland. Another example is Group 7 (mainly Siluro-Devonian granitic rocks and appinite). In this case our predicted values ranged from 2.41% to 50.07%, whereas in England the granite and igneous rock have average values of approximately 6% to values up to 68% (or even higher in granites with high U concentrations; Scheib et al., 2013). In Northern Ireland, the granodiorite plutons of the late Silurian early Devonian have values of 0.1 to 21% and the Mourne Mountains Granites from 7% to 24% (Appleton et al., 2015).

Since Irish building regulations request that all new dwellings built in a High Radon Risk Area (HRA) have a radon barrier, the model may be evaluated dividing the country in two groups; (1) areas where the probability of having an indoor radon concentration above the reference level is 10%, or higher (HRA), and (2) areas where this probability is lower than 10% (NoHRA). A confusion matrix (Table 10) shows that both maps (the EPA Map and the New Radon Map) correctly classify about 60% of the dwellings, providing a better estimation than if the entire country were classified as high risk, since the overall probability of having an indoor radon concentration above 200 Bq m⁻³ is 12%.

The false positive (error type I) from logistic regression is slightly reduced (from 41.4% to 41.1%) with respect to the EPA Map, while the false negative (error type II) is increased (from 25.3% to 28.4%). However, there are several grids with little or no indoor radon data and since the predominant source of radon is the soil and subsoil, it seems more appropriate to evaluate the radon risk by also taking into account the geological information, rather than basing this only on the data of neighbouring grids (i.e. Fennell et al., 2002).

The total area of Ireland represented by the five risk bands (i.e. Very Low, Prob. < 1%; Low, 1% ≤ Prob. < 5%; Moderate, 5% ≤ Prob. < 10%; High, 10% ≤ Prob. < 20%; and Very High, Prob. ≥ 20%) is shown in Table 1. According to such a categorisation, almost 30% of the country is classified as a High Radon Area (Prob[InRn>200] ≥ 10%). These results from the binomial distribution analysis are in agreement with the predicted probabilities of each risk band (Table 1), supporting the hypothesis that the data are well adjusted by logistic regression models.

The high uncertainty in the first group (i.e. Very Low; Table 1) is probably related to the low number of data points (n=44), and since the geographic area is also very small (16 km², or 0.02% of the country), it seems more reasonable to merge this group with the Low Risk classification. If “Very Low Risk” and “Low Risk” areas are merged, the probability is 3.23% (2.79% - 3.73%), more in keeping with a group in which the estimated probability should be less than 5% (Table 1). Similarly, when the High and Very High risk categories are merged, into what the EPA define as a High Radon Area (Fennell et al., 2002), the probability of having an indoor radon concentration above the reference level is 19.28% (18.63% - 19.93%). It is therefore possible to divide the country into three radon risk areas: Low, Medium and High Risk. The average probability of having an indoor radon concentration above the reference level of 200 Bqm⁻³ within these three risk areas (low, medium and high) is therefore 3%, 8% and 19% respectively.

The total population living in these three categorised radon risk areas corresponds to 1.08 million (24%), 2.13 million (46%), and 1.37 million (30%); in Low, Moderate and High risk areas respectively (Table 1). Assuming that the calculated probabilities in these areas also reflects the probability of an inhabitant to live in a house with an indoor radon concentration higher than 200 Bq m⁻³ (i.e. Low: 3.2%; Medium: 7.6%; and High 19.3%), the total population that may live with a high indoor radon concentration in each area is 35k (30k - 40k), 162k (152k - 172k), and 264k (255k - 273k); respectively. This means that approximately 10% of the total population of the Republic of Ireland (461k) may be exposed to high indoor radon concentrations. Estimates of the population living in each categorised risk area is based on total figures from census data, which calculated the inhabitants who live in Ireland by grids of 1 km² but not the number of dwellings. These figures should therefore be treated as an estimate and further research should be carried

out to assist with a more accurate calculation and spatial distribution of the number of dwellings and the total population that is living within each of the categorised radon risk areas.

3.3. Limitations of the model

Results show that the four geological datasets evaluated in this study are statistically significant. However, the model required some simplification of the geological information due to the number of geo-referenced indoor radon data currently available. Although such a simplification was deemed necessary to generate a robust statistical model and was carried out carefully, and when possible based only on geological criteria (e.g. lithology and/or age), this treatment may influence the results. As the number of available indoor radon measurements increases, it will be possible use more accurate information (BG scale of 1:500k or 1:100k) and enhance the predictive capacity of the model, since the percentage of variance explained by geology increase with the scale (Appleton et al., 2015).

High accuracy of geological information and a correct geo-reference of the datasets are also fundamental for developing a good model. Errors in the geo-referenced datasets may generate misclassifications in the dwelling location (e.g. assigned the house to one BG, e.g. Group 1; when it is another BG, e.g. Group 3) and errors in the combination of the explanatory variables (e.g. assigned to an area for example BG: Group 1, QG: Limestone till, SP: Low, AT: Productive fissured; when it is another, e.g. BG: Group 3, QG: Sandstone and shale till, SP: Low, AT: Productive fissured) that generate uncertainties in the predicted probabilities. These effects are difficult to evaluate and, for example, the reduction in the number of data measurements when the indoor radon were located with the geological information (from 32,108 to 31,910) indicate that our model is affected to some degree by these errors.

Another limitation of the present model is that possible building construction factors that may influence indoor radon concentrations were not recorded (e.g. type of house or if a remediation/preventative measure was installed) and, therefore it was not possible to adjust the indoor radon concentrations or remove dwellings which are not appropriate for radon mapping (e.g. houses with a radon barrier installed). We do not expect this latter factor to considerably affect our results, since most of the indoor radon measurements used in this data sets were carried out before the introduction in 1998 of the Irish building regulation that requests that all new houses in High Radon Areas must be built with a radon prevention measure (Dowdall et al., 2017). We recommend that information on the building type and presence or absence of remediation/preventative measures should be recorded as part of future indoor radon measurements for a national dataset.

Finally, the predicted probabilities of our map, and generally of all radon risk maps, could be of limited use if dramatic changes to soil properties are made during building construction. In this case, it could be possible that an area classified as low risk should be classified as high risk after construction, and therefore preventive measures should be implemented. For example, if the superficial geology is removed others factors will have a greater influence on the risk, and if these other factors are characterized by a higher radon risk (e.g. BG: Group 7 or AT: Karstified aquifer) the classification of the local-scale building site may change. One option in such a situation would be to carry out a site-specific radon risk assessment before building construction commences in order to evaluate if the site should be classified as a High Risk, and a radon barrier needs to be installed.

4. Conclusions

A new radon risk map of Ireland has been developed using a logistic regression model. The classification of the areas are based on available geo-referenced indoor radon measurements and four separate geological datasets (i.e. Bedrock Geology, scale 1:1 million; Quaternary Geology; Soil Permeability; and Aquifer Type), are used as explanatory variables. The procedure has been validated as a good tool to predict the probability of having an indoor radon concentration above the reference level, even when there were no indoor radon data for a specific geographic region. The use of logistic regression models using such a combination of datasets results in an improvement of the spatial resolution of the Indoor Radon Map of Ireland (Fennell et al., 2002) which was previously based solely on indoor radon measurements and represented at grids of 100 km².

According to our new model, the Republic of Ireland may be divided into three main radon risk areas: Low, Medium and High Risk; with probabilities of having an indoor radon concentration above the reference level (200 Bq m⁻³) of approximately 3%, 8% and 19% respectively. The total population that live in these areas are 1.08 million (24%), 2.13 million (46%), and 1.37 million (30%) respectively, and the inhabitants that may be directly affected by radon levels above 200 Bq m⁻³ are approximately 460,000, about 10% of the total population of the Republic of Ireland.

A radon risk map is essential to target geographic areas where radon preventive or remediation measures should be implemented. Therefore, a better prediction of radon prone areas will ultimately assist regulators and local authorities to design a proper strategy to minimise radon exposure in the built environment. Normally, such strategies (not only in Ireland) are focussed in predicted High Risk Areas (e.g. implementing of the building regulations that all new houses must be built with radon barrier, or increasing public awareness on radon in these areas). However, results obtained in this study indicate that the population affected by a high indoor radon

concentration may also be large in areas classified as Non-High Risk. Therefore, it is recommended to include some type of radon guideline in all areas (e.g. active radon soil-gas measurements for new buildings), and also further increase public awareness of radon at national scale.

In our new radon risk map of Ireland only information available at a national scale was used. However, a national survey to analyse the geochemical and geophysical properties of the soils and exposed geology of Ireland (i.e. Tellus Project, www.tellus.ie) is currently being developed. When this project is completed, uranium concentration in the soil will be available (as well as several other elemental concentrations, e.g. thorium and potassium). Therefore, as such geochemical data are useful for radon mapping (Appleton et al., 2011a, 2011b, 2008), they will likely be incorporated into an additional national scale geogenic model.

Acknowledgement

This work has been financed by the Irish Research Council (IRC), through the Enterprise Partnership Scheme Postdoctoral Fellowship 2015 (EPSPD/2015/46), and co-financed by the Geological Survey Ireland (GSI). The sole responsibility of this publication lies with the authors. The IRC is not responsible for any use that may be made of the information contained therein. The data used in this project have been collated by the Environmental Protection Agency of Ireland, and the Geological Survey of Ireland.

References

- Adepelumi, A.A., Ajayi, T.R., Ako, B.D., Ojo, A.O., 2005. Radon soil-gas as a geological mapping tool: case study from basement complex of Nigeria. *Environ. Geol.* 48, 762–770. doi:10.1007/s00254-005-0016-0
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi:10.1109/TAC.1974.1100705
- Appleton, J.D., Daraktchieva, Z., Young, M.E., 2015. Geological controls on radon potential in Northern Ireland. *Proc. Geol. Assoc.* 126, 328–345. doi:10.1016/j.pgeola.2014.07.001
- Appleton, J.D., Doyle, E., Fenton, D., Organo, C., 2011a. Radon potential mapping of the Tralee–Castleisland and Cavan areas (Ireland) based on airborne gamma-ray spectrometry and geology. *J. Radiol. Prot.* 31, 221–235. doi:10.1088/0952-4746/31/2/002
- Appleton, J.D., Miles, J.C.H., 2010. A statistical evaluation of the geogenic controls on indoor radon concentrations and radon risk. *J. Environ. Radioact.* 101, 799–803. doi:10.1016/j.jenvrad.2009.06.002
- Appleton, J.D., Miles, J.C.H., Green, B.M.R., Larmour, R., 2008. Pilot study of the application of

- 563 Tellus airborne radiometric and soil geochemical data for radon mapping. *J. Environ.*
 564 *Radioact.* 99, 1687–97. doi:10.1016/j.jenvrad.2008.03.011
- 565 Appleton, J.D., Miles, J.C.H., Young, M., 2011b. Comparison of Northern Ireland radon maps
 566 based on indoor radon measurements and geology with maps derived by predictive
 567 modelling of airborne radiometric and ground permeability data. *Sci. Total Environ.* 409,
 568 1572–83. doi:10.1016/j.scitotenv.2011.01.023
- 569 Bossew, P., 2015a. Mapping the Geogenic Radon Potential and Estimation of Radon Prone Areas
 570 in Germany. *Radiat. Emerg. Med.* 4, 13–20.
- 571 Bossew, P., 2015b. Estimation of radon prone areas through binary classification, part 2: radon
 572 prone geologies. *J. Environ. Radioact.* 141, 44–50. doi:10.1016/j.jenvrad.2014.11.020
- 573 Bossew, P., 2014. Determination of radon prone areas by optimized binary classification. *J.*
 574 *Environ. Radioact.* 129, 121–132. doi:10.1016/j.jenvrad.2013.12.015
- 575 Bossew, P., Tollefsen, T., Gruber, V., M, D.C., 2013. The european radon mapping project, in: IX
 576 Latin American IRPA Regional Congress on Radiation Protection and Safety - IRPA 2013. Rio
 577 de Janeiro, Brazil.
- 578 BSI, 1999. Code of practice for site investigations. BS 5930:1999. Br. Stand.
- 579 Burke, Ó., Long, S., Murphy, P., Organo, C., Fenton, D., Colgan, P.A., 2010. Estimation of seasonal
 580 correction factors through Fourier decomposition analysis—a new model for indoor radon
 581 levels in Irish homes. *J. Radiol. Prot.* 30, 433–443. doi:10.1088/0952-4746/30/3/002
- 582 Burke, O., Murphy, P., 2011. The use of volunteer radon measurements for radon mapping
 583 purposes: an examination of sampling bias issues. *J. Radiol. Prot.* 31, 319–28.
 584 doi:10.1088/0952-4746/31/3/001
- 585 Casey, J.A., Ogburn, E.L., Rasmussen, S.G., Irving, J.K., Pollak, J., Locke, P.A., Schwartz, B.S., 2015.
 586 Predictors of indoor radon concentrations in Pennsylvania, 1989–2013. *Environ. Health*
 587 *Perspect.* 123, 1130–1137. doi:10.1289/ehp.1409014
- 588 Cothorn, R., Smith, J., 1987. Environmental radon, *Environmental science research* ; v. 35. Plenum
 589 Press, New York.
- 590 CSO, 2013. Census of Population of Ireland 2011, 1 Km² Grid dataset, User Guide.
- 591 Dong, J.-J., Tung, Y.-H., Chen, C.-C., Liao, J.-J., Pan, Y.-W., 2011. Logistic regression model for
 592 predicting the failure probability of a landslide dam. *Eng. Geol.* 117, 52–61.
 593 doi:10.1016/j.enggeo.2010.10.004
- 594 Dowdall, A., Murphy, P., Pollard, D., Fenton, D., 2017. Update of Ireland’s national average indoor
 595 radon concentration – Application of a new survey protocol. *J. Environ. Radioact.* 169, 1–8.
 596 doi:10.1016/j.jenvrad.2016.11.034
- 597 Ettinger, S., Mounaud, L., Magill, C., Yao-Lafourcade, A.-F., Thouret, J.-C., Manville, V., Negulescu,
 598 C., Zuccaro, G., De Gregorio, D., Nardone, S., Uchuchoque, J.A.L., Arguedas, A., Macedo, L.,
 599 Manrique Llerena, N., 2016. Building vulnerability to hydro-geomorphic hazards: Estimating
 600 damage probability from qualitative vulnerability assessment using logistic regression. *J.*

- Hydrol. 541, 563–581. doi:10.1016/j.jhydrol.2015.04.017
- Fennell, S.G., Mackin, G.M., Madden, J.S., McGarry, A.T., Duffy, J.T., Colgan, P.A., Pollard, D., 2002. Radon in Dwellings The Irish National Radon Survey.
- Ferreira, A., Daraktchieva, Z., Beamish, D., Kirkwood, C., Lister, T.R., Cave, M., Wragg, J., Lee, K., 2016. Indoor radon measurements in south west England explained by topsoil and stream sediment geochemistry, airborne gamma-ray spectroscopy and geology. *J. Environ. Radioact.*
- Field, R.W., 2015. Radon: An Overview of Health Effects, in: *Reference Module in Earth Systems and Environmental Sciences*. Elsevier. doi:10.1016/B978-0-12-409548-9.09545-2
- Gray, A., Read, S., McGale, P., Darby, S., 2009. Lung cancer deaths from indoor radon and the cost effectiveness and potential of policies to reduce them. *BMJ* 338, a3110.
- GSI, 2005. Water Framework Directive (WFD) River Basin District Management Systems. Approach to Delineation of groundwater bodies. Guidance document no. GW2.
- Gunby, J.A., Darby, S.C., Miles, J.C.H., Green, B.M.R., Cox, D.R., 1993. Factors Affecting Indoor Radon Concentrations in the United Kingdom. *Health Phys.* 64, 2–12. doi:10.1097/00004032-199301000-00001
- Gunning, G.A., Pollard, D., Finch, E.C., 2014. An outdoor radon survey and minimizing the uncertainties in low level measurements using CR-39 detectors. *J. Radiol. Prot.* 34, 457–467. doi:10.1088/0952-4746/34/2/457
- Hodgson, J., Carey, S., Scanlon, R., 2014. Developing a new National Radon Risk Map. Dublin, Ireland.
- HSE, 2013. Pattern of radon levels - Ireland Data source : RPII. Ireland.
- Hunter Williams, N.H., Misstear, B.D.R., Daly, D., Lee, M., 2013. Development of a national groundwater recharge map for the Republic of Ireland. *Q. J. Eng. Geol. Hydrogeol.* 46, 493–506. doi:10.1144/qjegh2012-016
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*, Springer Texts in Statistics. Springer New York, New York, NY. doi:10.1007/978-1-4614-7138-7
- Kim, S.-H., Hwang, W.J., Cho, J.-S., Kang, D.R., 2016. Attributable risk of lung cancer deaths due to indoor radon exposure. *Ann. Occup. Environ. Med.* 28, 8. doi:10.1186/s40557-016-0093-4
- Kreienbrock, L., 2001. Case-Control Study on Lung Cancer and Residential Radon in Western Germany. *Am. J. Epidemiol.* 153, 42–52. doi:10.1093/aje/153.1.42
- Kropat, G., Bochud, F., Murith, C., Palacios (Gruson), M., Baechler, S., 2017. Modeling of geogenic radon in Switzerland based on ordered logistic regression. *J. Environ. Radioact.* 166, 376–381. doi:10.1016/j.jenvrad.2016.06.007
- Lee, M., Hunter-Williams, N., Meehan, R., Kelly, C., Kabza, M., Murphy, O., Spillane, M., 2008. Groundwater vulnerability mapping, in: *Irish National Hydrology Conference*. Tullamore,

- 639 Ireland.
- 640 Long, S., Fenton, D., Cremin, M., Morgan, a, 2013. The effectiveness of radon preventive and
 641 remedial measures in Irish homes. *J. Radiol. Prot.* 33, 141–9. doi:10.1088/0952-
 642 4746/33/1/141
- 643 Masterson, S., Kelly, C., Lee, M., 2008. County Cavan Groundwater Protection Scheme Volume I :
 644 Main Report Final.
- 645 McColl, N., Auvinen, A., Kesminiene, A., Espina, C., Erdmann, F., de Vries, E., Greinert, R., Harrison,
 646 J., Schüz, J., 2015. European Code against Cancer 4th Edition: Ionising and non-ionising
 647 radiation and cancer. *Cancer Epidemiol.* 39 Suppl 1, S93–100.
 648 doi:10.1016/j.canep.2015.03.016
- 649 Miksova, J., Barnet, I., 2002. Geological support to the national radon programme (Czech
 650 Republic). *Bull. Czech Geol. Surv.* 77, 13–22.
- 651 Miles, J.C.H., Appleton, J.D., 2005. Mapping variation in radon potential both between and within
 652 geological units. *J. Radiol. Prot.* 25, 257–276. doi:10.1088/0952-4746/25/3/003
- 653 Notario del Pino, J.S., Ruiz-Gallardo, J.-R., 2015. Modelling post-fire soil erosion hazard using
 654 ordinal logistic regression: A case study in South-eastern Spain. *Geomorphology* 232, 117–
 655 124. doi:10.1016/j.geomorph.2014.12.005
- 656 Pásztor, L., Szabó, K.Z., Szatmári, G., Laborczi, A., Horváth, Á., 2016. Mapping geogenic radon
 657 potential by regression kriging. *Sci. Total Environ.* 544, 883–891.
 658 doi:10.1016/j.scitotenv.2015.11.175
- 659 R Core Team, 2016. R: A language and environment for statistical computing.
- 660 Rochford, H., Fegan, M., Fenton, D., 2010. An Assessment of Radon Exposure to CR-39 Alpha
 661 Track Detectors during Postal Transit in Ireland, in: Third European IRPA Congress. p. 39.
- 662 Scheib, C., Appleton, J., Miles, J., Hodgkinson, E., 2013. Geological controls on radon potential in
 663 England. *Proc. Geol. Assoc.* 124, 910–928. doi:10.1016/j.pgeola.2013.03.004
- 664 Tollefsen, T., Cinelli, G., Bossew, P., Gruber, V., De Cort, M., 2014. From the European indoor
 665 radon map towards an atlas of natural radiation. *Radiat. Prot. Dosimetry* 162, 129–134.
 666 doi:10.1093/rpd/ncu244
- 667 UNSCEAR, 2000a. Report of the united nations scientific committee on the effects of atomic
 668 radiation to the general assembly. *Sources Eff. Ioniz. Radiat.*
- 669 UNSCEAR, 2000b. Exposures from natural radiation sources. Rep. 2000, Annex B 84–141.
- 670 US-EPA, 2003. EPA Assessment of Risks from Radon in Homes, Office of Radiation and Indoor Air.
 671 Washington, DC.
- 672 US Cancer Statistic Working Group, 2016. United States Cancer Statistics: 1999–2013 Incidence
 673 and Mortality Web-based Report. Atlanta: U.S. Department of Health and Human Services,
 674 Centers for Disease Control and Prevention and National Cancer Institute. Available at:
 675 www.cdc.gov/uscs [WWW Document].
- 676 US EPA, 2001. Building Radon Out. A Step-by-Step Guide On How To Build Radon-Resistant

Homes.

- Wang, Y., Song, C., Lin, Q., Li, J., 2016. Occurrence probability assessment of earthquake-triggered landslides with Newmark displacement values and logistic regression: The Wenchuan earthquake, China. *Geomorphology* 258, 108–119. doi:10.1016/j.geomorph.2016.01.004
- Wang, Z., Lubin, J.H., Wang, L., Zhang, S., Boice, J.D., Cui, H., Zhang, S., Conrath, S., Xia, Y., Shang, B., Brenner, A., Lei, S., Metayer, C., Cao, J., Chen, K.W., Lei, S., Kleinerman, R.A., 2002. Residential radon and lung cancer risk in a high-exposure area of Gansu Province, China. *Am. J. Epidemiol.* 155, 554–564. doi:10.1093/aje/155.6.554
- Watson, R.J., Smethurst, M.A., Ganerød, G. V., Finne, I., Rudjord, A.L., 2017. The use of mapped geology as a predictor of radon potential in Norway. *J. Environ. Radioact.* 166, 341–354. doi:10.1016/j.jenvrad.2016.05.031
- WHO, 2010. WHO Handbook on Indoor Radon: A Public Health Perspective. World Health Organization, France.
- Yang, Q., Smitherman, P., Hess, C.T., Culbertson, C.W., Marvinney, R.G., Smith, A.E., Zheng, Y., 2014. Uranium and radon in private bedrock well water in Maine: Geospatial analysis at two scales. *Environ. Sci. Technol.* 48, 4298–4306. doi:10.1021/es405020k

Figure captions

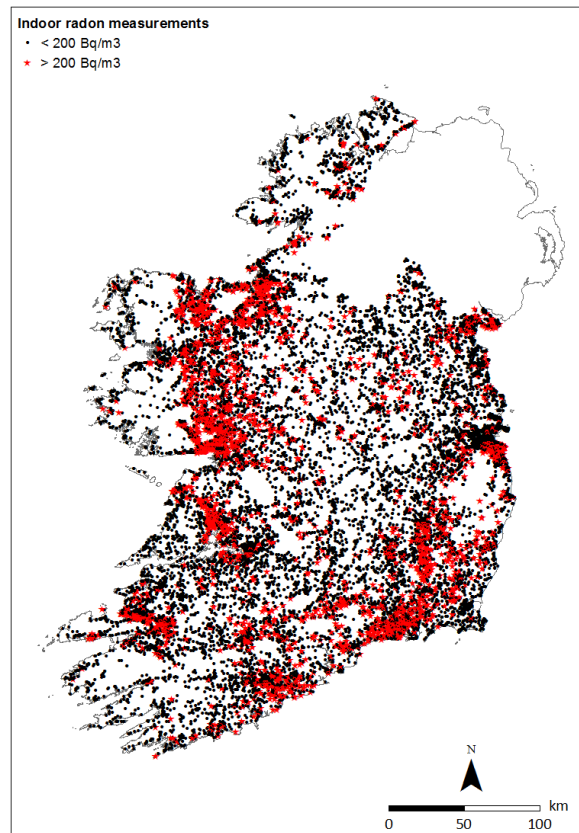


Figure 1: Indoor radon measurements in Ireland

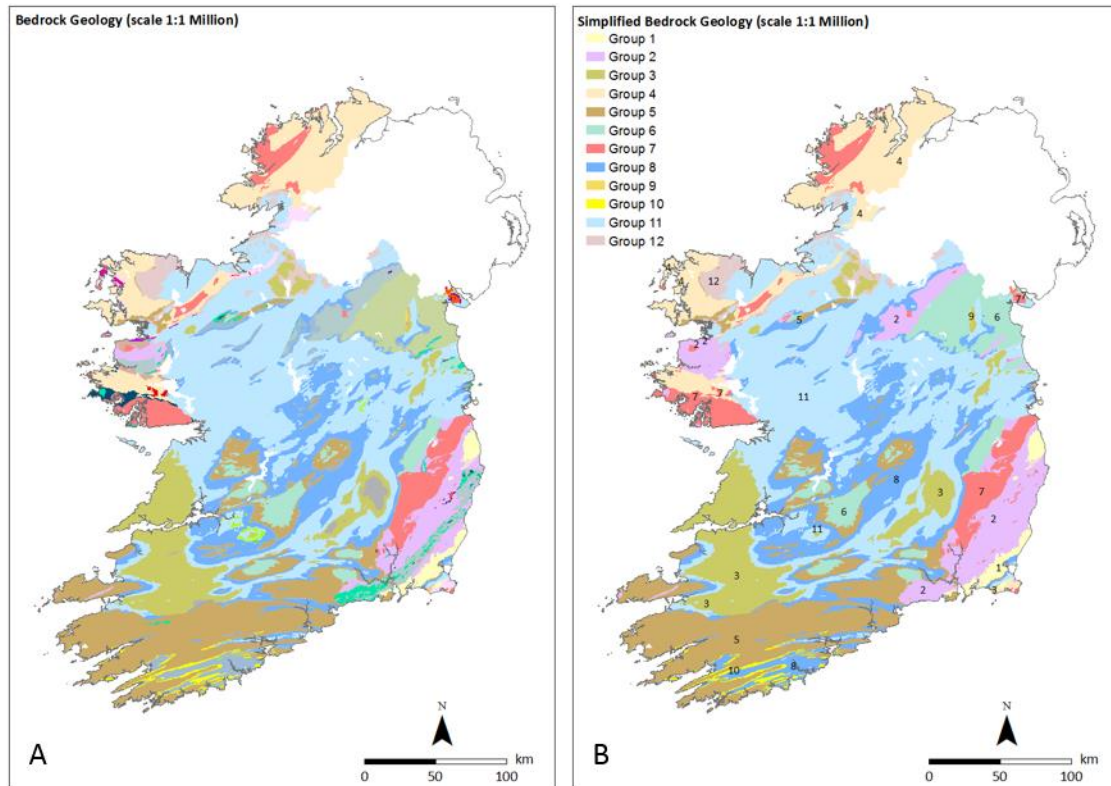


Figure 2: a) Bedrock Geology Map of Ireland (scale 1:1 million, the map with a complete legend can be viewed in the GSI Spatial Resources Viewer portal; www.gsi.ie); and b) Simplified Bedrock Geology Map of Ireland

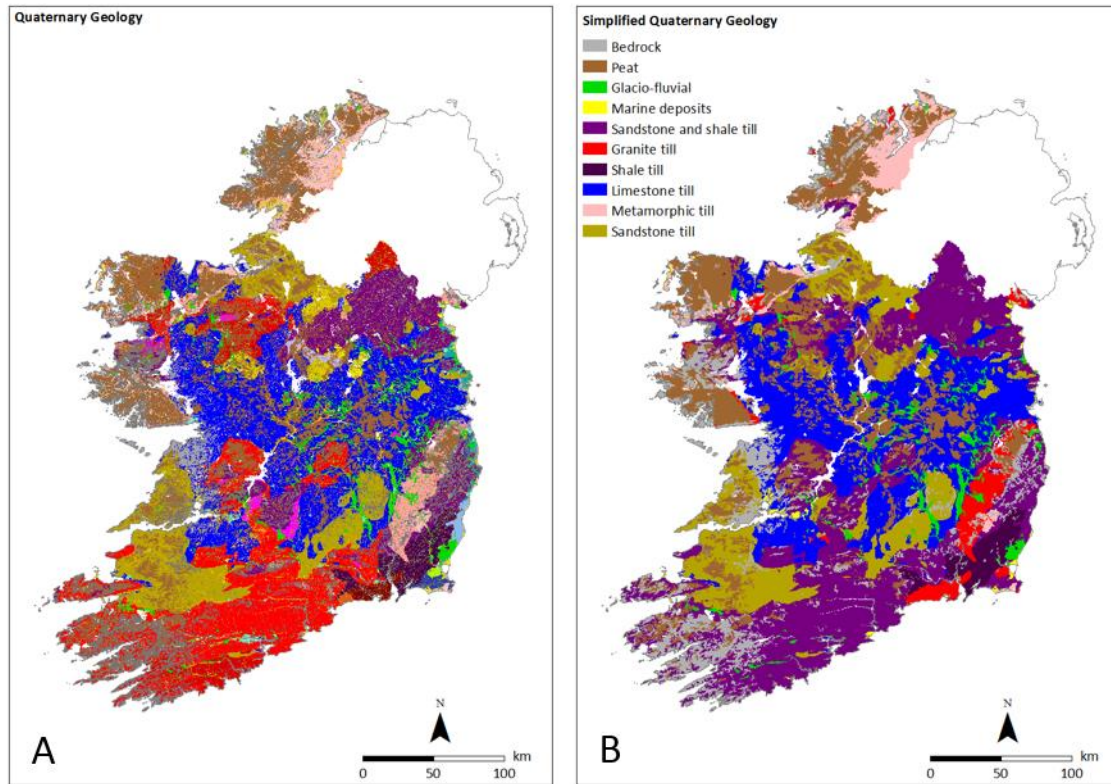


Figure 3: a) Original Quaternary Geology Map (the map with a complete legend can be viewed in the GSI Spatial Resources Viewer portal; www.gsi.ie); and b) simplified Quaternary Geology Map

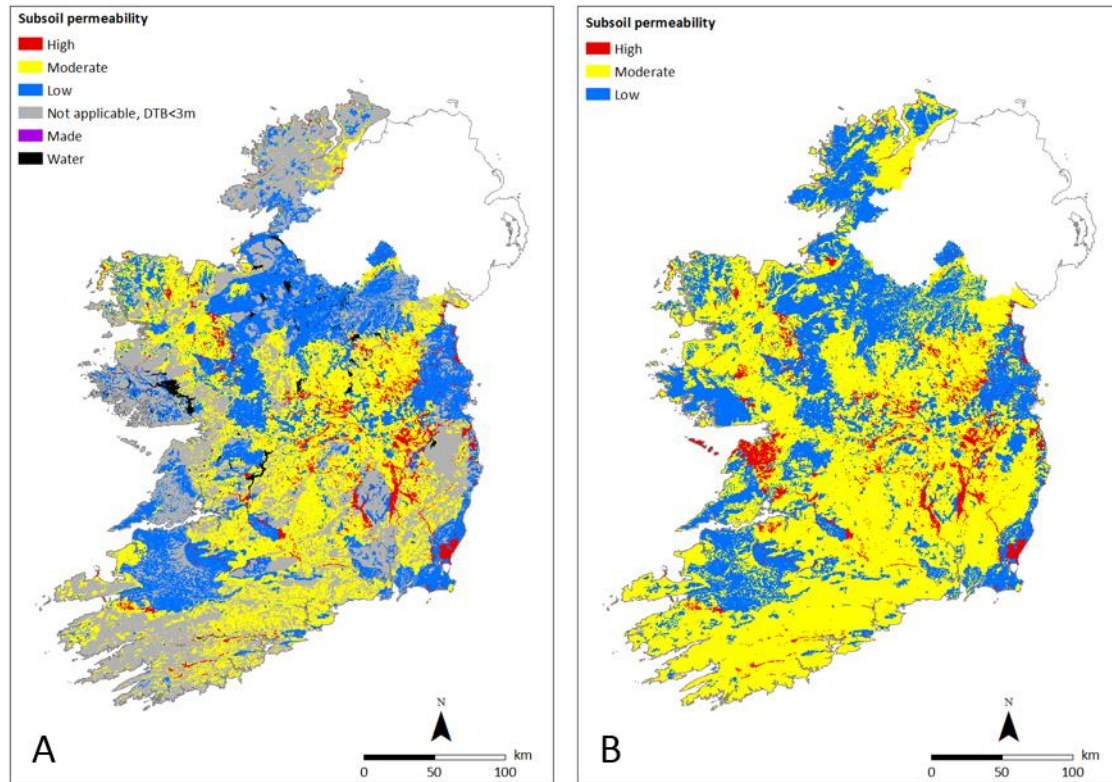


Figure 4: a) Subsoil permeability from the National Groundwater Recharge Map (GSI; www.gdi.ie); and b) subsoil permeability map of Ireland based on the National Groundwater Recharge Map and Quaternary geology

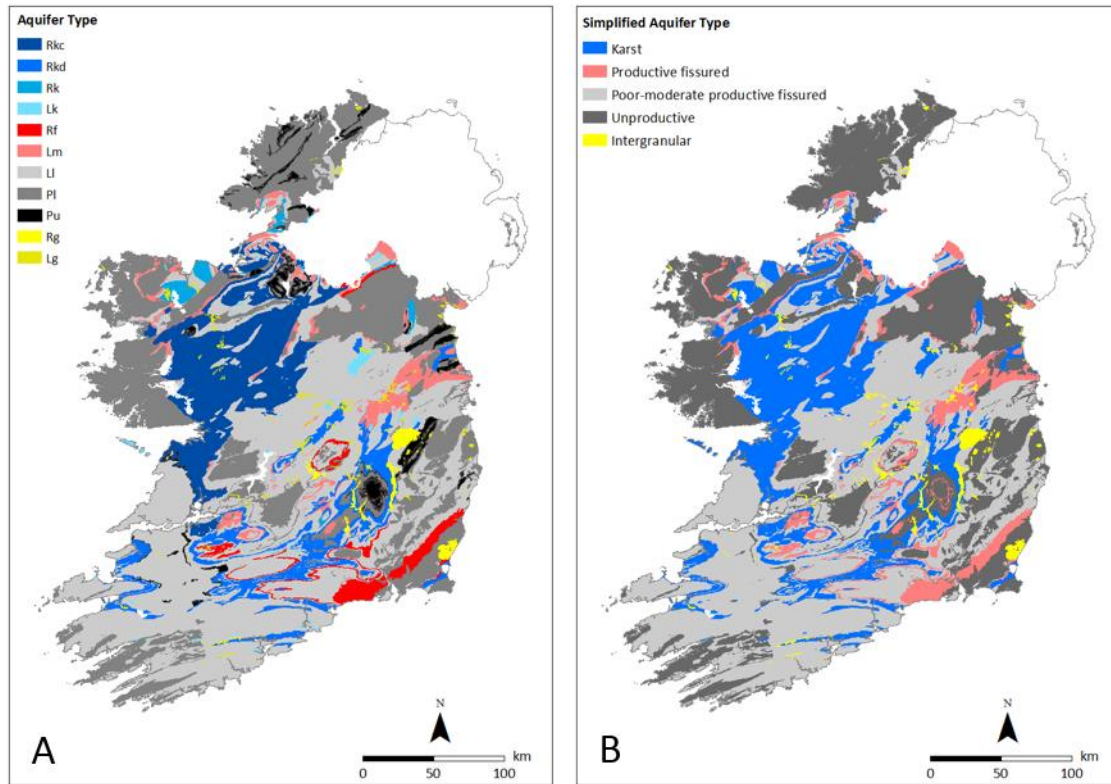


Figure 5: a) Groundwater Aquifer Map of Ireland (GSI; www.gsi.ie), and b) simplified aquifer type

Histogram

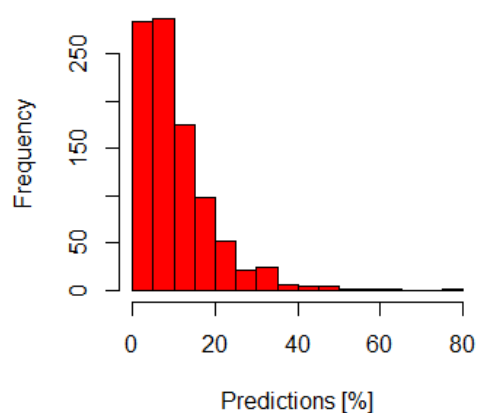


Figure 6: Histogram of the predicted probabilities of having an indoor radon concentration above 200 Bq m⁻³ in the 960 different polygons from the intersection of the four geological variables

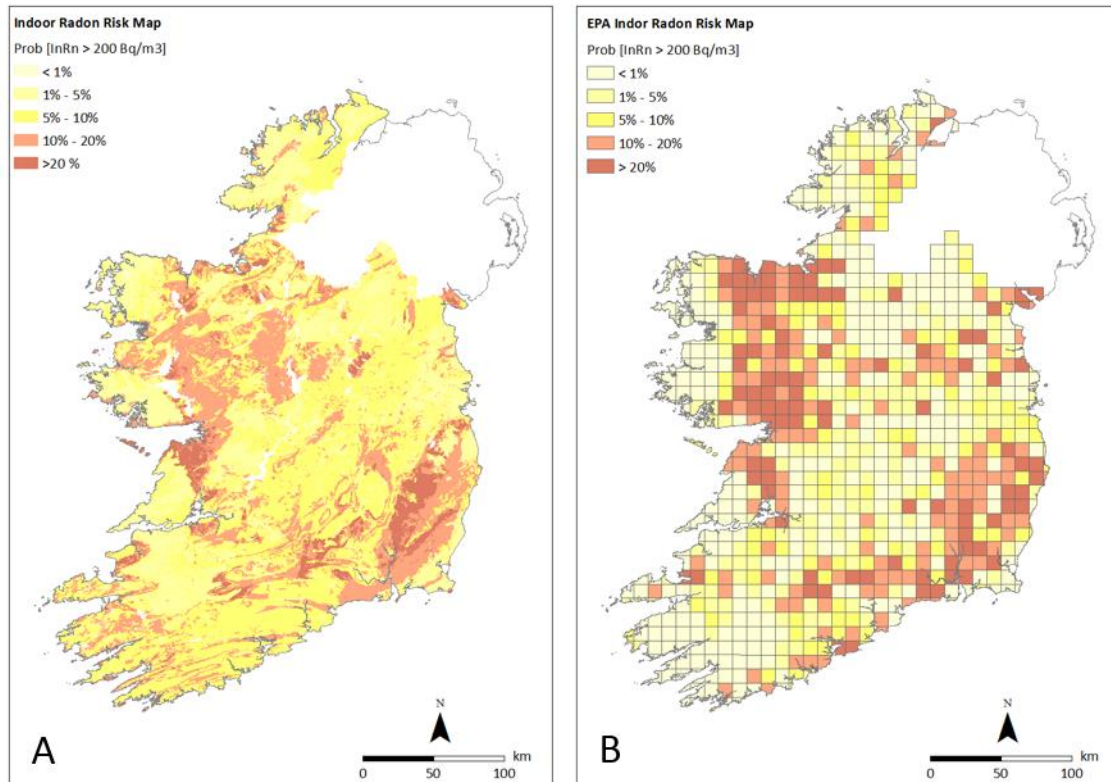


Figure 7: a) A new Indoor Radon Map for the Republic of Ireland. Predictions based on existing indoor radon measurements, bedrock geology (Scale 1:1 Million), Quaternary geology, subsoil permeability and aquifer type; b) EPA Indoor Radon Risk Map (Grids 100 km²; source: www.epa.ie/radon)

Table captions

Table 1: Grouping of the bedrock geological units (scale 1:1 Million)

Group	Unit Name	Area (km ²)	% total	Indoor Radon					
				N _i	N _T	N _T >RL	GM	GSD	Prob
1	Cambrian greywacke, slate, quartzite	702	1.02	400	400	13	46.17	2.36	4.37
2	Lower-Middle Ordovician slate, sandstone, greywacke, conglomerate	2,637	3.83	963	2,946	437	65.70	3.01	15.63
	Middle-Upper Ordovician slate, sandstone, greywacke, conglomerate	1,998	2.90	1,141					
	Ordovician volcanic rocks	587	0.85	769					
	Silurian terrestrial - shallow marine sandstone, siltstone, conglomerate	259	0.38	48					
	Serpentinite and sedimentary melange (Palaeozoic)	33	0.05	25					
3	Namurian shale, sandstone, siltstone & coal	5,493	7.97	1793	1,852	110	38.98	2.86	6.01
	Westphalian shale, sandstone, siltstone & coal	344	0.50	50					
	Permian sandstone, conglomerate, evaporite	25	0.04	9					
	Cretaceous chalk, flint, glauconitic sandstone	1	0.00	0					
4	Neoproterozoic metasedimentary rocks - Dalradian	5,519	8.01	1,407	1,556	77	44.48	2.76	6.97
	Neoproterozoic schist and gneiss	438	0.64	123					
	Mesoproterozoic gneiss	47	0.07	15					
	Palaeoproterozoic gneiss	16	0.02	11					
5	ORS, sandstone, conglomerate & mudstone	10,240	14.8	2,019	2,025	155	46.83	2.66	6.86
	Devonian volcanic rocks	41	0.06	6					
6	Late Ordovician-Silurian deep marine greywacke, mudstone (Longford-Down)	2,355	3.42	1,405	1,890	119	43.74	2.75	6.64
	Silurian deep marine mudstone, greywacke & conglomerate	1,950	2.83	485					
7	Siluro-Devonian granitic rocks & appinite	3,325	4.82	2,551	2,816	497	74.88	2.95	18.19
	Lower Palaeozoic basic-intermediate intrusion	249	0.36	175					
	Ordovician granitic rocks	54	0.08	0					
	Palaeogene granitic rocks	48	0.07	23					
	Palaeogene basic intrusive rocks	34	0.05	67					
8	Tournaisian limestone	7,903	11.46	4,007	5,120	498	53.47	2.72	9.35
	Tournaisian sandstone, mudstone, limestone	2,730	3.96	1,113					
9	Triassic sandstone, mudstone, evaporite	13	0.02	6	6	3	126.46	2.69	32.18
10	Up. Devonian sandstone & mudstone (Old Head Sandstone Formation)	485	0.70	134	134	18	65.29	2.60	12.08
11	Visean limestone & calcareous shale	19,858	28.81	12,714	12,777	1,887	62.65	3.11	15.33
	Carboniferous volcanic rocks	135	0.20	63					
12	Visean sandstone, mudstone & evaporite	1,416	2.05	388	388	37	41.23	3.02	7.64

N_i: number of samples in each geological unit; N_T: number of samples in each geological group; RL: Reference Level (200 Bq m⁻³), GM and GSD in Bq m⁻³. Prob: Probability (%) of having an indoor radon concentration above 200 Bq m⁻³ in a log-normal distribution

737

Table 2: Grouping of the Quaternary Sediments

Quaternary Sediment			Indoor Radon				
New group	Sediments*		N	N > RL	GM	GSD	Prob
Peats:							
Peat	All types		1,197	71	35.98	2.93	5.54
Glacio-fluvial sands and gravels:							
Glacio-fluvial	All types		1,679	116	52.87	2.50	7.35
Marine deposits:							
Marine deposits	All types		628	20	31.44	2.60	2.62
Tills:							
Granite till	TA _v TGr	TQz	3,070	643	83.22	3.11	21.95
Limestone till	TL _s		9,745	1140	57.99	2.91	12.32
Metamorphic till	TM _p		1,842	195	58.10	2.86	11.98
Sandstone and shale till	IrSTLPSS TCSs TDCSs	TLCSS TLPSS	7,788	701	50.89	2.73	8.62
Sandstone till	TCSsCh TLPDSs	TNSSs	3,813	541	56.69	3.36	14.92
Shale till	TLPS		676	124	81.82	2.69	18.32
Other categories:							
Bedrock	KaRck	Rck	1472	300	66.96	3.41	18.63

*Teagasc Subsoil Classes (GSI, 2013). N: number of samples; RL: Reference Level (200 Bq m⁻³), GM and GSD in Bq m⁻³. Prob: Probability (%) of having an indoor radon concentration above 200 Bq m⁻³ in a log-normal distribution

738

739

740

Table 3: Quaternary geology and subsoil permeability

Description*	Subsoil Permeability
Peat	Low
Glacio-fluvial sands and gravels	High
Alluvium (silt-clay)	Low
Alluvium (sand-gravel-undifferentiated)	Moderate
Lacustrine sediments (silt-clay)	Low
Lacustrine sediments (sand-gravel)	Moderate
Tills	Moderate
Bedrock at surface	Moderate
Bedrock at surface (karstified)	High
Marine deposits (gravel)	High
Marine deposits (sand-gravel)	Moderate
Marine deposits (silts-clays)	Low

*Source: Geological Survey Ireland (GSI, 2013)

741

742

Table 4: Summary statistics of indoor radon measurements by subsoil permeability

Subsoil Permeability	N	N > RL	GM	GSD	Prob
High (H)	2,531	356	63.18	2.89	13.87
Moderate (M)	21,436	3,005	62.18	3.03	14.58
Low (L)	7,943	490	43.04	2.72	6.23

N: number of samples; RL: Reference Level (200 Bq m⁻³), GM and GSD in Bq m⁻³. Prob: Probability (%) of having an indoor radon concentration above 200 Bq m⁻³ in a log-normal distribution

Table 5: Grouping of the Aquifer Types

Group	Aquifer Type	Bedrock Aquifer Description	Indoor Radon				
			N	N > RL	GM	GSD	Prob
1	Karst	Regionally Important Aquifer - Karstified; conduit (Rkc)	9,609	1,808	73.42	3.18	19.32
		Regionally Important Aquifer - Karstified; diffuse (Rkd)					
		Regionally Important Aquifer - Karstified (Rk)					
		Locally Important Aquifer - Karstified (Lk)					
2	Productive fissured	Regionally Important Aquifer - Fissured bedrock (Rf)	3,224	377	53.57	3.09	12.12
		Locally Important Aquifer - Bedrock which is Generally Moderately Productive (Lm)					
3	Poor-Moderate productive fissured	Locally Important Aquifer - Bedrock which is Moderately Productive only in Local Zones (LI)	10,290	854	48.88	2.76	8.27
4	Unproductive	Poor Aquifer - Bedrock which is Generally Unproductive except for Local Zones (PI)	7,863	744	52.25	2.85	10.03
		Poor Aquifer - Bedrock which is Generally Unproductive (Pu)					
5	Intergranular	Regionally Important Sand/Gravel Aquifer (Rg)	924	68	52.60	2.49	7.11
		Locally Important Sand/Gravel Aquifer (Lg)					

N: number of samples; RL: Reference Level (200 Bq m⁻³), GM and GSD in Bq m⁻³. Prob: Probability (%) of having an indoor radon concentration above 200 Bq m⁻³ in a log-normal distribution

749 Table 6: Statistics of all possible models from combinations of the four explanatory variables

Model	AIC	logLik	P (df)	Pr(>Chi)*
NULL	23505.88	-11751.94	1	-
AT	22924.74	-11457.37	5	< 2e-16
QG	22968.57	-11474.28	10	< 2E-16
BG	22991.48	-11483.74	12	< 2e-16
SP	23116.97	-11555.48	3	< 2e-16
QG+AT	22184.46	-11078.23	14	< 2e-16
BG+QG	22476.04	-11217.02	21	< 2e-16
BG+AT	22479.44	-11223.72	16	< 2e-16
QG+SP	22574.36	-11275.18	12	< 2e-16
BG+SP	22595.39	-11283.7	14	< 2e-16
SP+AT	22635.26	-11310.63	7	< 2e-16
QG+SP+AT	21977.94	-10972.97	16	< 2e-16
BG+QG+AT	22060.94	-11005.47	25	< 2e-16
BG+QG+SP	22140.74	-11047.37	23	< 2e-16
BG+SP+AT	22212.37	-11088.18	18	< 2e-16
BG+QG+SP+AT	21847.64	-10896.82	27	< 2e-16

(*)Likelihood ratio tests; significance level = 0.01

750

751

Table 7: Regression coefficients (β) of the logistic regression

Coefficients:	Estimate	Std. Error	z value	Pr(> z)	
β_0 (Intercept)	-4.699	0.338	-13.910	< 2e-16	***
β_1 BG: Group 1	0.000	-	-	-	-
BG: Group 9	3.559	0.874	4.073	0.000	***
BG: Group 7	1.457	0.291	5.006	0.000	***
BG: Group 12	1.421	0.348	4.082	0.000	***
BG: Group 10	1.352	0.395	3.425	0.001	***
BG: Group 2	1.123	0.297	3.783	0.000	***
BG: Group 11	0.981	0.303	3.238	0.001	**
BG: Group 6	0.971	0.304	3.195	0.001	**
BG: Group 8	0.738	0.304	2.427	0.015	*
BG: Group5	0.730	0.310	2.352	0.019	*
BG: Group 4	0.450	0.319	1.408	0.159	
BG: Group 3	0.350	0.316	1.107	0.268	
β_2 QG: Glacio-fluvial	0.000	-	-	-	-
QG: Shale till	1.858	0.172	10.810	< 2e-16	***
QG: Granite till	1.789	0.141	12.730	< 2e-16	***
QG: Bedrock	1.419	0.132	10.735	< 2e-16	***
QG: Sandstone till	1.314	0.134	9.811	< 2e-16	***
QG: Metamorphic till	1.054	0.153	6.898	0.000	***
QG: Sandstone and shale till	0.813	0.129	6.325	0.000	***
QG: Limestone till	0.772	0.126	6.151	0.000	***
QG: Peat	0.531	0.174	3.053	0.002	**
QG: Marine deposits	-0.075	0.255	-0.293	0.770	
β_3 SP: Low	0.000	-	-	-	-
SP: High	1.057	0.095	11.170	< 2e-16	***
SP: Moderate	0.713	0.054	13.130	< 2e-16	***
β_4 AT: Intergranular	0.000	-	-	-	-
AQ: Karstified	0.743	0.150	4.946	0.000	***
AT: Poor-moderate productive	-0.034	0.154	-0.224	0.823	
AT: Productive fissured	-0.386	0.169	-2.283	0.022	*
AT: Unproductive	-0.387	0.168	-2.296	0.022	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

752

Table 8: Top ten and bottom ten predicted probabilities ($P[Y=1|X]$) and 95% confidence intervals of the 960 different polygons from the intersection of the four geological variables

Explanatory variable				Area	Predictions (%)		
Bedrock (Group)	Quaternary Geology	Subsoil Perm.	Aquifer (Group)	km ²	Prob.	Lower	Upper
Top ten							
9	Sandstone and shale till	M	1	0.08	75.57	38.21	93.93
9	Sandstone and shale till	L	1	0.31	60.26	23.18	88.40
9	Sandstone and shale till	H	2	1.49	58.52	22.03	87.57
9	Sandstone and shale till	H	4	0.00	58.51	21.81	87.70
7	Granite till	M	1	0.00	50.07	45.20	54.94
9	Sandstone and shale till	M	2	8.87	50.00	16.79	83.21
9	Sandstone and shale till	M	4	0.00	49.99	16.62	83.37
12	Granite till	M	1	0.13	49.17	39.02	59.40
11	Granite till	H	1	0.00	46.79	41.48	52.17
12	Sandstone till	H	1	0.00	45.92	36.11	56.06
Bottom ten							
3	Marine deposits	L	3	5.41	1.15	0.69	1.89
4	Glacio-fluvial	L	4	0.70	0.96	0.67	1.39
1	Glacio-fluvial	L	5	1.64	0.90	0.47	1.73
4	Marine deposits	L	4	1.41	0.89	0.52	1.52
1	Glacio-fluvial	L	3	0.00	0.87	0.47	1.63
3	Glacio-fluvial	L	2	0.12	0.87	0.61	1.25
3	Glacio-fluvial	L	4	2.28	0.87	0.61	1.24
3	Marine deposits	L	4	0.54	0.81	0.48	1.37
1	Glacio-fluvial	L	4	6.05	0.61	0.33	1.13
1	Marine deposits	L	4	3.10	0.57	0.28	1.17

Table 9: Predicted probability ($P[Y=1|X]$) and 95% confidence intervals in the 20 polygons that represent 50% of the total area of the country

Explanatory variable				Predictions (%)			Area		
Bedrock (Group)	Quaternary Geology	Subsoil Perm.	Aquifer (Group)	Prob.	Lower	Upper	km ²	%	Accu.
11	Sandstone till	M	1	27.94	25.77	30.21	812	1.19	1.2
11	Limestone Till	M	1	18.40	17.33	19.52	4,204	6.15	7.3
5	Bedrock	M	3	13.34	10.98	16.11	1,482	2.17	9.5
5	Bedrock	M	4	9.77	7.71	12.30	807	1.18	10.7
11	Limestone Till	M	3	9.39	8.48	10.38	2,289	3.35	14.0
3	Sandstone till	M	3	8.66	7.19	10.39	870	1.27	15.3
11	Peat	L	1	7.99	6.27	10.14	1,565	2.29	17.6
8	Sandstone and shale till	M	3	7.80	6.91	8.80	1,531	2.24	19.8
5	Sandstone and shale till	M	3	7.74	6.62	9.03	4,568	6.68	26.5
8	Limestone Till	M	3	7.52	6.67	8.46	2,448	3.58	30.1
6	Sandstone and shale till	M	4	6.99	5.82	8.36	1,986	2.90	33.0
4	Metamorphic till	M	4	5.37	4.29	6.70	1,550	2.27	35.3
11	Limestone Till	L	3	4.84	4.26	5.48	869	1.27	36.5
3	Sandstone till	L	3	4.44	3.63	5.42	1,600	2.34	38.9
7	Peat	L	4	4.32	3.27	5.70	1,439	2.11	41.0
11	Peat	L	3	3.84	2.96	4.96	876	1.28	42.3
6	Sandstone and shale till	L	4	3.55	2.90	4.35	1,094	1.60	43.9
8	Peat	L	3	3.04	2.31	3.97	817	1.20	45.1
3	Peat	L	3	2.08	1.51	2.85	831	1.21	46.3
4	Peat	L	4	1.62	1.15	2.29	2,519	3.68	50.0

Total area cover by the Indoor Radon Map 68,365 km². Subsoil Permeability: Moderate (M), Low (L)

Table 10: Confusion matrix of the new radon risk map (logistic regression model) and the EPA map (Grids 100 km²)

Radon Risk Map	Actual InRn	Number dwellings	Predicted group	
			NoHRA	HRA
EPA	≤200 Bq m ⁻³	28,058	16,436 (58.6%)	11,623 (41.4%)
	>200 Bq m ⁻³	3,851	975 (25.3%)	2,876 (74.7%)
	Percentage of dwellings correctly classified: 60.5%			
New Map	≤200 Bq m ⁻³	28,058	16,513 (58.9%)	11,546 (41.1%)
	>200 Bq m ⁻³	3,851	1,094 (28.4%)	2,757 (71.6%)
	Percentage of dwellings correctly classified: 60.4%			

Table 11: Area, total population, number of sampled dwellings, and probability of success (InRn>200 Bq m⁻³) and 95% confidence interval in each radon risk zone

Radon Risk		Area		Sampled dwellings			Binomial distribution			Population	
Class.	Prob.	km ²	%	< R.L.	> R.L.	Total	Prob.	LCI	UCI	Total	%
VL	< 1	16	0.02	42	2	44	4.55	0.56	15.47	10	0.23
L	1 - 5	17,887	26.16	5,402	180	5,582	3.22	2.78	3.72	1,069	23.36
M	5 - 10	30,089	44.01	11,069	912	11,981	7.61	7.14	8.10	2,126	46.48
H	10 - 20	16,373	23.95	8,661	1,747	10,408	16.79	16.07	17.52	1,128	24.66
VH	≥ 20	4,001	5.85	2,885	1,010	3,895	25.93	24.56	27.34	241	5.28
Total		68,365	100	28,059	3,851	31,910				4,574	100

Reference Level (R.L.) = 200 Bq m⁻³; Probabilities (Prob.) in %; Population in thousand; LCI, UCI are the 95% lower and upper confidence limit, respectively; VL: Very Low, L: Low, M: Moderate, H: High, VH: Very high