A new perspective in radon risk assessment: Mapping the geological hazard as a first step to define the collective radon risk exposure

Eleonora Benà [a,*], Giancarlo Ciotoli [b,c], Eric Petermann [d], Peter Bossew [d,1], Livio Ruggiero [e], Luca Verdi [f], Paul Huber [g], Federico Mori [b], Claudio Mazzoli [a], Raffaele Sassi [a]

[a] Dipartimento di Geoscienze, Università di Padova, Padova, Italy
[b] Istituto di Geologia Ambientale e Geoingegneria (IGAG), Consiglio Nazionale delle Ricerche (CNR), Roma, Italy
[c] Istituto Nazionale di Geofisica e Vulcanologia (INGV), Roma, Italy
[d] Federal Office for Radiation Protection (BfS), Section Radon and NORM, Berlin, Germany
[e] Istituto Superiore per la Ricerca e la Protezione Ambientale (ISPRA), Roma, Italy
[f] Provincia Autonoma di Bolzano, Laboratorio analisi aria e radioprotezione, Bolzano, Italy
[g] Azienda Sanitaria dell'Alto Adige, Bressanone, Italy

HIGHLIGHTS

- Mapping the Geogenic Radon Potential using a robust machine learning technique
- Apply the risk definition to define the Collective Radon Risk Areas
- Construction of the geological-based Collective Radon Risk Areas map
- Mapping the Collective Risk to define areas subject of territorial planning

GRAPHICAL ABSTRACT
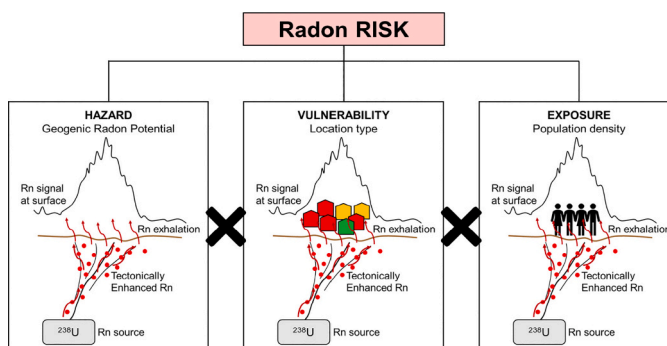
ARTICLE INFO

ABSTRACT

Radon is a radioactive gas and a major source of ionizing radiation exposure for humans. Consequently, it can pose serious health threats when it accumulates in confined environments. In Europe, recent legislation has been adopted to address radon exposure in dwellings; this law establishes national reference levels and guidelines for defining Radon Priority Areas (RPAs). This study focuses on mapping the Geogenic Radon Potential (GRP) as a foundation for identifying RPAs and, consequently, assessing radon risk in indoor environments. Here, GRP is proposed as a hazard indicator, indicating the potential for radon to enter buildings from geological sources. Various approaches, including multivariate geospatial analysis and the application of artificial intelligence

---

algorithms, have been utilised to generate continuous spatial maps of GRP based on point measurements. In this study, we employed a robust multivariate machine learning algorithm (Random Forest) to create the GRP map of the central sector of the Pusteria Valley, incorporating other variables from census tracts such as land use as a vulnerability factor, and population as an exposure factor to create the risk map. The Pusteria Valley in northern Italy was chosen as the pilot site due to its well-known geological, structural, and geochemical features. The results indicate that high Rn risk areas are associated with high GRP values, as well as residential areas and high population density. Starting with the GRP map (e.g., Rn hazard), a new geological-based definition of the RPAs is proposed as fundamental tool for mapping Collective Radon Risk Areas in line with the main objective of European regulations, which is to differentiate them from Individual Risk Areas.

1. Introduction

Radon (^{222}Rn, hereafter referred to as Rn) is a radioactive gas considered the primary source of ionizing radiation exposure for the population. Its potentially harmful effects on human health have been extensively documented (Zeeb and Shannoun, 2009). Rn gas represents a significant hazard when it accumulates in indoor environments, such as residential houses and workplaces (Indoor Radon).

Exposure to indoor radon is a serious issue that has prompted Europe to introduce legislation (Basic Safety Standards Directive 2013/59/ EURATOM). This legislation, on one hand, establishes maximal national reference levels aimed at reducing Indoor Radon Concentration (IRC) exposure. On the other hand, it encourages public administrations to define Radon Priority Areas (RPAs). Therefore, it is essential to identify areas characterised by the highest Rn hazard for the population.

Rn gas has a natural origin and its concentration in the environment can vary depending on the geological characteristics of an area. In particular, the Geogenic Radon Potential (GRP) can be considered an optimal Rn hazard indicator as it conceptualizes "*what Earth delivers in terms of radon*" from geogenic sources (e.g., radionuclides content in rocks and soils, faults and fractures) to the atmosphere (Bossew, 2015; Bossew et al., 2020).

The GRP is characterised by the interaction of three natural processes:

- Background Radon Source (BRS) represents the process that produces Rn through the natural decay of uranium (U) and thorium (Th) (^{220}Rn), which are present in rocks and soils.
- Tectonically Enhanced Radon (TER, Benà et al., 2022) accounts for processes that allow radon to migrate more easily towards the surface through permeable pathways (e.g., faults and fractures in the crust) from deeper sources, caused by increased stress and pressure conditions associated with tectonic activity.
- Surface Radon Exhalation (SRE) is the process by which radon gas is released from the ground into the atmosphere. SRE considers variables affecting radon movement in the shallow soil up to the soil/ atmosphere interface (e.g., land morphology, soil permeability, atmospheric pressure, humidity, and temperature). This quantity of radon represents the amount that could potentially enter buildings, although BRS and TER are the dominant geological radon sources.

Over the years, several approaches have been applied to estimate the GRP over an area (e.g., Neznal et al., 2004; Bossew, 2015; Pásztor et al., 2016; Ciotoli et al., 2017; Giustini et al., 2019; Petermann et al., 2021; Coletti et al., 2022).

A widely used method to define the GRP, due to its simplicity, was proposed by Neznal et al. (2004). Neznal's approach is based on the measurement of two quantities: Rn concentration in the soil and soil permeability.

More recently, Pásztor et al. (2016) and Ciotoli et al. (2017) applied multivariate geospatial analysis (regression kriging and geographical weighted regression, respectively) to model GRP, using Soil Gas Radon Concentration (SGRC) and selected environmental proxies for the first time.

In the past three years, researchers have developed more advanced multivariate techniques, such as regression kriging (Coletti et al., 2022) and Machine Learning (ML, Petermann et al., 2021) algorithms which include several predictor variables associated with the geogenic Rn component. However, it is important to emphasise that all these techniques require the measurement of SGRC or IRC as a response variable.

An issue faced by many European nations is the lack of sufficient SGRC measurements to support GRP mapping. As a consequence, the Geogenic Radon Hazard Index was proposed (Bossew et al., 2020). The Geogenic Radon Hazard Index concept arose from the need to determine a particular indicator using regionally accessible geological variables. The Geogenic Radon Hazard Index values for regions with consistent geogenic factors but different data sources should be comparable.

Cinelli et al. (2015) proposed a method for achieving consistency. It involved assigning weights to continuous or categorical input variables (covariates) based on their contribution to the index. Another way to enhance data consistency is the application of techniques that do not require a response variable (i.e., Spatial Multi Criteria Decision Analysis, Ciotoli et al., 2020).

It is crucial to map GRP as accurately as possible using a robust methodology, as GRP represents the amount of radon that could potentially enter buildings and is considered the most significant spatial predictor of IRC. In this context, the Basic Safety Standards European Directive 59/2013, transposed into Italian law by Legislative Decree n.101/2020, further emphasises the importance of identifying RPAs. Originally, RPAs were defined as areas where the annual average IRC in a significant number of dwellings is expected to exceed the reference level of 300 Bq m^{-3}. However, the concept and interpretation of "significant number of buildings" in the European Directive remained unclear.

Recently, Petermann et al. (2022) highlighted that the interpretation of "significant number" of buildings is factually based on the concept of geogenic hazard, and it relates to a relevant percentage of buildings within an area, regardless the number of houses or people affected. This concept does not encompass the collective concept of geogenic risk.

On the other hand, there is no uniform decision at the regional scale regarding the selection of the reference level and the threshold of probability percentage (p_0) of buildings exceeding the reference level. In general, the majority of European nations (including Finland, Germany, Greece, Montenegro, and Spain) adhere to the European Directive, adopting the recommended reference level of 300 Bq m^{-3} and a probability threshold of 10 % (Bossew, 2018). For instance, Italy has a reference level of 300 Bq m^{-3} and a p_0 of 15 % (D. Lgs. n. 101/2020). A map illustrating the confusing diversity of RPA definitions across Europe has been compiled in Bossew and Suhr (2023, see Fig. 2 in the cited paper).

As reported in Bossew and Petermann (2021), the goals of the legislation in term of radiation protection from Rn indoor are twofold: (i) protect people from high Rn exposure to reduce individual risk (even if few people are involved); (ii) avoid high exposure to the community: the harm to society is proportional to the collective risk.

European legislation aims to reduce the detriment from Rn exposure (i.e., the number of lung cancer deaths) and, as a consequence, reduce collective exposure. Collective exposure can be assessed by introducing

the concept of collective risk, which complements the individual risk concept (the "classical" RPA). Collective risk can be understood as consisting of many small individual risk zones.

Based on these considerations, we propose mapping Collective Radon Risk Areas (CRAs) by applying the definition of risk, which combines hazard, vulnerability and exposure factors. We use this concept as complementary to mapping Individual Risk Areas (IRAs) associated with IRC (i.e., "classical" RPA).

The main goal of this research is to demonstrate in a test area the efficacy of CRAs maps in defining Rn risk areas. We base this on the GRP map (the hazard factor) elaborated using a ML technique (i.e., Random Forest, RF). The test area chosen is the Pusteria Valley, which has been selected because of its well-known geological, structural, and geochemical characteristics, and the availability of numerous additional data (Benà et al., 2022). The obtained GRP map (hazard) was merged with the land use type (vulnerability) and population (exposure) data from census tracts available from the ISTAT (*Istituto Nazionale di Statistica*) website to identify CRAs.

2. Methods

2.1. Test area

The test area is located in the Pustertal/Pusteria Valley in the Italian Eastern Alps (Bolzano province, see Fig. 1 in Benà et al., 2022) and it has been selected because of its well-known geological and structural characteristics. Into details, the study area is located across the complex Pusteria Fault System which comprises three main faults: the Defereggen-Anterselva-Valles (DAV) mylonitic zone (Müller et al., 2000); the Kalkstein-Vallarga (KV) fault (Borsi et al., 1978) and the Pusteria fault (Schmid et al., 1989). The Pusteria fault separates two different domains in the study area: (i) the Austroalpine crystalline basement (to the north) that is mainly composed by micashists and paragneiss (locally grading to migmatites) (Sassi et al., 2004); (ii) the Southalpine crystalline basement (to the south) which consists of a thick phyllitic sequence (Spiess et al., 2010) with Permian granite intrusions. The test area and the fault system are also well known from a geochemical point of view since its high degassing has been previously documented in Benà et al. (2022).

Geographically, the study area includes a part of three main municipalities: Terenten/Terento (~1800 inhabitants) to the western part, Kiens/Chienes (~3000 inhabitants) to the central part and Pfalzen/Falzes (~2900 inhabitants) to the eastern part. In the mentioned municipalities, the population is mainly grouped in residential areas and housing units; however, most of the study area is mountainous and sparsely populated. In general, the residential houses have 3 to 4 floors spanning from the basement (−1) and the second floor (2) where present; the building materials are mainly brick and concrete, occasionally local rocks and wood (Verdi et al., 2004).

The test zone has been also chosen due to the availability of numerous additional data (see for example the in-situ measurements from Benà et al., 2022; the primary base maps available online in the Bolzano Province Geo-catalogue; the census variables and the demographic data available on the online ISTAT-*Istituto Nazionale di Statistica* website; indoor radon surveys reported in Minach et al., 1999 and Verdi et al., 2004, but not utilised in the current work to define the CRAs).

2.2. Experimental strategy

A dataset including different variables (e.g., response and predictors) was used to elaborate the GRP map for the study area. This was achieved by employing a ML technique, specifically the Random Forest (Breiman, 2001), to predict radon values at grid points. The resulting GRP map served as the hazard factor in the risk equation (Eq. (2), as detailed in Section 2.4) and was multiplied by census tract data for land use and population density, representing the vulnerability and the exposure factors, respectively.

Fig. 1 shows the flowchart of the applied procedures. Data processing was carried out using ArcGIS Pro 3.1.2 (copyright 2023@ESRI Inc.) and Scikit-learn library in Python PyCharm 2023.1.2 (Copyright © 2010–2023 JetBrains s.r.o.).

2.3. Dataset

The dataset comprises one response variable (SGRC) and ten independent variables, which were either measured on-site or derived from primary base maps available online through the Bolzano Province Geo-catalogue (http://geokatalog.buergernetz.bz.it/geokatalog/#!). These ten variables were selected as potential predictors for ML regression models.

Soil gas surveys (^{222}Rn, ^{220}Rn, CO_2) (Benà et al., 2022), Terrestrial Gamma Dose Rate (TGDR), and permeability measurements were collected on-site during two separate field campaigns in the summers 2021 and 2022, conducted under similar and stable climatic conditions. The Digital Terrain Model (DTM) at a 2.5 m resolution and fault density data were obtained from the base maps available in the Bolzano Province Geo-catalogue.

The potential predictors underwent pre-processing using geospatial analysis to generate 50×50 m raster maps (refer to Fig. S1 in Supplementary materials). We used the "Extract multi-value to point" tool in ArcGIS Pro to assign values of the predictors to each node of the raster maps, ensuring that we had a value for each predictor variable for each Rn observation. The resulting dataset, containing the predictors and the response variable (SGRC), was used to train the Random Forest (RF) model. Once the best model was identified, it was applied to a regular 50x50m point fishnet, corresponding to the centroids of the predictor raster grid. The final dataset consists of 27,758 points, encompassing complete information for all predictors. The following sections provide a detailed description of the response variable and the predictors.

2.3.1. Response variable

SGRC (kBq m^{-3}) was used as the response (dependent) variable in the Random Forest Regression algorithm to create the GRP map. The original Rn dataset is composed by soil gas radon surveys obtained in the field according to the methodology and sample pattern described in detail by Benà et al. (2022). In particular, 278 SGRC values have been collected in the study area (~60 km^2) and correspond to a sampling density of 4–5 points of measurements per km^2. This sampling density, in case of SGRC measurements, is perfectly in line with other known studies in the literature carrying out a similar methodological approach (see for example: Ciotoli et al., 2007, 2017; Pásztor et al., 2016; Iovine et al., 2018; Coletti et al., 2022; Petermann et al., 2021).

2.3.2. On-site predictor variables

Five predictors were measured in the field: thoron (^{220}Rn) and carbon dioxide (CO_2) concentration in soil gas, TGDR, permeability, and ^{222}Rn dissolved in groundwater. Thoron and carbon dioxide (CO_2) were measured using the same method and at the same sampling locations as described in Benà et al. (2022).

2.3.2.1. TGDR measurements.
The TGDR has been utilised as a proxy for the BRS contribution in the geogenic radon component. In particular, this variable is referred to the Rn produce by the decay of the principal radionuclides content in rocks (238U, 232Th) and it simulates the lithological background; in fact, terrestrial radiation also varies according to the lithology, genesis and age of rocks and by adsorption at the Earth's surface (Cinelli et al., 2019).

TGDR measurement were performed in-situ at 76 sampling points using a NaI γ-ray portable scintillometer (Scintrex GRS-500) pre-set to a total count rate window corresponding to the energy interval range
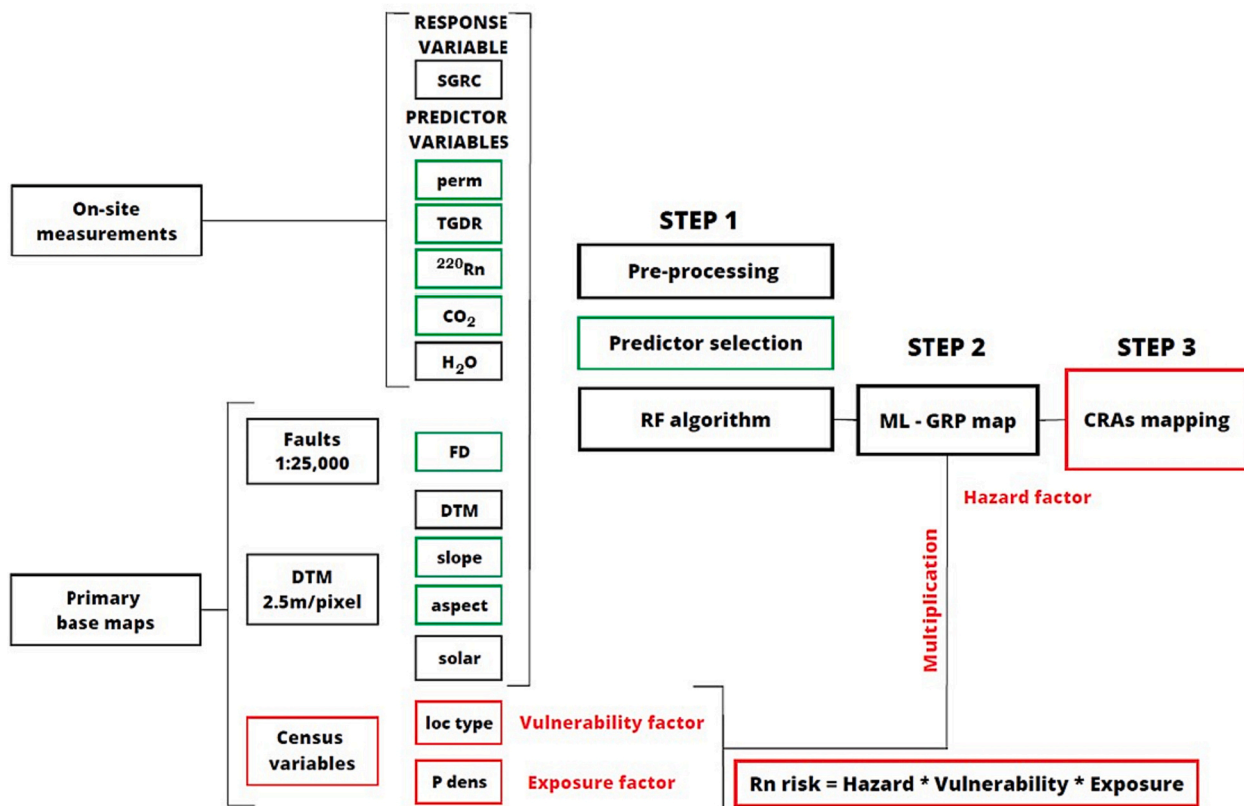
Fig. 1. Flowchart of the mapping process and procedures. SGRC = Soil Gas Radon Concentration; perm = soil permeability; TGDR = Terrestrial Gamma Dose Rate; ^{220}Rn = thoron; CO_2 = carbon dioxide concentration in soil gas; H_2O = concentration of radon dissolved in water; FD = fault density; DTM = digital terrain model; slope = slope; solar = solar radiation; loc type = location type: P dens = population density; GRP map = Geogenic Radon Potential map; RPAs = Radon Priority Areas. SGRC, permeability, TGDR, thoron, carbon dioxide, radon dissolved in water, faults, DTM were pre-processed in order to apply the RF (Random Forest) algorithm (first step) to construct the GRP map (hazard factor). The GRP map was then multiplied by the location type (vulnerability factor) and population density (exposure factor) to construct the CRAs map.

between 80 and 3000 keV. The device was held 1 m above the ground for a measuring time required to achieve a 3 % accuracy. The sensitivity factor of the Scintrex GRS-500 is 3.40 cps/nGyh, allowing the counting rates to be converted into the International System unit of gamma dose rate (μSv/h, Giustini et al., 2019, 2022). Geostatistical analysis, including experimental variogram calculation, modelling, and kriging, was employed to generate a prediction map of the TGDR (see Fig. S2a and b in the Supplementary materials).

2.3.2.2. Permeability. Soil gas permeability directly affects the migration of radon gas from the ground, primarily through advection along fractures and faults, as well as in the shallow soil primarily through diffusion mechanisms (Nuhu et al., 2021; Neznal and Neznal, 2005). The radon concentration in soil gas is directly dependent on the geological characteristics of the area, such as the radionuclides concentration in rocks and soils, and the presence of fractures and faults. It can also be strongly influenced by soil permeability, which relates to soil pore dimensions and soil moisture content (Benavente et al., 2019; Lara et al., 2015). Additionally, other physical characteristics of soils, including soil texture and grain size, significantly impact the mechanisms of radon emanation and exhalation in the soil environment (Huynh Nguyen et al., 2018; Yang et al., 2019).

In the study area, soil permeability was measured at 76 sampling points using a custom permeameter developed by the University of Roma Tre, designed to connect directly to the same probe used for soil gas sampling (Castelluccio et al., 2015). The soil is assumed to be homogeneous and isotropic, and standard state conditions are considered. Air is assumed to be incompressible. The calculation of the soil permeability (k) is based on Darcy's equation and expressed in square meters

(m^2). Geostatistical analysis, including experimental variogram calculation, modelling, and kriging, was employed to obtain a prediction map of soil permeability (see Fig. S3a and b in the Supplementary material).

2.3.2.3. Radon dissolved in groundwater. Dissolved ^{222}Rn was measured at 22 water springs in the study area. Water samples from the selected springs had already undergone chemical-physical analysis by the *Agenzia provinciale per l'ambiente e la tutela del clima - Laboratorio analisi acque e cromatografia* (Bolzano Province) in 2022.

Rn in groundwater is significant since an amount of the Indoor Radon Concentrations derive from the groundwater through the drinking water supply systems. This variable has the potential to affect the IRC, especially in cases where groundwater levels are close to the surface. In Italy, the reference level regarding the radon dissolved in groundwater intended to serve as drinking water is 100 Bq/L (see 59/2013/EURATOM and the transposed D. Lgs. n. 101/2020 in the Italian regulations). In the study area, all the municipalities use drinking water from captured springs located at high altitude (>1600 m.a.s.l.) and these are therefore analysed for the study. Water was directly sampled from the springs using glass bottles. Rn concentrations were measured using the RAD7 system (Durridge Company Inc.) in sniff mode, connected to the Big Bottle RAD H_2O and drystick (drierite desiccant) accessories. Before measurements, the system was purged to ensure that the moisture content inside the system was reduced to <10 % humidity. The sampled bottle was then connected in a closed air-loop mode to the RAD7 system. During system operation, continuous circulation gradually enriched the air within the closed loop with the dissolved Rn from the water sample. Each measurement was conducted with a 5-minute integration period and repeated until the difference between the last two readings was

<5–10 %. The final result was calculated by averaging the previous two integrations. Thiessen polygons were constructed to create a map of areas of influence around the water springs. The water springs represented the centroid of the Thiessen polygons, where the measured dissolved radon value (i.e., the centroid) was assumed to be representative of the entire area underlying the polygon.

2.3.3. Derived predictor variables

2.3.3.1. Fault density.
Faults and fractures serve as the primary pathways for the migration of radon and other gases, such as CO_2 and methane (CH_4), in the subsoil from deep sources (Ciotoli et al., 2007, 2014, 2017, 2020; Giustini et al., 2019). Therefore, the network of the fractured zones characterising the study area has been used as a proxy for secondary permeability. The distribution of the main faults in the study area (Keim et al., 2013) was converted into a fault density (FD) map using the quadratic kernel density function (Silverman, 1986), as described in Benà et al. (2022).

2.3.3.2. Digital terrain model (DTM).
DTM of the study area, representing elevation, was used as a proxy for meteorological conditions that can strongly affect radon migration and exhalation mechanisms. The Radon mobility can be impacted by the presence of slopes, hills, and depressions, which can alter air flow and soil pressure (Gundersen et al., 1992). Radon may not accumulate as much in areas with rough terrain because air circulation and groundwater drainage may be higher in such terrains. Conversely, low-lying areas and depressions may act as radon traps, resulting in higher concentrations in soil gas (Sukanya et al., 2021). Furthermore, Griffiths et al. (2014) emphasised the importance of considering topographic interactions when estimating radon concentrations across different geographical areas. The DTM (2.5 m/pixel) of the Bolzano Province is available from the Geo-catalogue of the Bolzano Province (http://geokatalog.buergernetz.bz.it/geokatalog/#!). The "Surface Parameters tool of Spatial Analyst" in ArcGIS Pro was applied to the DTM to create maps of additional potential proxies: slope, solar radiation (e.g., Areal Solar Radiation) and aspect. Slope can be used as a proxy for soil moisture and shallow soil meteorological conditions, while solar radiation serves as a proxy for microclimate and temperature. Aspect (i.e., slope exposure) refers to the compass direction of the downhill slope faces in relation to the sun. In detail, slope conditions, such as the angle, aspect, and elevation of a land surface, can strongly influence local weather patterns and microclimates, acting as proxies for various meteorological conditions (e.g., sun exposure, rainfall distribution, wind patterns, temperature gradients), all of which may impact radon generation and movement (Zalloni et al., 2018).

2.4. Predictor selection

Predictor selection was carried out using Least Absolute Shrinkage and Selection Operator (LASSO) regression. LASSO regression is an extension of ordinary least squares (OLS) regression used in statistical modelling and ML to estimate the relationships between variables and make predictions (Tibshirani, 1996, 2011; Durrant et al., 2021). This technique aims to strike a balance between model simplicity and accuracy by introducing a penalty term into the traditional linear regression model, which enables sparse solutions in which some coefficients are forced to be exactly zero. LASSO is especially useful for variable selection because it can automatically identify only the most relevant variables while discarding irrelevant or redundant ones, particularly when we assume that many of the features do not significantly contribute to the target variable (Durrant et al., 2021; Handorf et al., 2020). It also helps prevent overfitting by removing variables with low predictive value, potentially making the model more robust across datasets. Furthermore, because it can choose between correlated explanatory variables, it can aid in optimising models with high multicollinearity. In

simple terms, LASSO regression adds a penalty term to the Mean Squared Error (MSE) used in linear regressions. This penalty term is proportional to the sum of the absolute values of the variable coefficients. LASSO regression seeks coefficient values that minimize the sum of the MSE and the penalty.

The LASSO regression cost function is defined as follows (Eq. (1)):

$$J(\beta) = \left(\frac{1}{n}\right) * \sum (y_i - \widehat{y}_i)^2 + l * \sum |\beta_j| \tag{1}$$

where

- $J(\beta)$ is the cost function
- n is the number of data or physical samples (statistically, the sample size)
- y_i is the actual output for the i-th sample
- $\widehat{y}_i$ is the predicted output for the i-th sample
- βj represents the coefficients (weights) associated with each feature
- l is the regularization parameter that controls the amount of regularization applied to the model. Higher values of l lead to more regularization, resulting in a more pronounced feature shrinkage and potentially some coefficients becoming exactly zero.

In this work, LASSO regression was implemented in Python code using the Scikit-learn module (sklearn.linear_model.Lasso).

2.5. Machine learning and Geogenic Radon Potential mapping

ML algorithms enable the solution of highly complex problems. They involve the creation of a model by processing a dataset and subsequently predicting the values of new input data points by executing the established model, typically referred as supervised ML (Rebala et al., 2019). In recent literature within the field of environmental science, ML techniques have gained prominence for spatial prediction tasks. These applications include landslide prediction (Micheletti et al., 2014; Tehrani et al., 2022), soil mapping (Hengl et al., 2017), GRP mapping (Petermann et al., 2021), and time series analysis (Janik et al., 2018). ML can effectively handle complex, multi-dimensional, non-linear relationships and often does not make strong assumptions about the underlying data distribution (Fouedjio and Klump, 2019). Moreover, ML-based approaches have demonstrated superior performance compared to classical geostatistical models in various prediction tasks involving highly complex systems (e.g. Nussbaum et al., 2018; Hengl and MacMillan, 2019; Li et al., 2019). ML models excel in capturing the influence and interplay of numerous factors. In this specific study, we applied a supervised ML method, Random Forest (RF), to model the relationships between the SGRC (the response variable) and the seven selected predictors described in Sections 2.3.2 and 2.3.3 (^{220}Rn, CO_2, TGDR, permeability, fault density, slope, and aspect). Random Forest (RF) is an ensemble classifier algorithm developed by Breiman (2001), commonly used for classification and regression problems. It provides an output based on Decision Trees. A Decision Tree is a regression model built through a series of decisions based on variable values, with splitting values chosen to best separate subsets of data along various paths. Random Forest mitigates overfitting by combining multiple Decision Trees created from bootstrap samples of the full training dataset, using subsets of predictors at each split (Rebala et al., 2019).

In this study, we implemented the Random Forest algorithm by using the Scikit-learn module in Python code.

2.6. Radon risk mapping

2.6.1. Risk concept
The risk is commonly defined as the product of three factors: hazard, vulnerability and exposure, as expressed Eq. (2):

$$Risk = Hazard*Vulnerability*Exposure \tag{2}$$

Hazard represents a specific property that can potentially cause harm or damage to humans. Vulnerability refers to the susceptibility to experience harm or damage or harm due to the stress induced by an event of a certain intensity. Exposure quantifies the number of elements or individuals exposed to the risk (e.g., the number of individuals).

In this study, we identify the GRP as the hazard factor, while land use and population density serve as vulnerability and exposure factors, respectively. Notably, we focus on the concept of collective risk, which pertains to the risk faced by the general public due to geological factors. Applying this risk definition to map CRAs represents an initial and straightforward method to assess collective Rn exposure in the study area. Initially, the GRP mapping serves as a valuable tool for hazard assessment. Furthermore, merging the GRP map with vulnerability and exposure factors, is crucial for evaluating the collective risk.

2.6.2. Construction of Collective Risk Areas (CRAs) map

Following Eq. (2), we designated the GRP as the hazard factor, while location types and the total population of the census tracts within the study area served as vulnerability and exposure factors, respectively.

In the ISTAT dataset, location types are denoted by numbers representing specific type of building areas, ranging from 1 (residential areas) to 4 (sparse houses). To standardise these location types, reclassified them to assign the highest weight (4) to the area with the highest expected mean population density. The reclassification was as follows: (i) location type 4 = residential areas; (ii) location type 3 = housing unit; (iii) location type 2 = industrial areas; (iv) location type 1 = sparse houses.

Next, we utilised the total population and the location type information to compute population density. This involved calculating the ratio of the total population living in a specific location type to the total area (in km^2) of the census tract. Subsequently, we converted the location type and population density maps into a 50×50 m raster grid and normalised them to their respective maximum value. The GRP map was also normalised similarly.

These three factors were then multiplied using the Raster Calculator tool in ArcGIS Pro, following Eq. (2). The resulting risk map underwent further standardisation. To assign a risk value to each polygon of the census tract, we applied the Zonal Statistic tool in Spatial Analyst within ArcGIS Pro. We considered the maximum risk value assigned to the polygon to visualise the risk map and create risk classes. The final risk map is categorised into three risk classes (i.e., low, medium and high) using the natural breaks method. It is numerically expressed as a percentage of risk.

3. Results

3.1. Selected predictors, RF modelling and predictors importance

The LASSO regression successfully identified 7 predictors out of the initial 10 candidates: TGDR, CO$_2$, FD, ^{220}Rn, slope, aspect and soil permeability. These selected predictors, along with their coefficients, are presented in table S1 in the Supplementary materials. Notably, three predictors – DTM, solar radiation and Rn dissolved in groundwater – were excluded from the model due to their coefficients equalling 0, indicating their non-informative nature. The final set of predictors includes a geophysical parameter (TGDR), geochemical parameters (^{220}Rn and CO$_2$), geological parameters (fault density and soil permeability), and geomorphological parameters (slope and aspect). Each of these parameters plays a crucial role in understanding the processes underlying Rn production, migration, and behaviour in shallow soil, as well as interactions at the soil-atmosphere interface.

To ensure there was no redundancy among the selected predictors, we calculated the Variance Inflation Factor (VIF) with all the selected predictors showing VIF < 7 (refer to table S2 in the Supplementary materials).

Before executing the RF model, we set the number of decision trees to 1000. The model performance analysis yielded an R^2 value of 0.93 for training data and 0.47 for the test data. The Root Mean Square Error (RMSE) was found to be 0.30 for the training data and 0.83 for the test data. These results are visualised in Fig. S4 in the Supplementary materials, which displays predicted vs. observed values for training and test data.

To understand the relative influence of individual predictors on model performance, we assessed predictor importance using the RF model (see Fig. 2). The variable importance shows that TGDR, CO$_2$, fault density, ^{220}Rn, slope, aspect and soil permeability have progressively decreased influence on the model performance. In particular, TGDR, a proxy for the Rn source in rocks and soils, and CO$_2$, the main carrier gas in the study area (Benà et al., 2022), have importance exceeding 30 %, emerging as the most influential predictors. Fault density (FD), a proxy for secondary permeability, showed importance in 10–15 % range. ^{220}Rn and slope had an importance of <10 %, followed by aspect and soil permeability, with <5 % importance.

Furthermore, we constructed SHAP diagrams (SHapley Additive exPlanations) using the "shap" library in Python code to highlight the impact of each selected predictor on model predictions (refer to Fig. 3). These diagrams revealed that positive values of TGDR, CO$_2$, FD, ^{220}Rn, slope, and permeability had the most substantial influence on model output, while aspect was the sole variable exerting a negative impact on the model output.

To examine the relationship between each predictor and the model output more deeply, we generated Partial Dependent Plots using the "pdpbox" library in Python (see Fig. 5Sa–g and related Supplementary materials for explanation). PDPs allowed us to visualise the relationship between a target feature and the model outcome while holding all other features constant, aiding in interpreting how each predictor affected the model predictions. The PDPs considered the average effect of other predictors in the model when analysing the relationship of each predictor with the model outcome (Petermann et al., 2021).

3.2. Geogenic Radon Potential map

The RF algorithm was employed to create the GRP map of the study area, using SGRC as the response variable and the 7 selected predictors (i.e., TGDR, CO$_2$, fault density, ^{220}Rn, slope, aspect, soil permeability). The resultant GRP map exhibits a range of values, with a minimum of 7.21 kBq m^{-3} and a maximum of 182 kBq m^{-3} (as illustrated in Fig. 4). In accordance with results presented in Benà et al. (2022), we consider high GRP values those exceeding 50 kBq m^{-3}. This threshold corresponds to the local background level in the study area.

The high GRP values delineate a zone extending along the east-west direction from Falzes/Pfalzen to Terento/Terenten, aligning to the direction of the wide fracture zone associated with the Pusteria fault system. Within this area, the high GRP values are attributed to the presence of the Tectonically Enhanced Radon (TER) quantity, as elucidated by Benà et al. (2022).

3.3. The Collective Risk Areas (CRAs) map

Fig. 5 shows the CRA map of the study area, illustrating the density of collective risk obtained by multiplying the GRP map, the location type (vulnerability factor), and population density (exposure factor).

The map has been divided into three risk classes as follow: i) Risk < 5 % (depicted as low risk, indicated by white areas in Fig. 5); ii) risk between 5 % and 50 % (designated as medium risk, represented in yellow in Fig. 5); iii) risk >50 % (considered as high risk, marked in red in Fig. 5).

The CRA map is correlated with Table 1, which summarises certain parameters characterising the three defined risk classes: (i) the average GRP value in kBq m^{-3}; (ii) the average population density expressed as
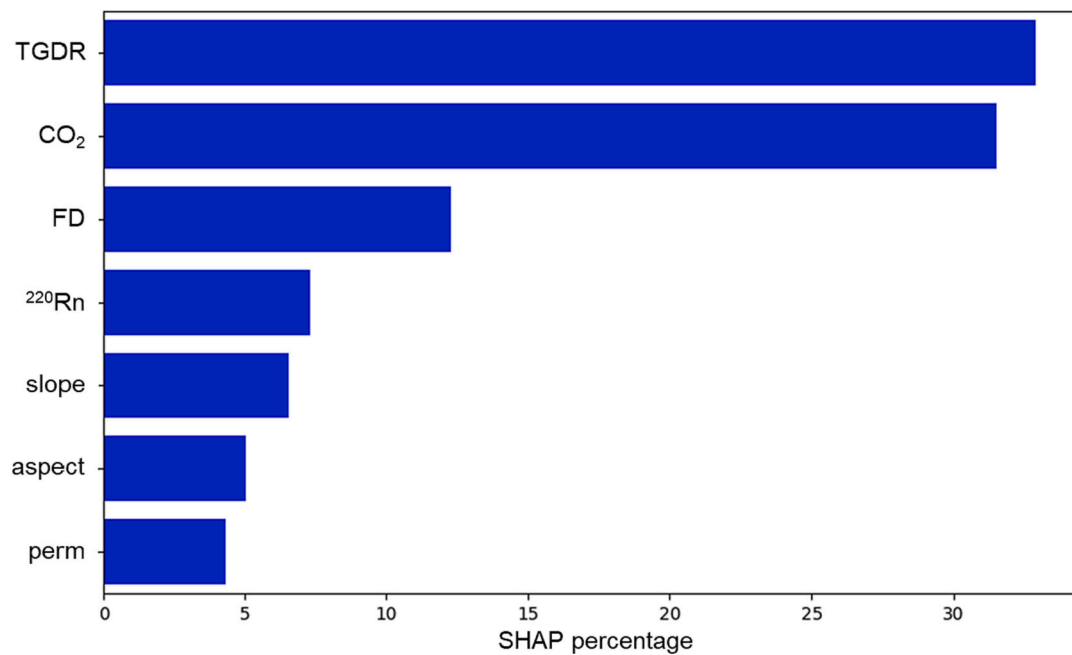
Fig. 2. Feature importance based on SHAP value percentage in the RF model. The predictors are ordered by decreasing importance.; X-axis: SHAP percentage; Y-axis = selected predictors. TGDR = Terrestrial Gamma Dose Rate; CO_2 = carbon dioxide; FD = fault density; ^{220}Rn = thoron; perm = soil permeability.

Fig. 3. SHAP diagram. The Y-axis of the SHAP diagram reports the 7 selected predictors in descending order of importance in the RF model from TGDR (the most influent) to the soil permeability (the less influent). The X-axis of the SHAP diagram represents the SHAP values quantifying the impact of a single feature on the model's output: positive SHAP values indicate that the feature positively contributes to the output, while negative values suggest a negative contribution. Red and blue dots represent the contribution of individual features to the prediction compared to a reference value. Red dots represent positive contributions and indicate that the feature is increasing the predicted output. Blue dots represent negative contributions and indicate that the feature is decreasing the predicted output. TGDR = Terrestrial Gamma Dose Rate; CO_2 = carbon dioxide; FD = fault density; ^{220}Rn = thoron; perm = soil permeability.

the number of people per km^2; (iii) the location type (i.e., 4, 3, 2, 1); (iv) the total area covered by the respective risk class.

Most of the study area (68.51 km^2) falls within the low-risk category, consistent with the mountainous terrain where the majority of the population resides in residential areas of the main municipalities, such as Terento/Terenten, Chienes/Kiens and Falzes/Pfalzen.

In general, the mean GRP values (indicating hazard) exceed the local background value of 50 kBq m^{-3} in all the three risk classes, with a slight increase from low risk (63 kBq m^{-3}) to high risk (76 kBq m^{-3}). The progressive increase in mean population density (representing exposure) from low- to high-risk areas is closely related to the location type (reflecting the vulnerability factor): (i) in the low-risk areas, most of the

census tracts (33) are described as residential areas (location type = 4) and sparse houses (location type = 1, 43 census tracts); (ii) in the medium-risk areas, the majority of the census tracts are designated as residential areas (location type 4, 22 census tracts); (iii) all census tracts falling within high-risk areas are described as residential areas (location type = 4) with the highest population density. As a result, the population density increases proportionally from low- to high-risk areas.
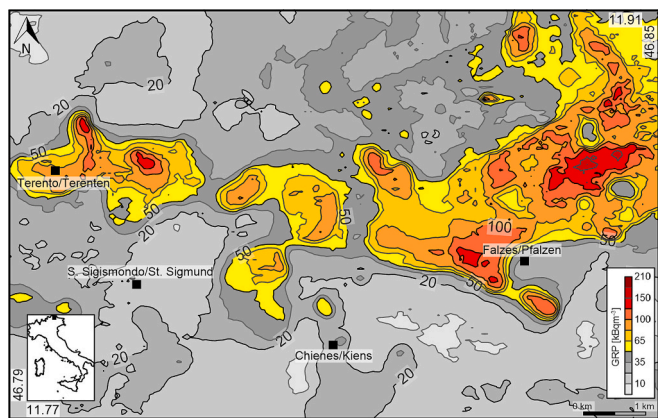
Fig. 4. GRP (kBq m^{-3}) map in the study area.

4. Discussion

4.1. Interpretation of predictors in the RF model

The RF model demonstrates that all of the selected predictors influence Rn concentrations and movement in the subsoil. This result is consistent with the dependence of Rn on the geochemical and structural characteristics of the study area, mainly linked to the generation and transport of Rn in the geological environment (i.e., from deep sources towards the subsoil) (Benà et al., 2022). The variable importance clearly shows that GRP is primarily affected by TGDR (35 %, Fig. 2), representing the BRS contribution, such as the content of ^{238}U and ^{232}Th radionuclides, from the main outcropping rocks, including gneiss, granite, and phyllite (Tchorz-Trzeciakiewicz and Rysiukiewicz, 2021; Giustini et al., 2019, 2022). Since the TGDR surveys were conducted at the ground level, its correlation with SGRC is stronger than with atmospheric concentrations. In the literature, also Bossew et al. (2017), Cinelli et al. (2019), Melintescu et al. (2018), and Sainz Fernández et al. (2017) have reported a positive correlation between TGDR and GRP.

The BRS contribution to the Rn amount in soil gas generates a relatively high spatial variability of Rn concentration in the soil gas, reflecting the homogeneous characteristics of the soil or rock environment at the local scale (BRS). However, Rn spatial variability can increase (in particular at local scale) near fault zones (TER), where Rn migration from deeper sources can be enhanced by intense fracturing and the presence of carrier gases, which play a dominant role in the advective transport and redistribution of trace gases at surface (Wilkening, 1980; Ciotoli et al., 2007, 2014; Prasetio et al., 2023, and reference therein). This is observable in the study area along the Pusteria fault system, where radon concentrations in soil gas have a positive correlation with CO_2 concentrations (importance of about 30 %, Fig. 2),
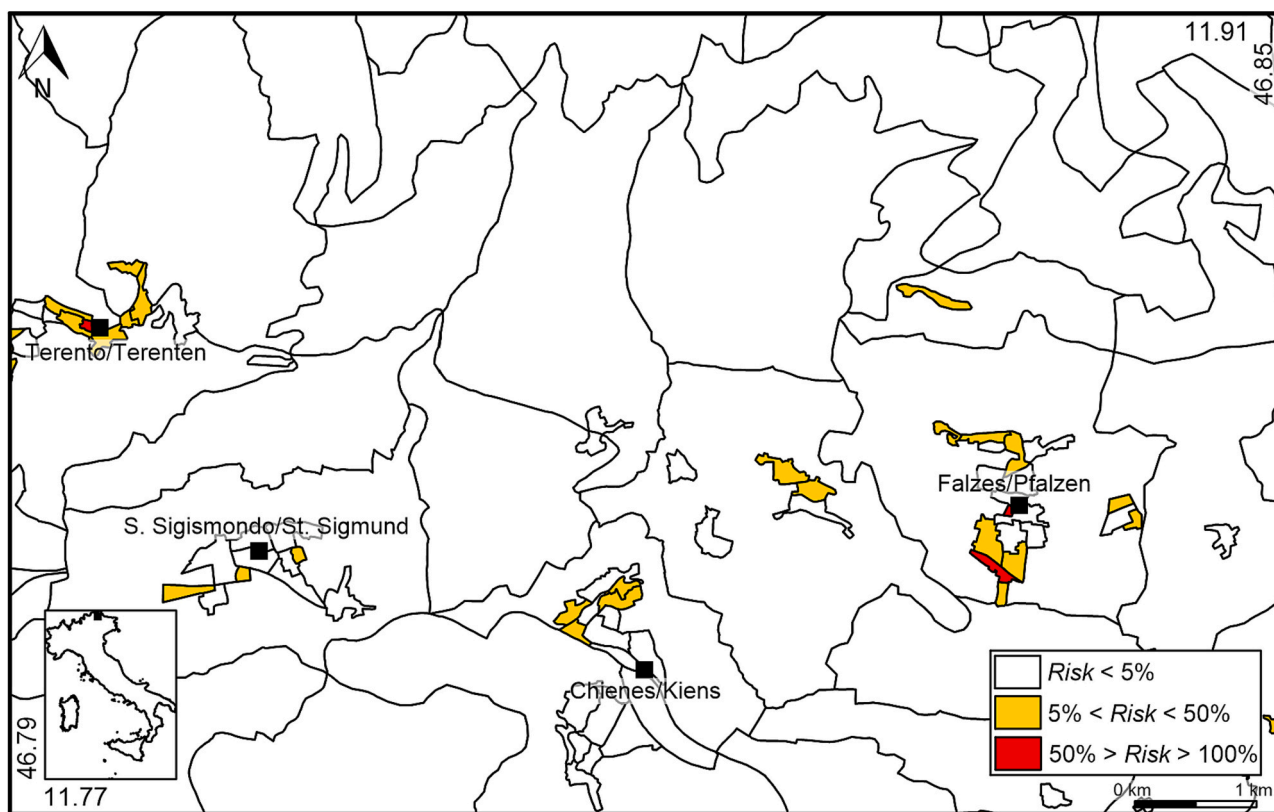

Fig. 5. Map of the CRAs in the study area.

Table 1

The table reports the risk class and the correspondent percentage of risk, the mean GRP value, the population density, the location type (denoted by numbers representing specific type of building areas) and the extension of the area covered by the considered risk class.

Collective risk class	Risk level (%)	GRP mean (kBq m^{-3})	Population density (people km^{-2})	Population (people)	Location type	Area (km^2)
Low	<5	63	546	5927	4, 3, 2, 1	68.51
Medium	5–50	65	6116	3072	4, 3, 2, 1	0.75
High	>50	76	17,549	622	4	0.05

suggesting a possible upward advective flow caused by pressure gradients.

The high importance (about 15 %, Fig. 2) of the fault density (interpreted as secondary permeability due to the fault zone) in the RF model confirms the effect of the Pusteria fault system on Rn migration (as well as other gases); this predictor is strictly related to the TER component (Benà et al., 2022). Indeed, damage zones related to high fracturing zones (fault areas) often exhibit a high permeability compared to the surrounding rocks, which may facilitate the advective transport of fluids for SGRC, potentially increasing radon release towards the surface, and consequently, Rn availability to enter buildings (IRC) (Ciotoli et al., 2007, 2014, 2016; Seminsky et al., 2014; Chen et al., 2018; Banrion et al., 2022; Zhou et al., 2023).

Similar importance of the other predictors (i.e., ^{220}Rn, slope, aspect and soil permeability) ranging from 4 % to 8 % can be explained by processes affecting Rn movement in the soil layer and at the soil-atmosphere interface (SRE) (Fig. 2). In the shallow environment, the influence of meteorological conditions can be complex, and the literature results are controversial. The influence of air temperature and pressure on soil radon concentration is small in comparison with the total seasonal variability of this gas. In any case, the influence of these two variables is further reduced by conducting soil gas measurement campaigns during periods of stable and good weather conditions (Ciotoli et al., 2014; Beaubien et al., 2015).

The principal drivers governing diurnal and seasonal changes in radon concentration in the soil are the water saturation and moisture retention in the soil pores (i.e., rainfall) (King and Minissale, 1994). These two parameters directly decrease soil permeability thus preventing radon gas diffusion in the shallow soil layers (Nazaroff, 1992; Alonso et al., 2019; Beltrán-Torres et al., 2023). High soil permeability allows ^{220}Rn to be detected at surface despite its short decay time (56 s).

In addition, the slope can be used as a proxy for soil moisture and meteorological conditions in absence of any other meteorological variables. High slopes also constitute zones characterised by increased soil permeability because they do not promote the retention of water and moisture in the soil pores. On the contrary, flat zones are characterised by low soil permeability because they favour the accumulation of water and moisture in the soil pores. In this regard, the SHAP diagram shows that high values of ^{220}Rn, slope and soil permeability are positively correlated with high GRP (Fig. 3). Soil permeability may be linked to the ability of radon to migrate and escape towards the Earth surface. In fact, where permeability is low, radon escapes more easily; this variable is also linked to the fault density representing the secondary permeability.

All these predictors, except for the aspect, have an impact on the GRP values prediction for positive values and show an increasing trend up to the expected average radon value (see PDPs, in Fig. S5 in the Supplementary materials). On the contrary, low values of the GRP are correlated with high values of the aspect (i.e., inverse correlation). The aspect identifies the compass direction that the downhill slope faces for each location; therefore, radon accumulation is easier in flat areas.

The model confirmed the correlations between geology and GRP and also provided insight into the utility and significance of other predictors that reflect the physical, chemical, and hydraulic properties of soil, as well as climatic predictors.

4.2. Map of the Collective Risk Areas (CRAs)

The GRP map obtained through RF regression represents radon hazard due to geological features in a specific region. It is closely related to Rn gas directly measured in the soil and to all geological predictors that significantly influence its concentration in the shallow environment. GRP maps are essential for evaluating Rn risk, as they represent the most significant spatial predictor of IRC (Bossew, 2015; Bossew et al., 2020).

As previously mentioned, the European regulations aim to identify RPAs and implement mitigation plans to limiting radon exposure,

thereby reducing the risk of lung cancer for the population. In undeveloped and uninhabited areas, high Rn values represent only a high hazard (i.e., GRP), but not an immediate risk. This concept is well-established and applied in the case of other natural phenomena, such as seismic micro zonation studies.

European legislation seeks to reduce the harm caused by Rn exposure (i.e., the number of lung cancer deaths) and, consequently, reduce collective exposure. Fig. 6 illustrates how GRP is a key factor in recognising CRAs.

In this paper, for the first time, we introduce the concept and define CRAs by applying the risk definition (Section 2.4). Mapping the GRP is undoubtedly the crucial first step in defining the Rn hazard, a specific property that cannot be mitigated. For this reason, it is important to map it as accurately as possible, considering multiple geological variables and employing robust mapping techniques.

As reported in Benà et al., 2022, Rn values exceeding the lithological background (50 kBq m^{-3}) are considered anomalous and linked to the wide fracturing zone of the Pusteria fault system, which represents Rn enhanced due to tectonics (TER). However, in Benà et al. (2022), this aspect is not discussed in terms of GRP, and it does not consider other important geological factors, such as gas permeability and indicators of deep circulation (e.g., Rn in groundwater), as well as the shallow effects governed by the morphological parameters (e.g., DTM).

The identification of a specific threshold value of GRP is not significant for delineating CRAs since the indoor radon risk exists even for "very low" concentrations of radon in the soil and, consequently, for very low GRP values. In fact, radon measured in the soil (GRP) is generally three orders of magnitude higher than indoor radon. It is evident that every area can be affected by potential indoor risk, and all dwellings are considered vulnerable.

However, GRP plays a key role in defining CRAs, primarily occurring along the Pusteria fault system where Rn degassing is enhanced by intense fracturing, resulting in high GRP values. This aligns with the fact that all the GRP values contribute to the risk. Therefore, the CRAs map highlights areas with low, medium and high collective risk, where IRC values may be high for residential areas.

5. Conclusions

The mapping and analysis of GRP (e.g., Rn hazard) serve as fundamental tools for delineating CRAs according to a new, more geological interpretation of RPAs, as compared to that outlined in the BSS directive (2013/59/EURATOM).

We used a risk formula to combine the GRP map, obtained through ML approaches, with characteristics of the census tracts as location type (the vulnerability factor) and population density (the exposure factor).

In alignment with a geological-based interpretation of RPAs, we can recognise hazard-based RPAs (CRAs) and detriment-based RPAs (IRAs) as complementary concepts within territorial planning and remediation actions, respectively, rather than alternatives.

Our findings lead to the following conclusions:

- The use of the random forest algorithm as a ML model proved to be robust and highly effective for generating a GRP map of the study area. This GRP map incorporated seven predictors, reflecting geological factors (BRS and TER), soil characteristics (groundwater circulation, soil permeability), and meteorological conditions (DTM derivatives). The variable importance analysis highlighted the dominant impact of the Rn source while still showing significant contributions from other predictors.
- Since GRP is considered the most critical spatial predictor of IRC, it is evident that an accurate mapping of this hazard factor effectively represents the total amount of radon that could potentially enter buildings.
- Given that GRP (e.g., soil gas concentration) values are three orders of magnitude higher than IRC, there is no reason to establish a
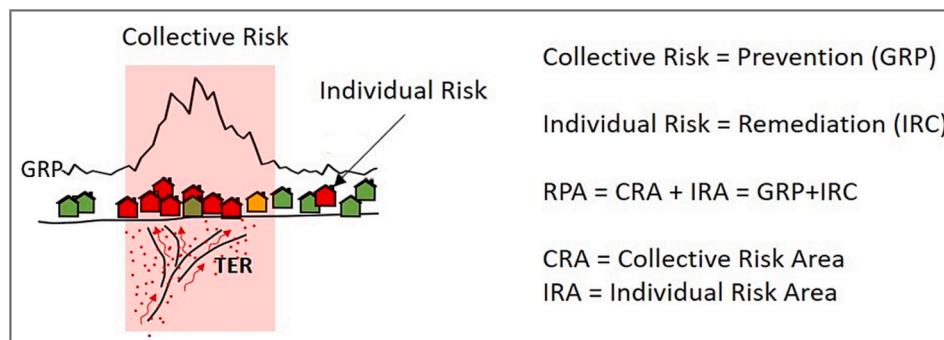
Fig. 6. Summary sketch of the CRAs concept.

specific GRP threshold. Indoor radon risk can exist even for "very low" concentrations of radon in the soil and, consequently, for very low GRP values. Qualitative GRP classes can serve as delineation of zones, akin to how they are used in seismic micro zonation studies, guiding land-use planning strategies, construction types, and remediation actions.

- The construction of GRP maps is a crucial tool for both Rn hazard and risk analysis. It forms the foundation for identifying RPAs, particularly under a new, more geological perspective. This is essential for collective risk assessment, including land-use planning and prevention, as well as individual risk assessment, aiding in strategic planning for indoor surveys, and specific remediation actions.
- The absence of clear guidelines for defining RPAs necessitates the geological-based conceptualization of a complementary approach to mapping both CRAs (for prevention), and IRAs (for building remediation actions).

This study may assist policymakers in implementing preventive measures in areas where new buildings are planned and in taking remediation actions in RPAs sensu stricto. Future studies could aim to define effective individual risk by constructing statistical models that also consider IRC measurements and anthropogenic factors.

CRediT authorship contribution statement

Eleonora Benà: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Giancarlo Ciotoli:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. **Eric Petermann:** Methodology, Software, Writing – review & editing. **Peter Bossew:** Conceptualization, Writing – review & editing. **Livio Ruggiero:** Formal analysis, Investigation, Writing – review & editing. **Luca Verdi:** Writing – review & editing. **Paul Huber:** Investigation. **Federico Mori:** Software. **Claudio Mazzoli:** Conceptualization, Writing – review & editing. **Raffaele Sassi:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.169569.

References

Alonso, H., Rubiano, J.G., Guerra, J.G., Arnedo, M.A., Tejera, A., Martel, P., 2019. Assessment of radon risk areas in the Eastern Canary Islands using soil radon gas concentration and gas permeability of soils. Sci. Total Environ. 664, 449–460. https://doi.org/10.1016/j.scitotenv.2019.01.411.

Banrion, M.H., Elio, J., Crowley, Q.G., 2022. Using geogenic radon potential to assess radon priority area designation, a case studyaround Castleisland, Co. Kerry, Ireland. J. Environ. Radioact. 251 (252), 106956. https://doi.org/10.1016/j.jenvrad.2022.106956.

Beltrán-Torres, S., Szabó, K.Z., Tóth, G., Tóth-Bodrogi, E., Kovács, T., Szabó, C., 2023. Estimated versus field measured soil gas radon concentration and soil gas permeability. J. Environ. Radioact. 265, 107224. https://doi.org/10.1016/j.jenvrad.2023.107224.

Benà, E., Ciotoli, G., Ruggiero, L., Coletti, C., Bossew, P., Massironi, M., Mazzoli, C., Mair, V., Morelli, C., Galgaro, A., Morozzi, P., Tositti, L., Sassi, R., 2022. Evaluation of tectonically enhanced radon in fault zones by quantification of the radon activity index. Sci. Rep. 12, 21586 https://doi.org/10.1038/s41598-022-26124-y.

Benavente, D., Valdés-Abellán, J., Pla, C., Sanz-Rubio, E., 2019. Estimation of soil gas permeability for assessing radon risk using Rosetta pedotransfer function based on soil texture and water content. J. Environ. Radioact. 208-209, 105992. https://doi.org/10.1016/j.jenvrad.2019.105992.

Borsi, S., Del Moro, A., Sassi, F.P., Zanferrari, A., Zirpoli, G., 1978. New geopetrologic and radiometric data on the Alpine history of the Austridic continental margin south of the Tauern Window (Eastern Alps). Mem. Sci. Geol. Univ. Padova 32, 1–17.

Bossew, P., 2015. Mapping theGeogenic Radon Potential and Estimation of Radon Prone Areas in Germany. Radiation EmergencyMedicine 4 (2), 13–20.

Bossew, P., 2018. Radon priority areas definition, estimation and uncertainty. Nucl. Technol. Radiat. Prot. 33 (3), 286–292. https://doi.org/10.2298/NTRP180515011B.

Bossew, P., Petermann, E., 2021. What is the objective of radon abatement policy? Revisiting the oncept of radon priority areas. In: 15th GARRM. International Workshop on the Geological Aspects of Radon Risk Mapping (21–24 Sept, Prague, Czech Republic).

Bossew, P., Suhr, N., 2023. European radon abatement policy: state and ongoing discussion. Braz. J. Radiat. Sci. 11 (1A), 1–12. https://doi.org/10.15392/2319-0612.2023.2162 (Suppl.). https://bjrs.org.br/revista/index.php/REVISTA/article/view/2162.

Bossew, P., Cinelli, G., Hernández-Ceballos, M., Cernohlawek, N., Gruber, V., Dehandschutter, B., Menneson, F., Bleher, M., Stöhlker, U., Hellmann, I., Weiler, F., Tollefsen, T., Tognoli, P.V., de Cort, M., 2017. Estimating the terrestrial gamma dose rate by decomposition of the ambient dose equivalent rate. J. Environ. Radioact. 166, 296–308. https://doi.org/10.1016/j.jenvrad.2016.02.013.

Bossew, P., Cinelli, G., Ciotoli, C., Crowley, Q.G., De Cort, M., Medina, J.E., Gruber, V., Petermann, E., Tollefsen, T., 2020. Development of a geogenic radon hazard index – concept, history, experiences. Int. J. Environ. Res. Public Health 17 (11), 4134. https://doi.org/10.3390/ijerph17114134.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Castelluccio, M., De Simone, M., Lucchetti, C., Moroni, M., Salvati, F., Tuccimei, P., 2015. A new technique to measure in situ soil gas permeability. J. Geochem. Explor. 148, 56–59. https://doi.org/10.1016/j.gexplo.2014.08.002.

Chen, Z., Li, Y., Liu, Z., Wang, J., Zhou, X., Du, J., 2018. Radon emission from soil gases in the active fault zones in the capital of China and its environmental effects. Sci. Rep. 8, 16772. https://doi.org/10.1038/s41598-018-35262-1.

Cinelli, G., Tositti, L., Capaccioni, B., Brattich, E., Mostacci, D., 2015. Soil gas radon assessment and development of a radon risk map in Bolsena. Central Italy. Environ. Geochem. Health 37, 305–319. https://doi.org/10.1007/s10653-014-9649-9.

Cinelli, G., Tollefsen, T., Bossew, P., Gruber, V., Bogucarskis, K., De Felice, L., De Cort, M., 2019. Digital version of the European Atlas of natural radiation. J. Environ. Radioact. 196, 240–252. https://doi.org/10.1016/j.jenvrad.2018.02.008.

Ciotoli, G., Lombardi, S., Annunziatellis, A., 2007. Geostatistical analysis of soil gas data in a high seismic intermontane basin: Fucino Plain, central Italy. J. Geophys. 112, B05407 https://doi.org/10.1029/2005JB004044.

Ciotoli, G., Bigi, S., Tartarello, C., Sacco, P., Lombardi, S., Ascione, A., Mazzoli, S., 2014. Soil gas distribution in the main coseismic surface rupture zone of the 1980, Ms = 6.9, Irpinia earthquake (southern Italy). J. Geophys. Res. Solid Earth 119, 2440–2461. https://doi.org/10.1002/2013JB010508.

Ciotoli, G., Sciarra, A., Ruggiero, L., Annunziatellis, A., Bigi, S., 2016. Soil gas geochemical behaviour across buried and exposed faults during the 24 August 2016 central Italy earthquake. Ann. Geophys. https://doi.org/10.4401/ag-7242.

Ciotoli, G., Voltaggio, M., Tuccimei, P., Soligo, M., Pasculli, A., Beaubien, S.E., Bigi, S., 2017. Geographically weighted regression and geostatistical techniques to construct the geogenic radon potential map of the Lazio region: a methodological proposal for the European atlas of natural radiation. J. Environ. Radioact. 166 (2), 355–375. https://doi.org/10.1016/j.jenvrad.2016.05.010.

Ciotoli, G., Procesi, M., Etiope, G., Fracassi, U., Ventura, G., 2020. Influence of tectonics on global scale distribution of geological methane emissions. Nat. Commun. 11, 2305. https://doi.org/10.1038/s41467-020-16229-1.

Beaubien, S.E., Ruggiero, L., Annunziatellis, A., Bigi, S., Ciotoli, G., Deiana, P., Graziani, S., Lombardi, S., Tartarello, C., 2015. The importance of baseline surveys of near-surface gas geochemistry for CCS monitoring, as shown from onshore case studies in northern and southern Europe. Oil & Gas Science and Technology – Rev. IFP Energies nouvelles 70 (4), 615–663. https://doi.org/10.2516/ogst/2014009.

Decreto Legislativo 31 luglio 2020, n. 101 (Attuazione Direttiva 2013/59/EURATOM). Gazzetta Ufficiale della Repubblica Italiana, Serie Generale n. 201 del 12-08-2020-Suppl. Ordinario n. 29.

Coletti, C., Ciotoli, G., Benà, E., Brattich, E., Cinelli, G., Galgaro, A., Massironi, M., Mazzoli, C., Mostacci, D., Morozzi, P., Mozzi, P., Nava, J., Ruggiero, L., Sciarra, A., Tositti, L., Sassi, R., 2022. The assessment of local geological factors for the construction of a Geogenic Radon Potential map using regression kriging. A case study from the Euganean Hills volcanic district (Italy). Sci. Total Environ. 808, 152064. https://doi.org/10.1016/j.scitotenv.2021.152064.

Durrant, J.T., Gee, K.L., Anderson, M.C., Cook, M.R., Loubeau, A., 2021. Preliminary lasso regression analysis of environmental effects on sonic boom metric variability. J. Acoust. Soc. Am. 150, A259. https://doi.org/10.1121/10.0008215.

European Union, 2013. Council Directive 2013/59/Euratom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom. Off. J. Eur. Union OJ L13, 1–73 (17.01.2014).

Fouedjio, F., Klump, J., 2019. Exploring prediction uncertainty of spatial data in geostatistical and machine learning approaches. Environ. Earth Sci. 78 (1), 38. https://doi.org/10.1007/s12665-018-8032-z.

Giustini, F., Ciotoli, G., Rinaldini, A., Ruggiero, L., Voltaggio, M., 2019. Mapping the geogenic radon potential and radon risk by using empirical Bayesian kriging regression: a case study from a volcanic area of Central Italy. Sci. Total Environ. 661, 449–464. https://doi.org/10.1016/j.scitotenv.2019.01.146.

Giustini, F., Ruggiero, L., Sciarra, A., Beaubien, S., Graziani, S., Galli, G., Pizzino, L., Tartarello, M.C., Lucchetti, C., Sirianni, P., Tuccimei, P., Voltaggio, M., Bigi, S., Ciotoli, G., 2022. Radon hazard in central Italy: comparison among areas with different geogenic radon potential. Int. J. Environ. Res. Public Health 19, 666. https://doi.org/10.3390/ijerph19020666.

Griffiths, A., Conen, F., Weingartner, E., Zimmermann, L., Chambers, S., Williams, A., Steinbacher, M., 2014. Surface-to-mountaintop transport characterised by radon observations at the Jungfraujoch. Atmos. Chem. Phys. 14, 12763–12779. https://doi.org/10.5194/acp-14-12763-2014.

Gundersen, L.C., Schumann, R.R., Otton, J.K., Dubiel, R.F., Owen, D.E., Dickinson, K.A., 1992. Geology of radon in the United States. Geol. Soc. Am. Spec. Pap. 271, 1–16. https://doi.org/10.1130/SPE271-p1.

Handorf, E., Yin, V., Slifker, M., Lynch, S., 2020. Variable selection in social-environmental data: sparse regression and tree ensemble machine learning approaches. BMC Med. Res. Methodol. 20, 302. https://doi.org/10.1186/s12874-020-01183-9.

Hengl, T., MacMillan, R.A., 2019. Predictive Soil Mapping With R. Wageningen, the Netherlands, OpenGeoHub Foundation, p. 370.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. Soil Grids 250m: global gridded soil information based on machine learning. PLoS One 12, e0169748. https://doi.org/10.1371/journal.pone.0169748.

Huynh Nguyen, P.T., Nguyen, V.T., Vu, N.B., Nguyen, V.D., Le Cong, H., 2018. Soil radon gas in some soil types in the rainy season in Ho Chi Minh City, Vietnam. J. Environ. Radioact. 193–194, 27–35. https://doi.org/10.1016/j.jenvrad.2018.08.017.

Iovine, G., Guagliardi, I., Bruno, C., et al., 2018. Soil-gas radon anomalies in three study areas of Central-Northern Calabria (Southern Italy). Nat. Hazards 91 (Suppl. 1), 193–219. https://doi.org/10.1007/s11069-017-2839-x.

Janik, M., Bossew, P., Kurihara, O., 2018. Machine learning methods as a tool to analyse incomplete or irregularly sampled radon time series data. Sci. Total Environ. 630, 1155–1167. https://doi.org/10.1016/j.scitotenv.2018.02.233.

Keim, L., Mair, V., Morelli, C., 2013. Inquadramento geologico regionale. In: Guida ai percorsi geologici Foglio 026 Appiano. Ufficio Geologia e prove materiali, Provincia Autonoma di Bolzano, LAC Firenze, ISBN 978-88-7914-678-4.

King, C.-Y., Minissale, A., 1994. Seasonal variability of soil-gas radon concentration in central California (San Andreas fault). Radiat. Meas. 23, 683–692. https://doi.org/10.1016/1350-4487(94)90004-3.

Lara, E., Rocha, Z., Santos, T.O., Rios, F.J., Oliveira, A.H., 2015. Soil features and indoor radon concentration prediction: radon in soil gas, pedology, permeability and 226Ra content. Radiat. Prot. Dosim. 167 (1–3), 126–129. https://doi.org/10.1093/rpd/ncv228.

Li, J., Siwabessy, J., Huang, Nichol, 2019. Developing an optimal spatial predictive model for seabed sand content using machine learning, geostatistics, and their hybrid methods. Geosciences 9, 180. https://doi.org/10.3390/geosciences9040180.

Melintescu, A., Chambers, S.D., Crawford, J., Williams, A.G., Zorila, B., Galeriu, D., 2018. Radon-222 related influence on ambient gamma dose. J. Environ. Radioact. 189, 67–78. https://doi.org/10.1016/j.jenvrad.2018.03.012.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., Kanevski, M., 2014. Machine learning feature selection methods for landslide susceptibility mapping. Math. Geosci. 46, 33–57. https://doi.org/10.1007/s11004-013-9511-0.

Minach, L., Verdi, L., Marchesoni, C., Amadori, C., 1999. Radon in Alto Adige. Agenzia per la Protezione dell'Ambiente di Bolzano, Laboratorio di chimica fisica.

Müller, W., Mancktelow, N.S., Meier, M., 2000. Rb-Sr microchrons of synkinematic mica in mylonites: an example from the DAV fault in Eastern Alps. Planet. Sci. Lett. 1880, 385–397. https://doi.org/10.1016/S0012-821X(00)00167-9.

Nazaroff, W.W., 1992. Radon transport from soil to air. Rev. Geophys. 30, 137–160. https://doi.org/10.1029/92RG00055.

Neznal, M., Neznal, M., Matolin, M., Barnet, I., Miksova, J., 2004. The new method for assessing the radonrisk of building sites. Czech Geol. SurveySpecial Papers. Czech Geol. Survey 16.

Neznal, M., Neznal, M., 2005. Permeability as an important parameter for radon risk classification of foundation soils. Ann. Geophys. 48 (1), 175–180. https://doi.org/10.4401/ag-3192.

Nuhu, H., Hashim, S., Aziz Saleh, M., Syazwan Mohd Sanusi, M., Hussein Alomari, A., Jamal, M.H., Abdullah, R.A., Hassan, S.A., 2021. Soil gas radon and soil permeability assessment: mapping radon risk areas in Perak State. Malaysia. PLoS ONE 16 (7), e0254099. https://doi.org/10.1371/journal.pone.0254099.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E., Papritz, A., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil 4, 1–22. https://doi.org/10.5194/soil-4-1-2018.

Pásztor, L., Szabó, K.Z., Szatmári, G., Laborczi, A., Horváth, Á., 2016. Mapping geogenic radon potential by regression kriging. Sci. Total Environ. 544, 883–891. https://doi.org/10.1016/j.scitotenv.2015.11.175.

Petermann, E., Meyer, H., Nussbaum, M., Bossew, P., 2021. Mapping the geogenic radon potential for Germany by machine learning. Sci. Total Environ. 754, 142291. https://doi.org/10.1016/j.scitotenv.2020.142291.

Petermann, E., Bossew, P., Hoffmann, B., 2022. Radon hazard vs. radon risk – on the effectiveness of radon priority areas. J. Environ. Radioact. 244–245, 106833. https://doi.org/10.1016/j.jenvrad.2022.106833.

Prasetio, R., Laksminingpuri, N., Satrio, Satrio, Pujiindiyati, E.R., Pratikno, B., Sidauruk, P., 2023. The 222Rn and CO2 soil gas distribution at Lembang Fault Zone, West Java – Indonesia. J. Environ. Radioact. 257, 107079 https://doi.org/10.1016/j.jenvrad.2022.107079 (February 2023).

Rebala, G., Ravi, A., Churiwala, S., 2019. Machine learning definition and basics. In: An Introduction to Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-030-15729-6_1.

Sainz Fernández, C., Quindós Poncela, L.S., Fernández Villar, A., Fuente Merino, I., Gutierrez-Villanueva, J.L., Celaya González, S., Quindós López, L., Quindós López, J., Fernández, E., Remondo Tejerina, J., Martín Matarranz, J.L., García Talavera, M., 2017. Spanish experience on the design of radon surveys based on the use of geogenic information. J. Environ. Radioact. 166, 390–397. https://doi.org/10.1016/j.jenvrad.2016.07.007.

Sassi, F.P., Cesare, B., Mazzoli, C., Peruzzo, L., Sassi, R., Spiess, R., 2004. The crystalline basements of the Italian Eastern Alps: a review of the metamorphic features. In: Castelli, D., Cesare, B. (Eds.), Advances in Metamorphic Petrology: Browsing Through Italian Classic Areas, Case Studies, and Approaches, Per. Mineral. Special Issue, vol. 2, pp. 23–42 (ISSN: 0369-8963).

Schmid, S.M., Aebli, H.R., Heller, F., Zingg, A., 1989. The role of the Periadriatic Line in the tectonic evolution of the Alps. In. Geol. Soc. London, Spec. Publ. 45, 153–171. https://doi.org/10.1144/GSL.SP.1989.045.01.08.

Seminsky, Kzh., Bobrov, A.A., Demberel, S., 2014. Variations in radon activity in the crustal fault zones: spatial characteristics. Phys. Solid Earth 50 (6), 795–881. https://doi.org/10.1134/S1069351314060081.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London. https://doi.org/10.1007/978-1-4899-3324-9.

Spiess, R., Cesare, B., Mazzoli, C., Sassi, R., Sassi, F.P., 2010. The crystalline basement of the Adria microplate in the eastern Alps: a review of the palaeostructural evolution from the Neoproterozoic to the Cenozoic. Rend. Fis. Acc. Lincei 21 (Suppl. 1), 31–50. https://doi.org/10.1007/s12210-010-0100-6.

Sukanya, S., Noble, J., Joseph, S., 2021. Factors controlling the distribution of radon (222Rn) in groundwater of a tropical mountainous river basin in southwest India. Chemosphere 263, 128096. https://doi.org/10.1016/j.chemosphere.2020.128096.

Tchorz-Trzeciakiewicz, D.E., Rysiukiewicz, M., 2021. Ambient gamma dose rate as an indicator of geogenic radon potential. Sci. Total Environ. 755, 142771 https://doi.org/10.1016/j.scitotenv.2020.142771.

Tehrani, F.S., Calvello, M., Liu, Z., Zhang, L., Lacasse, S., 2022. Machine learning and landslide studies: recent advances and applications. Nat. Hazards 114, 1197–1245. https://doi.org/10.1007/s11069-022-05423-7.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B 58 (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. J. R. Statist. Soc. B 73 (Part 3), 273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x.

Verdi, L., Weber, A., Stoppa, G., 2004. Indoor radon concentration forecasting in South Tyrol. Radiat. Prot. Dosim. 111 (4), 435–438. https://doi.org/10.1093/rpd/nch069.

Wilkening, M., 1980. Radon transport processes below the Earth's surface. In: The Natural Radiation Environment, III. US Department of Energy, Washington, DC, USA, pp. 90–104.

Yang, J., Busen, H., Scherb, H., Hürkamp, K., Guo, Q., Tschiersch, J., 2019. Modeling of radon exhalation from soil influenced by environmental parameters. Sci. Total Environ. 656, 1304–1311. https://doi.org/10.1016/j.scitotenv.2018.11.464.

Zalloni, E., Battipaglia, G., Cherubini, P., De Micco, V., 2018. Site conditions influence the climate signal of intra-annual density fluctuations in tree rings of *Q. ilex* L. Ann. For. Sci. 75, 68. https://doi.org/10.1007/s13595-018-0748-0.

Zeeb, H. & Shannoun, F. (eds.). WHO Handbook on Indoor Radon - A Public Health Perspective (2009). ISBNS 978 92 4 154767. https://www.who.int/publications/i/item/9789241547673.

Zhou, H., Su, H., Li, C., Wan, Y., 2023. Geochemical precursory characteristics of soil gas Rn, Hg, H2, and CO2 related to the 2019 Xiahe Ms5.7 earthquake across the northern margin of West Qinling fault zone, Central China. J. Environ. Rad. 264, 107190. https://doi.org/10.1016/j.jenvrad.2023.107190.