Science of the Total Environment 544 (2016) 883–891

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

# Mapping geogenic radon potential by regression kriging

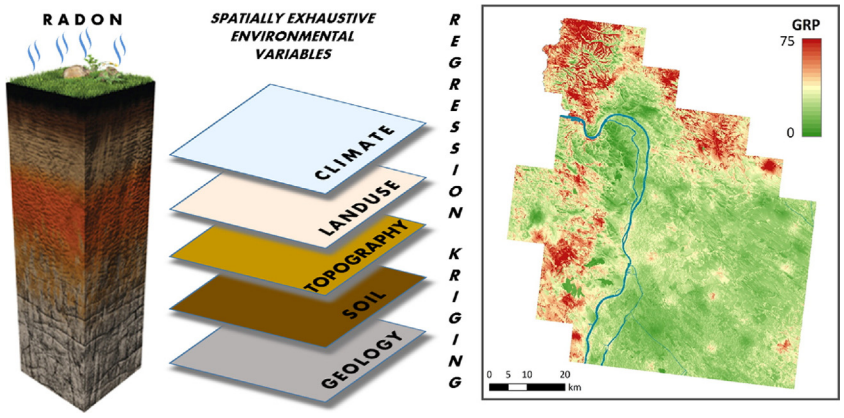
László Pásztor [a], Katalin Zsuzsanna Szabó [b,*], Gábor Szatmári [a], Annamária Laborczi [a], Ákos Horváth [c]

[a] Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences, Department of Environmental Informatics, Herman Ottó út 15, 1022 Budapest, Hungary
[b] Department of Chemistry, Institute of Environmental Science, Szent István University, Páter Károly u. 1, Gödöllő 2100, Hungary
[c] Department of Atomic Physics, Eötvös University, Pázmány Péter sétány 1/A, 1117 Budapest, Hungary

## HIGHLIGHTS

- A new method, regression-kriging (RK) was tested for GRP mapping.
- Usage of spatially exhaustive, auxiliary data on soil, geology, topography, land use and climate.
- Inherent accuracy assessment (both global and local).
- Interval estimation for the spatial extension of the areas of GRP risk categories.
- Significance of fluvial sedimentary rock, pyroclast and land use properties on radon risk.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Article history:*
Received 26 August 2015
Received in revised form 27 November 2015
Accepted 30 November 2015
Available online xxxx

Editor: Tack F.M.

*Keywords:*
Soil gas radon
Spatial prediction
Auxiliary environmental variables
Regression kriging
Physical soil properties

## ABSTRACT

Radon ($^{222}$Rn) gas is produced in the radioactive decay chain of uranium ($^{238}$U) which is an element that is naturally present in soils. Radon is transported mainly by diffusion and convection mechanisms through the soil depending mainly on the physical and meteorological parameters of the soil and can enter and accumulate in buildings. Health risks originating from indoor radon concentration can be attributed to natural factors and is characterized by geogenic radon potential (GRP). Identification of areas with high health risks require spatial modeling, that is, mapping of radon risk. In addition to geology and meteorology, physical soil properties play a significant role in the determination of GRP. In order to compile a reliable GRP map for a model area in Central-Hungary, spatial auxiliary information representing GRP forming environmental factors were taken into account to support the spatial inference of the locally measured GRP values. Since the number of measured sites was limited, efficient spatial prediction methodologies were searched for to construct a reliable map for a larger area. Regression kriging (RK) was applied for the interpolation using spatially exhaustive auxiliary data on soil, geology, topography, land use and climate. RK divides the spatial inference into two parts. Firstly, the deterministic component of the target variable is determined by a regression model. The residuals of the multiple linear regression analysis represent the spatially varying but dependent stochastic component, which are interpolated by kriging. The final map is the sum of the two component predictions. Overall accuracy of the map was tested by Leave-One-Out Cross-Validation. Furthermore the spatial reliability of the resultant map is also estimated by the calculation of the 90% prediction interval of the local prediction values. The applicability of the applied method as well as that of the map is discussed briefly.



* Corresponding author.
*E-mail address:* sz_k_zs@yahoo.de (K.Z. Szabó).

http://dx.doi.org/10.1016/j.scitotenv.2015.11.175
0048-9697/

## 1. Introduction

Natural radon ($^{222}$Rn) is a radioactive noble gas occurring in every soil due to the radium ($^{226}$Ra) and uranium ($^{238}$U) content of the lithosphere. Since radon is an inert gas it can easily enter from the soil into buildings and its daughter isotopes can cause damage to lung tissue. This process produces more than half the average natural dose for humans (2.4 mSv $y^{-1}$) (UNSCEAR, 2008). The impact of radon on health is highlighted by EU data which state that about 20,000 people die every year in the EU (Darby et al., 2005) due to elevated indoor radon concentration. Radon concentration shows a wide distribution in the soil. In some areas as high as ten times the average readings can be measured.

In several countries a number of regulations were adopted and countrywide surveys were launched to identify radon prone areas (Gue, 2015). The idea of the geogenic map is to visualize the purely natural radon hazard, i.e. independently of anthropogenic factors which are subject to secular changes, as building styles and living habits change with time and also vary regionally, while the geogenic radon potential (GRP) is constant over geological eras (Bossew et al., 2013). The basic concept of a geogenic radon prone area is a region where for natural, i.e. geogenic reasons, elevated indoor radon levels and elevated probability of their occurrence must be expected (Bossew, 2014). The geogenic source of this hazard (or potential risk) at a location or over an area is described by its radon potential. Knowing the GRP of an area one can decide whether the area should be investigated in detail or the assessment of the site of new buildings is necessary.

There are a lot of methods to define and map the physical quantity geogenic radon potential (GRP) of an area. These methods are based on several measured parameters as numerical (such as permeability, radon concentration in soil air, and $^{226}$Ra concentration) or geological, lithological data as categorical controls (Bossew et al., 2013). One of the internationally recognized approaches to quantify the GRP for the geogenic radon map of Europe is the continuous variable (formerly radon index) developed by Neznal et al. (2004), which is based on field measurements of the radon concentration in soil gas and the gas permeability of soils. This conventional continuous variable approach was applied for several areas of Hungary and the Czech Republic, too (Neznal et al., 2004; Szabó et al., 2014). In a different approach, multivariate classification, one cross-tabulates physical, mostly categorical factors which control the concept of the RP. The entries of the (possibly multi-dimensional) table are classified into RP classes. These factors are typically base and surface geology, geology, granulometry (as a proxy of permeability), hydrological properties, tectonics, and occurrence of "special features" such as caves, mines or other anthropogenically modified conditions which may enhance or reduce the natural RP. One elaborated example of this approach has been presented by Friedmann (2005).

At the same time several studies have shown the importance of the influence of soil physical parameters on the soil gas radon concentration. The effect of moisture content and grain size on radon emanation (radon escape from grain to the pore space) is well investigated with macroscopic soil models (Sakoda et al., 2010, Schumann and Gundersen, 1996). The emanation coefficient is higher where the moisture content is higher because the radon diffusion length is about 600 times lower (0.1 μm) in water than in air (63 μm) (Tanner, 1980). Thus radon atoms will remain in the pores (in the water) and could not reach another grain. Faheem and Matiullah (2008) investigated the moisture dependence of radon exhalation of several soils in laboratory measurements. Radon exhalation rate (radon escape from grain to the surface) was found to increase with an increase in moisture, reached its maximum value and then decreased with further increase in the water content. Schweikani et al. (1995) found that the increase in the moisture content causes a reduction in the radon diffusion since the pores through which radon diffuses is filled with water. Also they concluded that the degree of moisture saturation of the interstitial void space is the important factor rather than the moisture content as a percentage of the dry weight and/or the porosity of soil. The degree of moisture saturation is closely related to physical soil properties such as soil texture.

It is possible to establish accurate and robust estimation of specific derivatives by selected soil parameters using suitable pedo-radon transfer functions (PRTF). In this case more easily available soil maps and spatial databases can be used for the compilation of radon potential maps. Kemski et al. (2001) used an empirical ranking classification for the classification of geogenic radon potential due to the lack of exact functional relationship between radon concentration in soil gas and air permeability. Ielsch et al. (2002) emphasized the complicated interactions between the different pedological factors and radon exhalation, which preferably leads to statistically based models. Sun et al. (2004) found that radon exhalation from soil and soil radon concentration are more easily impacted by soil characters and they change in a rather large range. Winkler et al. (2001) investigated the spatial and temporal variability of the soil $^{222}$Rn concentration at field scale for rather small pilot areas. They found significant differences in the case of various soil conditions. Oliver and Khayrat (2001) focused their investigation on the spatial variation of radon concentration in soil. They emphasized that appropriate information about the spatial scales of radon variation in soil is needed to effectively sample for its spatial prediction, that is, mapping. Buttafuoco et al. (2007) tested various geostatistical methods (ordinary kriging, lognormal kriging, ordinary multi-Gaussian kriging, and ordinary indicator cokriging) to study spatial structure of radon concentration for mapping purposes. The tested methods did not use environmental co-variables for the spatial inference. Multi-Gaussian kriging proved to be the most accurate method of the considered interpolation techniques.

Sampling based mapping is inherently predictive, the value or class of the mapped variable can only be estimated at unvisited locations (Gessler et al., 1995; Scull et al., 2003). Spatial prediction can be carried out (i) taking exclusively the mapped variable into consideration based on its spatial features; (ii) also based on the mapped variable, but the constraints of spatial validity are provided by further spatial, ancillary information; (iii) in every predicted location supported by environmental, auxiliary co-variables (McKenzie and Ryan, 1999).

Geology, climate, physical soil properties and radiological data are the main GRP forming environmental factors. Spatial exhaustive information on them is available relatively more easily. Thus spatial inference of locally measured GRP values can rely on methods which exploit their existence.

Regression Kriging (RK; Hengl, 2009) is a spatial prediction technique that combines the regression of the dependent variable on auxiliary variables with kriging of the regression residuals. It is mathematically equivalent to the interpolation method variously called universal kriging and kriging with external drift. (Hengl et al., 2004). Essentially, RK respects this fact, neither environmental correlation nor pure geostatistical interpolation (simple, ordinary kriging) alone is able to account for the whole spatial variation that is to produce approximately perfect map products. They can be used as complementary spatial inference approaches where one can improve the other's drawbacks.

The main objective of the study has been to test a new method of GRP spatial prediction provided by regression-kriging (RK) using spatially exhaustive auxiliary environmental variables (geology, soil physical properties, topography, land use and climate). The expected result has been a more detailed map than previous maps based on spatial resolution of the selected auxiliary variables. A further aim has been to determine the performance and uncertainty of the method.

## 2. Materials and methods

### 2.1. Study area

The study area is located in the Pannonian Basin, in Central Hungary and includes Budapest, the majority of Pest County and some surrounding

areas (Fig. 1). The study area encompasses 5400 km$^2$ covering 6.5% of the country. This part of Hungary has the highest population density. 28% (2.83 million) of the population of the country (9.9 million) live in the 220 settlements of the study area. The area is also characterized by diverse geological and pedological environment, thus providing excellent conditions for radon risk mapping and geological modeling research.

The diverse geological background can be related to the joining point of the two highest Hungarian mountain ranges, the North Hungarian Mountains and the Transdanubian Mountains with the Great Hungarian Plain. Accordingly, there are hills (the highest elevation is 938 m asl.) along the longitudinal extension (N–S direction) in the western part of the area. The River Danube enters the study area from the west and sharply turns to the south at the Danube Bend, an intense tourist area. The northeast part is hilly (Cserhát Mts. and Gödöllő Hills with the highest elevation 652 and 345 m asl., respectively). The northwestern part of the Great Hungarian Plain, the Pest Plain covers the middle and southern parts (100–150 m asl.) of the study area (Fig. 1).

Hungary has a temperate continental climate with a long-term annual average temperature of 11 °C with warm summers (20 °C), cold winters (0 °C) and mild springs and autumns (10 °C). The annual average precipitation is 500–550 mm. The precipitation falls mostly during the summer months (especially June) and the least in the winter season (especially February) based on the Hungarian Meteorological Service (OMSZ) database.

### 2.2. Data collection

In the present mapping approach we used a subset of our former study (Szabó et al., 2014). Original locations of measurements contain two densely sampled areas (two settlements). In order to apply regression kriging efficiently, we filtered the original dataset. We only included two representative locations per settlement. Additionally two locations were left out which were outside Hungary, since the auxiliary environmental data were available only for Hungary. Altogether 145 locations were involved in the analyses. These GRP data are based on field measured soil gas radon activity concentration and soil gas permeability.

Soil gas radon activity concentration was measured in situ with a RAD7 Electronic Radon Detector (Durridge Company Inc., 2000) coupled with a soil probe. Soil gas was pumped out from (if possible) generally 0.8 m depth of the soil. The standard "Grab" sampling lasted 20 min. Soil gas radon activity concentration was determined from the 3-min half-life $^{218}$Po alpha peak at 6.0 MeV and gives radon activity concentration in Bq m$^{-3}$ units. The rate of flow of the pump is 1 L min$^{-1}$. The RAD7 was calibrated in 2009 and the calibration is highly stable according to the manufacturer's specifications (Durridge Company Inc., 2000). Typical drift is less than 2% per year. The uncertainty (two-sigma) of RAD7 with a 5-min counting cycle is about 40%, 10%, 8% and 5% in the case of 1, 10, 20, 40 kBq m$^{-3}$ radon activity concentration values, respectively.

Soil gas permeability measurement was performed immediately after the soil gas radon measurement by Radon-JOK equipment (Radon v.o.s.) using the same soil probe. Calculation of the gas permeability was based on Darcy's equation (Koorevaar et al., 1983) according to the equipment manual (Radon v.o.s.). The equipment works by withdrawing air by the application of negative pressure. Air is pumped from the soil under constant pressure through the probe with a constant contact between the probe head and the soil. The soil is assumed to be homogeneous and isotropic with standard conditions prevailing. Furthermore, the air is assumed to be incompressible (pressure differences are very much smaller than atmospheric pressure).

Site selection followed stratified, conditional random sampling. The internationally suggested 10 × 10 km grid for the European indoor radon map (Dubois et al., 2010; Tollefsen et al., 2011) represented the uniform strata of the sampling (Fig. 2). The conditions of site selection were the extension of geological formations and the distribution of settlements (built-up areas). On average, three measurement sites were assigned in each cell sampling the three dominating geological formations of the cell, also taking the locality of the towns into account. We preferred geological formations inside and around the settlements since these are the target areas for planned building developments.

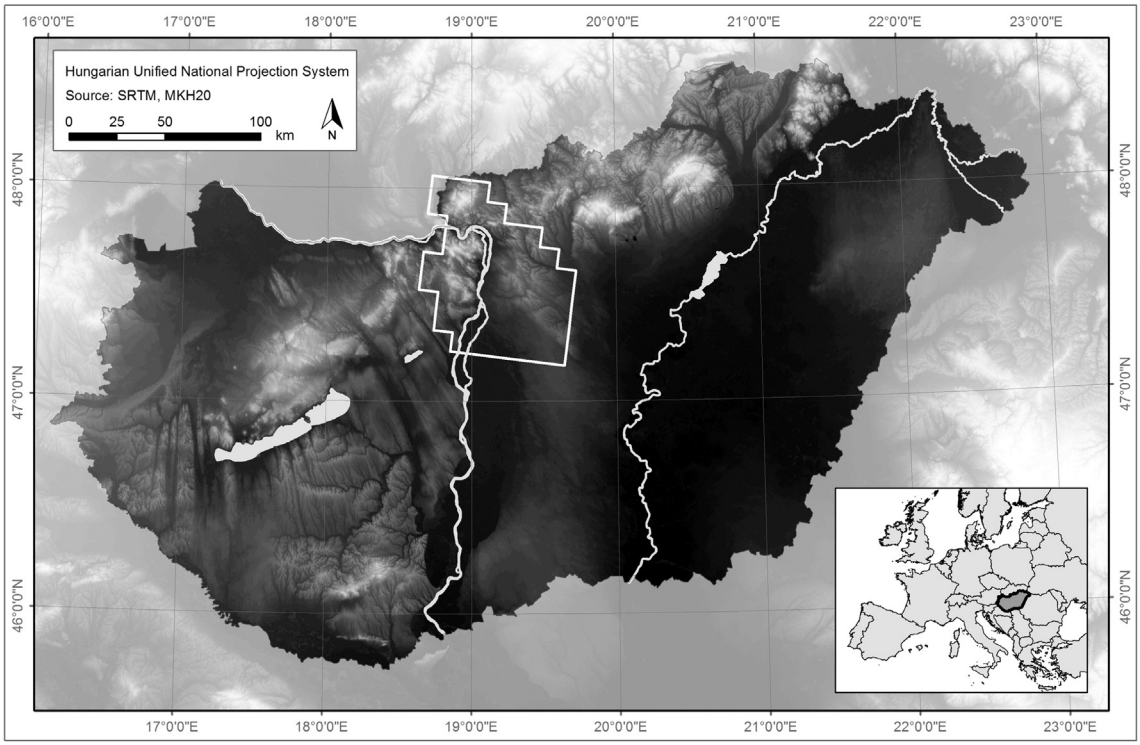



**Fig. 1.** The location of the study area.

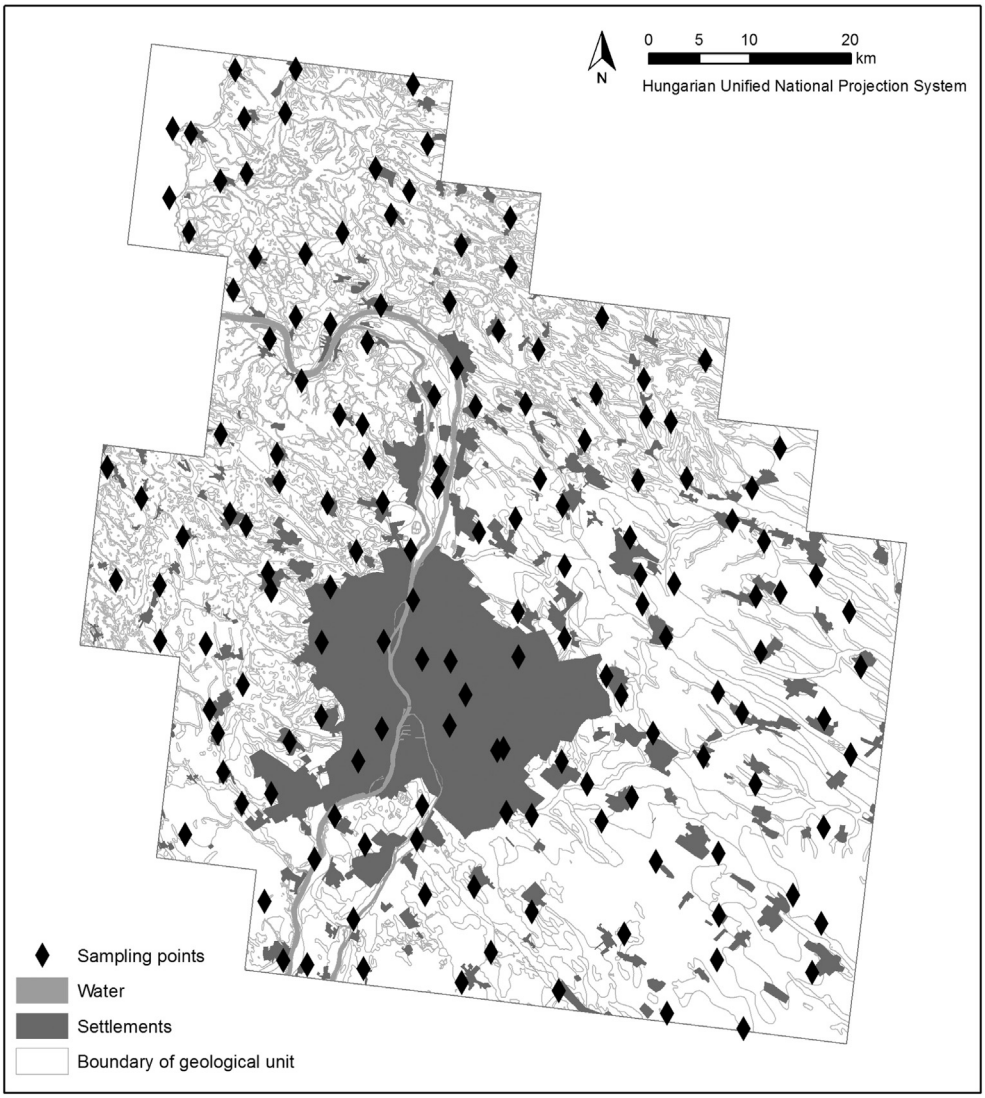



**Fig. 2.** Spatial distribution of the sampling points within the pilot area.

The average nearest-neighbor distance between the measurement sites is 4.15 km. At each site, one "Grab" soil gas radon and one soil gas permeability measurement were made.

Measurements were performed in all months (i.e., from May 2010 to December 2011), but mostly in summer, from May to September (85% of the measurements), to reduce the possible effect of the seasonality. Measurements were performed between 7:30 am and 9:00 pm.

### 2.3. Database

For the compilation of a proper regression kriging interpolation model we have used spatially exhaustive ancillary information on soil, geology, topography, climate and land use (Table 1).

The Digital Kreybig Soil Information System (DKSIS; Pásztor et al., 2010, 2012) consists of soil mapping units, which were delineated based on overall chemical and physical soil properties of the soil root zone at a scale of 1:50,000. Combined texture and water management categories attributed to SMUs were elaborated according to water retention capability, permeability and infiltration rate of soils. As a consequence physical soil property layer of DKSIS provides regionalized, that is spatial ancillary information related to GRP behavior, which could be involved in the spatial inference.

Geology was represented by proper segment of the Geological Map of Hungary 1:100,000 (Gyalog & Síkhegyi, 2005). In order to simplify the huge amount of geology and facies categories of the map, they were correlated with the nomenclature of parent material defined in the FAO Guidelines for soil description (Bakacsi et al., 2014; FAO, 2006). The applied dataset describes the geological environment by 15 categories.

Topography was also taken into consideration for the potential refinement of the spatial inference. On the one hand, spatially exhaustive information on topography in the form of digital elevation models (DEM) is relatively easily available as compared to other relevant thematic themes (like soil or geology). On the other hand DEM and its joint, specific morphometric derivatives are highly informative on the latter. These are the reasons why DEMs are used in digital soil mapping. As a consequence, a detailed DEM may represent the unmapped variability of the less detailed environmental data, on the whole improving the spatial prediction of the target variable, GRP in our case. Topography was characterized by a 25 m Digital Elevation Model (EU-DEM dataset) and its various morphometric derivatives (Aspect, Diurnal Anisotropic Heating, Elevation, General Curvature, Multiresolution Index of Valley Bottom Flatness — MRVBF, SAGA Wetness Index, Slope, Topographic Position Index, Topographic Wetness Index and Vertical Distance to Channel Network). The secondary terrain features were calculated from the DEM within SAGA GIS (Bock et al., 2007).

Climate was represented by four relevant layers: average annual precipitation, average annual temperature, annual evaporation and

**Table 1**
Set of potential auxiliary variables. Column labeled 'Observed significance' shows the significant parameters and the calculated significance levels from to the stepwise regression method. Parameters which are indicated by at least one asterisk were entered into the mapping procedure. Three asterisks indicate 0.1%, two asterisks indicate 1% and one asterisk indicates 5% significance level. Column labeled 'Effect on GRP' shows the positive or negative correlation of the variable with the GRP.

| GRP factor | Representing auxiliary variables | Type | Category values | Observed significance | Effect on GRP |
|---|---|---|---|---|---|
| Geology | Geological formation according to the Geological Map of Hungary (1:100,000), correlated with the FAO nomenclature of parent material | Categorical | Acid igneous rock | >0.05 | |
| | | | Intermediate igneous rock | >0.05 | |
| | | | Pyroclastic rock | 0.0429* | − |
| | | | Acid metamorphic rock | >0.05 | |
| | | | Clastic sediments (consolidated sedimentary rock) | >0.05 | |
| | | | Carbonatic, organic (consolidated) sedimentary rock | >0.05 | |
| | | | Anthropogenic/technogenic (unconsolidated) sedimentary rock | >0.05 | |
| | | | Colluvial (unconsolidated) sedimentary rock | >0.05 | |
| | | | Eolian (unconsolidated) sedimentary rock | >0.05 | |
| | | | Hydroeolic (unconsolidated) sedimentary rock | >0.05 | |
| | | | Fluvial (unconsolidated) sedimentary rock | 0.0158* | − |
| | | | Lacustrine (unconsolidated) sedimentary rock | >0.05 | |
| | | | Marine, estuarine (unconsolidated) sedimentary rock | >0.05 | |
| | | | Organic (unconsolidated) sedimentary rock | >0.05 | |
| | | | Weathered residuum (unconsolidated sedimentary rock) | >0.05 | |
| Land use/land cover | Land cover categories according to CORINE Land Cover 1:50,000 (CLC50) | Categorical | Arable land | 0.0084** | + |
| | | | Orchard, vineyard | >0.05 | |
| | | | Grassland | 0.0109* | + |
| | | | Sparse vegetation | >0.05 | |
| | | | Waterlogged area | >0.05 | |
| | | | Forest | 0.0003*** | + |
| | | | Water | >0.05 | |
| | | | Sealed soil | 0.0002*** | + |
| Soil | Soil physical categories according to the Digital Kreybig Soil Information System (DKSIS) | Categorical | Good water retention and good permeability, and infiltration rate | >0.05 | |
| | | | High water retention and moderate permeability, and infiltration rate | >0.05 | |
| | | | High water retention and poor permeability, and infiltration rate, vertic properties | >0.05 | |
| | | | Moderate water retention and high permeability, and infiltration rate | >0.05 | |
| | | | Poor water retention and very high permeability, and infiltration rate | >0.05 | |
| | | | Very high water retention and good permeability, and infiltration rate | >0.05 | |
| | | | Stony, gravelly soils | 0.0049** | − |
| | | | Peaty soils | >0.05 | |
| | Average annual precipitation | | | 0.0167* | + |
| | Average annual temperature | | | >0.05 | |
| | Annual evaporation | | | 0.0112* | + |
| Climate | Average annual evapotranspiration | Continuous | | 0.0129* | + |
| | Aspect | | | <0.0001*** | + |
| | Diurnal Anisotropic Heating | | | <0.0001*** | − |
| | Elevation | | | >0.05 | |
| | General curvature | | | 0.0046** | − |
| | Multiresolution Index of Valley Bottom Flatness — MRVBF | | | 0.0129* | + |
| | SAGA Wetness Index | | | <0.0500 | |
| | Slope | | | 0.0028** | − |
| | Topographic Position Index | | | 0.0046** | − |
| | Topographic Wetness Index | | | 0.0148* | − |
| Topography | Vertical Distance to Channel Network | Continuous | | <0.0001*** | − |

average annual evapotranspiration. The spatial layers are compiled using the MISH method elaborated for the spatial interpolation of surface meteorological elements (Szentimrey et al., 2011; Szentimrey and Bihari, 2015) by which the 30 year observations of the Hungarian Meteorological Service were interpolated with 0.5′ resolution.

Land use was taken into account by the CORINE Land Cover 1:50.000 (CLC50; Büttner et al., 2004). CLC50 is a national land cover database elaborated on the basis of the CORINE nomenclature of the European Environment Agency (EEA), and fitted to the characteristics of Hungary. Merged categories of CLC50 were used (arable land, orchard & vineyard, grassland, sparse vegetation, forest, waterlogged area, water, sealed soil) to distinguish regions with different land cover.

### 2.4. Regression Kriging

RK assumes that, the deterministic component of the target variable is accounted for by the regression model, while the model residuals represent the spatially varying but dependent stochastic component. The estimation for Z variable at an unvisited location $\mathbf{s}_0$ is:

$$Z(\mathbf{s}_0) = \mathbf{q}_0^{\mathbf{T}} \cdot \boldsymbol{\beta} + \boldsymbol{\lambda}_0^{\mathbf{T}} \cdot (\mathbf{z} - \mathbf{q} \cdot \boldsymbol{\beta}), \tag{1}$$

where $\boldsymbol{\beta}$ is the vector of the regression coefficients, $\mathbf{q_0}$ is the vector of the covariates at the unvisited location, $\boldsymbol{\lambda_0}$ is the vector of the kriging weights, $\mathbf{z}$ is the vector of the observations and $\mathbf{q}$ is the matrix of covariates at the sampling locations. Regression Kriging Error Variance (RKV) at $\mathbf{s}_0$ is given by:

$$\sigma^2(\mathbf{s}_0) = c(0) - \mathbf{c}_0^{\mathbf{T}} \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0 + \\ + \left(\mathbf{q}_0 - \mathbf{q}^{\mathbf{T}} \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0\right)^{\mathbf{T}} \cdot \left(\mathbf{q}^{\mathbf{T}} \cdot \mathbf{C}^{-1} \cdot \mathbf{q}\right)^{-1} \cdot \left(\mathbf{q}_0 - \mathbf{q}^{\mathbf{T}} \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0\right), \tag{2}$$

where $c(0)$ is the variance of the residuals, $\mathbf{c_0}$ is the vector of covariances between the residuals at the observed and unvisited locations and $\mathbf{C}$ is

the variance–covariance matrix of the residuals. RKV is independent from the observed values.

The RK algorithm, especially the inherent matrix calculus, requires the same spatial resolution of the predictor variables. Topographical features were derived from a 25 m DEM, while climatic parameters were originally predicted with 0.5′ resolution. In order to harmonize the former was upscaled, while the latter was downscaled to a common 100 m grid system. The corresponding predictors were resampled via SAGA GIS (Bock et al., 2007). The nominal scales of the polygon-based maps on soil, geology and land use are also different, but their real spatial information density only slightly differs from that corresponding to a 100 m raster resolution. They were also converted to the common raster format with proper raster-to-vector conversion. The common 100 m grid system also defines the spatial resolution of the result map.

Before regression kriging, all of the auxiliary variables were normalized into a common scale (0–255). Category variables were taken apart into indicator variables according to their categories. Every single category has become a layer with a 0 or 255 value, where 255 is assigned to the presence of the given category, while 0 is assigned to the absence of the given category, according to the 8 bit coding.

Multiple linear regression analysis (MLRA) was used for modeling the joint effect of the selected environmental factors on GRP. In order to decrease the multicollinearity carried by the applied variables principal component analysis (PCA) was performed on the continuous environmental auxiliary variables and the resulting principal components (PCs) were used in further procedures. Since PCs are orthogonal and independent, they satisfy the requirements of MLRA. The PCs and the categorical variables were used as explanatory variables and the natural logarithm values of GRP as response variable in MLRA. Selection of the most affecting environmental covariates was carried out by applying stepwise selection and 0.05 significance level. The coefficient of determination proved to be 0.41% and the model was significant, indicating that there is real functional correlation between the dependent and independent variables.

MLRA only partly explains the spatial variability (pattern) of the distribution of GRP. On the other hand, taking the role of environmental factors into account by the linear model eliminates the trend which used to conflict with geostatistical interpolation. Kriging of the MLRA residuals represents the stochastic factor of RK. The semivariogram of the residuals was estimated to model the spatial auto-correlation of the stochastic component. Nested semivariogram model was fitted composed of two components: (i) spherical, isotropic, partial sill = 0.1836, nugget = 0, range = 3000 m; (ii) spherical, isotropic, partial sill = 0.185, range = 24,000 m. The fitted model was applied to calculate the kriging weights in the spatial interpolation. Superimposing the regression and interpolation results has provided the overall RK prediction map.

### 2.5. Validation of the prediction map

The accuracy of the prediction was tested both globally and locally. Overall performance of the map was tested by Leave-One-Out Cross-Validation (LOOCV). The estimated spatial reliability of the result map has been represented by the 90% prediction interval (PI) of the predicted GRP.

In the frame of LOOCV RK is carried out *n-1* times, leaving out each time one of the samples. The predicted and measured GRP values of the left-out sample are then compared. The series of the results of these comparisons is used for the estimation of the overall accuracy by the following indicators: mean error (ME), mean absolute error (MAE), root mean squared error (RMSE) (Eqs. (1)–(3)).

$$ME = \frac{1}{n}\cdot\sum_{i=1}^{n}[z(s_i) - z(s_i)] \quad (1)$$

$$MAE = \frac{1}{n}\cdot\sum_{i=1}^{n}|z(s_i) - z(s_i)| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}[z(s_i) - z(s_i)]^2} \quad (3)$$

where $z(s_i)$ and $z(s_i)$ the predicted and measured values in $s_i$ place.

The results of RK can be further utilized for local accuracy assessment by proper quantification of the uncertainty. Assuming the normality of the predicted variable, the lower and upper limits of the 90% prediction interval can be calculated by subtracting and adding 1.64 times the kriging standard deviation to the prediction values provided by regression kriging (Heuvelink, 2014), thus providing an estimation for the spatial reliability of the result map.

## 3. Results

Summary statistics of the observed and predicted GRP for 145 sites and the overall prediction for the whole area are shown in Table 2.

The predicted GRP map is displayed in Fig. 3. The estimated spatial reliability of the result map is represented by the 90% prediction interval of the predicted GRP. The map of lower and upper estimate shows the lower and upper boundary of the 90% prediction interval respectively for each cell of prediction.

Turning continuous GRP values into classes according to Neznal et al. (2004) risk categorization, the areas characterized by small, medium and large values are presented in Fig. 4 (middle panel). Low risk means GRP lower than 10, medium risk means GRP between 10 and 35 and high risk means GRP higher than 35. The left and right panels show the categorization according to the 90% lower and upper predictions (as above).

We summed up the small, medium and large risk areas on all three categorized maps. These values can give a 90% interval estimation for the areal extension of the three risk categories (Table 3). Our estimate for the large risk area gave 1110 ha and the 90% interval estimate is (719 ha, 1749 ha).

The results of the Leave-One-Out-Cross-Validation process are summarized in Table 4. The mean difference of the estimated and the observed values resulted in −1.47. This means that the process underestimates the GRP values in general. The root means square value of the differences (Eq. (3)) was found to be 7.5 (see Table 4). This average value is 21% of the large risk limit (35).

The stepwise regression method provides the possibility for evaluating the importance of the applied environmental covariables using the significance of the entering variables. The auxiliary category variables, which proved to be important in the analysis are indicated with asterisks in Table 1. The fifth column of Table 1. shows the significance of the mapping parameters. Parameters which are indicated by at least one asterisk were entered into the stepwise multiple linear regression analysis. Three asterisks indicate 99.9%, two asterisks indicate 99%, one asterisk indicates 95% significance level. The last column labeled 'Effect on GRP' shows the positive or negative correlation of the variable with the GRP.

**Table 2**
Summary statistics for the observed and predicted 145 GRP sites and for the whole area.

| Variable | Mean | Minimum | Maximum | Std. dev. |
|---|---|---|---|---|
| Observed GRP for 145 sites | 10.8 | 1.0 | 43.9 | 8.6 |
| Estimated GRP for 145 sites | 9.4 | 0.8 | 36.2 | 5.5 |
| Estimated GRP for the whole area | 8.6 | 0.2 | 75 | 4.9 |

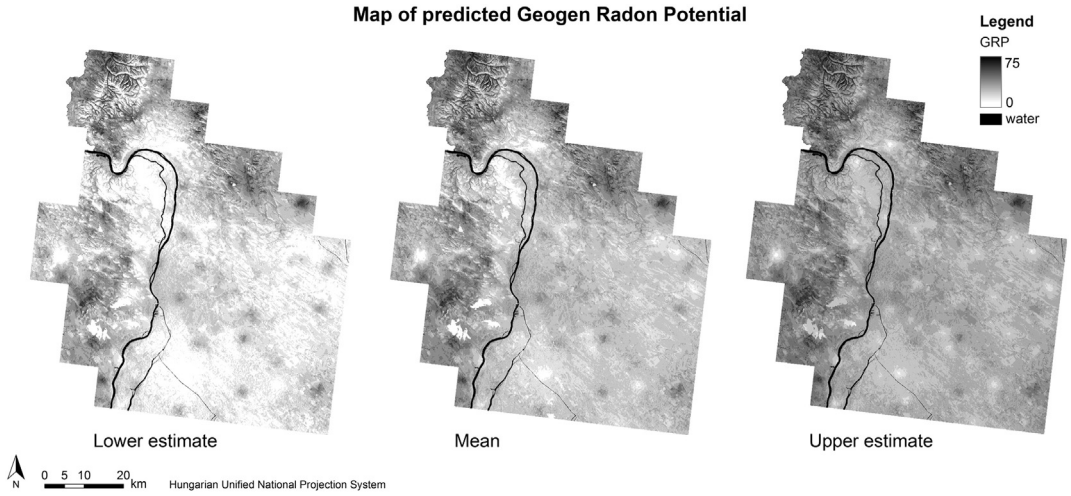



**Fig. 3.** RK-based predicted GRP map of the pilot area with 90% prediction interval boundaries.

## 4. Discussion

We tested a new method for GRP mapping, namely regression-kriging (RK) using spatially exhaustive auxiliary environmental variables related to geogenic radon potential. Our hypothesis on the applicability of environmental covariables in the mapping process is supported by the fact that all the five selected environmental factors (topography, land use, soil physical properties, climate and geology) are represented in the final MLRA model thus contributing to the compilation of the GRP map. Another useful feature of the presented approach is the furnishing of the estimated uncertainty. This type of mapping may represent an efficient contribution to the European radon mapping strategy.

GRP is a complex environmental phenomenon. Its behavior is related to various environmental factors. The present approach enables the inclusion of several, spatially exhaustive auxiliary variables that are available for mapping. Increasing the number of the applied parameters would increase the performance of the map. Recent expansion of available spatial data infrastructure might ensure the further improvement of GRP maps. Applying the method at different locations can result in other important parameters.

The spatial resolution of the recent map is better than that of the previous radon map of Pest County. Besides the geological information this map is built on more relevant spatial information, as a consequence the accuracy of the present map outperforms that of the former, purely geology based map.

Medium (10–35) and low (<10) GRP values characterizes the study area (Fig. 4). High GRP values (>35) are located only locally in the hilly areas (Fig. 4). Estimated GRP values show similar spatial patterns found in the geogenic mapping procedure presented in Szabó et al. (2014). Szabó et al. (2014) created the GRP risk map by attributing GRP median values to geological formations based on field measurements. The two independent studies gave accordant results. The high risk areas are located in the hilly side of the area, and the plain at the southern eastern region corresponds to low risk. The acid and intermediate rocks (e.g. rhyolite, dacite and andesite) which usually contain enhanced concentration of uranium and radium are principally located in the hilly areas. The detailed locations of the radon risk areas, however, differ on the two maps according to the differences in the methods.

In addition to map construction, which was the main aim of our study, the procedure resulted in how significant the co-variables are in the determination of GRP. The two most significant lithological categories were the fluvial sedimentary rock and the pyroclast. It is interesting that the geological parameters do not play the most important role, but this result does not mean the insignificance of the geological information. Some geological information is hidden in certain topographical variables. For example the identified significance of Vertical Distance to Channel Network may be related to mass movement along the slopes

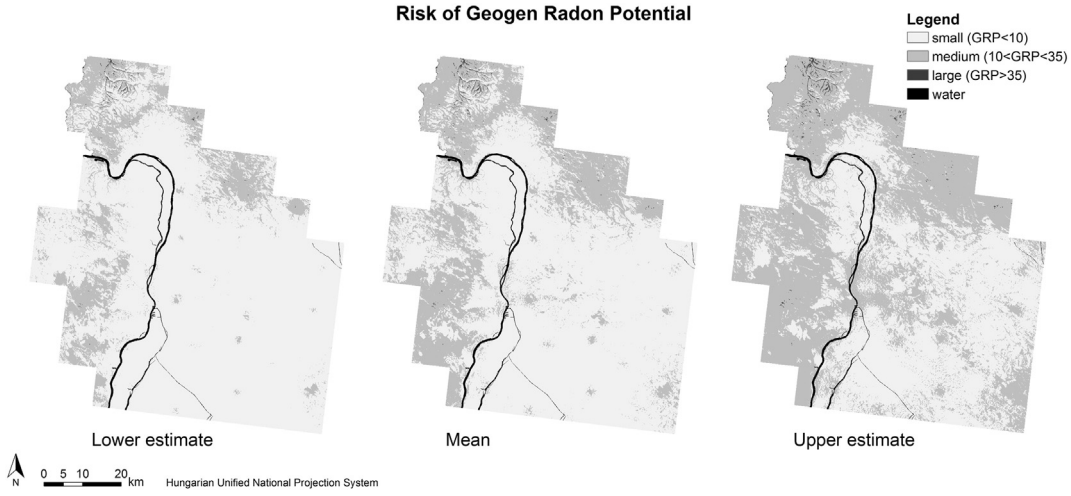



**Fig. 4.** RK-based predicted GRP map according to the Neznal et al. (2004) risk categorization. Predicted categories according to 90% prediction interval lower and upper boundaries are also indicated.

**Table 3**
90% interval estimate for the areal extension of the three GRP risk categories.

| GRP risk category | Area affected [ha] | | |
|---|---|---|---|
| | Lower estimate | Mean estimate | Upper estimate |
| Small | 444,529 | 383,087 | 265,974 |
| Medium | 82,626 | 143,677 | 260,151 |
| Large | 719 | 1110 | 1749 |

**Table 4**
Result of the Leave-One-Out-Cross-Validation (LOOCV). (ME = mean error, MAE = mean absolute error, RMSE = root mean squared error, AVRD = average Value of the relative difference).

| | |
|---|---|
| ME | −1.47 |
| MAE | 5.42 |
| RMSE | 7.50 |

mixing on-site origin and transported material. More generally the erosion on the hilly/mountainous areas as well as the deposition on the plains is partly driven by topography, while they determine some geological and pedological features like thickness of soil – that is distance to bedrock – which can have significant effects on GRP. The emanation coefficient of soils can also be affected by diurnal asymmetric heating on slopes with differing aspect. Land use properties also have significance in radon risk. This could be attributed to their effect on the permeability of soils similarly to that of soil physical categories.

A further improvement of the RK-based GRP mapping would undoubtedly be the introduction of scanned images provided by airborne gamma measurements. The gamma activity reflects the potassium, uranium or the thorium content of the soil in the first 30 cm from the surface. One of these, the uranium concentration, is a main factor in the soil air radon concentration, the main factor of geogenic radon potential.

## 5. Conclusions

We investigated a new approach (regression kriging) for mapping geogenic radon potential, which has several advantages such as application of environmental co-variables (used in spatially continuous form), refined spatial resolution and inherent accuracy assessment. Our results demonstrated that not only pure geological information should be relied on in GRP mapping, but further environmental variables (related to climate, land use, soil and topography) also play important roles. The resulting map is characterized by global and local measures of its accuracy. The method also provides interval estimation for the spatial extension of the areas of risk categories. All of these outputs provide useful contribution to spatial planning, radon action planning and decision making.

## Acknowledgements

Our work has been partly supported by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167 and OTKA, Grant No. PD115810). Special thanks go to Csaba Szabó, Lithosphere Fluid Research Lab (LRG), for his valuable support.

## References

Bakacsi, Zs., Laborczi, A., Szabó, J., Takács, K., Pasztor, L. (2014). Proposed correlation between the legend of the 1:100,000 scale geological map and the FAO code system for soil parent material. (in Hungarian) Agrokémia és Talajtan 63(2), 189–202.

Bock, M., Böhner, J., Conrad, O., Köthe, R., Ringeler, A., 2007. Methods for creating functional soil databases and applying Digital Soil Mapping with SAGA GIS. In: Hengl, T., Panagos, P., Jones, A., Toth, G. (Eds.), Status and Prospect of Soil Information in South-eastern Europe: Soil Databases, Projects and ApplicationsEUR 22646 EN Scientific and Technical Research series. Office for Official Publications of the European Communities, Luxemburg, pp. 149–162.

Bossew, P., 2014. Determination of radon prone areas by optimized binary classification. J. Environ. Radioact. 129 (2014), 121–132.

Bossew, P., Tollefsen, T., Gruber, V., De Cort, M., 2013. The European radon mapping project. IX. Latin American IRPA Regional Congress on Radiation Protection and Safety — IRPA 2013. Sociedade Brasileira De Protecao Radiologica - SBPR, Rio de Janeiro, Brazil (April 15–19, 2013).

Buttafuoco, G., Tallarico, A., Falcone, G., 2007. Mapping soil gas radon concentration: a comparative study of geostatistical methods. Environ. Monit. Assess. 131, 135–151.

Büttner, Gy., Maucha, G., Bíró, M., Kosztra, B., Pataki, R., Petrik, O. (2004). National land cover database at scale 1:50,000 in Hungary. EARSeL eProceedings, 3(3), 323–330.

Darby, S., Hill, D., Auvinen, A., Barrios-Dios, J.M., Baysson, H., Bochicchio, F., Deo, H., Falk, R., Forastiere, F., Hakama, M., Heid, I., Kreienbrock, L., Kreuzer, M., Lagarde, F., Makelainen, I., Muirgead, C., Oberaigner, W., Pershagen, G., Ruani-Ravina, A., Ruosteenoja, E., Rosario, A.S., Tirmarche, M., Tomasek, L., Whitley, E., Wichmann, H.E., Doll, R., 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case–control studies. Br. Med. J. 330 (7485), 223–226.

Dubois, G., Bossew, P., Tollefsen, T., De Cort, M., 2010. First steps towards a European atlas of natural radiation: status of the European indoor radon map. J. Environ. Radioact. 101, 786–798.

Durridge Company Inc., 2000. RAD7 electronic radon detector, User Manual.

EU-DEM dataset. http://www.eea.europa.eu/data-and-maps/data/eu-dem#tab-european-data European Environment Agency.

Faheem, M., Matiullah, 2008. Radon exhalation and its dependence on moisture content from samples of soil and building materials. Radiat. Meas. 43, 1458–1462.

FAO, 2006. Guidelines for soil description. fourth ed. p. 97 (Rome, FAO).

Friedmann, H., 2005. Final results of the Austrian radon project. Health Phys. 89 (4), 339–348.

Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil–landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Inf. Syst. 9, 421–432.

Gue, L., 2015. David Suzuki Foundation Report: Revisiting Canada's Radon Guideline (ISBN digital: 978–1-897375-91-4 print: 978–1-897375-90-7).

Gyalog, L., Síkhegyi, F. (Eds.), 2005. Geological Map of Hungary, 1:100 000. Geological Institute of Hungary, Budapest (In Hungarian). Digital version. Retrieved December, 1, 2008 from http://loczy.mfgi.hu/fdt100/.

Hengl, T., Heuvelink, G.M.B., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 122 (1-2), 75–93.

Hengl, T., 2009. A practical guide to geostatistical mapping. University of Amsterdam, Amsterdam.

Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products. In: Arrouays, et al. (Eds.), GlobalSoilMap. Taylor & Francis Group, London, pp. 335–340.

Ielsch, G., Ferry, C., Tymen, G., Robé, M.C., 2002. Study of a predictive methodology for quantification and mapping of the radon-222 exhalation rate. J. Environ. Radioact. 63 (1), 15–33.

Kemski, J., Siehl, A., Stegemann, R., Valdivia-Manchego, M., 2001. Mapping the geogenic radon potential in Germany. Sci. Total Environ. 272, 217–230.

Koorevaar, P., Menelik, G., Dirksen, C., 1983. Elements of Soil Physics. Developments in Soil Science vol. 13. Elsevier, Amsterdam, New York.

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.

Neznal, M., Neznal, M., Matolin, M., Barnet, I., Miksova, J., 2004. The New Method for Assessing the Radon Risk of Building Sites. Czech Geol. Survey Special Papers vol. 16. Czech Geol. Survey, Prague (http://www.radon-vos.cz/pdf/metodika.pdf).

Oliver, M.A., Khayrat, A.L., 2001. A geostatistical investigation of the spatial variation of radon in soil. Comput. Geosci. 27, 939–957.

Pásztor, L., Szabó, J., Bakacsi, Z., 2010. Digital processing and upgrading of legacy data collected during the 1:25.000 scale kreybig soil survey. Acta Geodaet. Geophys. Hung. 45, 127–136.

Pásztor, L., Szabó, J., Bakacsi, Zs., Matus, J. & Laborczi, A. 2012. Compilation of 1:50,000 scale digital soil maps for Hungary based on the digital kreybig soil information system. J. Maps 8(3): 215–219.

Sakoda, A., Ishimori, Y., Hanamoto, K., Kataoka, T., Kawabe, A., Yamaoka, K., 2010. Experimental and modeling studies of grain size and moisture content effects on radon emanation. Radiat. Meas. 45, 204–210.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography 27, 171–197.

Schumann, R., Gundersen, L.C.S., 1996. Geologic and climatic controls on the radon emanation coefficient. Environ. Int. 22 (Suppl. I), S439–S446.

Schweikani, R., Giaddui, T.G., Durrani, S.A., 1995. The effect of soil parameters on the radon concentration values in the environment. Radiat. Meas. 25 (1-4), 581–584.

Sun, K., Guo, Q., Cheng, J., 2004. The effect of some soil characteristics on soil radon concentration and radon exhalation from soil surface. J. Nucl. Sci. Technol. 41 (11), 1113–1117.

Szabó, K.Z., Jordan, G., Horváth, Á., Szabó, C., 2014. Mapping the geogenic radon potential: methodology and spatial analysis for central Hungary. J. Environ. Radioact. 129, 107–120.

Szentimrey, T., Bihari, Z., 2015. Meteorological Interpolation based on Surface Homogenized Data Basis (MISH v1.02). Hungarian Meteorological Service (http://www.dmcsee.org/uploads/file/330_1_mishmanual.pdf p.33).

Szentimrey, T., Bihari, Z., Lakatos, M., Szalai, S., 2011. Mathematical, methodological questions concerning the spatial interpolation of climate elements. Proceedings from the Second Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2009. Időjárás 115, pp. 1-2–1-11.

Tanner, A.B., 1980. Radon migration in the ground: supplementary review. In: Gesell, T.F., Lowder, W.M. (Eds.), Proc. Natural Radiation Environment III, Conf-780422. US Dept. of Commerce, National Technical Information Service, Springfield, VA, p. 5 (1980).
Tollefsen, T., Gruber, V., Bossew, P., De Cort, M., 2011. Status of the European indoor radon map. Radiat. Prot. Dosim. 145, 110–116.
UNSCEAR, 2008. Effects of Ionizing Radiation. Vol. I. Annex B. United Nations, New York.
Winkler, R., Ruckerbauer, F., Bunzl, K., 2001. Radon concentration in soil gas: a comparison of the variability resulting from different methods, spatial heterogeneity and seasonal fluctuations. Sci. Total Environ. 272, 273–282.