Journal of Environmental Radioactivity xxx (2016) 1–6

Contents lists available at ScienceDirect

Journal of Environmental Radioactivity

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

Modeling of geogenic radon in Switzerland based on ordered logistic regression

Georg Kropat [a,*], François Bochud [a], Christophe Murith [b], Martha Palacios (Gruson) [b], Sébastien Baechler [b]

[a] Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland
[b] Swiss Federal Office of Public Health, Schwarzenburgstrasse 157, 3003 Berne, Switzerland

ARTICLE INFO

Article history:
Received 9 November 2015
Received in revised form
20 April 2016
Accepted 6 June 2016
Available online xxx

Keywords:
Geogenic radon
European atlas of natural radiation
Ordered logistic regression
Terrestrial gamma dose rate
Fault lines
Soil permeability

ABSTRACT

Purpose: The estimation of the radon hazard of a future construction site should ideally be based on the geogenic radon potential (GRP), since this estimate is free of anthropogenic influences and building characteristics. The goal of this study was to evaluate terrestrial gamma dose rate (TGD), geology, fault lines and topsoil permeability as predictors for the creation of a GRP map based on logistic regression.
Method: Soil gas radon measurements (SRC) are more suited for the estimation of GRP than indoor radon measurements (IRC) since the former do not depend on ventilation and heating habits or building characteristics. However, SRC have only been measured at a few locations in Switzerland. In former studies a good correlation between spatial aggregates of IRC and SRC has been observed. That's why we used IRC measurements aggregated on a 10 km × 10 km grid to calibrate an ordered logistic regression model for geogenic radon potential (GRP). As predictors we took into account terrestrial gamma doserate, regrouped geological units, fault line density and the permeability of the soil.
Results: The classification success rate of the model results to 56% in case of the inclusion of all 4 predictor variables. Our results suggest that terrestrial gamma doserate and regrouped geological units are more suited to model GRP than fault line density and soil permeability.
Conclusion: Ordered logistic regression is a promising tool for the modeling of GRP maps due to its simplicity and fast computation time. Future studies should account for additional variables to improve the modeling of high radon hazard in the Jura Mountains of Switzerland.

1. Introduction

Radon is a radioactive gas that is known to be the second leading cause of lung cancer after smoking (Zeeb and Shannoun, 2009). As a daughter product of Uranium, radon appears everywhere in nature and occurs in substantial concentrations especially in buildings. Hence, the major exposure of the population to radon takes place at home. In order to manage this health risk, many countries have developed maps based on indoor radon measurements (Dubois, 2005). The drawback of maps based on indoor radon measurements is that they are dependent on building characteristics and anthropogenic influences. This limits the generalizability of the resulting estimates. In order to get rid of the anthropogenic influences on potential estimates of radon, the Radioactivity Environmental Monitoring group at the Joint Research Center of the European Commission is working on the development of a Europe-wide map of Geogenic Radon Potential (GRP). Geogenic radon refers to "what the earth delivers" in terms of radon. A GRP map is very important when making decisions for the construction of new buildings. Since each house has a unique characteristic of radon entry, an initial radon estimate used to determine radon preventive measures should be free of building related influences. The goal of this study was to evaluate the following variables as predictors for the creation of a GRP map based on logistic regression: terrestrial gamma dose rate (TGD), geology, fault lines and topsoil permeability.

* Corresponding author. Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland.
E-mail address: georg.kropat@chuv.ch (G. Kropat).

http://dx.doi.org/10.1016/j.jenvrad.2016.06.007

2. Material and methods

2.1. The radon data

The GRP can be described as the geogenic source of the radon hazard at a location or over an area (Szabó et al., 2014). Several studies proposed different definitions of a geogenic radon potential (Kemski et al., 2001; Neznal et al., 2004; Szabo et al., 2014). Ideally a GRP estimation is derived from radon soil gas concentration (SRC) measurements. However, in Switzerland nearly no SRC measurements are available. SRC and indoor radon concentrations (IRC) often show only a weak correlation for single buildings. Nevertheless, spatial aggregation results in a stronger association between both quantities (Chen and Ford, 2016; Kemski et al., 2009). The Europe-wide GRP map is planned to be based on a spatial support of a 10 km × 10 km grid. On this spatial resolution we expect a considerable correlation between aggregated SRC and IRC. Therefore, we used IRC mean values on a spatial grid of 10 km × 10 km in order to calibrate our GRP model.

The WHO proposes an indoor radon concentration of 100 Bq/m^3 as a reference level to minimize health hazards (Zeeb and Shannoun, 2009). For countries for which this reference level is not feasible, the WHO recommends not to exceed a reference level of 300 Bq/m^3, which corresponds to a dose of 10 mSv per year at an equilibrium factor of 0.4 and an annual occupancy rate of 7000 h (IAEA, 2014). Therefore, we assigned the GRP classes to the grid cells based on the following scheme: Low < 100 Bq/m^3, Medium 100–300 Bq/m^3, High > 300 Bq/m^3 After sub-setting of the IRC data from the Swiss radon survey to inhabited rooms at the ground floor and removal of missing data, 72,638 IRC measurements remained for analysis. For each building we calculated the average mean of all measurements taken in inhabited rooms on the ground floor of one house.

2.2. The predictor variables

As predictor variables for the GRP model we considered the terrestrial gamma dose rate, geology, fault line density and the permeability of the soil.

2.3. Terrestrial gamma dose rate

Under the assumption, that terrestrial gamma dose rate (TGD) is proportional to the uranium content in the ground, we included this variable into our model. The TGD data observations are based on a method to extract the terrestrial component of ambient dose equivalent rate from EURDEP routine monitoring data (Cinelli et al., 2014). In order to interpolate the data, we used support vector regression using the coordinates as predictors. Support vector regression is an approach to determine a function $f(\overrightarrow{x})$ for the estimation of a dependent variable y (in our case the TGD) in dependence of a set of predictor variables $\overrightarrow{x} = x_1, \ldots, x_k$. The optimization problem is to find the optimum between flatness and over-fitting of $f(\overrightarrow{x})$. This tradeoff is described by the cost parameter C. A small C represents under-fitting of the data and a high C over-fitting. The interested reader will find more detail about support vector regression in (Cherkassky and Mulier, 2007; Smola and Schölkopf, 2004). We performed training and validation of the support vectors in R using the package e1071 (Meyer et al., 2014). The cost parameters were determined via 5-fold-cross validation.

The aggregation of the TGD estimations on a 10 km × 10 km grid was performed by predicting the TGD on a 100 m × 100 m grid and then calculating the average per 10 km × 10 km grid cell. We used this method to obtain a more representative estimate for the average TGD in a 10 km × 10 km cell instead of just predicting the TGD at the center of a 10 km × 10 km cell.

2.4. Grouping of geological units

Numerous studies on radon hazard mentioned the association between IRC or SRC and geology (Appleton and Miles, 2010; Appleton et al., 2011; Drolet et al., 2014; Friedmann and Gröller, 2010; Kemski et al. 2006, 2009; Smethurst et al., 2008). To take advantage of this predictor in our model we grouped geological units into 6 classes based on a data driven approach that considers the similarity of IRC distributions between geological units. This method basically performs k-medoids clustering based on the pairwise Kolmogorov distances between the IRC distributions of the geological units. A detailed description of this method can be found in (Kropat et al., 2015). We used Swiss geological data that can be obtained from the OneGeology project (Federal Office of Topography swisstopo, 2013). To take account of the lithological properties as well as the age of the rock, we classified the geology by the fields “description” and “urn_litho1”. Only geological classes that covered more than 30 IRC measurements were included in the analysis to provide a stable estimate. This resulted in 58 combined classes of both geological age and lithology.

The conversion of the polygon data to the 10 km × 10 km grid was performed by sampling the regrouped geological units on a 1 km × 1 km grid all over Switzerland. Then the most prevalent regrouped geological unit of the 1 km × 1 km grid within a 10 km × 10 km grid cell was assigned to the corresponding 10 km × 10 km grid cell.

2.5. Fault lines

The Jura Mountains in Switzerland are known to be subject to high IRC. Former studies suggested that the abundant karstification of the Jura Mountains could explain the higher IRCs measured in this region (Savoy et al., 2011). To get an estimate of local karstification in the Jura Mountains we analyzed the tectonic accidents (fault lines) using a geological map of Switzerland on a scale of 1:500,000 (swisstopo, 2005).

To obtain an aggregated estimate of the density of fault lines within a 10 km × 10 km cell, we calculated the total length of lines within a cell and divided it by the surface area of the cell.

2.6. Soil permeability

Many studies include soil permeability as predictor for GRP, since this variable is supposed to influence flux of radon in the soil (Neznal et al., 2004; Szabo et al., 2014; Kemski et al., 2001). Following the Czech approach it is possible to derive a rough estimation of the permeability very easily from the weight percentage of fine fraction (<63 μm) (Barnet et al., 2008). Soils with the weight percentage of the fine fraction <15% were designed as high permeable soils, in the range 15–65% as medium permeable and in the case of the fine fraction above 65% as low permeable ones. The soil erodibility, k-factor parameter, was used as proxy of the weight percentage of fine fraction (Panagos et al., 2014). The great advantage of this parameter is the fact that it has already been modelled by soil experts for all Europe (with a grid size of 500 m) using LUCAS data (Tóth et al., 2013). We used an extrapolated dataset for Switzerland. The correlation between k-factor value and weight percentage of fine fraction have been calculated using data of LUCAS samples. Hence the value of k-factor corresponding to high, median and low permeability have been estimated and the permeability class assigned to each cell. Similar to the geological data we aggregated the permeability classes by assigning the most prevalent permeability within a 10 km × 10 km cell to the