EGU2020-8501
https://doi.org/10.5194/egusphere-egu2020-8501
EGU General Assembly 2020

Mapping the geogenic radon potential for Germany by machine learning

Eric Petermann[1], Hanna Meyer[2], Madlene Nussbaum[3], and Peter Bossew[1]
[1]Section Radon and NORM, Federal Office for Radiation Protection (BfS), Berlin, Germany
[2]Institute of Landscape Ecology, Westfälische Wilhelms-Universität Münster, Münster, Germany
[3]School of Agricultural, Forest and Food Sciences, Berne University of Applied Sciences (BFH), Berne, Switzerland

The radioactive gas radon (Rn) is considered as an indoor air pollutant due to its detrimental effects on human health. Radon is known as the second most important cause for lung cancer after tobacco smoking. The dominant source of indoor Rn is the ground beneath the building in most cases. Following the European Basic Safety Standards, all EU Member States are required to delineate Rn priority areas, i.e. areas with increased risk of high indoor radon concentrations. One possibility to this end is using the “geogenic Rn potential” (GRP), which quantifies the availability of geogenic Rn for infiltration into buildings. The GRP is defined as a function of Rn concentration in soil gas and soil gas permeability.

In this study we used > 4,000 point measurements across Germany in combination with ~50 environmental co-variables (predictors). We fitted machine learning regression models to the target variables Rn concentration in soil and soil gas permeability. Subsequently, the GRP is calculated from both quantities. We compared the performance of three algorithms: Multivariate Adaptive Regression Splines (MARS), Random Forest (RF) and Support Vector Machines (SVM). Potential candidate predictors are geological, hydrogeological and soil landscape units, soil physical properties, soil chemical properties, soil hydraulic properties, climatic data, tectonic fault data, and geomorphological parameters.

The identification of informative predictors, tuning the model hyperparameters and estimation of the model performance was conducted using a spatial 10-fold cross-validation, where the folds were split by spatial blocks of 40*40 km. This procedure counteracts spatial autocorrelation of predictor and response data and is expected to ensure independence of training and test data. MARS, RF and SVM were evaluated in terms of its prediction accuracy and prediction variance. The results revealed that RF provided the most accurate predictions so far. The effect of the selected predictors on the final map was assessed in a quantitative way using partial dependence plots and spatial dependence maps. The RF model included 8 and 14 informative predictors for radon and permeability, respectively. The most important predictors in the RF model were geological and hydrogeological units as well as field capacity for radon and soil landscape, geological and hydrogeological units for soil gas permeability.