Science of the Total Environment 754 (2021) 142291

Contents lists available at ScienceDirect

Science of the Total Environment

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

Mapping the geogenic radon potential for Germany by machine learning

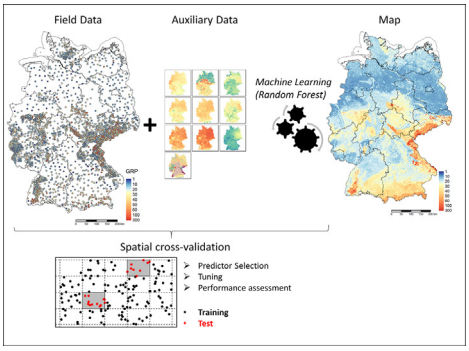
Eric Petermann [a,*], Hanna Meyer [b], Madlene Nussbaum [c], Peter Bossew [a]

[a] Federal Office for Radiation Protection (BfS), Section Radon and NORM, Berlin, Germany
[b] Westfälische Wilhelms-Universität Münster, Institute of Landscape Ecology, Münster, Germany
[c] Bern University of Applied Sciences (BFH), School of Agricultural, Forest and Food Sciences, (HAFL), Zollikofen, Switzerland

HIGHLIGHTS

- Mapping of the geogenic radon potential as hazard indicator for indoor radon
- Comparison of three machine learning algorithms
- Application of spatial cross-validation using spatial blocks to split the data
- Partial and spatial dependence plots reveal predictor-response relationship
- Random forest GRP map outperforms previous maps using geostatistics

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 25 May 2020
Received in revised form 12 August 2020
Accepted 7 September 2020
Available online 14 September 2020

Editor: Pavlos Kassomenos

Keywords:
Geogenic radon potential
Machine learning
Spatial cross-validation
Digital soil mapping
Soil radon
Partial dependence

ABSTRACT

The radioactive gas radon (Rn) is considered as an indoor air pollutant due to its detrimental effects on human health. In fact, exposure to Rn belongs to the most important causes for lung cancer after tobacco smoking. The dominant source of indoor Rn is the ground beneath the house. The geogenic Rn potential (GRP) - a function of soil gas Rn concentration and soil gas permeability - quantifies what "earth delivers in terms of Rn" and represents a hazard indicator for elevated indoor Rn concentration. In this study, we aim at developing an improved spatial continuous GRP map based on 4448 field measurements of GRP distributed across Germany. We fitted three different machine learning algorithms, multivariate adaptive regression splines, random forest and support vector machines utilizing 36 candidate predictors. Predictor selection, hyperparameter tuning and performance assessment were conducted using a spatial cross-validation where the data was iteratively left out by spatial blocks of 40 km*40 km. This procedure counteracts the effect of spatial auto-correlation in predictor and response data and minimizes dependence of training and test data. The spatial cross-validated performance statistics revealed that random forest provided the most accurate predictions. The predictors selected as informative reflect geology, climate (temperature, precipitation and soil moisture), soil hydraulic, soil physical (field capacity, coarse fraction) and soil chemical properties (potassium and nitrogen concentration). Model interpretation techniques such as predictor importance as well as partial and spatial dependence plots confirmed the hypothesized dominant effect of geology on GRP, but also revealed significant contributions of the other predictors. Partial and spatial dependence plots gave further valuable insight into the quantitative predictor-response relationship and its spatial distribution. A comparison with a previous version of the German GRP map using 1359 independent test data indicates a significantly better performance of the random forest based map.

* Corresponding author.
E-mail address: epetermann@bfs.de (E. Petermann).

https://doi.org/10.1016/j.scitotenv.2020.142291
0048-9697/

1. Introduction

The natural occurring radioactive gas radon (^{222}Rn, hereafter referred to as Rn) is considered as an indoor air pollutant due to its detrimental effects on human health (e.g. WHO (2009)). On a global average, exposure to Rn is the most important component of the natural radiation budget that the public receives. This results on average in an effective dose of ~1 mSv/a according to the dose conversion of ICRP (2007). For comparison, this is roughly ten times higher than the dose received by cosmic radiation during an intercontinental flight, e.g. from Frankfurt to Tokyo (BfS, 2020). In high-Rn areas the dose received by an individual can be easily an order of magnitude higher than the global average. Epidemiological studies have shown that a significant fraction of lung cancer cases can be attributed to residential Rn exposure. The estimates of this attributable fraction are 9% for Europe (Darby et al., 2005), 10–15% for the US (Krewski et al., 2006) and 13.6 to 16.5% on a global scale (Gaskin et al., 2018).

Generally, people receive the highest exposure to Rn indoors, i.e. in homes and at workplaces. The dominant source of indoor Rn is in most cases the ground beneath the house. Other sources of indoor Rn are known (e.g., building material and tap water) but contribute usually only a minor fraction to the total exposure (Bruno, 1983). Rn enters houses mainly via advective flux from soil gas through cracks and fissures in the building's foundation. After entering the house, Rn accumulates due to limited air exchange.

The detrimental health effects of Rn have been acknowledged by international and national legislation. In Europe, for instance, the EU launched the Basic Safety Standards Directive 2013/59/EURATOM in 2013 (European Council, 2013) aiming at the reduction of human exposure to Rn in homes and at workplaces. According to this legislation, all EU member states are required to delineate Rn priority areas, i.e. areas where many houses exceed the national reference level. The delineation of Rn priority areas aims at prioritization of Rn protection measures. One possibility to this end is to map the so called "geogenic Rn potential" (GRP) as an indicator of the Rn related hazard. The GRP - a function of Rn concentration in soil gas and soil gas permeability - aims at quantifying the availability of geogenic Rn for infiltration into buildings.

In this study, we will focus on GRP mapping in Germany. However, the described mapping approach can be applied to any region, country or continent given that a reasonable amount of observational data and meaningful and exhaustive co-variable data are available. The current version of the German GRP map was produced by applying sequential Gaussian simulation using geology as deterministic predictor (Bossew, 2015). A former version of a German GRP map was proposed by Kemski et al. (2001) using inverse distance weighting under consideration of geology. However, using solely geology as co-variable might not be sufficient to model spatial GRP variability due to the high complexity of the system "geogenic Rn". In recent years, the availability of auxiliary spatial data has increased considerably in the course of the development of remote sensing platforms. Also computational capacities increased strongly. Additional predictors besides geology have been rarely applied for GRP mapping (Pásztor et al., 2016). We hypothesize that the inclusion of more co-variables (hereafter referred to as predictors) into the predictive model, as well as the application of algorithms able to account for complex relationships, will increase the accuracy of the end-product. Therefore, we fitted machine learning regression models which utilize a multitude of potential predictors. We applied three different machine learning techniques: multivariate adaptive regression splines, random forest and support vector machines. These three algorithms are all capable of building complex predictor-response relationships, cope with cross-correlation in predictors, but rely on fundamentally different adaptation strategies which provides a range of different ways to extract information from observational and predictor data.

The objective of this study is to develop a machine learning based modelling framework for producing an accurate and high resolution (1 km*1 km) GRP map. The numeric response variable is the geogenic Rn potential which is modelled as a function of environmental co-variables (predictors). The individual steps of the model building process (predictor selection, hyperparameter tuning, and performance assessment) will be described with emphasis on the specific requirements of spatial data, i.e. applying a spatial cross-validation procedure which allows spatial predictor selection, spatial hyperparameter tuning and spatial performance assessment. In addition, this paper will deliver insights into the usefulness of individual predictors for GRP mapping by analyzing importance and partial dependence of the individual predictors which allows for guidance in other studies. Finally, a comparison with the current German GRP map will be made based on independent test data which were not used during any step of model building.

2. Geogenic Rn potential

The geogenic Rn potential (GRP) aims at quantifying "what earth delivers in terms of Rn" and represents a Rn hazard indicator with respect to Rn concentration in indoor air (see Bossew et al. (2020) for further discussion). Anthropogenic factors (e.g., building characteristics, living habits) that also affect the indoor Rn concentration are excluded on purpose. An advantage of using the GRP is that its concept is also applicable for areas that are not yet developed.

Several GRP definitions have been suggested during the last three decades (e.g., Tanner (1988), Wiegand (2001), Kemski et al. (2008), Ielsch et al. (2010) to name just a few). The most widely applied definition was proposed by Neznal et al. (2004) where GRP (dimensionless) is expressed as a function of Rn concentration in soil gas and soil gas permeability (see Eq. (1)). This GRP formula was developed semi-empirically to allow optimal prediction of indoor Rn concentration.

$$GRP\,[-] = \frac{C_{Rn}\left[\frac{Bq}{m^3}\right]}{-\,log_{10}k[m^2]-10} \tag{1}$$

The GRP varies spatially and temporally as a consequence of heterogeneous spatial distribution and temporal variability of factors controlling Rn built-up, migration and exhalation such as soil chemical and physical properties (spatial variability) and soil moisture (temporal variability).

2.1. Rn concentration in soil gas

Rn belongs to the ^{238}U decay chain and is generated by radioactive decay of its parent nuclide ^{226}Ra, which is present in virtually any mineral and organic material. Therefore, Rn is ubiquitous in the natural environment due to its constant production in rock and soil. Depending on the interaction of various factors such as grain size distribution or soil moisture (Markkanen and Arvela, 1992) Rn can be released (i.e., emanated) into the pore space, migrate through the pores and, finally, can be exhaled from the soil. The concentration of Rn (more correctly the "activity concentration") is measured in Becquerel (Bq) per m^3. 1 Bq means one disintegration of a Rn atom per second. The Rn concentration in soil usually has a considerable small-scale variability - vertically and horizontally (Neznal et al., 1996; Winkler et al., 2001) as well as temporally (e.g., Winkler et al. (2001), Font et al. (2008), Torkar et al. (2010), Szabó et al. (2013), Janik and Bossew (2016)). Variation in space can be an order of magnitude within only a few tens of meters if geological conditions (i.e., ^{226}Ra content) change (Neznal et al., 1996) but may even vary within the same geological unit by several factors (Kemski et al., 2008). Temporal variation was observed on day-to-day as well as on seasonal scales as a consequence of varying meteorological conditions. The long-term observations at an individual site can have a range from a factor of two up to a factor of ten (Neznal et al., 1996).

2.2. Soil gas permeability

The soil gas permeability can be seen as the counterpart of the soil hydraulic permeability. Both soil permeability characteristics are a function of soil texture which determines pore geometry and flow path topology. Another crucial factor governing soil permeability is soil moisture (water saturation of the pores). In a fully saturated soil, gas permeability is negligibly low. Consequently, soil gas permeability and soil hydraulic permeability are inversely proportional. However, the relationship between soil gas and soil hydraulic permeability reveals a non-linear pattern and, in addition to that, depends on whether the soil is wetting or drying (hysteresis effect) (see Scanlon et al. (2001)). Similarly to Rn concentration in soil, a considerable small-scale as well as temporal variability of soil gas permeability can be assumed. Spatial variability is expected to be driven by variation of soil physical properties or existence of preferential flow paths (e.g., tree roots, macrofauna burrow holes), whereas temporal variability is mainly governed by soil moisture fluctuations.

3. Material and methods

3.1. Field data

The German GRP database (status 30 November 2019) consists of 4448 locations with Rn in soil gas and soil gas permeability data (see Fig. 1, data is not publicly accessible). The soil gas measurements were mainly compiled in the period 1990–2003. The available data originates from different regional surveys which were initially not designed for producing a national GRP map. Therefore, the sampling design was non-representative in a way that measurements focused on potential high GRP areas and/or on areas with especially high small-scale variability of geological characteristics. Consequently, the measurements show a strong spatial clustering and do not represent the geological units by proportional share of their area (Fig. 1). Further, it has to be emphasized that these spatial point observations represent only a snapshot measurement within a most often highly dynamic time-series (see Sections 2.1 and 2.2).

The database consists of soil gas measurements of Rn concentration and permeability that were conducted at a depth of 1 m. Soil gas was extracted and analyzed on-site for Rn (after rejection of the first 10–15 L to avoid contamination) using alpha spectrometry or Lucas cells (Kemski et al., 1996). Gas permeability was measured in-situ by injecting a defined volume of air; the flow rate was measured at constant pressure, and permeability was calculated following Darcy's law under assumption of laminar transportation conditions (Kemski et al., 1996). At each location three boreholes (within a triangle of 5 m edge length) were drilled. In each borehole, Rn was measured twice, permeability once. The value attributed to the individual site was the maximum ($n = 6$) for Rn and the arithmetic mean ($n = 3$) for permeability (Kemski et al., 2001). Based on these observational data the GRP was calculated following Eq. (1). Log-transformed permeability ($\log_{10}(k)$) values > -10.5 ($n = 16$) were substituted by -10.5 to allow for numerical stability of GRP values (see Eq. (1)).

The GRP data is best described by a log-normal distribution and is characterized by an arithmetic mean of 36, a median of 21 and interquartile range from 11 to 40 (Fig. 1). However, due to the strong spatial

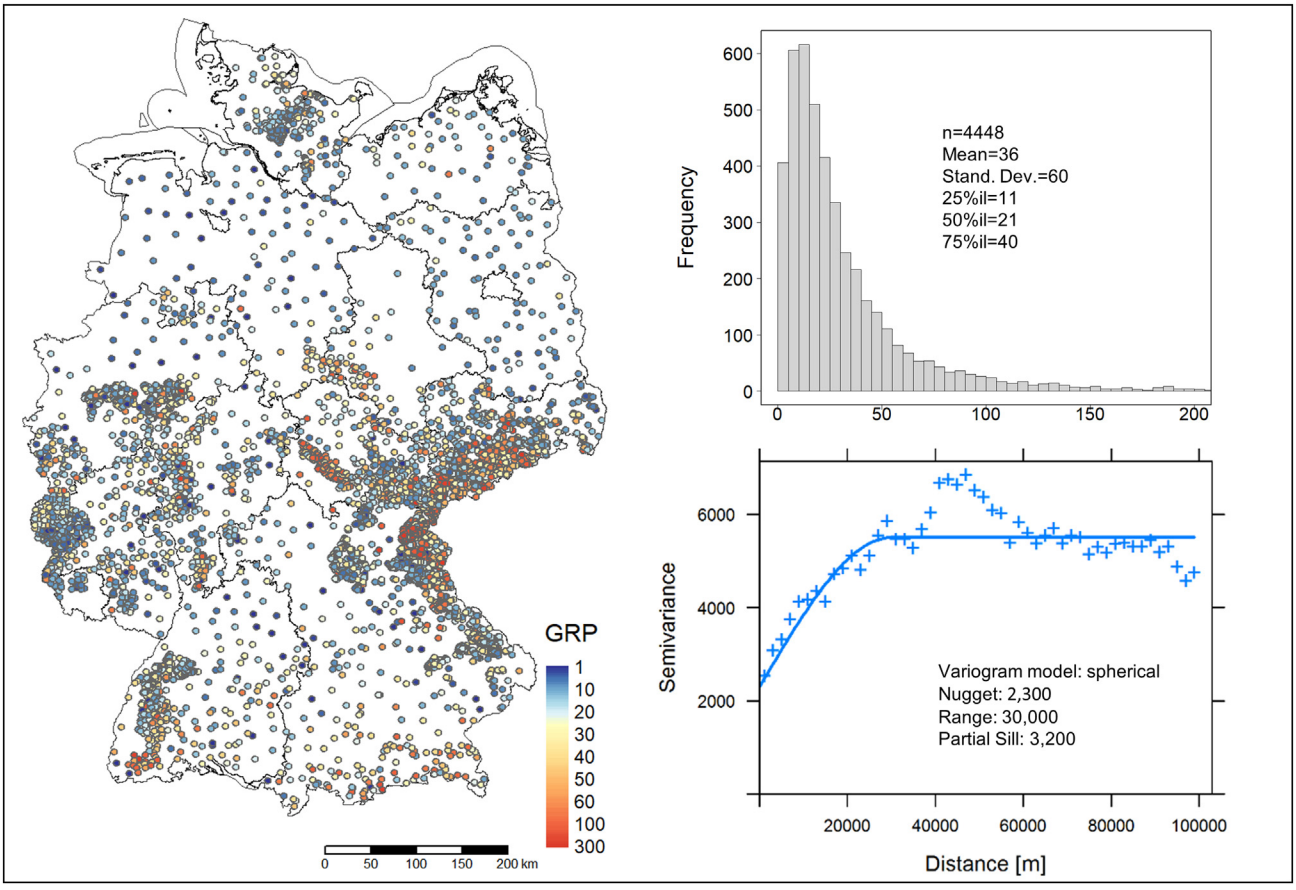


Fig. 1. Observations of the Geogenic Rn Potential (GRP) in Germany. (Left panel) Locations of the calculated GRP reveal a strong clustering in certain areas (e.g., parts of eastern and southeastern Germany). The black solid lines indicate the borders of the federal states. (Top right panel) Histogram and summary statistics of the geogenic Rn potential show a near log-normal distribution. (Bottom right panel) GRP semi-variogram model (visually fitted) and its parameters indicate a correlation length of ~30 km. The large nugget effect reveals a high small-scale variability.

clustering of the observations these distributional characteristics are likely to differ from the characteristics of the true GRP distribution in Germany. The data shows strong spatial auto-correlation with a correlation length of ~30 km (Fig. 1). The distinct nugget effect of 2300 (i.e., the semi-variance at distance 0) indicates a considerable small scale variability which makes up 40% of the total variance (5500).

New field campaigns have started in many federal states in 2018 to increase the available data before defining Rn priority areas. These data became only available at the end of the current study und were used as independent test data. Test data were not used at any step of model building but were applied for a final assessment of the ML based map in comparison to the current GRP map of Germany (Bossew, 2015).

3.2. Predictor data

36 co-variables were selected as possible candidate predictors for machine learning regression models. These predictors can be grouped as following:

- Geological class based on the geological map of Germany, scale 1:1,000,000 (BGR, 1993). Data was re-classified based on the classification used previously for the GRP map of Germany (Bossew, 2015) and further simplified into 30 classes. Classification was mainly done by geological criteria (stratigraphy, petrography and genesis). Further, classes with similar statistical properties were merged to reduce the number of classes for computational reasons and to allow a minimum number of observations in each class. For details see Table 2 (Appendix).
- Soil hydraulic properties in 1000 m resolution (Tóth et al., 2017):
  - o saturated hydraulic conductivity
  - o saturated water content
  - o field capacity
  - o wilting point
  - o parameter α of the hydraulic conductivity curve
- Soil physical properties in 500 m resolution (Ballabio et al., 2016):
  - o clay content
  - o silt content
  - o sand content
  - o coarse fraction
  - o available water capacity
  - o bulk density
- Soil chemical properties in 500 m resolution (Ballabio et al., 2019):
  - o pH in H_2O
  - o cation exchange capacity
  - o carbon:nitrogen ratio
  - o concentration of calcium carbonate
  - o concentration of nitrogen
  - o concentration of phosphorous
  - o concentration of potassium
- Soil uranium concentration in 10 km resolution (Cinelli et al., 2019)
- SAGA wetness index derived from the digital elevation model of Germany (resolution 25 m) (BKG, 2018)
- Climate data in 1000 m resolution (DWD, 2018a, 2018b, 2018c):
  - o Temperature: annual and seasonal means 1981–2010 (DWD, 2018a)
  - o Precipitation: annual and seasonal means 1981–2010 (DWD, 2018b)
  - o Soil moisture: annual and seasonal means 1991–2010 (DWD, 2018c)

3.2.1. Pre-processing

Original raster data was available at resolutions of 500 m, 1000 m or 10 km. Field data was attributed to the predictor data at the original resolution. The estimation grid (1 km cells) on which the mapping is based required harmonization of predictor data. For this purpose data at a resolution higher than 1 km was upscaled (arithmetic mean for numeric data, modal value for categorical data). The only predictor available at a lower resolution (uranium concentration) was downscaled using resampling. As further potential predictors, the SAGA wetness index was calculated (Böhner et al., 2002; Böhner and Selige, 2002). For soil hydraulic properties the weighted mean was calculated from data of seven depths ranging from 0 to 200 cm by using weights proportional to the layer thickness.

3.3. Machine learning

The number of studies using machine learning (ML) for spatial prediction in environmental sciences has increased considerably in recent years (e.g., Kanevski et al. (2009), Li et al. (2011), Lary et al. (2016), Hengl and MacMillan (2019)). ML applications for spatial settings include the assessment of landslide susceptibility (Micheletti et al., 2014), global air quality mapping (Lary et al., 2014) or soil mapping (Hengl et al., 2017), to mention just a few. Advantages of ML compared to geostatistical models are that ML can better handle complex multidimensional non-linear relationships and mostly makes no or just weak assumptions of the underlying distribution of the data (Fouedijo and Klump, 2019). However, the main reason for its increasing popularity is that ML based approaches have proven to outperform classical (geo)statistical models for many predictive tasks (e.g., Henderson et al. (2005), Nussbaum et al. (2018), Hengl and MacMillan (2019), Li et al. (2019)); especially when dealing with highly complex systems. The superior performance is seen as a consequence of their ability to reflect the influence and interplay of a multitude of factors. A drawback of ML is, however, that predictor-response relationships are sometimes more difficult to interpret in a physical way.

In this study multivariate adaptive regression splines, random forest and support vector machines have been used. These three ML algorithms were selected because they have provided promising results for spatial prediction tasks such as digital soil mapping (Ballabio et al., 2016; Hengl et al., 2018; Liess et al., 2016), landslide hazard mapping (Bui et al., 2016) or mapping of atmospheric particulate matter (Choubin et al., 2020).

In Rn research ML has been applied only sparsely. The few examples that we are aware of comprise the application of ensemble regression trees for predicting indoor Rn concentration in Switzerland (Kropat et al., 2015), the use of bagged neural networks for predicting the mean indoor Rn concentration in the Czech Republic (Timkova et al., 2017), as well as ML applications for time-series analysis (Torkar et al. (2010), Janik et al. (2018)).

For a deeper background on ML strategies in general as well as on individual algorithms the reader is referred to Hastie et al. (2009) and James et al. (2013) (both more theory-based) or for its practical implementation to Witten et al. (2016). A broad overview on ML using the R programming environment (R Core Team, 2019), with which all the modelling in this study was conducted, is given by Kuhn and Johnson (2013) and Boehmke and Greenwell (2019).

3.3.1. Multivariate adaptive regression splines

The algorithm multivariate adaptive regression splines (MARS) was developed by Friedman (1991). MARS creates a piecewise linear model where each predictor models an isolated part of the original data. For this purpose, each data point is evaluated for each predictor as a candidate knot point by creating linear regression models using the candidate predictor. The predictor/knot point combination with the lowest error is then added to the model. In a forward stepwise procedure the algorithm adds successively terms to the model that improve the model fit until a stopping criteria is met (Kuhn and Johnson, 2013). Then, terms that do not significantly contribute to the model performance are removed in a backward elimination process ("pruning"). The contribution of the individual terms in the model is evaluated based on the generalized cross-

validation (GCV) statistic. The MARS model has two hyperparameters: *degree* and *nprune*. The *degree* hyperparameter determines the maximum degree of interaction of the predictors: "1" means no interaction (additive model), "2" means interaction of up to two predictors (multiplicative model). The second tuning parameter *nprune* determines the maximum number of terms retained in the model after pruning. The optimal values for *degree* and *nprune* were estimated using a spatial cross-validation. In this study the MARS implementation in the *earth* package (Millborrow, 2019) was used.

3.3.2. Random forest

Random forest (RF) is an ensemble algorithm that was developed by Breiman (2001). The base learner of RF are classification or regression trees which partition the data into smaller, statistically more homogeneous subsets by finding predictors and values to split on in order to minimize an objective function (e.g., sum of squared errors). These individual trees are low bias, high variance techniques which fit the training data well, but are sensitive to small changes in the training set (overfitting). Therefore, in RF a large number of regression trees (e.g., n = 500) are built using bootstrap samples of the original data in order to reduce the correlation between the individual trees. This technique is known as "bagging" (bootstrap aggregation) and improves the predictive performance over a single tree by lowering the variance in the prediction. The decorrelation of the trees is further increased by injecting more randomness in the tree growing process by evaluating only a randomly selected subset of the available predictors at each split. RF is protected from overfitting in a sense that its predictive accuracy is not negatively affected by increasing the number of trees (Kuhn and Johnson, 2013). In this study the RF implementation in the *party* (Hothorn et al., 2006a) and *partykit* (Hothorn and Zeileis, 2015) packages were used. The implementation is slightly different from the classical RF by Breiman (2001): conditional inference trees (Hothorn et al., 2006b) are built using statistical test procedures for both predictor selection at the splits and definition of a stopping criteria (Strobl et al., 2007). Strobl et al. (2007) have proven that this procedure – in contrast to the original implementation - produces unbiased estimates if categorical predictors with many possible splits are involved. The RF implementation has one relevant tuning hyperparameter, *mtry*, which is the number of randomly chosen predictors to be evaluated at each split. The optimal value for *mtry* was estimated using a spatial cross-validation.

3.3.3. Support vector machines

Support vector machines (SVM) were originally developed in the context of classification by Vapnik in the 1960s, details can be found in Vapnik (2013). SVM regression is a type of robust regression aiming at the minimization of the effect of outliers on the regression equation. Therefore, SVM use an objective function that gives less weight to outliers (Kuhn and Johnson, 2013). SVM makes use of a so-called "kernel-trick" for non-linear problems with which input data is implicitly mapped into a higher dimensional feature space. After model fitting the data is back-transformed to the original feature space. Here, an epsilon-support vector regression with a radial basis function (as kernel) was applied using the *kernlab* package (Karatzoglou et al., 2004). The parameter *epsilon* defines a funnel around the regression line. All data points within this funnel do not contribute to the regression equation, data points outside but close to the funnel become so-called support vectors, i.e. data points which define the regression equation. The cost parameter *C* governs the flexibility of the model (the larger *C*, the more flexible) and is treated as only tuning hyperparameter in this study. The scaling parameter *sigma* of the radial basis function was estimated automatically within *kernlab* for the given data set with 0.006 and for epsilon the default value of 0.1 was used.

3.3.4. Model building

Spatial data usually does not fulfill the assumption of identically and independently distributed data due to its inherent spatial autocorrelation. This characteristic requires special consideration in all steps of model building when applying ML for spatial predictions as shown, e.g. by Roberts et al. (2017), Meyer et al. (2019) and Schratz et al. (2019).

For building a spatial ML model, the following steps were considered for each algorithm:

- Predictor selection
- Hyperparameter tuning
- Performance assessment
- Building a final model
- Mapping (i.e., applying the model for spatial prediction)

3.3.4.1. Performance metrics. The assessment of model performance can be done using several criteria. For a comprehensive assessment of the model performance, several statistic metrics focusing on different aspects of the goodness of fit were used in this study: the coefficient of determination (r^2, Eq. (2)), the root mean squared error (RMSE, Eq. (3)), the mean absolute error (MAE, Eq. (4)) the root relative squared error (RRSE, Eq. (5)) and the root mean squared log error (RMSLE, Eq. (6)). Here y_i and $\hat{y}_i$ are the observed and predicted values at location i, $\overline{y}$ and $\overline{Y}$ are the arithmetic mean of y_i and $\hat{y}_i$. The objective function for all steps of model building was the maximization of r^2.

RMSE, MAE, RRSE and RMSLE are given in the unit of the response variable. Since GRP is dimensionless, the used error metrics are dimensionless as well.

$$r^2 = \left(\frac{\sum_{i=1}^{n} (y_i - \overline{y})\,(\hat{y}_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}\,\sqrt{\sum_{i=1}^{n} (\hat{y}_i - \overline{Y})^2}} \right)^2 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - y_{;-})^2}} \tag{5}$$

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (log(y_i + 1) - log(\hat{y}_i + 1))^2} \tag{6}$$

3.3.4.2. Cross-validation strategy. A common way for validating ML models (especially with a high number of observations n, say n > 1000) is a k-fold cross-validation. For this procedure the data is split into k folds where k typically takes values of 5 or 10. The assignment of the data to the individual folds is usually conducted in a random or rarely in a stratified way. The model is trained using data form k-1 folds and tested using the remaining hold-out-fold, i.e. the trained model is applied for predicting values at the test locations using the local predictor setting. These predictions are then compared to the observational data in the test dataset. In total, the model is build k times, each time having a different composition of training and test data by iterating the fold used for testing from 1 to k (see Kuhn and Johnson (2013) for details). For the final cross-validation performance estimate, the performance of the individual folds was averaged for predictor selection and tuning. Performance assessment of the tuned model with the optimal predictor subset was done using the predictions for the individual test samples.

The existence of auto-correlation in response and predictor data (in our case spatial auto-correlation) requires special attention when utilizing ML. This is a consequence of a crucial assumption on which the reliability of cross-validation depends on; namely the independence of training and test data. The assumption of independence is likely to be violated if the individual locations are assigned to the folds in a random or a stratified way since test locations might simply be too close to training locations, i.e. within their respective correlation length. Therefore, a spatial cross-validation strategy is suggested for data with spatial dependencies (Pohjankukka et al. (2017), Roberts et al. (2017), Meyer et al. (2019)). In this study, spatial cross-validation was conducted by creating spatial blocks using the package *blockCV* (Valavi et al., 2019). Definition of the block size requires a trade-off between having large enough blocks to maximally reduce auto-correlation between training and test data and not having too large blocks that could result in an insufficient coverage of the predictor parameter space. Roberts et al. (2017) suggested a block size that is significantly larger than the autocorrelation observed in the model residuals. In this study, a block size of 40 km seemed reasonable since autocorrelation in the model residuals was 15–20 km during test model runs. The spatial blocking resulted in a total of 251 40 km*40 km blocks (Fig. 2). After block creation, the blocks were randomly split into 10 folds. Hence, a 10 fold cross-validation was performed, where each fold consists of 9/10 of the spatial blocks, and the performance was estimated using the respective remaining 1/10 of spatial blocks. For repeated cross-validation the splitting of the blocks into folds was repeated in a random way.

3.3.4.3. Predictor selection. The motivation of predictor selection is to remove non-informative predictors from the model which is expected to improve model performance. Non-informative predictors can be either just irrelevant (in the best case) or even cause overfitting. Although the ML algorithms applied in this study have an implemented predictor selection and should, at least in theory, be robust against non-informative predictors (Kuhn and Johnson, 2013), others have shown that even for these algorithms predictor selection improved the performance (e.g., Li et al. (2011), Liess et al. (2016)). Another advantage of predictor selection is that a model with less predictors fulfills the parsimony principle which also decreases computing time significantly.

In this study, a forward predictor selection approach using the *CAST* package (Meyer, 2018) was used which has proven to provide a suitable predictor selection strategy for spatial modelling tasks when combined with spatial cross-validation (Meyer et al., 2018). In short, models are built for every predictor pair combination. The predictor pair with the best performance is selected. Then, the remaining predictors are individually tested by adding them to the previously selected best predictor pair. The three predictor subset with the best performance is selected. This procedure continues as long as none of the remaining predictors results in an improvement of performance. The predictor selection was conducted using a 10-fold spatial cross-validation (see Section 3.3.4.2), hence predictors are selected with respect to improving the spatial prediction performance of the model. The hyperparameter setting was held at constant values (MARS: $degree = 1$, $nprune = 100$; RF: $mtry = 5$, $ntrees = 200$; SVM: $C = 256$, $epsilon = 0.1$, $sigma = 0.006$) during predictor selection to save computational costs, but were tuned extensively during final model fitting.

3.3.4.4. Hyperparameter tuning. Hyperparameter tuning was conducted using an iterative grid search for finding the optimal hyperparameter set for the previously chosen optimal predictor subset. First, a rather coarse initial grid with a wide value range was tested, followed by a second search using a refined grid centered on the optimal values from step one. Tuning was conducted within the framework provided by the *caret* package (Kuhn et al., 2019) using a repeated 10-fold spatial cross-validation (see Section 3.3.4.2).

The final model was finally trained using the ideal predictors revealed during spatial variable selection and the ideal hyperparameters revealed during spatial parameter tuning and all of the available GRP data. The trained model was then applied to make predictions for entire Germany.

4. Results

4.1. Selected predictors, tuning and predictor importance

The predictor selection procedure identified a similar number of informative predictors for each ML algorithm: 8 predictors for MARS, 10

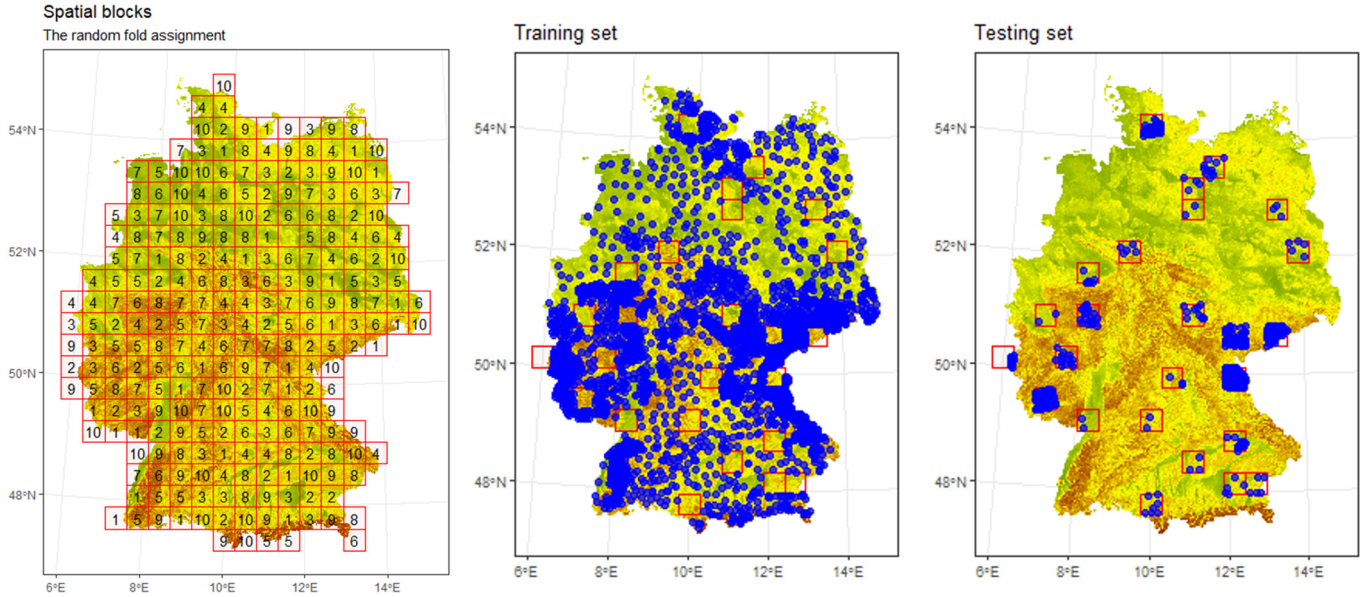


Fig. 2. Spatial blocking for fold creation with the package blockCV (Valavi et al., 2019). (Left) The study area is separated into 251 squares of equal size (40 km*40 km). To each fold an integer between 1 and 10 is randomly assigned. All fold numbers have a roughly equal frequency. (Middle) Example of a training data set (blue dots) containing data from fold numbers 1 as well as 3 to 10. The empty red squares indicate areas of testing folds. (Right) Example of a testing data set (blue dots in red squares) containing data from fold number 2. (Background layer of the maps is the elevation with greenish colors referring to lowland and brownish to mountainous terrain). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

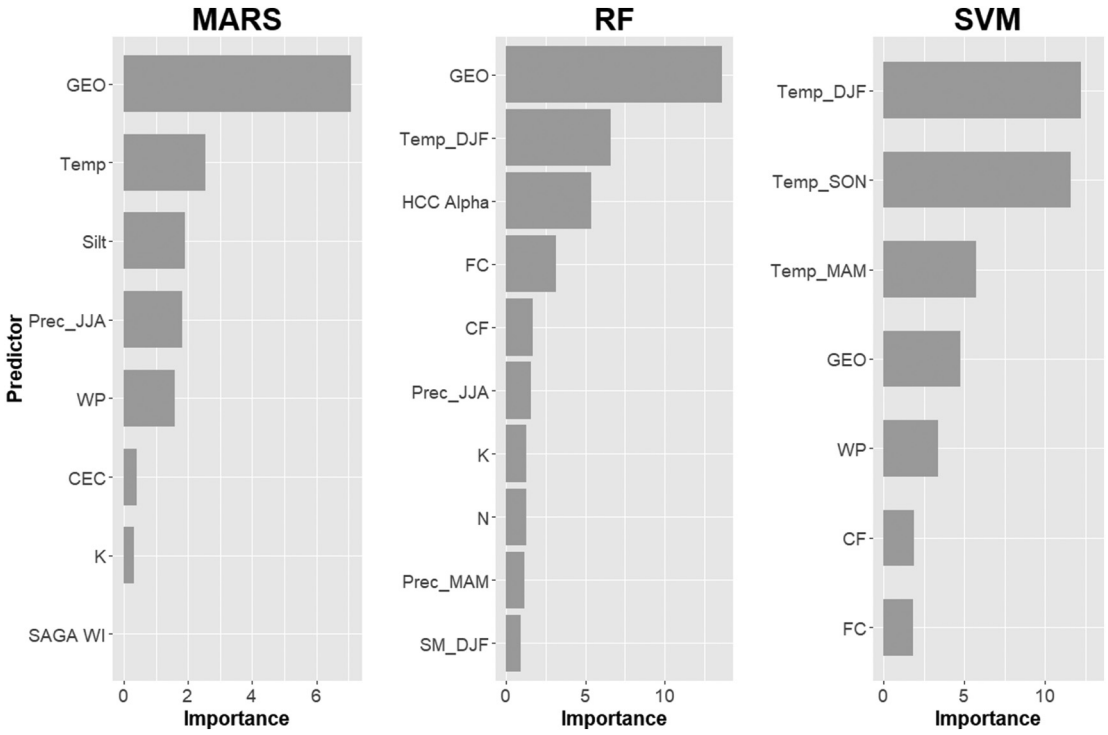


Fig. 3. Variable importance plots for multivariate adaptive regression splines (MARS), random forest (RF) and support vector machines (SVM). The predictors are ordered by decreasing importance. Overall, geological class ("GEO") was the most influential predictor (top predictor for MARS and RF, number 4 for SVM). Predictor abbreviations are explained in Section 4.1.

for RF and 7 for SVM (Fig. 3). In total 17 predictors out of the 36 candidates were selected at least for one algorithm. Only one predictor, namely geological class ("GEO"), was selected by every algorithm. At least two predictors describing the soil physical properties were selected for each algorithm: silt content ("Silt") for MARS, wilting point ("WP") for MARS and SVM as well as field capacity ("FC") and coarse fraction ("CF") for RF and SVM. The parameter alpha of the hydraulic conductivity curve ("HCC Alpha") was the only soil hydraulic property that was selected (only for RF). The soil chemical predictor Potassium ("K") was selected for MARS and RF, cation exchange capacity ("CEC") for MARS only and Nitrogen ("N") for RF only. Climate predictors were selected for all algorithms: the annual mean temperature ("Temp") for MARS, summer precipitation ("Prec_JJA") for MARS and RF, winter temperature ("Temp_DJF") for RF and SVM, spring precipitation ("Prec_MAM") for RF, winter soil moisture ("SM_DJF") for RF as well as autumn temperature ("Temp_SON") and spring temperature ("Temp_MAM") for SVM. The SAGA wetness index ("SAGA WI") was only selected for MARS.

The optimal hyperparameter setting for the individual algorithms with the identified optimal predictor subset was *degree* = 1 and *nprune* = 21 for MARS, *mtry* = 5 for RF and $C = 1500$, for SVM. For RF *ntree* was set at 1000 since no increase in performance was found for higher values.

The importance of the individual predictors in each model was analyzed using the *vip* package (Greenwell et al., 2020). In this case, importance is defined as the relative influence of an individual predictor on the model performance. The importance was measured by calculating the increase of the model's prediction error after permuting (i.e., randomly shuffling values) of the predictor of interest. The results are shown in Fig. 3. For both, MARS and RF, geological class was by far the most influential predictor. However, for SVM geological class ranks as number four only. Further, climate related predictors (especially temperature) turned out to be very important predictors (number one for SVM). Indeed, for MARS and RF a temperature related predictor was the most important after geological class. The importance of individual soil physical ("Silt", "WP", "FC","CF") and soil chemical predictors ("CEC", "K", "N") is smaller, their cumulative importance is, however, still substantial. The SAGA wetness index was selected as informative for MARS, though, the influence on model performance was negligible.

4.2. Model performance

Model performance was assessed using the criteria described in Section 3.3.4.1 using the individual predictions from the hold-out-folds of a 100-times repeated spatial 10-fold cross-validation.

An analysis of the performance metrics (Table 1) reveals that RF and SVM have a similar performance with respect to r^2 of 0.22 (which was the objective function during model training). However, regarding RMSE and RRSE RF clearly outperformed the other algorithms, while SVM had the lowest MAE. The RMSLE could not be calculated for MARS and SVM due to negative values in the predictions. Considering the performance metrics, we can state that RF produces the best predictions followed by SVM and MARS. Furthermore, RF predictions seem to be more stable in a sense that in contrast to MARS and SVM no predictions occur that are very far off the observed value which is seen as a consequence of RF being an ensemble algorithm (i.e., an aggregate of 1000 individual regression trees). A characteristic that is common to all three algorithms is that the predictions are smoothed relative to the observations (i.e. the distribution of the predictions is narrower compared to the observed values). This manifests with overestimation of very low observations (<5) and a tendency of underestimation of high observations (>200). This behavior is indicated by the smoothed conditional mean (red line in Fig. 4) having a much shallower slope than the 1:1 line.

Table 1
Spatial performance metrics of multivariate adaptive regression splines (MARS), random forest (RF) and support vector machines (SVM).

	RMSE [−]	r^2	RMSLE [−]	MAE [−]	RRSE [−]
MARS	55.19	0.156	NA	25.64	0.92
RF	52.76	0.218	0.92	23.69	0.88
SVM	54.39	0.216	NA	22.17	0.91

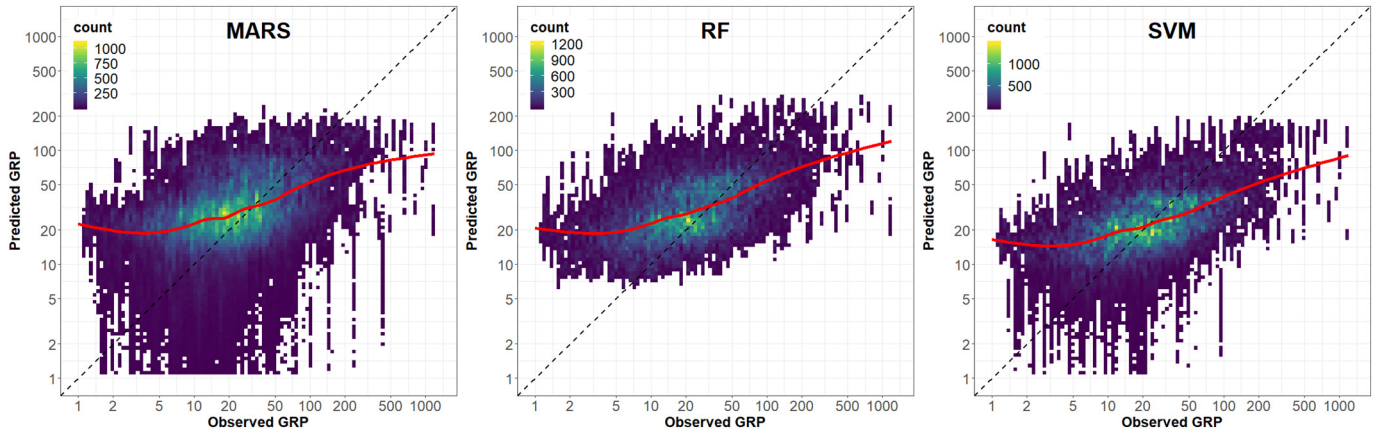


Fig. 4. Performance on test data (hold-out-folds of a 100 times repeated spatial cross-validation) for multivariate adaptive regression splines (MARS), random forest (RF) and support vector machines (SVM). The dashed line indicates the 1:1 line. Please note that both axes are on a logarithmic scale. Models were fitted on the non-transformed response. Due to overlapping of values, the plots show point density. The red line shows the smoothed conditional mean estimated by a generalized additive model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Residual analysis

The spatial structure of the residuals for the ML models were further analyzed. Variograms were processed using the *gstat* package (Gräler et al., 2016; Pebesma, 2004); variogram fitting was done manually using a spherical variogram model. Fig. 5 indicates that the nugget of the residuals of all three models (2100 to 2600) is close to the nugget in the observed data ("Measurements" in Fig. 5). The range of the spatial auto-correlation (i.e., the correlation length) was reduced slightly from ~30 km for the observational data in the observed data down to 20–25 km in the residuals of the ML models. However, the most striking feature when comparing the residual variogram models is that the residuals in RF model expose a significantly lower partial sill (i.e., the difference between the nugget and the semivariance beyond the correlation length) compared to MARS and SVM. Overall, the RF model performed best in reducing the spatial structure in the residuals.

4.4. GRP map based on random forest

Sections 4.2 and 4.3 have shown that RF has provided the best GRP predictions determined by its overall best performance statistics and lowest spatial auto-correlation in the residuals. Therefore, the focus of the upcoming sections will be on the discussion of the RF model.

The estimated GRP in Germany (Fig. 6) using RF show a value range from 6 to 314 with an arithmetic mean of 22 and an interquartile range (25%ile to 75%ile) from 13 to 25. The map reveals significant spatial variability. Regions with high GRP values (>50) comprise almost exclusively regions in the Central German Uplands (Ore mountains, Thuringian Forest, Harz mountains, Fichtel mountains, Upper palatine Forest, Bavarian Forest, Black Forest, parts of Upper Lusatia, Odenwald and Sauerland) as well as parts of the German Alps. Values within an intermediate GRP range (25 to 50) can be predominantly found in parts of central and southern Germany. Low GRP values below 25 are located in large parts of northern Germany and some parts in the south-west.

$$\text{rel unc} = \frac{\hat{y}_{75\%ile} - \hat{y}_{25\%ile}}{\hat{y}_{\mu}} \tag{7}$$

The relative uncertainty *rel unc* (Eq. (7)) expresses the prediction variability of the interquartile range of the predictions made by the individual 1000 trees normalized to the GRP estimate ($\hat{y}_{\mu}$), hence indicates the stability of the predictions. The quantiles were calculated using the

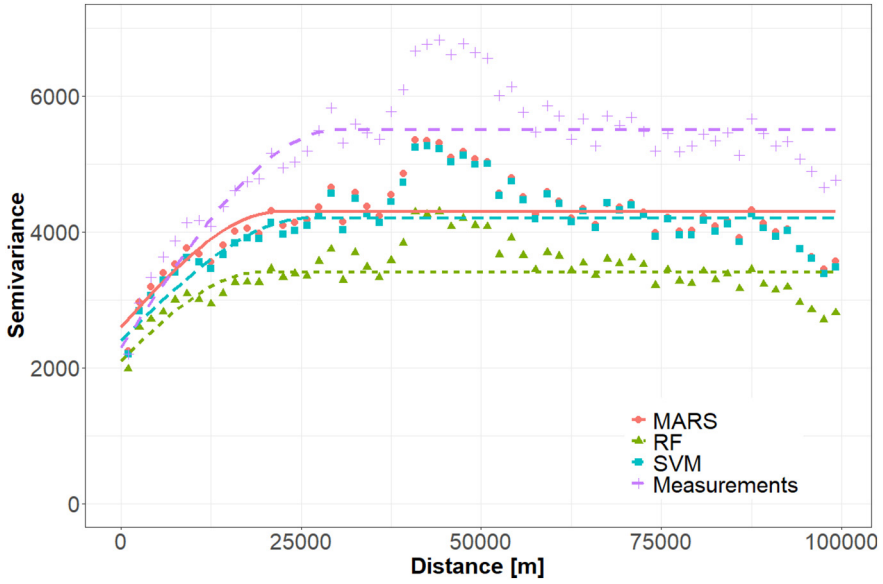


Fig. 5. Variograms of residuals from MARS, RF and SVM models. Residuals were calculated using predictions on the hold-out-folds. For comparison, the variogram of the observed data ("Measurements") is shown as well.

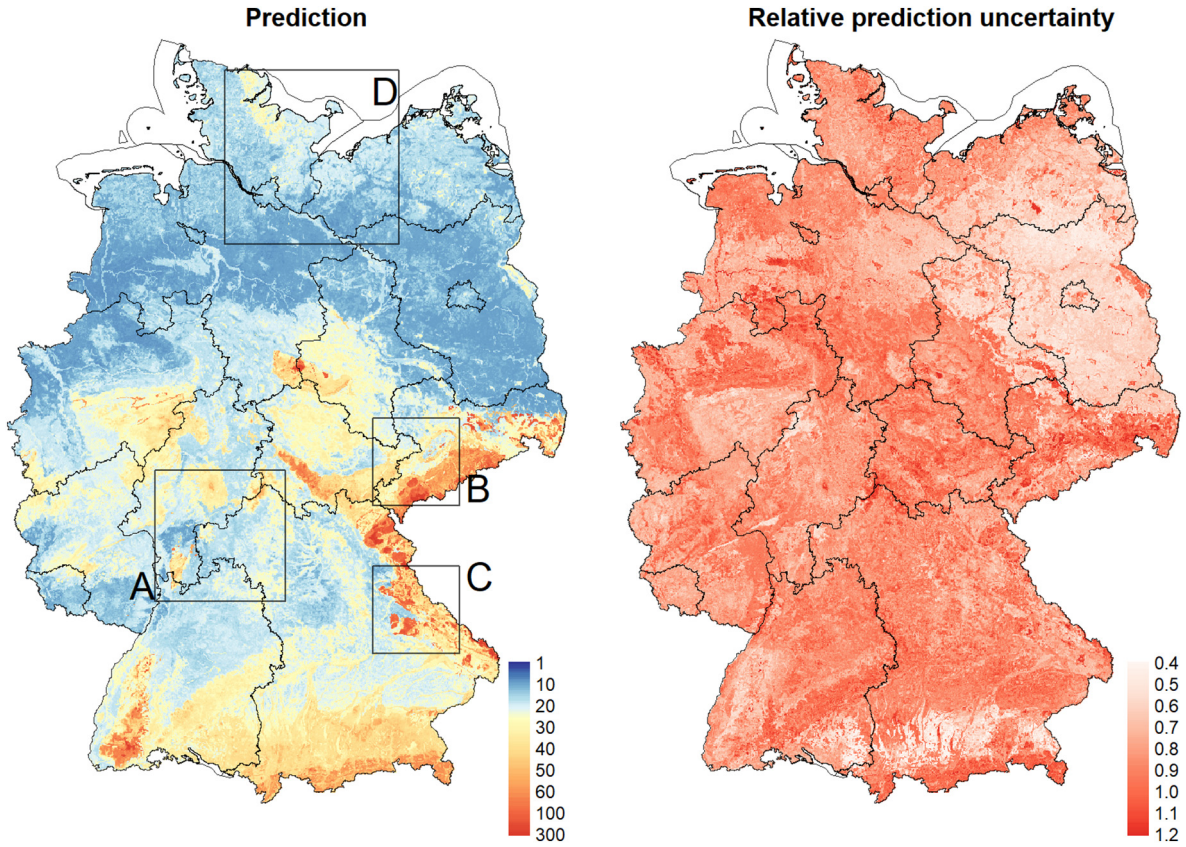


Fig. 6. Predictions of the Geogenic Radon Potential (GRP) based on random forest (left panel) and its respective relative prediction uncertainty estimated from the variability of predictions made by the individual trees of the forest (right panel). The black lines indicate the borders of the federal states. The black squares A – D in the left figure refer to exemplary regions which will be compared with the current German GRP map in 4.6.

partykit package (Hothorn and Zeileis, 2015) which adapts the idea of quantile regression forest (Meinshausen, 2006) to the *cforest* framework. The variability of predictions ranged from 0.4 to 1.2. Relative uncertainty exhibits generally a patchy pattern with only minor regional trends. Higher relative prediction uncertainty can be found in parts of central and eastern Germany. Northeastern Germany and the Alpine foreland exhibit the lowest relative uncertainty.

4.5. Model interpretation

The next step in model interpretation is to gain a deeper quantitative understanding of the effect of an individual predictor on the output. This can be realized by partial dependence plots (PDPs). This technique is well-suited to assess the predictors-response-relationship in ML, especially for models often characterized as black-boxes such as RF. The partial dependence of the predictor of interest is calculated by accounting for the average effect of the other predictors in the model. Hence, the effect of the other predictors is averaged out. We utilized the package *pdp* (Greenwell, 2017) for constructing PDPs for the RF model (Fig. 7). Further, the partial dependence of the predictors were extended to spatial dependence maps (Fig. 8) by simply applying the predictor-response function to the predictor rasters (Behrens et al., 2014). Maps of the ten predictor rasters selected for the RF model can be found in Fig. 12 (Appendix).

The most important predictor in the RF model is geological class (Fig. 3). Geological classes that increase GRP predictions the most are, in decreasing order, ID 912 (granites from the Carboniferous to Permian), ID 911 (granites from the Devonian to Carboniferous), ID 930 (intermediate to felsic plutonites from the Proterozoic to Paleozoic), ID 770 (high metamorphic rocks from the Proterozoic to Paleozoic) and ID 720 (low metamorphic rocks from the Upper Proterozoic to Middle Paleozoic) to name just the top five. These geological classes can be mainly found in the Central German Uplands. The geological class with the lowest GRP predictions is ID 300 (sedimentary rocks mainly from the Mesozoic) which is found in large parts of central and southern Germany. The second most important predictor, winter temperature, reveals a clear trend with higher GRP predictions for lower temperatures. Therefore, areas where winter temperature increases GRP predictions can be found in the mountainous areas in eastern and southern Germany. The third most important predictor, parameter HCC Alpha, is associated with much higher GRP predictions for values above 500 which are predominantly found in northern Germany. For field capacity, lower GRP predictions are related to values of around 30. The spatial pattern of the effect of field capacity on the GRP prediction is similar to HCC Alpha with higher values mainly in northern Germany. For summer precipitation, a trend with a small magnitude can be detected: increasing summer precipitation is connected to higher GRP predictions which applies mainly for mountainous areas in southern Germany. The reverse can be observed for spring precipitation: decreasing GRP predictions for spring precipitation between 100 and 250 mm. Lower spring temperature and associated GRP predictions can be found in eastern and north-eastern Germany. A similar geographical area is also affected by low winter soil moisture which is related to higher GRP predictions. Nitrogen reveals only minor fluctuations with respect to its effect on GRP predictions without a notable spatial pattern. For coarse fraction higher GRP predictions can be observed for increasing predictor values which were observed in all mountainous areas. Finally, potassium concentrations of about 200 mg/kg are associated with minimum GRP predictions, which are located in the northern German Lowland, whereas an effect of potassium on higher GRP can be found in the mountainous areas only.

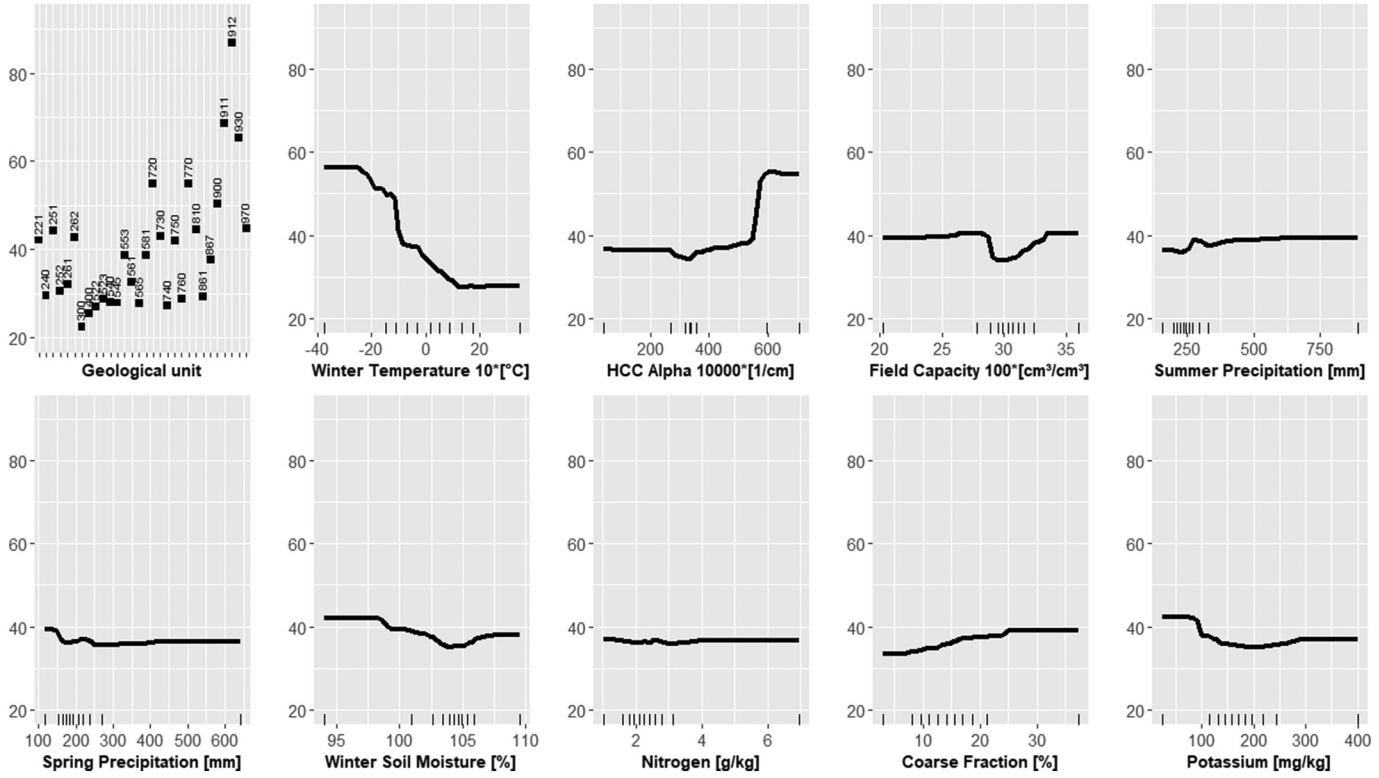


Fig. 7. Partial dependence plots for the ten predictors in the random forest model. The vertical axis shows the predicted values of the Geogenic Radon Potential (dimensionless). The small vertical bars on the x axis indicate minimum, maximum and deciles of the predictor value range for numeric predictors. The numbers in the plot for geological class refer to its ID number (see Table 2 (Appendix)).

The physical explanation of the partial dependence of the individual predictors is, however, not straightforward for many predictors and remains speculative which is certainly a drawback of the ML approach. For instance, the large effect of winter temperature is somewhat surprising since its direct effect on GRP is likely to be marginal. However, winter temperature could act as a proxy for soil moisture (probably higher quality of the winter temperature data compared to the soil moisture data itself) or reflect the tendency of higher uranium content in soil in mountainous areas where winter temperature is lower. If the latter would be the case, winter temperature can be seen as complementary to geological data, which might not be sufficient to explain all the variability of uranium in soil as a consequence of its limited resolution (1:1,000,000 map). A detailed interpretation of the effect of all individual predictors is beyond the scope of this study.

4.6. Comparison with current German GRP map

Two previous attempts of a German GRP map were mentioned in this paper. However, since Kemski et al. (2001) use a GRP definition based on six classes it is not directly comparable to our study. The study of Bossew (2015), which is the current German GRP map, applies the same GRP definition as our study which makes it suitable for a comparison. For this purpose, 1359 new samples which were not used during any step of model building were used to evaluate predictive performance of the GRP maps based on RF (this study) and on sequential Gaussian simulation Bossew (2015). Test data were provided by six out of 16 federal states which, consequently, results in an incomplete areal coverage of the mapping area. In addition to that, test data are spatially clustered within these areas (Fig. 9).

Further, test data were mainly sampled in areas with predominantly low to medium GRP values – with only a few exceptions. This is illustrated by an arithmetic mean of 23 and an interquartile range (IQR) of 10 to 29 of test data in comparison data used for model building with a mean of 36 and an IQR of 11 to 40 (Fig. 1). Therefore, the GRP value range is also incompletely covered. Despite the described limitations, these data provide the opportunity for a preliminary evaluation of the GRP map performance being valid at least for the areas and value ranges for which test data are available.

Because of the limited value range of the test data compared to the predicted values we grouped the predictions into seven GRP classes (<10, ≥10 to <15, ≥15 to <20, ≥20 to <25, ≥25 to <30, ≥30 to <35, ≥35). Fig. 10 presents the arithmetic mean and the interquartile range of the test data (GRP observations) within the respective GRP class of predictions. The means of the test data are well in accordance with the mapped values for each GRP class indicated by their proximity to the 1:1 line. Regarding the SGS map, however, test data and mapped values deviate notably; in particular for classes <10, ≥30 to <35 and ≥35. This visual observation of superior performance of the RF map in terms of its lower average error is manifested by the RMSE (based on the arithmetic mean of the mapped values and the arithmetic mean of the test data within the respective GRP class on the map): 3.6 for the RF map and 9.3 for the SGS map. The width of the interquartile range indicates, however, that individual observations may deviate significantly from the map.

The RF map reveals the desirable strictly monotonic behavior of higher GRP observations for increasing predicted values as proven by the test data. This is not always the case for the SGS map, especially for the value range between 15 and 35, where discrimination of RF map is much better.

For further comparison of both maps four exemplary regions in Germany were analyzed in detail (Fig. 11). The location of the four regions within Germany is shown in Fig. 6. Region A comprises areas with high GRP in the southwest, the Odenwald (parts of it consist of felsic plutonites) as well as in the north and northeast with the Vogelsberg and the Rhoen mountains (largely made of mafic volcanic rocks). Region B shows the Ore mountains (partly felsic plutonites and

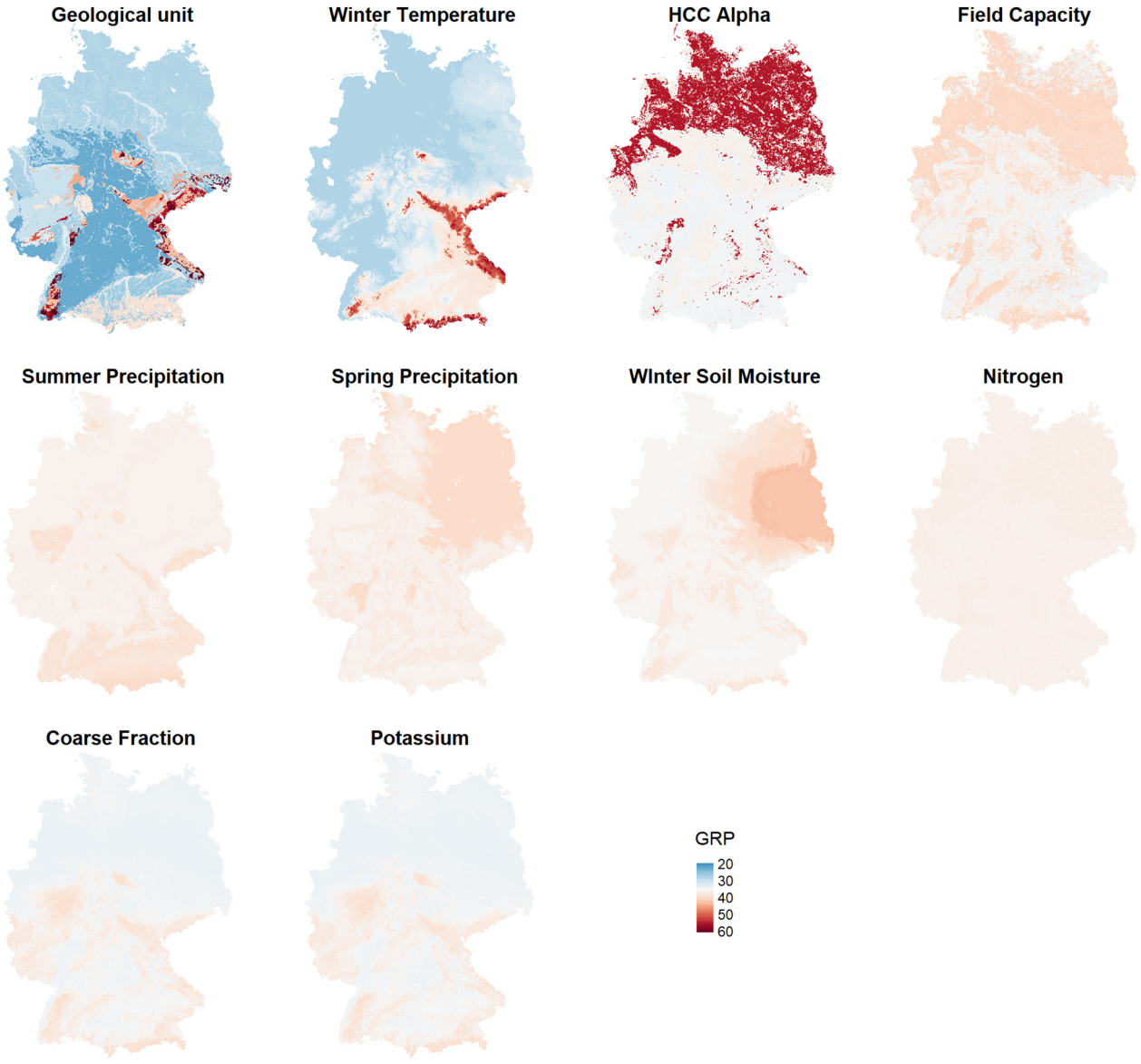


Fig. 8. Spatial dependence plots for the ten predictors in the random forest model. Each map shows the spatial distribution of the partial dependence on GRP predictions (dimensionless) (Fig. 7) for each predictor.

metamorphic rocks with plutonic educts) in the south with dominantly high to very high GRP values. Region C covers large parts of the Bavarian forest (partly felsic plutonites and metamorphic rocks with plutonic educts) along the German border with also partly very high GRP values. Region D shows parts of the German Baltic sea coast with generally low values but some areas covered with young glacial sediments associated with higher GRP values. The comparison between both maps reveals a higher degree of detail for RF which generally better reflects the natural structures of geology and landscape. Nonetheless, both approaches are consistent in a way that they found similar hotspot regions, however, with the RF based map these areas can be delineated much more precisely.

5. Discussion

5.1. Performance of machine learning models

By evaluating all the performance metrics for the different ML algorithms RF has achieved the best results (highest r^2, lowest RMSE and RRSE), the lowest variation in the predictions and, in addition, the lowest magnitude of spatial structure in the model residuals. The superior performance of RF relative to MARS and SVM is seen as a result of its ensemble character, which has been proven to be robust against outliers and noisy data (Kuhn and Johnson, 2013).

Our hypothesis that utilizing additional predictors besides geology will improve the predictive accuracy is supported by the superior test performance of the RF based map compared to the SGS based map. Further, the large sets of predictors (7–10) for all ML algorithms indicates that these additional predictors helped to improve the mapping performance. Due to the limitations of the test data such as incomplete spatial coverage of the mapping area, spatial clustering and incomplete coverage of the GRP value range, this performance analysis remains somewhat preliminary and valid only for the area and the value range where test data are available. However, the strikingly better performance of the RF map gives strong evidence of its overall better performance. The superior performance of the RF GRP map is seen as the result of applying a more powerful algorithm (that utilizes more auxiliary data) as well as mapping at a higher spatial resolution.

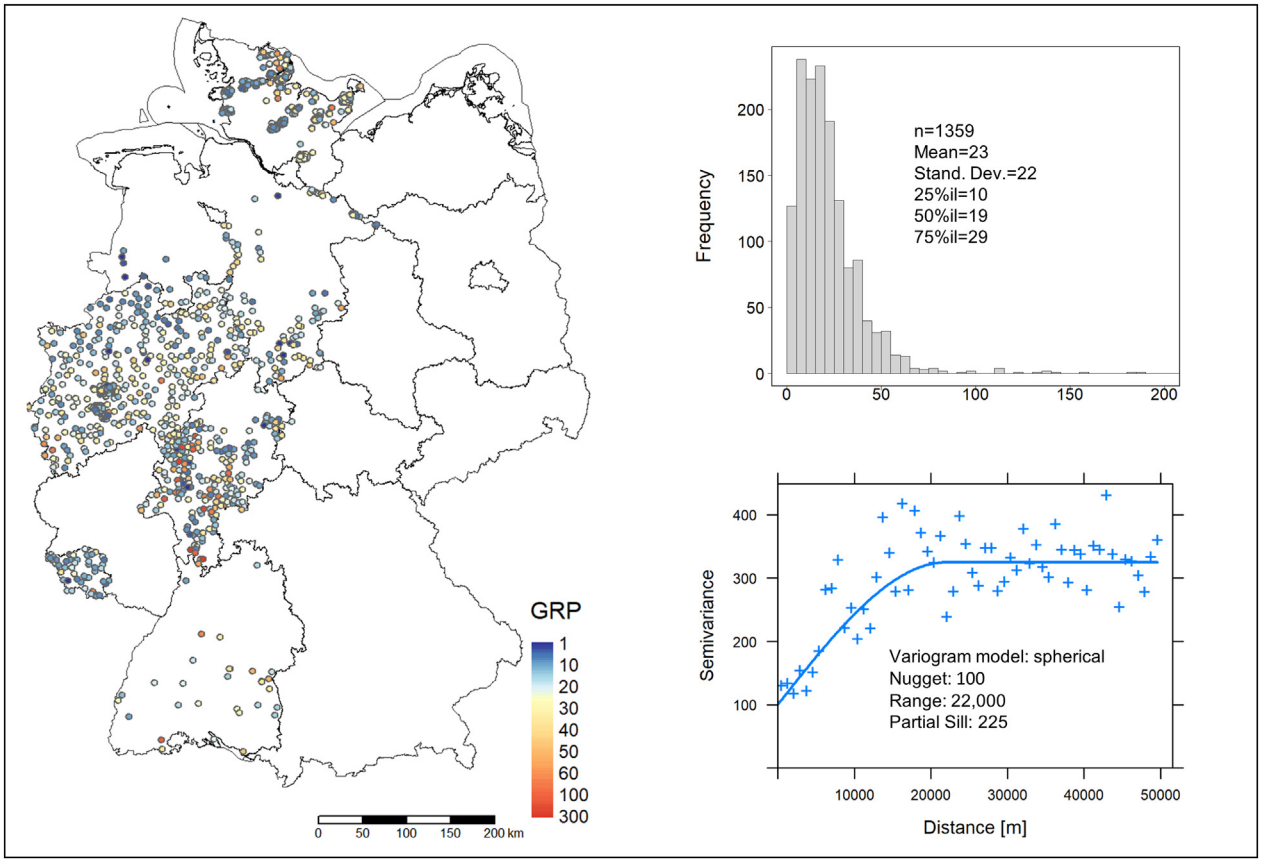


Fig. 9. Characteristics of GRP test data. (Left panel) Test data are located in six federal states with varying sampling density. The black solid lines indicate the borders of the federal states. (Top right panel) Histogram and summary statistics indicate a lower value range in comparison to the data used for model building (Fig. 1). (Bottom right panel) GRP test data semi-variogram model (visually fitted) and its parameters.

Fig. 10 revealed that the RF map is accurate to predict the mean value over the value range of 5 to ~40 for which sufficient test data are available. However, the large range of test data values within each GRP class indicated by the wide interquartile range shows that an individual GRP observation may deviate significantly from the predicted value on the map. This behavior is seen as a consequence of the high complexity of the system Rn in soil as well as the high spatial heterogeneity of relevant environmental variables on the sub-grid scale which cannot be resolved with the available predictor data.

Further, a comparison of different GRP maps would also be feasible by assessing their predictive power for modelling indoor Rn concentration (Bossew, 2017) since GRP is intended to be the predictor of indoor Rn hazard in German legislation. However, such an analysis is beyond the scope of this study.

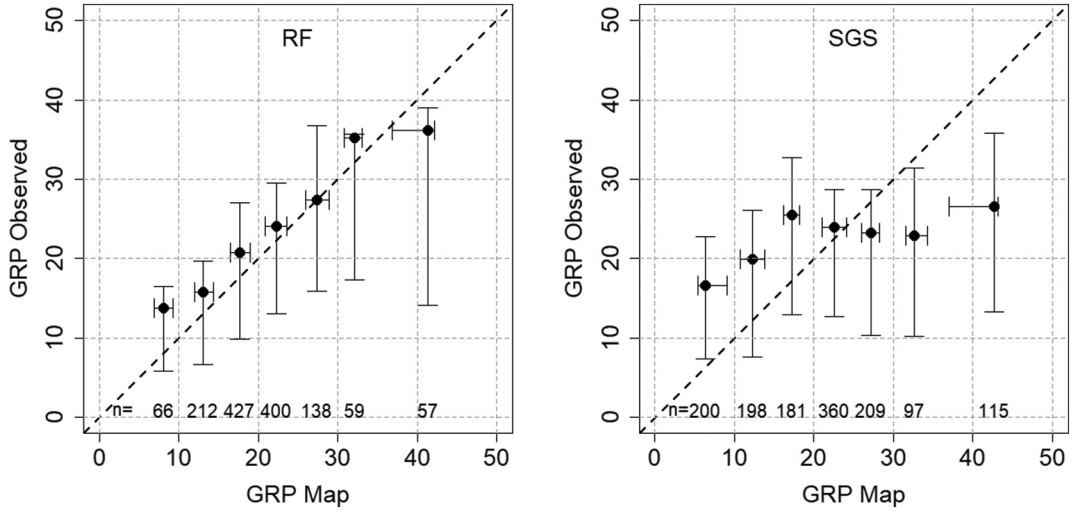


Fig. 10. Test performance of GRP maps: random forest (RF, this study, left panel) vs. sequential Gaussian simulation (SGS, Bossew (2015), right panel). GRP predictions (mapped values) were grouped into seven classes (<10, ≥10 to <15, ≥15 to <20, ≥20 to <25, ≥25 to <30, ≥30 to <35, ≥35) and related to test data at test locations. The black dots indicate the arithmetic mean, the bars the interquartile range of test data (vertical bars) and predicted values (horizontal bars) within the respective predicted GRP class. (n: number of GRP test data within respective GRP class).

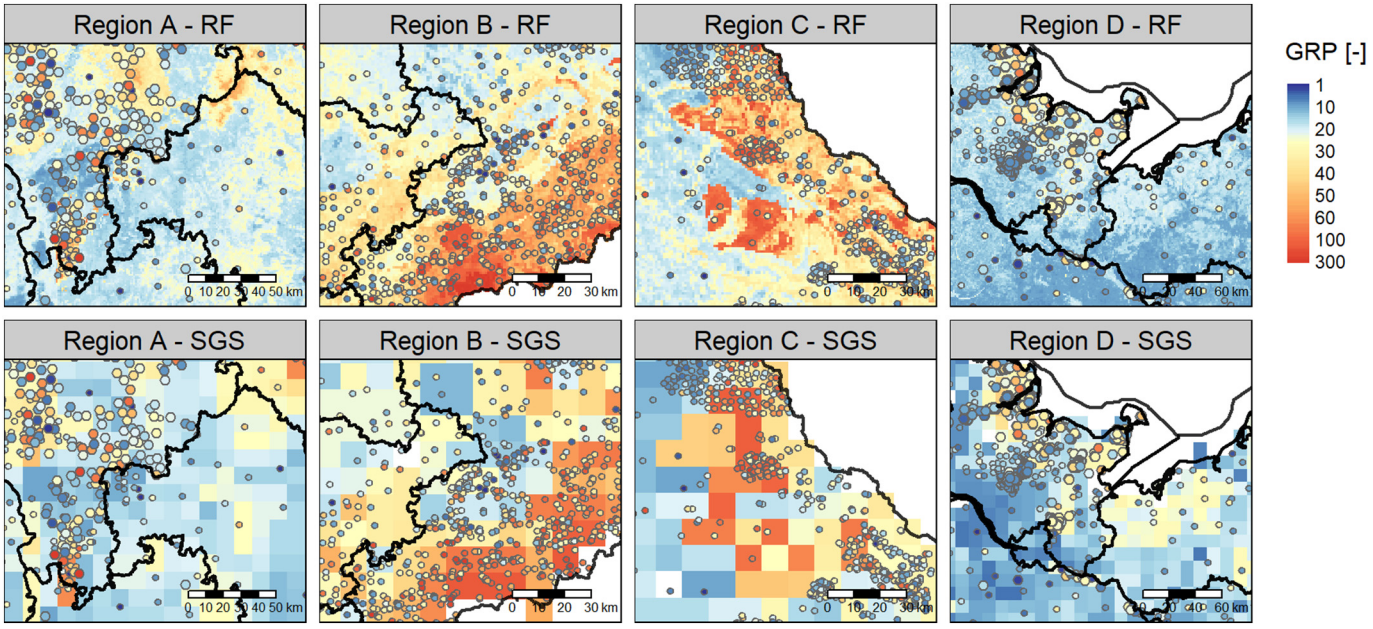


Fig. 11. Comparison of random forest based GRP map (this study) with a sequential Gaussian simulation (SGS) based map (Bossew, 2015) for exemplary regions A–D. The RF map has a spatial resolution of 1 km * 1 km, the SGS map of 10 km * 10 km. The locations of the regions A–D within Germany are given in Fig. 6. Small dots indicate GRP field observations which were used for building the RF model, large dots indicate test data (only applicable for region A and D). The color coding refers to both, the continuous GRP maps as well as the point samples. Maps were produced with tmap (Tennekes, 2018). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The spatial cross-validated r^2 of 0.22 achieved by RF is a comparably low value. However, for a judgement of this number, several aspects require consideration: (1) the target variable GRP was not transformed, (2) GRP field observations are extremely noisy, and (3) the performance assessment using a spatial cross-validation is rather conservative. First, a transformation (e.g., log-transformation) would have expectedly resulted in a higher r^2 due to the near log-normal distribution of the GRP. However, such a transformation (not common in ML since no Gaussian distributions are assumed), would have implied the necessity of back-transformation, which is not a standard procedure for ML regression models. Second, the significant temporal variability of GRP (see Section 2), e.g., as a consequence of soil moisture fluctuations, implies a large characterization uncertainty for the individual locations, i.e. the measured values might not reflect the long-term average GRP. Therefore, in our case the total uncertainty budget contains not only the modelling uncertainty but also a significant contribution from the site characterization uncertainty. Third, the performance assessment using a spatial cross-validation provides a rather conservative (i.e., over-pessimistic) performance estimate because the predictor space is likely to be insufficiently covered during model training. This could - but this requires further investigation - be especially relevant when categorical predictors with many levels rank amongst the most important predictors. For illustration, if an important level of a categorical predictor (geological classes in our case), that extends only over a small area, is present in the testing fold, data on this respective categorical predictor level might be only limited available for model training which complicates establishing a proper predictor-response-relationship.

Fig. 4 revealed a strong smoothing of the predicted values relative to the observations. The reasons for this characteristic could be manifold such as 1) incompleteness of the set of candidate predictors, 2) uncertainty in the predictor data (e.g., potentially insufficient spatial resolution of the geological map) or 3) high site characterization uncertainty of GRP observations. Observations with very low values (GRP < 5) are generally over-estimated by the model, very high values (GRP > 200) are generally under-estimated by the model. However, this misfit at both tails of the distribution could also be a consequence of the fact that these extreme values do not necessarily reflect the average characteristics of the respective site, whereas the model is intended to predict the average characteristics.

The presented RF model is only capable to explain 22% of the variability as indicated by the spatial cross-validated r^2. However, it is very likely that the performance of the RF based map is significantly better than the performance of the SGS approach as shown by the evaluation on test data.

5.2. Random Forest model interpretation

Regarding model interpretation, it is not surprising that geology turned out to be the overall most important predictor. The geological classes associated with higher GRP predictions (Fig. 8) confirm the understanding of the dependence of Rn on geology. The geological classes with felsic magmatites (granites and rhyolites) and intermediate to felsic magmatites (e.g., monzogranites, granodiorites) were those associated with highest GRP predictions. Younger granites are related to even higher GRP in comparison to older granites – a characteristic which has already been acknowledged by Kemski et al. (2001). Further, important geological classes are metamorphic rocks either originating from intermediate to felsic magmatites (e.g., orthogneiss) or organic rich metamorphic rocks (e.g. alum shale). In areas with glacial sediments those areas affected by the most recent glacial advance (e.g., near the coast of the Baltic Sea in northern Germany) reveal higher GRP values due to lower mineral weathering compared to glacial deposits from older glacial advances.

Beyond the confirmation of the geology-GRP relationship, the RF model revealed insight into the usefulness and the impact of other predictors reflecting climate as well as physical, chemical and hydraulic soil properties for and on GRP predictions. These predictors directly describe or represent proxies for the physical processes related to Rn generation, emanation, transport in and exhalation from the ground. Whereas geology reflects mainly the content of the Rn parent nuclide ^{226}Ra in the soil which determines Rn generation, the soil chemical predictors can be correlated with ^{226}Ra (e.g., potassium). Soil physical and hydraulic properties determine the gas permeability of the soil and govern Rn transport and exhalation. Further, these predictors may also affect Rn emanation rates which depends on grain size. Climate

predictors are expected mainly to be proxies for soil moisture (interestingly, soil moisture itself was less frequently selected than temperature and precipitation) which determines Rn emanation rates as well as transport and exhalation: high soil moisture inhibits Rn transport, which causes higher soil gas Rn concentrations and, consequently, lower Rn exhalation rates. Curiously, uranium concentration was not selected as an informative predictor. This is surprising because Rn belongs to the ^{238}U decay chain and both elements should exhibit spatial covariance. Possibly, the spatial resolution of uranium as a predictor (10 km * 10 km) is not sufficient and/or the information in the uranium data is already provided by the geological data. Some of the predictors are highly correlated (e.g., seasonal patterns of temperature or precipitation) which results in these cases in an exchangeability (e.g., spring temperature instead of autumn temperature) of the predictors which would have only a marginal effect on predictive performance.

The estimated relative uncertainty of GRP predictions coincides to some extent with the respective sampling density, therefore, future sampling campaigns should focus on these under-sampled areas to establish a more stable predictor-response relationship. Nonetheless, certain areas with low sampling density exhibit a noticeable low relative uncertainty such as parts of the north-eastern German Lowland. This characteristic reflects most likely the relatively low spatial geogenic heterogeneity in these areas yielding a more stable predictor-response relationship.

In comparison to previous attempts (Bossew, 2015; Kemski et al., 2001) the RF derived GRP map has a much higher spatial resolution. The general pattern with higher GRP in the majority of the Central German Uplands and lower GRP in northern Germany is - to a large extent - in accordance. However, the effect of the higher spatial resolution and higher mapping accuracy reveals more details on geological and landscape structures. Hence, the RF based map draws a spatially more differentiated picture allowing a more precise delineation of high GRP areas.

5.3. Modelling strategy

The spatial cross-validation procedure applied in this study is expected to reduce the effect of spatial auto-correlation in predictor and response data on the model building process. Consequently, predictor selection, hyperparameter tuning and performance assessment can be assumed to be more reliable compared to a random or stratified cross-validation procedure. However, a critical choice remains the definition of the size for the spatial blocks (40 km in this study) as discussed in Roberts et al. (2017). Anyway, independently from the size of the blocks it is always possible that data points from neighboring blocks belonging to different folds are within a spatial distance smaller than the correlation length of response and/or predictor data.

The existence of spatial structure in the residuals of the RF model might encourage to conduct an additional modelling step on top of the RF model by adding the interpolated residuals (e.g., using kriging) to the RF estimates. In this study, we deliberately rejected this approach, known as regression kriging, because (1) it would impede model interpretability, i.e. the insight gained from variable importance as well as partial and spatial dependence plots would be diluted, and (2) it might give a probable disproportionately high weight to individual observations, which is critical due to the assumed high site characterization uncertainty. This applies particularly to regions with low sampling density where a single unrepresentative measurement could unjustifiably increase/decrease the GRP estimate over the whole correlation length.

5.4. Outlook

Further improvements will be presumably achieved by implementation of more field data and additional predictor data (e.g., a higher resolved geological map). From a methodological point of view the effect of block size during spatial cross-validation, the effect of alternative objective functions during model building, e.g. minimization of RMSE or RMSLE instead of r^2, and other ML algorithms require further investigation. Regarding ML algorithms gradient boosting machines (Janik et al., 2018) and deep learning neural networks (Behrens et al., 2018) seem particularly promising alternatives. Moreover, techniques such as model averaging (Nussbaum et al., 2018) and model stacking (Taghizadeh-Mehrjardi et al., 2020) have shown that they can outperform individual ML algorithms.

Finally, it need to be stated that the GRP map pictures the average situation what earth delivers in terms of Rn. However, the GRP is not intended to provide a definite site assessment for an individual house in terms of its indoor Rn situation. This is not only a consequence of the high small-scale variability of geogenic factors but also of the dependence of indoor Rn on anthropogenic factors (e.g., building type, resident behavior). Consequently, the Rn concentration of an individual house can only be determined by an indoor Rn measurement.

6. Conclusion

In this study, three ML algorithms were compared regarding their performance for predicting the GRP for Germany. Model building was done using a spatial cross-validation using spatial blocks for fold creation which is believed to have minimized dependence of training and test data. The best results were clearly achieved by RF providing the best predictive accuracy (highest r^2, lowest RMSE and RRSE), the lowest scatter in the predictions and, in addition, the lowest magnitude of spatial structure in the model residuals.

The GRP map based on RF utilizes ten predictors which reflect geology, climate (winter temperature, summer precipitation, spring precipitation and winter soil moisture), soil hydraulic (parameter alpha of the hydraulic conductivity curve), soil physical (field capacity, coarse fraction) and soil chemical properties (potassium, nitrogen). The map has a spatial resolution of 1 km*1 km. Calculations of the variable importance have shown the dominant impact of geology but still significant contributions of the other predictors. Partial and spatial dependence plots have revealed further valuable insights into the quantitative predictor-response relationship. A comparison with the current German GRP map (Bossew, 2015) indicated a better mapping accuracy of the RF based approach using independent test data ($n = 1359$). However, this preliminary conclusion would be further supported with test data from other parts of the country.

This study is intended to provide a machine learning based modelling framework for GRP mapping rather than to present a “final” GRP map for Germany. Additional observational and predictor data as well as methodological progress are expected to further improve this map.

CRediT authorship contribution statement

Eric Petermann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Hanna Meyer:** Methodology, Software, Validation, Writing - review & editing. **Madlene Nussbaum:** Methodology, Software, Validation, Writing - review & editing. **Peter Bossew:** Methodology, Validation, 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.

Acknowledgements

We would like to thank the Bern University of Applied Sciences (BFH) for providing access to their computational resources which significantly contributed to the success of this study.

Appendix A

Table 2
Stratigraphical and lithological description and summary statistics of the classified geological units.

ID	Stratigraphy	Lithology (genesis)	25% ile	Mean	75% ile	Max	n
221	Proterozoic - Upper Paleozoic	Slate, quartzite, sandstone, phyllitic schist, greywacke	16.4	41.8	51.1	360.0	136
240	Devonian	Slate, sand stone, greywacke, quartzite, limestone	12.2	25.0	31.1	293.0	405
251	Carboniferous	Greywacke, slate, Plattenkalk, siliceous limestone, alum shale	17.8	44.2	42.9	811.2	138
252	Carboniferous	Quartzite, greywacke, clayey shale	15.2	24.5	28.2	100.0	97
261	Permian	shale clay, sand stone, conglomerate, black coal, porphyric breccia, tuff	8.6	25.3	35.3	167.1	160
262	Permian - Lower carboniferous	Dolomite, clay stone, limestone, olisthostrome	17.2	38.0	54.6	94.6	49
300	Triassic, Jurassic, Cretaceous, Upper Carboniferous	Clastic and evaporitic sedimentary rock	7.1	18.6	24.0	237.6	850
400	Tertiary	Sand, clay, marl, occasional conglomerate, etc.	8.3	20.8	27.8	96.9	131
522	Quaternary (Elster-Saale)	Sand and gravel (periglacial-fluviatile)	6.2	13.3	17.1	48.5	117
523	Quaternary (Elster-Saale)	Silt, clay, sand, gravel and stones (ground-, terminal moraine)	6.5	15.4	18.5	95.0	110
540	Quaternary (Weichsel)	Sand and gravel (periglacial-fluviatile)	5.7	13.4	14.9	119.1	171
545	Quaternary (Weichsel)	Silt, clay, sand, gravel and stones (ground-, terminal moraine)	9.1	20.5	27.2	152.6	253
553	Quaternary (Würm-Riss) -only southern Germany	Sand and gravel (periglacial-fluviatile)	22.8	37.2	42.7	105.0	82
561	Quaternary (holocene)	Sand, gravel, rubble (fluviatile-limnic)	10.5	26.8	36.9	215.9	125
565	Quaternary (holocene)	Sand, dunes (perimarine or aeolian); peat	2.9	10.7	14.2	58.5	42
581	Quaternary	Limestone, dolomite, marl, sand, gravel (derived from triassic source)	21.1	46.1	70.1	165.7	26
720	Upper Proterozoic - Middle Paleozoic	Phyllite, quartzite, serizite	29.0	60.5	70.4	915.0	145
730	Upper Proterozoic - Cambrian	Mica schist	18.3	46.9	56.0	273.3	101
740	Upper Proterozoic - Devonian	Medium-high metamorphic (e.g., amphibolite, gneiss, mylonite, hornfels)	14.5	23.7	29.2	105.5	74
750	Upper Proterozoic - Lower Paleozoic	Paragneiss	17.1	43.6	54.3	659.2	317
760	Upper Proterozoic - Paleozoic	High metamorphic (e.g., migmatite, diatexite, palite, granulite)	6.1	25.9	44.0	108.5	56
770	Proterozoic - Paleozoic	High metamorphic (e.g., orthogneiss, gneiss-anatexit)	25.2	57.2	71.7	338.9	110
810	Carboniferous - Permian (Rotliegend)	Felsic vulcanite (rhyolite)	10.5	48.0	73.4	199.5	57
861	Paleozoic - Lower Carboniferous	Mafic vulcanite (e.g., spilite, basalt, diabase, trachyte)	8.7	25.3	29.3	139.0	47
867	Tertiary - Quaternary	Mafic vulcanite (e.g., alkaline basalt, basanite, tephrite, nephelinite, tholeiitic)	10.5	34.1	39.0	232.0	61
900	Carboniferous - Permian (Rotliegend)	Mafic vulcanite + plutonite (e.g., andesite, latite, dacite and basalt, diorite, gabbro)	19.0	49.4	56.0	425.4	54
911	Upper Devonian - Carboniferous	Felsic plutonite (granite)	23.2	73.7	96.5	1027.9	307
912	Carboniferous - Permian (Rotliegend)	Felsic plutonite (granite)	28.7	127.0	169.4	1132.3	148
930	Upper Proterozoic - Paleozoic	Intermediate-felsic plutonite (granodiorite, monzogranite; meta-granodiorite)	13.8	90.5	64.8	1193.9	28
970	Upper Proterozoic	Mylonite, Ultramylonite	26.0	53.0	57.6	207.2	47

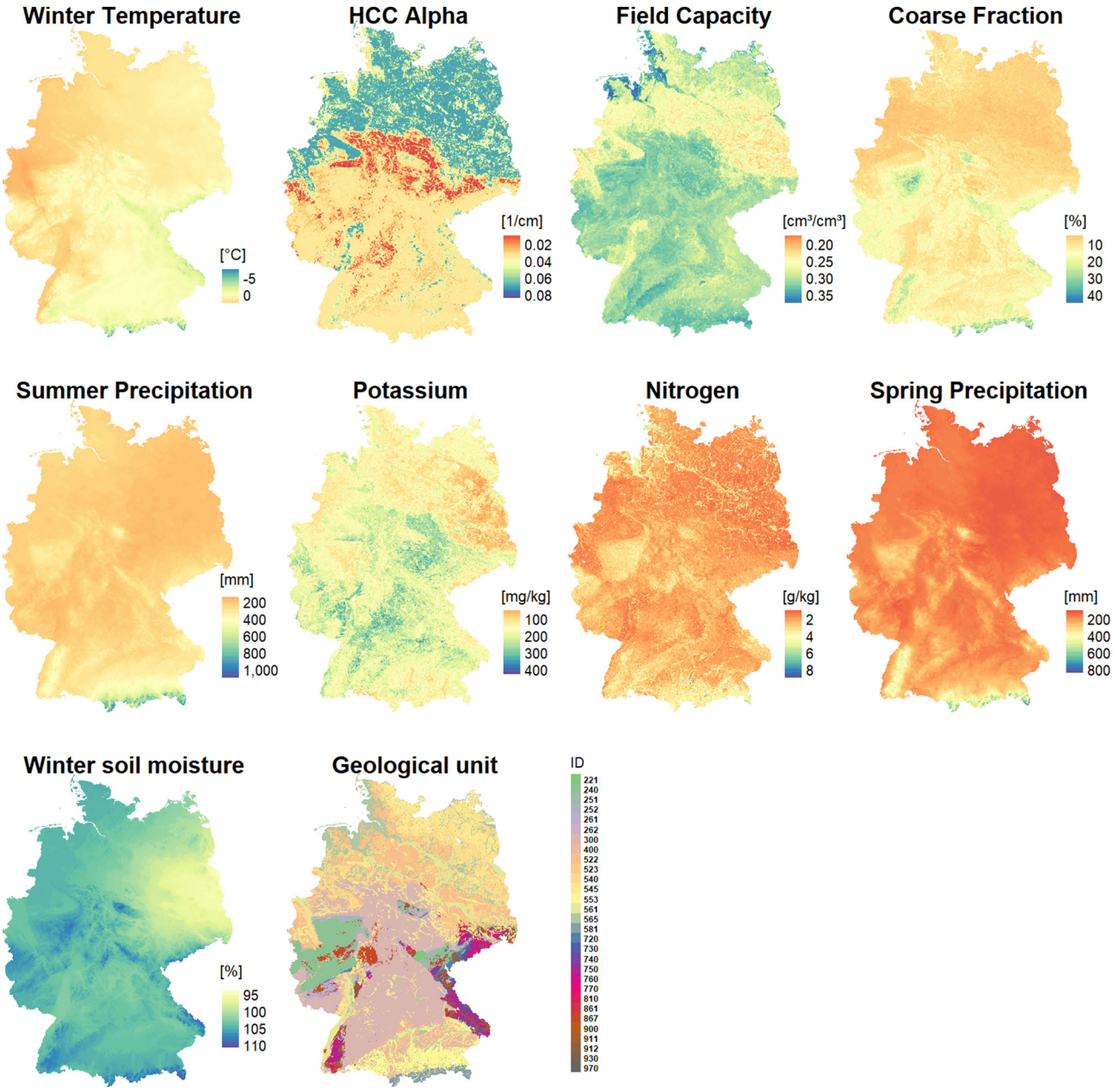


Fig. 12. Informative predictors in the random forest model.

References

Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261, 110–123.

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., et al., 2019. Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. Geoderma 355, 113912.

Behrens, T., Schmidt, K., Ramirez-Lopez, L., Gallant, J., Zhu, A.-X., Scholten, T., 2014. Hyper-scale digital soil mapping and soil formation analysis. Geoderma 213, 578–588.

Behrens, T., Schmidt, K., MacMillan, R.A., Viscarra Rossel, R.A., 2018. Multi-scale digital soil mapping with deep learning. Sci. Rep. 8, 15244.

BfS, 2020. Radiation exposure of airline passengers. Ionising Radiation. Federal Office for Radiation Protection (BfS), Salzgitter, Germany, p. 2019.

BGR, 1993. Geological Map of Germany Scale 1: 1 000 000. Federal Institute for Geosciences and Natural Resources, Hannover, Germany.

BKG, 2018. Digital Elevation Model of Germany 25 m Resolution. Federal Agency for Cartography and Geodesy, Frankfurt am Main, Germany.

Boehmke, B., Greenwell, B.M., 2019. Hands-on Machine Learning with R. CRC Press.

Böhner, J., Selige, T., 2002. Spatial prediction of soil attributes using terrain analysis and climate regionalization. Gottinger Geograpihsche Abhandlungen 115.

Böhner, J., Koethe, R., Conrad, O., Gross, J., Ringeler, A., Selige, T., 2002. Soil regionalisation by means of terrain analysis and process parameterisation. Soil Classification 2001, 213–222.

Bossew, P., 2015. Mapping the geogenic radon potential and estimation of radon prone areas in Germany. Radiat. Emerg. Med 4, 13–20.

Bossew, P., 2017. Local probability of indoor radon concentration to exceed the threshold estimated from geogenic radon potential. Nuclear Technology and Radiation Protection 32, 70–76.

Bossew, P., Cinelli, G., Ciotoli, G., Crowley, Q.G., De Cort, M., Elío Medina, J., et al., 2020. Development of a geogenic radon hazard index—concept, history, experiences. Int. J. Environ. Res. Public Health 17, 4134.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bruno, R.C., 1983. Sources of indoor radon in houses: a review. Journal of the Air Pollution Control Association 33, 105–109.

Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13, 361–378.

Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., et al., 2020. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Sci. Total Environ. 701, 134474.

Cinelli, G., Tollefsen, T., Bossew, P., Gruber, V., Bogucarskis, K., De Felice, L., et al., 2019. Digital version of the European Atlas of natural radiation. J. Environ. Radioact. 196, 240–252.
Darby, S., Hill, D., Auvinen, A., Barros-Dios, J.M., Baysson, H., Bochicchio, F., et al., 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ 330, 223.
DWD, 2018a. Multiannual Mean of Air Temperature (2m) in Germany, Period 1981–2010. German Weather Service (DWD) - Climate Data Center, Offenbach.
DWD, 2018b. Multiannual Mean of Precipitation in Germany, Period 1981–2010. German Weather Service (DWD) - Climate Data Center, Offenbach.
DWD, 2018c. Multiannual Mean of Soil Moisture Under Grass and Sandy Loam, Period 1991–2010. German Weather Service (DWD) - Climate Data Center, Offenbach.
European Council, 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. 2013/59/Euratom. 57(L13). Off. J. Eur. Union 73.
Font, L., Baixeras, C., Moreno, V., Bach, J., 2008. Soil radon levels across the Amer fault. Radiat. Meas. 43, S319–S323.
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.
Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 1–67.
Gaskin, J., Coyle, D., Whyte, J., Krewksi, D., 2018. Global estimate of lung cancer mortality attributable to residential radon. Environ. Health Perspect. 126, 057009.
Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. The R Journal 8, 204–218.
Greenwell, B.M., 2017. pdp: an R package for constructing partial dependence plots. The R Journal 9, 421–436.
Greenwell, B., Boehmke, B., Gray, B., 2020. vip: Variable Importance Plots.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398.
Hengl, T., MacMillan, R.A., 2019. Predictive Soil Mapping With R. OpenGeoHub foundation, Wageningen, the Netherlands, p. 370.
Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Graler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518.
Hothorn, T., Zeileis, A., 2015. partykit: a modular toolkit for recursive partytioning in R. J. Mach. Learn. Res. 16, 3905–3909.
Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., Laan, M., 2006a. Survival ensembles. Biostatistics (Oxford, England) 7, 355–373.
Hothorn, T., Hornik, K., Zeileis, A., 2006b. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Stat. 15, 651–674.
ICRP, 2007. The 2007 Recommendations of the International Commission on Radiological Protection. ICRP Publication. International Commission on Radiological Protection, p. 103.
Ielsch, G., Cushing, M.E., Combes, P., Cuney, M., 2010. Mapping of the geogenic radon potential in France to improve radon risk management: methodology and first application to region Bourgogne. J. Environ. Radioact. 101, 813–820.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. vol. 112. Springer.
Janik, M., Bossew, P., 2016. Analysis of simultaneous time series of indoor, outdoor and soil air radon concentrations, meteorological and seismic data. Nukleonika 61, 295–302.
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.
Kanevski, M., Pozdnoukhov, A., Timonin, V., 2009. Machine Learning for Spatial Environmental Data: Theory, Applications, and Software. EPFL press.
Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab - an S4 package for kernel methods in R. J. Stat. Softw. 11, 1–20.
Kemski, J., Klingel, R., Siehl, A., 1996. Classification and mapping of radon-affected areas in Germany. Environ. Int. 22, 789–798.
Kemski, J., Siehl, A., Stegemann, R., Valdivia-Manchego, M., 2001. Mapping the geogenic radon potential in Germany. Sci. Total Environ. 272, 217–230.
Kemski, J., Klingel, R., Siehl, A., Valdivia-Manchego, M., 2008. From radon hazard to risk prediction-based on geological maps, soil gas and indoor measurements in Germany. Environ. Geol. 56, 1269–1279.
Krewski, D., Lubin, J.H., Zielinski, J.M., Alavanja, M., Catalan, V.S., William Field, R., et al., 2006. A combined analysis of north American case-control studies of residential radon and lung cancer. J. Toxic. Environ. Health A 69, 533–597.
Kropat, G., Bochud, F., Jaboyedoff, M., Laedermann, J.P., Murith, C., Palacios Gruson, M., et al., 2015. Improved predictive mapping of indoor radon concentrations using ensemble regression trees based on automatic clustering of geological units. J. Environ. Radioact. 147, 51–62.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. vol. 26. Springer.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., et al., 2019. caret: Classification and Regression Training.
Lary, D., Faruque, F., Malakar, N., Moore, A., Roscoe, B., Adams, Z., et al., 2014. Estimating the global abundance of ground level presence of particulate matter (PM2.5). Geospat. Health 8, 292.
Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geosci. Front. 7, 3–10.
Li, J., Heap, A.D., Potter, A., Daniell, J.J., 2011. Application of machine learning methods to spatial interpolation of environmental variables. Environ. Model Softw. 26, 1647–1659.
Li, J., Siwabessy, P.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.
Liess, M., Schmidt, J., Glaser, B., 2016. Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. PLoS One 11, e0153673.
Markkanen, M., Arvela, H., 1992. Radon emanation from soils. Radiat. Prot. Dosim. 45, 269–272.
Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.
Meyer, H., 2018. CAST: 'caret' Applications for Spatial-Temporal Models.
Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model Softw. 101, 1–9.
Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction. Ecol. Model. 411, 108815.
Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., et al., 2014. Machine learning feature selection methods for landslide susceptibility mapping. Math. Geosci. 46, 33–57.
Millborrow, S., 2019. earth: Multivariate Adaptive Regression Splines.
Neznal, M., Neznal, M., Šmarda, J., 1996. Assessment of radon potential of soils—a five-year experience. Environ. Int. 22, 819–828.
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. 16. Czech Geol. Survey.
Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., et al., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil 4, 1–22.
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.
Pebesma, E., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.
Pohjankukka, J., Pahikkala, T., Nevalainen, P., Heikkonen, J., 2017. Estimating the prediction performance of spatial models via spatial k-fold cross validation. Int. J. Geogr. Inf. Sci. 31, 2001–2019.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., et al., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929.
Scanlon, B., Nicot, J.-P., Massmann, J., 2002. Soil gas movement in unsaturated systems. Soil Physics Companion 389, 297–341.
Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol. Model. 406, 109–120.
Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8, 25.
Szabó, K.Z., Jordan, G., Horváth, Á., Szabó, C., 2013. Dynamics of soil gas radon concentration in a highly permeable soil based on a long-term high temporal resolution observation series. J. Environ. Radioact. 124, 74–83.
Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., et al., 2020. Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning covariate space. Remote Sens. 12, 1095.
Tanner, A.B., 1988. A tentative protocol for measurement of radon availability from the ground. Radiat. Prot. Dosim. 24, 79–83.
Tennekes, M., 2018. tmap. Thematic Maps in R 2018 (84), 39.
Timkova, J., Fojtikova, I., Pacherova, P., 2017. Bagged neural network model for prediction of the mean indoor radon concentration in the municipalities in Czech Republic. J. Environ. Radioact. 166, 398–402.
Torkar, D., Zmazek, B., Vaupotič, J., Kobal, I., 2010. Application of artificial neural networks in simulating radon levels in soil gas. Chem. Geol. 270, 1–8.
Tóth, B., Weynants, M., Pásztor, L., Hengl, T., 2017. 3D soil hydraulic database of Europe at 250 m resolution. Hydrol. Process. 31, 2662–2666.
Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: an r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods Ecol. Evol. 10, 225–232.
Vapnik, V., 2013. The Nature of Statistical Learning Theory: Springer Science & Business Media.
WHO, 2009. Handbook on Indoor Radon - A Public Health Perspective. Geneva.
Wiegand, J., 2001. A guideline for the evaluation of the soil radon potential based on geogenic and anthropogenic parameters. Environ. Geol. 40, 949–963.
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.
Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016. Data Mining. Fourth edition. Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers Inc.