Machine Learning Techniques for the Prediction of Indoor Gamma-ray Dose Rates – Strengths, Weaknesses and implications for Epidemiology

Word count: abstract 176 words, main text (excluding references, Tables, Figures) 10,141 words

67 references

7 Tables, 3 Figures

GM Kendall[a], JD Appleton[b], P Chernyavskiy[c], A Arsham[d], MP Little[e, f]

[a]Cancer Epidemiology Unit, NDPH, University of Oxford, Richard Doll Building, Old Road Campus, Headington, Oxford OX3 7LF, UK

[b]British Geological Survey, Kingsley Dunham Centre, Nicker Hill, Keyworth, Nottingham NG12 5GG, UK

[c]Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville VA 22908-0717, USA

[d]Center for Data, Mathematical & Computational Sciences, Goucher College, Baltimore, Maryland, USA

[e]Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, DHHS, NIH, Bethesda, MD 20892-9778, USA

[f]Faculty of Health and Life Sciences, Oxford Brookes University, Headington Campus, Oxford, OX3 0BP, UK

Word count: abstract 176 words, main text (excluding references, Tables, Figures) 10,141 words

67 references

7 Tables, 3 Figures

Author roles

Conceptualization, Methodology: Mark Little, Gerald Kendall; Software: Mark Little; Data curation: Gerald Kendall; Writing - Original draft preparation: Gerald Kendall, Mark Little;

Supervision: Gerald Kendall; Software, Validation: Mark Little; Writing - Reviewing and Editing: all authors.

Acknowledgements

The authors are grateful to Dr Richard Wakeford and the two referees for detailed and helpful comments. The work of MPL was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics.

Abstract

We investigate methods that improve the estimation of indoor gamma ray dose rates at locations where measurements had not been made. These new predictions use a greater range of modelling techniques and larger variety of explanatory variables than our previous examinations of this subject. Specifically, we now employ three types of machine learning models in addition to the geostatistical, nearest neighbour and other earlier models. A large number of parameters, mostly describing the characteristics of dwellings in the area in question, have been added to the set of explanatory variables. The use of machine learning methods results in significantly improved predictions over earlier models. The machine learning models are noisy and there is some instability in the relative importance of particular explanatory variables although there are general and consistent tendencies supporting the importance of certain classes of variable. However, the range of predicted indoor gamma ray dose rates is much smaller than that of the measurements. It is probable that epidemiological studies using such predictions will have lower statistical power than those based on direct measurements.

Keywords: machine learning; natural background; gamma radiation; geostatistcs; neural networks; random forests

Highlights

- Direct estimates of background γ-ray doses are expensive and liable to bias.
- Machine learning and more standard models were fitted to γ data in Great Britain.
- Machine learning results in significantly improved predictions over other models.
- All machine learning models performed well, with boosting models performing best.
- Machine learning models are noisy, and may result in loss of power.

1. Introduction

Natural background gamma radiation from terrestrial sources produces a globally averaged annual effective dose of 0.48 mSv with a range of 0.3 to 1.0 mSv (United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2010). These doses are straightforward to measure on an individual basis, with the largest contribution coming from exposures within dwellings. Power calculations suggest (Little et al., 2010) that these doses might induce a detectable increase in the rate of childhood leukaemia for very large, but achievable, epidemiological studies. Direct demonstration of the effects of such low doses is of considerable importance in radiological protection. For example, low doses affect the appropriateness of the Linear No-Threshold (LNT) dose-response model for cancer as a basis for practical decision making. A number of epidemiological studies of the effects of this component of natural background radiation on childhood cancer rates have been published (Berlivet et al., 2021; Berlivet et al., 2020; Demoury et al., 2017; Kendall et al., 2013; Mazzei-Abba et al., 2021; Nikkilä et al., 2016; Spycher et al., 2015). Some suggest that such an effect has been detected (e.g., (Kendall et al., 2013; Mazzei-Abba et al., 2021; Spycher et al., 2015)), while others of apparently equivalent statistical power have been negative (Berlivet et al., 2021; Berlivet et al., 2020; Demoury et al., 2017).

Natural background external radiation as measured, for example on thermoluminescent detectors (TLDs), comprises a telluric and a cosmic ray component. The former consists of gamma rays from radionuclides in the local environment. As their name implies, cosmic rays originate beyond the earth and their interactions in the upper atmosphere are complex (United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 1977). However, at ground level, they consist mostly of muons; there are also cosmic ray neutrons but these are not directly ionising and will not be considered further here. For brevity, in the present paper we will refer to telluric and directly ionising cosmic ray components together as “gamma rays”.

The dose-rate from directly ionising cosmic rays varies with altitude. UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2000) give a

formula by which dose-rate at different heights above sea level may be calculated. At moderate altitudes, less than 500m, the predicted dose-rate is within 10% of that at sea level. However, this altitude variation is small compared with the variation in indoor dose-rates caused by the variation in shielding in different buildings (United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2000).

The dose of importance for epidemiological studies of radiation, whether a dose to individual body organs or an average whole body dose (International Commission on Radiological Protection (ICRP), 2021), is that incurred by study subjects over the whole period of study. In the case of studies of directly ionising external radiation, this includes both time-weighted in- and out-door components, with the former having contributions from the study subject's home and from other buildings. Occupancy factors for different buildings are of central importance. UNSCEAR gives authoritative discussions (United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2000) and Folly *et al* (Folly et al., 2022) conducted a detailed investigation, including individual dosimetry, in Switzerland. A general conclusion is that indoor doses in the home dominate, but that indoor doses depend on the characteristics of the building in question and on the age of the individual.

Ideally, an epidemiological study would obtain estimates of the gamma-ray doses incurred by study subjects through direct measurement. However, for large studies direct measurements would be prohibitively expensive and liable to participation and other biases (Kendall et al., 2015; Law et al., 2002; United Kingdom Childhood Cancer Study Investigators, 2002a). Therefore, modelling is required that permits the calculation of indoor gamma-ray dose rates at locations where no measurement has been made.

Epidemiological studies of childhood cancer and terrestrial gamma ray exposure (with or without a contribution from cosmic rays) have estimated doses in various ways. The published British study (Kendall et al., 2013) made use of mean dose rates in 459 County Districts (administrative units of ~110,000 population of all ages) to apply to study subjects. These means were estimated using 2283 measurements from the UK National Survey of indoor exposures from

natural radiation (Wrixon et al., 1988). Criticisms of the study were directed at the accuracy of these individual dose estimates (National Council on Radiation Protection and Measurements (NCRP), 2018; United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2017).

Accordingly, the study team set out to derive more accurate dose rate predictions using an enlarged measurement set and a variety of plausibly relevant explanatory variables (Chernyavskiy et al., 2016; Kendall et al., 2018; Kendall et al., 2016b). Several models were explored including, classical geostatistical models, simple averaging over a variety of administrative or geological boundaries, and various kinds of *K*-nearest neighbour models (which we shall for simplicity describe from now on as nearest neighbour models). However, the models were not entirely satisfactory for a number of reasons:

1) Precise predictions could not be made: the standard deviation (SD) on modelled *vs* predicted values at the measurement points being not much less than 20% (~20 nGy/h on a mean of ~95 nGy/h)

2) The most successful model, the "extended ordinary least squares" (E-OLS) involved a weighted mean of a number of simpler models. There are sub-optimal statistical aspects of such a model, specifically that the dependent variable (dose rate) was used in formulation of some of the explanatory independent variables. There was also no satisfactory basis for exploring the effect of errors in dose estimates on the results of subsequent epidemiological estimates of risks per unit dose.

3) Previous models did not include some variables that might have been most relevant in the prediction of gamma dose rates. The most powerful of these missing variables is information on the radionuclide content of building materials, which were virtually impossible to obtain. But other, rather less direct factors like building age and type (e.g., flat, detached house, etc.) were more realistic targets.

4) The available models could not incorporate inclusion of closely related explanatory variables.

However, the situation has changed with the advent of machine learning (ML) models. Classification and regression trees (Breiman, 1984) model subgroup assignments, similar to those used in many algorithmic settings. However, tree-based models like other ML models, are intrinsically noisy. Specifically, small changes in the training set or variables included in the model can result in large changes of the fitted model. In this respect ML models differ substantially from more standard statistical models, e.g., linear regression. Random forest (RF) models (Breiman, 2001) are based on averaging predictions from individual trees in a forest, these trees being constructed independently. Each tree is created from a bootstrap sample of the training dataset. Each subgroup (node) within a tree is then formed by optimising the reduction in prediction error from a random selection of explanatory variables. The RF prediction averages individual tree predictions, this aggregation being known as bagging (bootstrap aggregating) (Breiman, 1996). Stochastic gradient-boosted machine (GBM) models (Friedman, 2002) are also tree-based algorithms. Unlike RF algorithms which use bagging, GBM uses boosting, by creating a sequence of trees, in which each tree updates the previous tree model to reduce prediction error. Both types of models are flexible, due to their ability to incorporate partially correlated variables, allowing for interactions between variables, and control for overfitting (Hastie et al., 2017). Artificial Neural network (ANN) models (McCulloch and Pitts, 1943), which to some extent mimic the structure of low-level connections in the nervous system, can have equivalent performance, although their additional complexity makes them more prone to overfitting (Hastie et al., 2017). An overfitted model fits the training set (data used to create the model) well but does not generalise well to other datasets (datasets independent of those used to create the model). In this paper, we explore these ML models in more detail. Other workers have also used such models for estimating natural background radiation (Elío et al., 2023; Nikkilä et al., 2020; Petermann et al., 2021), albeit for radon rather than gamma-rays.

For the present calculations, a number of new explanatory variables were introduced. These are described in more detail in the Materials and Methods, but, in brief, they consisted of:

1) Summarised house characteristics data from the Valuations Agency's Council Tax Stock of Properties (CTSOP) database

2) Estimated outdoor gamma-ray dose rates broken down into components from the three main radioactive decay chains (Appleton and Kendall, 2022). The uranium component may be taken as an analogue of the "uranium potential" used in other calculations (Berlivet et al., 2021; Demoury et al., 2017; Warnery et al., 2015).

3) The altitude of the measurement location (in addition to its geographical location as specified by Easting and Northing on the British National Grid)

4) Various socioeconomic status (SES) variables from the National Statistics Postcode Lookup (NSPL) index.

It should be noted that, these variables have been chosen because they are plausibly related to factors affecting indoor gamma-ray dose rates. However, the models make use of whatever purely empirical correlations there may be in the data.

As reported previously, we now have a database of over 10,000 measurements of indoor gamma-ray dose rates. These have a mean value of 95.6 nGy/h with a SD of 22.7 and a range of 25-278 nGy/h (Kendall et al., 2016b). Measurement errors are smaller than the inherent natural variability (Kendall et al., 2018). Statistical models that are based on averaging will reduce the range of the predictions to some extent, but an ideal model would predict a distribution of indoor gamma-ray dose rates not much narrower than the distribution of measurements. The SD between observed and calculated values would be about 5 nGy/h, and the SD between the calculated values would be a little over 20 nGy/h. In fact, the distributions of predicted values from the optimum models previously published are much narrower than that of the measurements (Kendall et al., 2018).

Nearest neighbour models average training set measurements close to the location of interest. In circumstances where we do not have the variables needed to fit the data well, it is possible that they may preserve the variation between points better than ML models which give all measurements equal weight and which treat variation that they cannot model as noise (Elío et al.,

2023). We therefore also explore nearest neighbour models which are not optimised over the full range of parameters, but which are more akin to the observed distribution; we refer to them as "Local" models. Geostatistical models also give more weight to nearby measurements, but generally to a lesser extent than nearest neighbour models.

2. Materials and Methods

2.1 Indoor gamma ray measurements.

For our analysis we are interested in modelling indoor gamma ray measurements. A total of 10,184 measurements of indoor gamma-ray dose rates in dwellings were available. These included the contribution from directly ionising cosmic rays. The measurements were from two sources: 2283 were from the UK National Survey of natural radioactivity in homes (Wrixon et al., 1988) and 7901 were from the UK Childhood Cancer Study (United Kingdom Childhood Cancer Study Investigators, 2002b). Both sets of measurements had been carried out over a six-month period by the National Radiological Protection Board (now the UK Health Security Agency) using thermoluminescent dosemeters. In almost all instances, the measurement was the average of the readings of two dosemeters, one from a living room and one from an occupied bedroom. It is estimated that the measurement uncertainty was around 5% (Kendall et al., 2018). The measurement locations of both surveys may be taken to be a population-weighted sample of UK homes (Kendall et al., 2018). It should be noted that the number of measurements is slightly lower than in previous publications (Chernyavskiy et al., 2016; Kendall et al., 2018; Kendall et al., 2016b). This is because 15 duplicate measurements have been detected and removed.

2.2 Variables used in modelling indoor gamma-ray dose rates

In order to predict indoor gamma-ray dose rates at locations where no measurements have been made (specifically the locations of participants in epidemiological studies) models are required. These models involve explanatory variables which are likely to vary with gamma-ray level. In total, 157 such variables have been used in the calculations reported here. A list is given in Supplement C. As noted in the Introduction, these variables are categorised into four main classes.

These are now discussed in turn, with particular emphasis on new variables, that were not used in previous modelling.

2.2.1 Geographical variables

The locations (eastings and northings) of the indoor gamma-ray dose rate measurements are known reasonably accurately, typically to within 100 m or less (Kendall et al., 2016b). These locations can be used to assign measurements to various geological or administrative units such as "County Districts" or their equivalents. County districts are relatively small administrative areas, numbering 459 in Great Britain, with a mean all-age population of rather more than 110,000.

All geographical variables were potentially available to earlier modelling except altitude (height of bare ground above sea level). This was obtained from the Ordnance Survey Terrain 50 dataset.

https://www.ordnancesurvey.co.uk/business-government/products/terrain-50

The height values are given to the nearest 0.1 m and the root mean square error is 4 m. The measurement locations, and the birth locations of study subjects in the epidemiological study of naturally occurring radiation and childhood cancer(Kendall et al., 2013), are all relatively low lying, with median heights around 60 m and about three quarters being at 100 m or less. The highest locations are at around 400 m.

2.2.2 Geological variables

Two main types of geological variables were used in earlier modelling:

LEX-ROCK Bedrock codes

These describe the underlying geology as defined by the British Geological Survey (BGS) (Smith, 2013), provide the name of each rock unit (via its LEX or Lexicon code), and composition (via its ROCK or lithology code). More details on these variables are provided in Kendall *et al.*(Kendall et al., 2016b).

50K-BEDSUP combined simplified bedrock and superficial codes

These variables were developed by the BGS in the context of an investigation into the variation of indoor radon concentrations across the UK (Miles and Appleton, 2005).

242 Simplified bedrock and superficial geology classification systems were developed which
243 grouped some geological units with similar characteristics (Appleton, 2005). A number of further
244 groupings were introduced later (Chernyavskiy et al., 2016).

245 Our present analysis now incorporates new explanatory variables including outdoor gamma
246 ray dose rates from potassium, thorium and uranium (with their decay chains) estimated from
247 geochemical data by Appleton and Kendall (Appleton and Kendall, 2022). The calculated values
248 for uranium, in particular, provide an analogue for geological uranium potential as used in other
249 modelling (Warnery et al., 2015). Appleton and Kendall (Appleton and Kendall, 2022) used over
250 a quarter of a million geochemical measurements of concentrations of potassium, thorium and
251 uranium in soils and in stream sediments. The soil concentrations are generally at a depth of 5–20
252 cm with some at 35–50 cm. Soil measurements give relatively precise estimates spatially, but were
253 not available for much of Scotland, where stream sediment data were also used. Kriging was used
254 to estimate surface concentrations of K, Th and U across Great Britain and dose rates were imputed
255 from these concentrations. Fuller details are given by Appleton and Kendall (Appleton and
256 Kendall, 2022). We also make use of measured outdoor gamma ray dose rates, as discussed below.

257 ***2.2.3 Socio-Economic Status (SES) variables***

258 It is well known that the affluence of the household is correlated with the level of the natural
259 radioactive gas radon (Kendall et al., 2016a) and the same is true for indoor gamma rays (albeit
260 that the more affluent tend to have higher radon and lower gamma-ray exposures) (Kendall et al.,
261 2018). Previous work has used SES variables, in a rather wide sense, from the UK national
262 censuses of 1981 and 1991. The variables concerned were calculated for census wards (these wards
263 are of variable geographical size and population with a mean of about 5000 people of all ages).
264 They include the Carstairs index of Deprivation (Carstairs and Morris, 1991) together with the
265 variables from which it is derived:

266 a) male unemployment rates;
267 b) the proportion of households in which the head of the household is in social class 4 or 5;
268 c) non-car ownership; and

d) overcrowding in private households.

The epidemiological studies of childhood cancer and natural background gamma rays that use models of the kind developed here use not just area-based SES variables like Carstairs, but also individual-based measures based on paternal employment. These employment-based measures are not available for dose rate measurement locations and cannot be used in dosimetric modelling.

Under the same general heading of SES were also included, for each census ward:

a) Population

b) Population Density

c) Urban/Rural status

A further variable included under the SES category was not derived from the censuses. This is the Land Use code developed by Dudley Stamp (Appleton and Cave, 2018; Kendall et al., 2018; Stamp, 1931).

Some new SES variables have been added since Kendall *et al.* (Kendall et al., 2018). These come from the National Statistics Postcode Lookup File (NSPL) (Office for National Statistics (ONS), 2024).

The new variables are:

- The 2011 Census Group Output Area Classification (OAC), and its Super- and Sub-Group.

  These describe the predominant population of specified census areas. For example, subgroup 1A1 (Rural workers and families) belongs to Group 1A (Farming communities) and Supergroup 1 (Rural residents).

- Index of Multiple Deprivation (IMD); rank as percentage within country.
- The unified 2011 Census rural-urban classification.

These NSPL variables (and also most of the house characteristics variables described below) are assigned to a more recent description of census areas than that used for the 1981 or 1991 censuses. The new scheme involves combinations of census “output areas” (OAs). The scheme

described is for England and Wales, but that for Scotland is broadly similar. In order of increasing size:

- Lower Layer Super Output Areas (LSOA) have a mean population figure of 1500 and each consists of a grouping of (typically) five OAs. There are 34,753 LSOAs, with a minimum population size of 1000.
- Middle Layer Super Output Areas (MSOA) have a mean population figure of 7200 (minimum 5000) and consist of a grouping of LSOAs.
- Local Authority Districts or unitary authorities (LAUA) correspond roughly to the old County Districts. Above LAUA are Counties, except where these have been replaced by unitary authorities.

2.2.4 Variables describing House Characteristics

Work previously reported had made use of only one house characteristic. This was an approximate division into construction before or after 1940. As described by Appleton and Cave (Appleton and Cave, 2018) and Kendall *et al* (Kendall et al., 2018), this is based on the Dudley Stamp land use maps completed in 1941 but mainly based on survey work carried out between 1931 and 1934 (Stamp, 1931). Dwellings in areas described in those maps as urban or suburban were taken to have been in existence at that time; dwellings in other areas were taken to be constructed later.

An extensive range of other house characteristics for England and Wales have now been obtained from the Council Tax Stock of Properties (CTSOP) files maintained by the Valuations Office Agency.

https://www.gov.uk/government/statistics/council-tax-stock-of-properties-2019

The data available are percentages of different:

Property types (5 types, e.g., detached house, semi-detached house, etc.)

Property value (9 bands, as estimated for purposes of Council Tax)

Period of construction (11 periods)

Number of bedrooms (1, 2, 3, 4 and more)

From these, rough estimates of mean property value, mean year of construction and mean number of bedrooms were derived. These data are tabulated for the newer administrative units, LSOA, MSOA, LAUA etc., as described above. With a total of 32 house characteristics for each of these units, these account for 96 of the total 157 parameters. More details are in Supplement C.

Rather more restricted data, limited to the Council Tax valuation at the level of LAUA, are available for Scotland.

2.2.5 Outdoor gamma-ray dose variable

Information on outdoor gamma-ray dose-rates came from a survey by Green *et al.* (Green et al., 1989): 2398 measurements were made using an energy-compensated Geiger–Müller tube over a period of 10 min. The random uncertainties on the measurements were typically ~8%. The principal published results were the double-smoothed dose-rates tabulated for each 10 km-side square of the National Grid in GB. More details are to be found in Kendall *et al.* (Kendall et al., 2018; Kendall et al., 2016b).

2.3 Statistical methods

2.3.1 Machine Learning Models

We fitted three types of machine learning models, specifically:

(1) a variant type of random forest (RF) algorithms, a generalization of the RF model developed by Breiman (Breiman, 2001) to take account of modifications to the tree expansion rules developed by Little *et al* (Arsham et al., 2023; Little et al., 2022), and also using guided regularised random forest algorithm (Deng and Runger, 2013);

(2) the stochastic gradient boosting machine (GBM) model of Friedman (Friedman, 2001, 2002); and

(3) artificial neural network (ANN, sometimes just “neural network”) models (McCulloch and Pitts, 1943).

There are four hyperparameters (user specified parameters) in the standard RF algorithm (Breiman, 2001), specifically, the number of trees (i.e., the number of bootstrap samples, **ntree**), and the number of variables sampled per node (**mtry**) used to determine splitting of a node, the

maximum number of nodes per tree (**maxnodes**), and the minimum number of records available for splitting in the penultimate node of each tree (**n.minobsnode**). The version of the algorithm that we have implemented incorporates a number of additional parameters that determine whether tree generation is halted, specifically:

(a) the proportion of the total variance (in the total dataset) of the outcome variable in a given node used to determine whether to stop the further development of the tree from that node downwards;

(b) the proportion of the total range (= maximum – minimum) (in the total dataset) of the outcome variable in a given node used to determine whether to stop the further development of the tree from that node downwards;

(c) the proportion of the intercentile range [X%, 100 – X%] (in the total dataset) of the outcome variable in a given node used to determine whether to stop the further development of the tree from that node downwards. We used X=10% and X=25%.

(d) the minimum number of observations per parent node (corresponding to **n.minobsnode**).

(e) the minimum number of observations per terminal (leaf) node.

The tree generation at a particular RF node is halted if any of conditions (a)-(e) is triggered. In most implementations of the standard RF model (Breiman, 2001), for example the R **randomForest** package (Liaw, 2018), only criteria (d) is available; in some software, in particular in the **randomForestSRC** (Ishwaran and Kogalur, 2020) and **partykit** (Hothorn, 2021) R packages criteria (d) and (e) are available as options. Further details are given in the papers of Arsham *et al.* (Arsham et al., 2023) and Little *et al.* (Little et al., 2022). The guided regularised random forest algorithm (Deng and Runger, 2013) proceeds by means of a two-stage model fit, the first of which is a standard fit of the model, from which is calculated the normalised Gini importance score for each variable $\mathrm{ImpNorm}(X_i) = \mathrm{Imp}(X_i) / \max\{\mathrm{Imp}(X_j) : j = 1,..,p\}$. In the

second model fit the potential gain in MSE at each node $Gain(X_i)$ from use of a potential variable X_i is replaced by:

$$\begin{cases} [(1-\gamma)\lambda_0 + \gamma \mathrm{ImpNorm}(X_i)]Gain(X_i) \text{ if } X_i \notin F \\ Gain(X_i) \text{ if } X_i \in F \end{cases} \quad (1)$$

where F is the set of nodes already used for splitting in this tree. The variable chosen for splitting (from the **mtry** variables selected for that node) is the one that maximises this weighted gain in MSE. The Fortran 95-2003 code implementing the regression RF algorithm described above is given in Supplement A, along with a number of parameter steering files.

In fits of the GBM models there are five hyperparameters, the maximum interaction depth (**interaction.depth**), that is the highest level of variable interactions allowed (so **interaction.depth**=1 would correspond to a model without any interactions), the shrinkage or learning rate (**shrinkage**) controls the contribution of each tree iteration, the minimum node size in the terminal nodes (**n.minobsinnode**), the bagging fraction (**bag.fraction**), that is the fraction of the training set observations randomly selected to propose the next tree in the sequence, and the number of trees used (**n.trees**). In all fits we used **n.trees**=1000 and determined the number of trees ≤1000 that minimises the cross-validated mean square predictive error (MSPE). The model was fitted using the R **gbm** library (Greenwell et al., 2020).

In fits of the neural network models the main adjustable parameters are the number of layers, and the number of neurons in each layer. Additionally, one must specify the probability of dropout (**dropout**), that is to say, the probability of particular model weights for a node being set to 0, the number of epochs (**epoch**), that is to say, the number of iterations of the back-propagation algorithm, and the batch size (**batch_size**), that is the number of samples from the training set used per state update. The model was fitted using the R **keras** library (Kalinowski et al., 2023).

For all three machine learning models the categorical variables were treated as numeric (non-categorical) variables. All categorical variables were ordered by the mean gamma dose rate per group. We also performed additional model fits in which we used the method of coding categorical

399 variables of Breiman (Breiman, 2001), but as these generally yielded inferior model fits, as
400 measured by the MSPE, we do not report these further.

401 Variables which were missing for some measurement locations, such as the LSOA, MSOA,
402 and LAUA property data for Scotland, were coded as zero and were used at their nominal level.
403 For tree-based methods such as RF and GBM these missing values are almost always at the lower
404 extreme of the distribution.

405 ***2.3.2 Geostatistical models***

406 We also employed a spatial geostatistical model similar to that used by Kendall *et al.* (Kendall et
407 al., 2018), which assumed an underlying geologically determined spatial variation, superimposed
408 on which is a stochastic random variable, modelling the empirically observed correlation of dose-
409 rates in nearby houses, more than can be accounted for by purely geological factors (Diggle and
410 Ribeiro, 2007). Specifically, it was assumed that the gamma-ray dose-rate at a given location was
411 given by a sum of:

412 (a) a "mean" process, with distinct values given by a combination of various geological and
413 spatial variables;

414 (b) a stationary and isotropic stochastic spatial correlation process; and

415 (c) a Gaussian "noise" term.

416 The spatial correlation was assumed to result from unobserved variables that would be
417 correlated between nearby areas (with greater correlation the closer the areas are), and that might
418 affect gamma-ray dose-rate. We used the spatial correlation process developed by Matérn (Matérn,
419 1960). The model fitted differed from that of Kendall *et al.* (Kendall et al., 2018) only in that a
420 rather richer set of variables was used to model the mean process; details are given in Table 1.
421 Further details of the model structure are given by Kendall *et al.* (Kendall et al., 2018).

422 For the spatial Gaussian-Matérn model the categorical variables were replaced by a set of
423 binary indicator variables for the various levels, so that a variable with n levels contributed n-1
424 binary indicator variables.

425 ***2.3.3 Nearest neighbour models***

An inverse weighted power of distance model was also used to predict dose rates. The dose-rate at a point p in the test set was estimated based on measurements of dose-rate $\{d_i(p)\}_{i=1}^{N}$ among the N nearest neighbours in the training set, with spatial coordinates $\{x_i(p)\}_{i=1}^{N}$, $i=1,...,N$; an estimate of dose rate based on inverse power q of distance weighting was defined by:

$$D_{est}(p \mid (d_i(p), x_i(p))_{i=1}^{N}) = \sum_{i=1}^{N} d_i(p) / \max[\mid p - x_i(p) \mid, d_{\min}]^q \Big/ \sum_{i=1}^{N} 1 / \max[\mid p - x_i(p) \mid, d_{\min}]^q \quad (2)$$

Special cases include $q=0$ (constant), $q=1$ (inverse distance) and $q=2$ (inverse distance squared). There is inevitably some uncertainty in the spatial resolution of the data, and for this reason and because there are a few distinct points with the same grid coordinates, which would otherwise cause the expression to be indeterminate, the distance between candidate points was limited to some minimum distance, $d_{\min}$, generally taken to be 0.1 km. MSPE was evaluated in the standard way from this predicted dose rate.

For the optimised NN models the parameters N and q were varied to minimise the MSPE. This proved to be for $q=0$. The Fortran 95-2003 code and controlling parameter input files used to control the fit of this model are given in Supplement A.

Modelling based on averaging, including all models considered in this paper, will tend to reduce the range of predictions. However, the marked narrowing observed is probably a result of the limited data we can offer the models. To the extent that such compression of the dose scale is a result of randomisation, then the power of the study will be reduced. To further illuminate what may be going on, we employ a simple ad hoc index of "mixing", which we define as the percentage of predicted values that fall on the opposite side of the mean of the distribution from the observation.

2.3.4 Testing and training sets

For all four types of model, we trained the models on a set of *n*=8046 records. This was close to the *n*=8057 records previously used for training (Kendall et al., 2018), differing only in that 11

duplicate records were removed. The holdout validation test set used in all cases n=2138 records, differing from the n=2142 records previously employed (Kendall et al., 2018) via removal of 4 duplicate records. This training/testing division is, subject to the removal of 15 duplicate records (0.15% of all records), the same as that used in previous publications (Chernyavskiy et al., 2016; Kendall et al., 2018; Kendall et al., 2016b). In these publications it was referred to as a "70/30% split" because it reflected a division into seven and three internal groupings. However, in terms of number of records it is approximately 80/20%. For clarity, in the present publication we refer to it as the "2018 split"

For all machine learning models, we used quadratic (square loss) and evaluated MSPE via 5-fold cross validation in which the set of measurements were randomly divided into quintiles which were used successively as the test set with the remaining 80% used as the training set. The strategy used for each machine learning model was to choose the value of each hyperparameter to minimise the MSPE, holding the other hyperparameters constant; fixing that hyperparameter value one then proceeds to minimise the next hyperparameter. The spatial Gaussian-Matérn model was fitted to each of the 5 cyclically-rotated training sets, minimising all parameters simultaneously via maximum-likelihood, and mean MSPE for the respective test sets evaluated in the standard way.

For all models we compute Spearman and Pearson correlation coefficients between the model predicted and actual dose rate in the holdout test set. We also estimate the concordance correlation coefficient (CCC), a modified version of the coefficient of determination (R^2) (Lin, 1989). In addition to measuring the intensity of correlation, CCC also indicates the extent to which model predictions differ from the X=Y line through the origin. This feature makes it an effective tool for assessing both forecast accuracy and precision. Unlike the Pearson correlation coefficient, the CCC can detect bias in model forecasts.

3 Results

3.1 Main Results

Table 1 summarises the results of the best fitting machine learning (ML) and geostatistical models using the 2018 testing/training split. The three types of ML models used the full set of 157 descriptive variables and categorical variables ordered by mean dose rate per group. It can be seen that GBM models perform slightly better than the other two ML models but that all three ML models have MSPE in the approximate range 340 to 350, corresponding to a standard deviation of 18 or 19 on a mean value of about 95 nGy/h. The MSPE for the Gaussian-Matérn model given here are close to those reported earlier(Kendall et al., 2018). Detailed results relating to the fitting of ML models are presented in Supplement B.

Figure 1 compares the distributions of measured and predicted indoor gamma-ray dose rates at the measurement locations for all main types of model, specifically three ML models, the Gaussian-Matérn model and various instances of *K*-nearest neighbour model. As reported previously (Kendall et al., 2018), the predicted values have a bimodal distribution but are substantially narrower (population SD 12-13) than the measured values (SD 22.6). We have argued that the underlying distribution is probably bimodal (or multimodal) but that inter-house variations obscure this pattern. It is striking that the RF and GBM models produce exceptionally similar predictions (correlation coefficient 0.981). The correlation between RF or GBM and ANN predictions is also high (0.93-0.94) and indeed the correlations between all six optimised models exceed 0.8 (Supplement B Table B15). However, the correlations of all model predictions with the measured doserates are often lower, the CCC being particularly low (e.g. 0.46-0.50 for RF, GBM and ANN, 0.40 for GM) (Table 1).

Figure 2 shows scatter-plots of measured vs predicted dose-rates for the 2018 holdout set for the three ML models and for the Gaussian-Matérn model. The more restricted spread of the predicted values compared to that of the measurements is apparent. Further, bimodality of the predicted distributions can be discerned in most cases. However, the wide distributions of the scatterplots shows that predicted values often fall well above or below the measurements.

We investigated the possibility that most measurements were being fitted well, but that a small minority of "difficult" points was degrading the overall MSPE. No such set of difficult points could be identified.

Table 2 gives results for variations on the Gaussian-Matérn geostatistical model. Three such models are considered, with the mean process taken as 1) a constant; 2) the optimum model found by (Kendall et al., 2018) and 3) the latter with the addition of a term for the concentrations of potassium, uranium and thorium. In every case the MSPE is larger than for the ML Models. The last-mentioned Gaussian-Matérn model produces the lowest MSPE and it is this that is shown in Table 1. However, it is notable that the addition of the radionuclide potentials for potassium, thorium and uranium results in only minor improvements in MSPE.

3.2 Nearest neighbour models

We also considered a number of nearest neighbour models, defined by q, the inverse power of distance, and N, the number of nearest neighbours considered. Table 3 gives detailed results for the fitting of nearest neighbour models, exploring the goodness of fit (MSPE) of the models as q and N are varied. For q=0, an unweighted average of neighbouring measurements, averaging over N=50 nearest neighbours gave best results. For q=1, inverse distance weighting, that for N=100 was a little lower. However, to simplify comparisons we chose N=50 in both cases. For the "local" model, N was chosen to be 10 and q=2 so as to obtain a predicted distribution that was more similar to the set of measurements. We also considered a nearest neighbour model with N=1, i.e., the nearest available measurement is taken as the prediction for an unmeasured location. Although a simple and intuitive option, it gives very poor results. Parameters of the fits for nearest neighbour models corresponding to those of Table 1 are in Table 4. In all cases these results are not competitive with those obtained from ML models. However, the optimised nearest neighbour models have MSPE and correlation coefficients only a little worse than those of the best Gaussian-Matérn Model.

Figure 1 compares the distributions of measured and predicted indoor gamma-ray dose rates at the measurement locations for the four nearest neighbour models. The two optimised models

(q=0, N=50 and q=1, N=50) give very similar predictions (Pearson correlation between the two sets of predictions =0.927). The two other nearest neighbour models (q=2, N=10; N=1) give much broader distributions but with much larger MSPE (Table 4). The N=1 model has much the widest range of any of the predictions, very similar to that of the measurements.

Figure 3 shows scatterplots for the four nearest neighbour models. As with Figure 2, predicted value often fall above or below the measurements. The more compressed ranges of predicted values for the optimised models based on N=50 are similar to the RF, ANN, GBM and Gaussian-Matérn models (see Table 1, Table 4, Figure 2). This contrasts with the wider range of predicted values for the q=2, N=10 model (see Table 4, Figure 3) which also has a measured predicted dose rate trend closer to 1:1 than the N=50 models.

3.3 Calculations using restricted sets of explanatory variables

If the RF calculation of Table 1 is repeated using the 46 explanatory variables available to Kendall *et al* (Kendall et al., 2018), the MSPE is substantially higher, at 367.97 (result not shown), so both the new models and the extra variables are contributing significantly to the improved fits reported here compared with earlier results (Kendall et al., 2018). Likewise, if the GBM model is refitted using the original 46 variables then the test MSPE is much degraded, 363.38, as is also the case for the ANN model, 362.53 (results not shown).

Table 5 gives results for calculations using restricted parameter sets, the 30 parameters which appeared most important, as well as results for the full 157 parameter calculations. The results with the restricted parameter sets are similar to those using the full set, but can result in a slightly lower MSPE.

3.4 Five-fold cross validation

Table 6 compares results for the five components of the five-fold cross validation of the various predictive models. In a five-fold cross-validation, each observation in the available dataset is used for training four times and testing once. The reported cross-validation mean squared prediction error (MSPE) is the average of the MSPEs calculated from the five folds. There is substantial variation in MSPE between quintiles for the three ML models, from less than 330 to nearly 370.

As can be seen, the Gaussian-Matérn model performs consistently poorly in comparison with the three ML models. However, the variation in MSPE between the three ML models for a given holdout quintile is smaller than the variation between the five holdout quintiles for each model. Moreover, the distributions of predicted dose rates are similar for all five quintiles for each of the models (results not shown). The trends for the quintile results for the optimised (N=50) nearest neighbour models in Table 6 generally follow those for the Gaussian-Matérn model. The results for the other two nearest neighbour models are more variable.

Combining the predictions for the holdout sets for the five branches of the cross validation provides predicted values for all of the 10,184 measurement locations. Generalization error metrics such as MSPE, that are calculated using cross-validation allow each of the 10,184 measurements to be used for training and testing. In contrast, a hold-out approach randomly selects a proportion of data for training and the remaining data are held-out for testing. Arguably, cross-validation provides a more accurate measure of model generalizability than a hold-out approach, as each observation is used for model fitting and assessment. These results are shown in Table 6 which also include Pearson correlation coefficients calculated using all 10,184 measurements and also the index of mixing. These other measures of goodness of fit for the different models generally follow the MSPE with MSPE increasing as correlation decreases and mixing increases.

3.5 Relative importance of the explanatory variables

An important question is which of the various explanatory variables are most significant. The tables and discussion in Supplement C give information on this topic. Table 7 offers a simpler overview of the results in terms of the four types of variables: geographical, geological, SES or house characteristics. It can be seen that, for both GBM and RF models, the geographical variables account for 50% or more of the total MSPE reduction, with geological and house characteristic variables contributing very roughly 20%. However, it should be noted that the different types of variable are often correlated, for example it is difficult to separate the influence of geographical and geological variables. There are many more house characteristic variables than other types. Geographical variables contribute ~57% of the fit for RF calculations and slightly less, ~49%, for

582 GBM calculations. Amongst the individual variables, it is striking that much the most effective is
583 the County District, followed by the very similar Local or Unitary Authority code (LAUA) which
584 replaced it. The most important few variables contribute the majority of the MSPE reduction, but
585 significant contributions come from the tail. The importance of County District and LAUA is not
586 just a reflection of the general importance of geographical variables. If the GBM calculations are
587 repeated, using 155 variables excluding County District and LAUA, the MSPE is degraded, with
588 value 346.48 vs 337.44; the geographical variable Postcode District becomes more important, but
589 it is the geological variables that now dominate, contributing 45% of the MSPE reduction
590 (Supplement B Table B14).

591 Table 7 also examines the effect of reducing the number of explanatory variables from the
592 full set of 157 to 30. Two reduced sets were chosen, the 30 most important variables from the full
593 (157 variable) calculations for RF and for GBM models. These two sets are not identical, but
594 overlap considerably, 22 variables being common and with broadly similar order of importance.
595 For the RF calculations, reducing the variable set to the 30 most important from the full 157
596 variables increases the importance of the geographical variables, while the importance of
597 geological variables is little changed. For the corresponding GBM calculations, reducing the
598 parameter set from 157 to just the 30 most important, increases the weight of the geographical
599 parameters from 50% to 55%; again, the relative importance of the geological variables is little
600 changed.

601
602 ## 4. Discussion

603 The purpose of this paper is to investigate improved methods for estimating indoor gamma-ray
604 dose rates at locations where no measurements were made. The new calculations reported here,
605 using a set of just over 10,000 measurements, employ both a greater range of statistical techniques
606 and also a greater variety of explanatory variables than previously (Chernyavskiy et al., 2016;
607 Kendall et al., 2018; Kendall et al., 2016b). Specifically, we now employ three types of machine
608 learning (ML) models in addition to the geostatistical, nearest neighbour and other earlier models.

A number of variables, mostly describing the characteristics of dwellings in the area in question, have been added as explanatory variables. The measurements were divided into a "training" set (used for fitting the model) with the remainder (the "test set") used to estimate the goodness of the fit. The performance of the various models is estimated in terms of the mean square predictive error (MSPE), the mean square of the difference between predicted and measured dose rates on the test set. We also provide correlation coefficients between predicted and measured values. Models with the lowest MSPE tended to have the highest correlations. Calculations are reported using the split into training and test sets used in earlier publications, the "2018 split" (Chernyavskiy et al., 2016; Kendall et al., 2018; Kendall et al., 2016b) and also using five-fold cross validation. Models were generally fitted and their hyperparameters optimised using five-fold cross validation within the training set.

4.1 Main Results

Kendall *et al* (Kendall et al., 2018) reported MSPEs of 401.74 for the optimum Gaussian-Matérn model and 396.41 and 398.04 for two multi-resolution Gaussian process (MRGP) models with 5 and 3 covariates respectively, though this latter class of models subsequently proved to have insuperable numerical difficulties in error estimation. As described in the Introduction, the model of Kendall *et al* (Kendall et al., 2018) that produced the lowest MSPE, of 355.71, the so-called E-OLS model, also had serious statistical shortcomings. Our main results (Table 1) show a pronounced improvement in the goodness of fit (MSPE) over previous calculations using comparable models. All three of the ML models yield MSPEs between 349.67 and 355.65, with the GBM models also equal to or somewhat better than the other two for most of the criteria in Table 1. The best geostatistical Gaussian-Matérn model had an MSPE of about 404.

It is striking that the RF and GBM models produce very similar predicted distributions of dose rates at the measurement locations (Figure 1) with a Pearson correlation coefficient of 0.979. The correlation between RF or GBM and ANN predictions is also high (correlation 0.93-0.94 – see Supplement B Table B15). Nevertheless, these predicted distributions are very much narrower

than the measurements with standard deviations 12-13 for the predictions and 22.6 for the measurements.

To what extent are the improved results over those reported by Kendall *et al*(Kendall et al., 2018) a consequence of better models or of better explanatory variables? These calculations use 157 descriptive variables, a much larger set than the 46 variables that were used in the earlier paper (Kendall et al., 2018). These had a significant effect in improving the fits. For the GBM model with the 2018 testing/training split, the MSPE using the full 157 parameter set was 349.67 (Table 1), compared to 363.38 using just the 46 parameters. Likewise, for the RF and neural network models with the 2018 testing/training split, the MSE using the full 157 variable set was 355.65 and 352.61, respectively (Table 1), compared to 367.97 and 362.53, respectively, using just the 46 parameters. Even the fits using the restricted variable set show improvements over the best geostatistical results of Kendall *et al* (Kendall et al., 2018). Therefore, we conclude that both the new models and the expanded set of variables lie behind the improved fits.

ML models, particularly RF and GBM models, are intrinsically somewhat noisy, in that small changes in the training data or input variables can produce large changes in the resulting models, and hence differences in the variable importances. The ANN models are similarly noisy, resulting from the fact that convergence (via the standard back-propagation algorithm) is only partial, as a result of an iterative approach that is truncated; all fits given here used a fixed number of epochs (=50), because preliminary fits suggested MSPE was not appreciably changed by increasing this.

Our full set of explanatory variables has 157 variables. Many of the variables reflect similar factors resulting in some degree of redundancy. This can be illustrated by MSPE calculations that exclude some variables. For example, using just the most important 30 variables for the GBM model the MSPE was similar to, but slightly lower, 348.47, than that using the full set of variables 349.67. Likewise, using the most important 30 variables for the RF model yielded an MSPE of 354.06, again slightly lower than the MSPE using the full set, 355.65 (Table 5).

The results of Supplement C show that the most important explanatory variables were not exactly the same for different calculations. Nevertheless, there were general indications as to which classes of variables were the most important. Generally, it was the geographical variables that contributed most strongly to the model predictions (Table 4) with geological and house characteristics rather less useful. Amongst the geographical variables the administrative units "County District" and its successor "local or Unitary Authority" were much the most important. Kendall *et al* (Kendall et al., 2016b) found that mean values for County Districts gave the most accurate predictions of any areal or geological mean tested. Perhaps those drawing the boundaries of these administrative units were influenced by factors such as urbanicity that correlate with indoor gamma-ray doses, but the strength of the association is surprising.

It is not possible to divide the variables into subsets of the essential and the redundant. Nevertheless, the RF and GBM models had large overlap in the most influential variables, also the consistency in top-scoring variables when only the 30 most influential variables were used for each of these types of model (Supplement C and Supplement B Table B6, Table B7). One of the strengths of the ML models are their ability to incorporate partially correlated variables. However, the results offer little help in choosing between them. Arguably such a choice is not necessary, as the models perform well with all variables included. However, a useful area of future research would be to identify much reduced set of important variables.

4.2 Five-fold Cross Validation

For the ML models, the MSPEs were computed using five-fold cross-validation with results reported in Table 6. The MSPEs from each fold vary substantially, from less than 330 to more than 370. However, the three ML predictions for each quintile are generally similar, indicating that this is an intrinsic consequence of the nature of the data. Similar conclusions would be drawn about the relative merits of the models whichever quintile was examined

4.3 Nearest neighbour Models

We also considered a number of nearest neighbour models (Table 3, Table 4). The optimised nearest neighbour models have MSPEs in the range 400-410, comparable with the Gaussian-Matérn model. Estimating unknown values as simply the value of the one nearest available measurement yielded very much worse fit (MSPE=738.89).

Elio *et al* (Elío et al., 2023) pointed out that one of the shortcomings of some ML methods is that they "make no explicit consideration of local observations: observations are used for fitting/calibrating the algorithm, but observations close to a prediction location are not used explicitly. That means, close by observations have the same weight as far distant observations [if] they cover the same parameter space." This is, of course, a general feature of many statistical models, although geostatistical models and nearest neighbour models are possibly less affected than many. In order to explore this possibility we investigated a "Local" nearest neighbour inverse distance squared weighted nearest neighbour model where the number of neighbours considered was arbitrarily limited to 10. This produced a SD on the predicted values, rather closer to that of the measurements than that of the ML models (Table 1). However, the results were otherwise disappointing (MSPE for the combined test quintiles 475.4; Pearson correlation coefficient with the predicted values 0.42 compared to 0.57-0.58 for the ML models; index of mixing 33.7%, worse than any other model tested). The model estimating the dose at an unmeasured point as that at the nearest available measurement (N=1) reproduced the width of the measurements very closely. However, in every other respect it performed less well than any other model tested. The poor performance of these models does not invalidate the remarks of Elio *et al* (Elío et al., 2023), but our attempts to exploit their suggestion were unsuccessful. Nearest neighbour models have well known statistical shortcomings, specifically their high noise (when a low number of neighbours are selected) and poor predictive performance (when a large number of neighbours are selected) (Hastie et al., 2017).

4.4 General Observations

Nevertheless, although the general trends in the predicted doses mirror those in the measurements, particularly for the three ML models (Figure 2) none of the models produced close fits to the

714 measured points despite the large number of explanatory variables and the variety of models
715 (Figure 2, Figure 3). It seems that close fits to the measured points require methods, or perhaps
716 more probably, types of explanatory variables, that are not available to us. Although, as noted
717 below, some kinds of imperfect fits would not necessarily weaken the power of epidemiological
718 studies using modelled doses, it is unlikely that the results reported above are of this kind. As
719 shown by the index of mixing, even the best models predict over a quarter of the test points to lie
720 on the wrong side of the mean.

721 Very similar conclusions were drawn by Peterman *et al.* (Petermann et al., 2021), in a study
722 of machine learning methods for radon mapping. Peterman *et al.* (Petermann et al., 2021) noted
723 that the distribution of the predicted radon concentrations is narrower compared to the observed
724 values and that the best linear fit to the data was much shallower than 1:1. These characteristics
725 are also seen in our Figure 2 and Figure 3. Some compression of range is an unavoidable feature
726 of any statistical model, as all will to some extent average data over groups. However, here the
727 coefficient of variation of the predicted dose rates is about half that of the measurements.

728 The compression of the range and the induced error structure (see Supplement D) in the
729 predicted indoor gamma-ray dose rates has implications for epidemiological studies of this
730 component of natural background radiation and childhood cancer. Power calculations for studies
731 of childhood cancer and natural gamma rays (Little et al., 2010) have been widely cited (Hauri et
732 al., 2013; Nikkilä et al., 2016; Spix et al., 2017; Spycher et al., 2015). These power calculations
733 used two indoor gamma-ray dose distributions: that of the individual house measurements,
734 effectively the same as that in Figure 1, and that of means for County Districts, very much narrower
735 with SD 13 rather than 23 nGy/h and thus broadly similar to the fits in Table 1. The power
736 calculations for individual-based studies, cohort and case/control, used the broader distribution of
737 the individual measurements.

738 Epidemiological studies depend on differences in risk between those receiving higher doses
739 and those receiving lower doses. As the dose distribution narrows, the differences become smaller
740 and power typically decreases. It is true that such a compression would not necessarily result in

tests of difference between cases and controls being statistically less powerful. If every dose were multiplied by a factor of $k<1$ then tests for trend will produce exactly the same p-values as before, but the slope will be biased upwards by a factor $1/k$, thereby possibly increasing power, depending on the alternative hypothesis being tested. Such a distribution would have correlation coefficient approaching unity with respect to the observed values and a linear scatterplot. The scatter plots in Figure 2 and Figure 3, the correlation coefficients between measured and predicted values and the index of mixing (Table 1, Table 4, Table 6) show that we do not have this situation for any of our candidate models. However, if the compression of the dose scale is a result of induced random error, then the power of the study will almost certainly be reduced.

In general, classical measurement error will result in loss of power (Carroll et al., 2006; Tosteson et al., 2003), as also in most cases will Berkson error (Carroll et al., 2006; Stram and Kopecky, 2003; Zhang et al., 2017). However, if a Berkson error is assumed and one performs tests of the regression coefficient for departures from 0 the power is unchanged (Stram and Kopecky, 2003). It is not clear in our case what the form of measurement error is, but the gamma measurements themselves, used to fit the dose data, will have a small amount of classical error, while the various types of regression are likely to induce a mixed Berkson and classical type of error model. This question is examined in detail in Supplement D. It is likely that this would be associated with reduction in power.

The distribution of doses to study subjects will be the same as that of the house measurements if they are the result of direct measurements in the dwellings in question or if they are the result of modelling which fully reproduces the range of the distribution. However, the predictions reported here produce narrower distributions than those of direct measurements. Analysis of epidemiological data using these predicted doses will almost certainly have lower power than would be predicted by naïve application of the results of Little *et al* (Little et al., 2010).

4.5 Approaches by other researchers

A number of other papers in the literature have also reported modelled gamma-ray dose rates. Some excluded the contribution from directly ionising cosmic rays. At low altitudes this

768 contribution is often reasonably constant and to the extent that this is the case, it will make little
769 difference to the modelling of indoor dose-rates or to the analysis of possible associations between
770 radiation exposure and childhood cancer whether cosmic ray doses are included with telluric
771 gamma rays or treated separately. Where cosmic ray doses differ significantly between study
772 subjects and can be estimated, this should of course be done. None of the other studies, to our
773 knowledge have tested an extensive range of models in the way that we have. However, in some
774 cases estimates were of outdoor rather than indoor dose rates (Folly et al., 2021; Spycher et al.,
775 2015). For purposes of epidemiology this has the substantial disadvantage that indoor exposures
776 are generally greater than outdoor. The two are correlated, but far from perfectly. Nevertheless,
777 outdoor dose rates may be easier to predict because the important, but very variable and elusive
778 contribution from building materials and shielding are not included. The importance of building
779 materials has been generally recognised (Chernyavskiy et al., 2016; Folly et al., 2022; Kendall et
780 al., 2018; Kendall et al., 2016b; Nikkilä et al., 2016; Warnery et al., 2015).

781 Folly *et al* (Folly et al., 2021) updated the estimates of Rybach *et al* (Rybach et al., 2002) of
782 outdoor gamma-ray dose rates in Switzerland using more extensive data and more sophisticated
783 analytical methods. Folly *et al* (Folly et al., 2021) used detailed airborne gamma-ray spectrometry
784 data with geological and land coverage information via Bayesian spatial modelling. In a separate
785 study Folly *et al* (Folly et al., 2022) made direct short-term measurements of children's gamma-
786 ray exposures using personal dosemeters. Information on children's movements were also
787 collected, together with data on building materials and other variables. Folly et al concluded that
788 geographic variation of outdoor doses and construction materials of dwellings accounts for a
789 considerable proportion of the variability in children's exposure to background gamma radiation
790 in Switzerland.

791 Nikkilä *et al.* (Nikkilä et al., 2016) estimated indoor gamma ray dose rates in Finland using
792 a map of outdoor dose rates from a nationwide mobile survey (Arvela et al., 1995). Cosmic ray
793 doses were estimated by formula and subtracted from the measurements. An estimate of doses
794 from Chernobyl fallout was included. Conversion to indoor dose rates was based on indoor dose

rate measurements in 346 dwellings coupled with information on house characteristics from national datasets. Different conversion factors were used for houses and apartments. Nevertheless, this is a fairly limited sample and there are inevitably limitations in the conversion.

Warnery *et al.* (Warnery et al., 2015) estimated indoor exposures to gamma-rays and cosmic rays in France as the average dose rate in 1 x 1 km squares. The gamma-ray components were estimated using multi-collocated cokriging. This made use of 97,595 measurements in 17,404 dentist surgeries and veterinary clinics and also a map of geogenic uranium potential (as used for radon estimation) (Ielsch et al., 2017; Ielsch et al., 2010). Ielsch *et al* (Ielsch et al., 2017; Ielsch et al., 2010) investigated the uranium bedrock concentration in generalised units on the 1:1,000,000 geological map of France. The main source of information was 5092 geolocated rock analyses plus some hundred non-geolocated analyses and various other sources of information. The geological units were finally assigned to five categories of "uranium potential" based on their mean uranium content. The introduction of the uranium potential data resulted in a rather modest improvement in the mean square error from 409 to 407 $(nSv/h)^2$ compared to ordinary kriging using the same measurement set. Neither the uranium potential of Ielsch *et al* (Ielsch et al., 2017; Ielsch et al., 2010), based on consideration of bedrock at a scale of 1:1,000,000, nor the much more detailed equivalent used here (Appleton and Kendall, 2022) proved very effective as a predictive tool for indoor gamma-ray dose rates.

We were unable to test exactly the same methodology as Warnery *et al* (Warnery et al., 2015) because the details of the statistical models used were incompletely specified, but our closest equivalent is the Gaussian-Matérn model with uranium potential included amongst the parameters used to model the mean process. As shown in Table 1, this gives results inferior to those of the three ML models.

5 Conclusions

In this work, implementing several machine learning models and incorporating a greater variety of explanatory variables we achieved significant improvement in the modelling of indoor gamma-ray dose rates. The machine learning models outperformed other models that we have previously

explored in terms of prediction accuracy. The intrinsic noisiness of machine learning models causes some disquieting characteristics, in particular lack of stability in importance of particular variables. Nevertheless, the fitted machine learning models identified similar general classes variables as being most relevant in the prediction of indoor gamma dose rates.

However, the predicted dose distributions are substantially narrower than the observed distribution and may include a significant component of classical error, which will bias trends in dose. As a result, epidemiological analyses involving predicted gamma dose rates will likely have decreased statistical power compared to analyses involving direct measurements.

More accurate predictions could be made if more directly relevant explanatory variables, particularly details of individual house construction and the radionuclide content of building materials, were available.

6 Competing interests

The authors declare that they have no competing interests.

Table 1. Mean square prediction error (MSPE) with various other statistics for measured *vs* predicted indoor gamma-ray dose rates (nGy/h) for different models using all 157 descriptive variables and the 2018 testing/training split

Model type	MSPE	RMSPE SD on predictive error (nGy/h)	Spearman correlation coefficient	Pearson correlation coefficient	Concordance correlation coefficient (95% CI)	Coefficient of predicted *vs* actual dose via linear model (predicted dose ~ actual dose), with constant term = 0 (95% CI)	Range (nGy/h)	SD (nGy/h)
ndom Forest (RF) model	355.65	18.9	0.583	0.572	0.480 (0.454, 0.505)	0.960 (0.952, 0.968)	(61.5, 135.8)	12.488
rtificial Neural network (ANN) models	352.61	18.8	0.594	0.585	0.463 (0.438, 0.487)	0.926 (0.919, 0.934)	(67.4, 139.6)	11.657
adient boosting machine (GBM) models	349.67	18.7	0.592	0.582	0.497 (0.471, 0.522)	0.961 (0.953, 0.969)	(58.6, 155.3)	12.939
ssian-Matérn (GM) model[a]	403.86	20.1	0.511	0.499	0.395 (0.368, 0.421)	0.937 (0.928, 0.945)	(64.6, 129.5)	11.414

Notes
Random Forest and Gradient Boosting Machine models use tuning parameters from optimal models
MSPE = Test error evaluated with 2018 split using training data to train models
RMSPE = Root mean squared predictive error
Spearman rank correlation coefficient for predicted vs measured dose rate
Pearson correlation coefficient for predicted vs measured dose rate
Concordance correlation coefficient (Lin, 1989) for predicted vs measured dose rate
Coefficient predicted vs actual dose = Coefficient of predicted dose as function of actual dose via linear model (predicted dose ~ actual dose), with constant term = 0 (+95% CI)
SD = Standard Deviation within predicted values
[a]Optimal Gaussian-Matérn model from Kendall *et al* (Kendall et al., 2018) - Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + 1981 urban-rural (6-level) + ln[external gamma dose rate] + 50K-BEDSUP Surface (23-level) + ln[dose from potassium] +ln[dose from uranium] +ln[dose from thorium]

Table 2 Mean square prediction error (MSPE) for Gaussian-Matérn geostatistical model

Records with spatially coincident points are jittered, applying displacements using random uniform numbers in the interval [-0.5, 0.5] m.)

Model description	Parameters for Matérn covariance	Other model parameters	MSPE within 2018 split training data	MSPE within 2018 split holdout validation data	Mean absolute error (training data)	Mean absolute error (validation data)
Constant	2	1	397.26	412.99	15.456	15.334
Model 1[a]	2	36	395.65	403.94	15.406	15.167
Model 2[b]	2	39	395.64	403.86	15.384	15.164

Notes

[a]Model 1 : Optimal Gaussian-Matérn model from Kendall *et al* (Kendall et al., 2018) - Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + 1981 urban-rural (6-level) + ln[external gamma dose rate] + 50K-BEDSUP Surface (23-level)

[b]Model 2 : Optimal Gaussian- Matérn model from Kendall *et al* (Kendall et al., 2018) - Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + 1981 urban-rural (6-level) + ln[external gamma dose rate] + 50K-BEDSUP Surface (23-level) + ln[dose from potassium] +ln[dose from uranium] +ln[dose from thorium]

Table 3 Nearest Neighbour model fitting: Mean square prediction error (MSPE) evaluated via test quintile (Q) left out as a function of *q* (inverse power of distance) and the number of nearest neighbours. In each panel the lowest MSPE for the 2018 split is in bold and underlined

***q*=0**							
	Number of nearest neighbours						
Holdout	1	5	10	20	30	50	100
Q1	707.41	434.18	393.06	382.72	372.97	370.60	374.06
Q2	694.39	438.39	405.46	402.28	397.15	400.62	404.20
Q3	695.46	434.73	412.21	409.28	411.41	409.82	407.70
Q4	712.54	458.25	435.70	429.78	425.24	424.03	428.56
Q5	663.86	419.57	385.52	378.72	372.54	368.46	372.89
2018 Split	738.89	460.21	426.06	413.57	408.63	**407.55**	410.54
***q*=1**							
	Number of nearest neighbours						
Holdout	1	5	10	20	30	50	100
Q1	707.41	459.62	414.07	394.17	383.79	374.73	369.35
Q2	694.39	449.05	414.07	399.91	392.72	390.08	388.10
Q3	695.46	448.86	420.78	409.66	406.11	401.00	396.26
Q4	712.54	477.15	448.15	432.34	425.85	420.54	418.22
Q5	663.86	441.68	401.56	384.52	377.06	370.00	366.93
2018 split	738.89	483.34	445.04	425.76	416.97	410.08	**404.27**
***q*=2**							
	Number of nearest neighbours						
Holdout	1	5	10	20	30	50	100
Q1	707.40	511.69	479.41	464.43	457.67	450.30	443.84
Q2	694.39	492.15	464.32	449.98	443.15	437.71	432.23
Q3	695.46	493.10	470.89	460.57	456.06	450.60	444.91
Q4	712.54	522.88	501.53	489.52	484.68	479.55	474.71
Q5	663.86	487.39	457.30	444.01	438.25	432.42	427.29
2018 split	738.89	532.88	505.68	491.79	485.57	479.93	**473.39**

Table 4 Mean square prediction error (MSPE) with various other statistics for measured *vs* predicted indoor gamma-ray dose rates (nGy/h) for different nearest neighbour models using the 2018 testing/training split

Nearest neighbour Model type	MSPE	RMSPE SD on predictive error (nGy/h)	Spearman correlation coefficient	Pearson correlation coefficient	Coefficient of predicted *vs* actual dose via linear model (predicted dose ~ actual dose), with constant term = 0 (95% CI)	Range (nGy/h)	SD (nGy/h)
Optimised A (q=0, N=50)	407.55	20.2	0.490	0.478	0.961 (0.953, 0.970)	(66.9, 120.8)	11.478
Optimised B (q=1 N=50)	410.08	20.3	0.488	0.478	0.963 (0.955, 0.972)	(56.7, 136.9)	12.538
Single neighbour (N=1)	738.89	27.2	0.318	0.295	0.969 (0.958, 0.981)	(35.2, 210.7)	22.799
Local (q=2 N=10)	505.68	22.5	0.420	0.398	0.967 (0.957, 0.976)	(37.6, 167.7)	16.975

Notes
MSPE = Test error evaluated with 2018 split
RMSPE = Root mean squared predictive error
Spearman rank correlation coefficient for predicted vs measured dose rate
Pearson correlation coefficient for predicted vs measured dose rate
Coefficient predicted vs actual dose = Coefficient of predicted dose as function of actual dose via linear model (predicted dose ~ actual dose), with constant term = 0 (+95% CI)
SD = Standard Deviation within predicted values

Table 5 Mean square prediction error (MSPE) evaluated using all variables and also using the top 30 optimal variables for random forest (RF) and stochastic gradient boosting machine (GBM) models with the 2018 testing/training split. *MSPE evaluated using all variables, also top 30 optimal variables (in relation to variable importance metric) for random forest (RF) and stochastic gradient boosting machines (GBM) models, using tuning parameters from optimal RF and GBM models (with mtry=15 for the 30 variable models)*

Variable	Top 30 random forest model variables	Top 30 gradient boosting machine model variables	All variables
Random forest models			
5-fold cross validated MSPE evaluated within training set	344.14	344.52	345.20
MSPE evaluated with the 2018 test data, using the remaining data to train models	354.06	354.69	355.65
Gradient boosting machine models			
5-fold cross validated MSPE evaluated within training set	338.09	337.15	337.44
MSPE data to train models	348.14	348.47	349.67

Table 6 Mean square prediction error for predictive models evaluated via test quintile left out from five-fold validation. Total MSPE and the Pearson correlation coefficient with measured doserate are calculated over all 10,184 points, The mixing statistic is the percentage of predicted values on the opposite side of the mean of the distribution from the measured value.

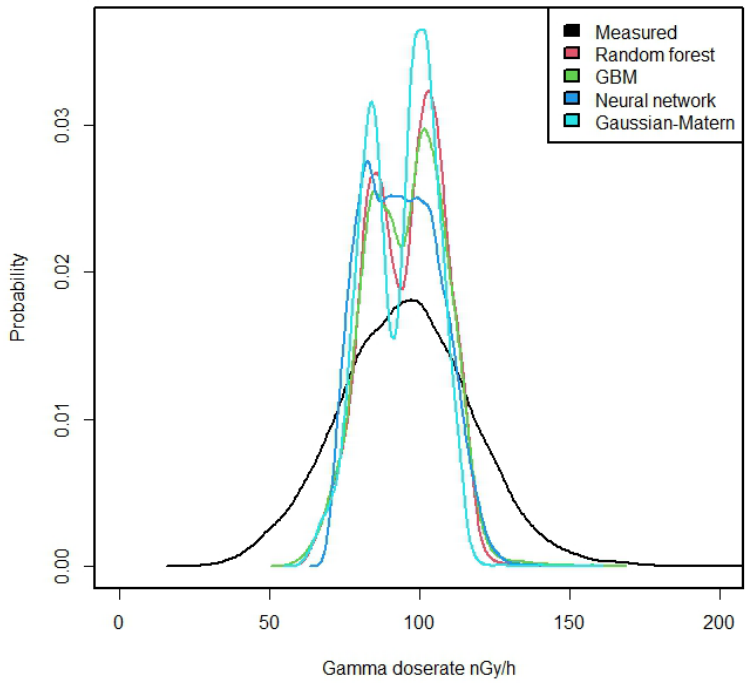
Holdout Quintile	Mean Square Prediction Error							
	RF	GBM	Artificial neural network	Gaussian-Matérn	Nearest neighbour q=0, N=50	Nearest neighbour q=1, N=50	Nearest neighbour q=2, N=10	Nearest neighbour q=1, N=1
1	333.9	334.5	336.1	373.8	370.6	374.7	479.4	707.4
2	342.3	336.5	357.4	403.1	400.6	390.1	464.3	694.4
3	354.0	346.9	356.1	397.8	409.8	401.0	470.9	695.5
4	370.1	362.0	370.5	418.9	424.0	420.5	501.5	712.5
5	329.1	323.7	335.0	367.1	368.5	370.0	457.3	663.9
Total MSPE	345.8	340.7	351.0	392.1	394.7	391.3	474.7	694.7
Pearson correlation coefficient with measured doserate	0.5709	0.5793	0.5661	0.4932	0.4808	0.4904	0.4191	0.3221
Mixing statistic	28.0	28.0	28.5	30.6	30.2	30.4	33.7	38.1

Table 7 Breakdown of variable importance (MSPE reduction) by type of parameter for RF and for GBM calculations with 157 and with 30 parameters. Calculations use the 2018 testing/training split. The 30 most important parameters are selected on the basis of calculations using the full 157 parameter set. Results are compared for GBM and for RF calculations using (a) all 157 parameters (b) the 30 most important parameters from the RF calculations and (c) the 30 most important parameters from the GBM calculations

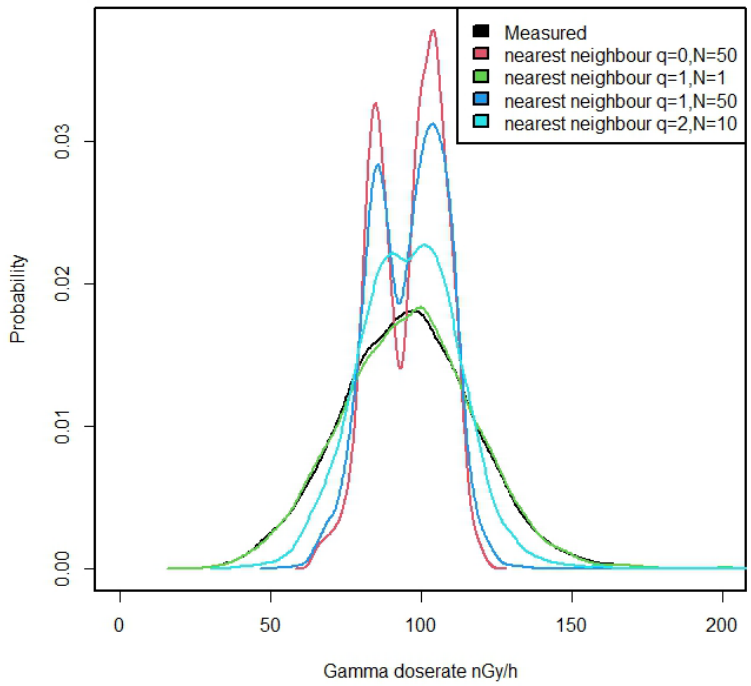
	Results for GBM Model			Results for RF Models		
Type of parameter	All 157	Top 30 RF	Top 30 GBM	All 157	Top 30 RF	Top 30 GBM
Geography	49%	60%	54%	57%	62%	62%
Geology	22%	22%	25%	20%	20%	21%
Socioeconomic status	7%	3%	6%	5%	3%	4%
House Characteristics	22%	16%	15%	18%	15%	13%
Total	100%	100%	100%	100%	100%	100%

Figure 1. Observed doserate and model predicted doserate for (a) main optimal machine learning models and Gaussian-Matérn model and (b) four example nearest neighbour models. All are derived using kernel density smoothing in R with default kernel selection (Sheather and Jones, 1991).

Dose rate observed and optimal random forest, gradient boosting machine, neural network, Gaussian-Matérn predicted

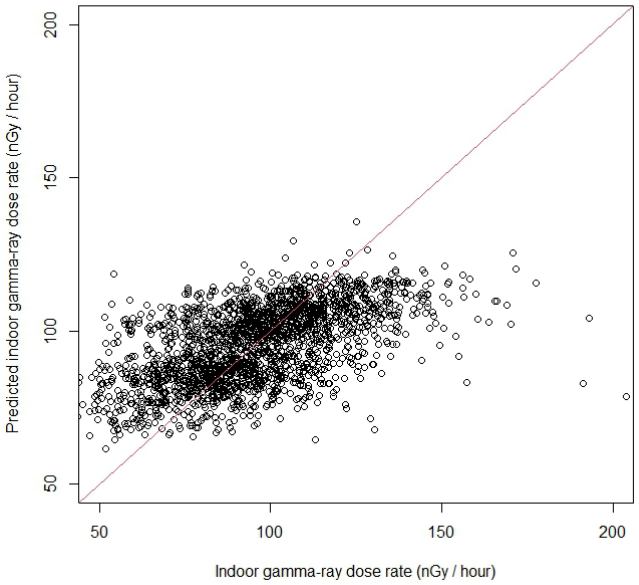


Dose rate observed and predicted by four nearest neighbour models (q=0+N=50, q=1+N=1, q=1+N=50, q=2+N=10)

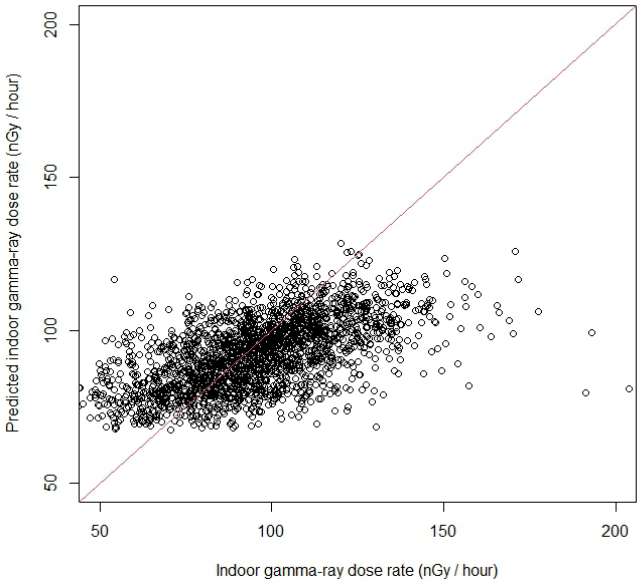


913 **Figure 2. Optimal model predicted vs observed indoor gamma dose rate for RF, neural**
914 **network, GBM, Gaussian-Matérn models**
915

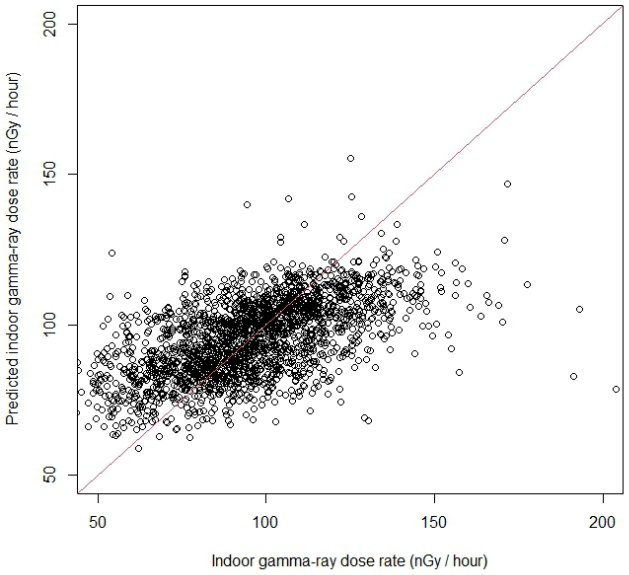
RF



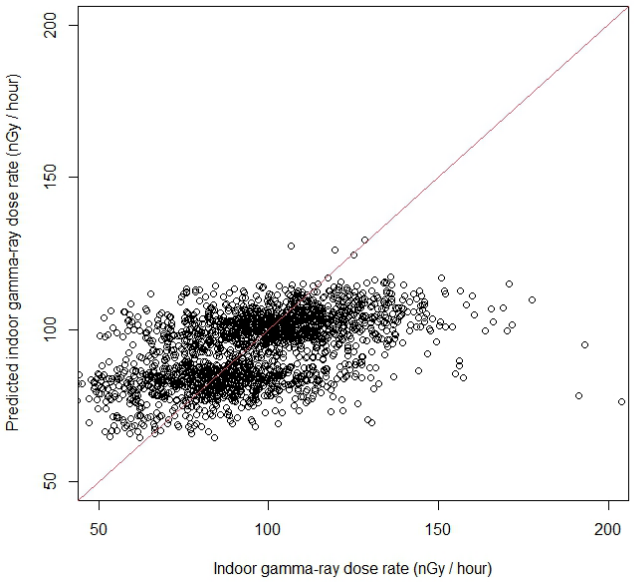
Neural network



GBM



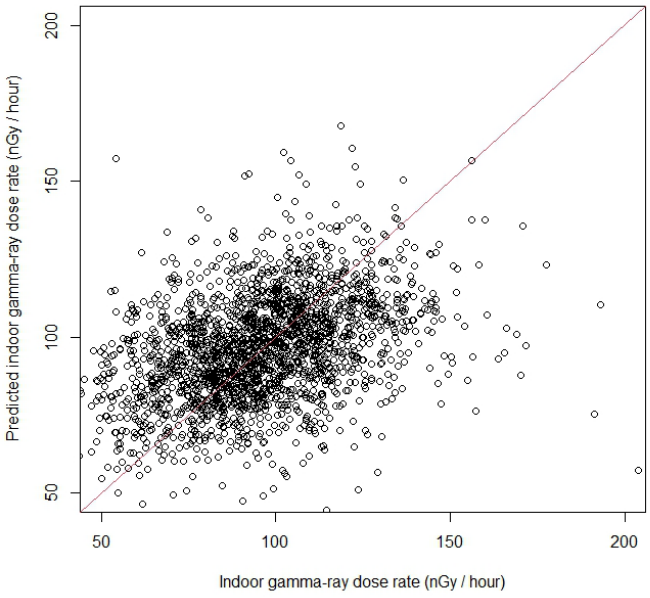
Gaussian-Matérn



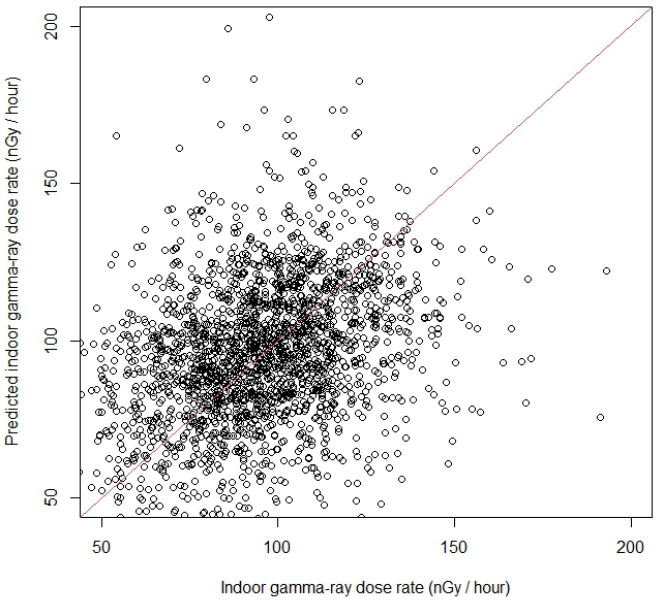
916

Figure 3. Optimal model predicted vs observed indoor gamma dose rate for various types of power of distance weighted nearest neighbour models

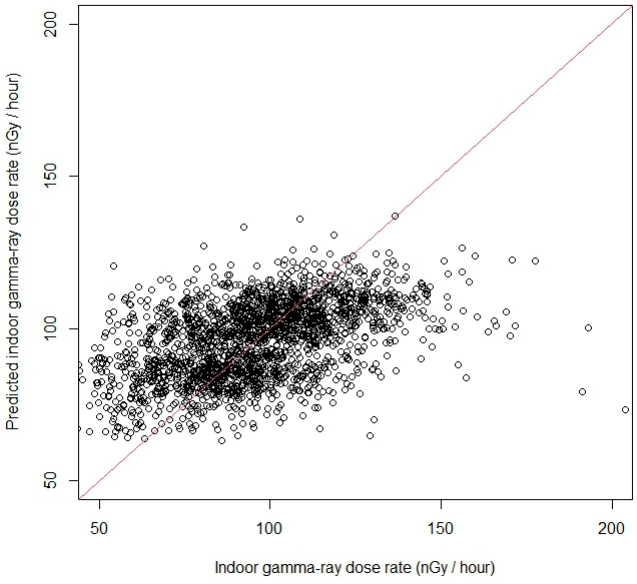
q=2, N=10



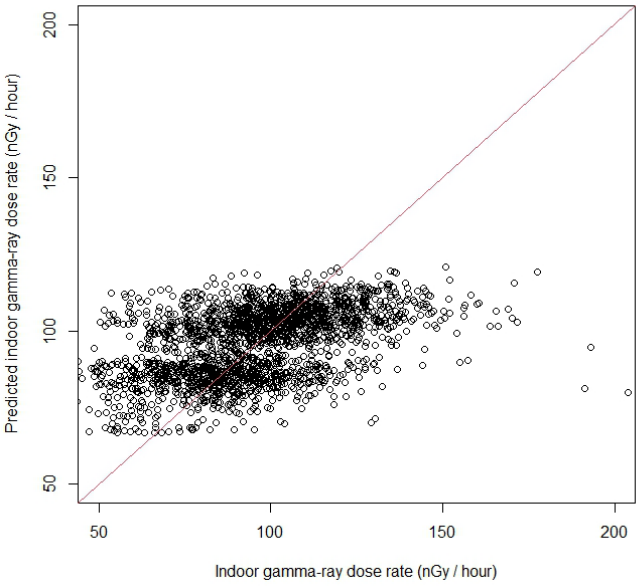
q=1, N=1



q=1, N=50



q=0, N=50



References

Appleton, J.D., 2005. Simplified geological classification for radon potential mapping in England and Wales (based on DiGMapGB-50 V3.12). British Geological Survey, Natural Environment Research Council, Keyworth, UK.

Appleton, J.D., Cave, M.R., 2018. Variation in soil chemistry related to different classes and eras of urbanisation in the London area. Applied Geochemistry 90, 13-24.

Appleton, J.D., Kendall, G.M., 2022. Gamma-radiation levels outdoors in Great Britain based on K, Th and U geochemical data. Journal of environmental radioactivity 251-252, 106948.

Arsham, A., Rosenberg, P., Little, M., 2023. Effects of stopping criterion on the growth of trees in regression random forests. N. Engl. J. Statist. Data Sci. 1, 46-61.

Arvela, H., Hyvönen, H., Lemmelä, H., Castrén, O., 1995. Indoor and Outdoor Gamma Radiation in Finland. Radiation Protection Dosimetry 59, 25-32.

Berlivet, J., Hémon, D., Cléro, E., Ielsch, G., Laurier, D., Faure, L., Clavel, J., Goujon, S., 2021. Residential exposure to natural background radiation at birth and risk of childhood acute leukemia in France, 1990-2009. J. Environ. Radioact. 233, 106613.

Berlivet, J., Hemon, D., Clero, E., Ielsch, G., Laurier, D., Guissou, S., Lacour, B., Clavel, J., Goujon, S., 2020. Ecological association between residential natural background radiation exposure and the incidence rate of childhood central nervous system tumors in France, 2000-2012. J. Environ. Radioact. 211, 106071.

Breiman, L., 1984. Classification and Regression Trees. Routledge, New York.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123-140.

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

Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement error in nonlinear models. A modern perspective. Chapman and Hall/CRC, Boca Raton, FL.

Carstairs, V., Morris, R., 1991. Deprivation and health in Scotland. Aberdeen University Press, Aberdeen.

Chernyavskiy, P., Kendall, G.M., Wakeford, R., Little, M.P., 2016. Spatial prediction of naturally occurring gamma radiation in Great Britain. Journal of environmental radioactivity 164, 300-311.

Demoury, C., Marquant, F., Ielsch, G., Goujon, S., Debayle, C., Faure, L., Coste, A., Laurent, O., Guillevic, J., Laurier, D., Hemon, D., Clavel, J., 2017. Residential exposure to natural background radiation and risk of childhood acute leukemia in France, 1990-2009. Environ. Health Perspect. 125, 714-720.

Deng, H., Runger, G., 2013. Gene selection with guided regularized random forest. Pattern Recognition 46, 3483-3489.

Diggle, P.J., Ribeiro, P.J., Jr, 2007. Model-based geostatistics, Springer Series in Statistics. Springer, 233 Spring Street, New York, NY 10013, pp. 1-228.

Elío, J., Petermann, E., Bossew, P., Janik, M., 2023. Machine learning in environmental radon science. Appl Radiat Isot 194, 110684.

Folly, C.L., Konstantinoudis, G., Mazzei-Abba, A., Kreis, C., Bucher, B., Furrer, R., Spycher, B.D., 2021. Bayesian spatial modelling of terrestrial radiation in Switzerland. Journal of environmental radioactivity 233, 106571.

Folly, C.L., Mazzei-Abba, A., Coste, A., Kreis, C., Spycher, B.D., 2022. Measurements and determinants of children's exposure to background gamma radiation in Switzerland. Journal of Radiation Research 63, 354-363.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Statist. 29, 1189-1232.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367-378.

Green, B.M.R., Lomas, P.R., Bradley, E.J., Wrixon, A.D., 1989. Gamma-radiation levels outdoors in Great Britain. National Radiological Protection Board (NRPB), Chilton, Didcot, Oxon, OX11 0RQ, pp. 1-39.

Greenwell, B., Boehmke, B., Cunningham, J., 2020. gbm: Generalized Boosted Regression Models. Version 2.1.8., 2.1.8. ed. CRAN - The Comprehensive R Archive Network.

Hastie, T., Tibshirani, R., Friedman, J., 2017. The elements of statistical learning. Data mining, inference, and prediction. 2nd edition (corrected at 12th printing). Springer, New York.

Hauri, D., Spycher, B., Huss, A., Zimmermann, F., Grotzer, M., von der Weid, N., Weber, D., Spoerri, A., Kuehni, C.E., Röösli, M., 2013. Domestic radon exposure and risk of childhood cancer: A prospective census-based cohort study. Environ. Health Perspect. 121, 1239-1244.

Hothorn, T., 2021. partykit. Version 1.2-15, 1.2-15 ed. CRAN - The Comprehensive R Archive Network.

Ielsch, G., Cuney, M., Buscail, F., Rossi, F., Leon, A., Cushing, M.E., 2017. Estimation and mapping of uranium content of geological units in France. Journal of environmental radioactivity 166, 210-219.

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. Journal of environmental radioactivity 101, 813-820.

International Commission on Radiological Protection (ICRP), 2021. Use of dose quantities in radiological protection. ICRP publication 147. Ann. ICRP 50, 1-82.

Ishwaran, H., Kogalur, U.B., 2020. randomForestSRC. Version 2.9.3, 2.9.3 ed. CRAN - The Comprehensive R Archive Network.

Kalinowski, T., Falbel, D., Allaire, J.J., Chollet, F., Tang, Y., van der Bijl, W., Studer, M., Keydana, S., 2023. keras: R interface to keras. Version 2.11.1, 2.11.1 ed. CRAN - The Comprehensive R Archive Network.

Kendall, G.M., Chernyavskiy, P., Appleton, J.D., Miles, J.C.H., Wakeford, R., Athanson, M., Vincent, T.J., McColl, N.P., Little, M.P., 2018. Modelling the bimodal distribution of indoor gamma-ray dose-rates in Great Britain. Radiat. Environ. Biophys. 57, 321-347.
Kendall, G.M., Little, M.P., Wakeford, R., Bunch, K.J., Miles, J.C.H., Vincent, T.J., Meara, J.R., Murphy, M.F.G., 2013. A record-based case-control study of natural background radiation and the incidence of childhood leukaemia and other cancers in Great Britain during 1980-2006. Leukemia 27, 3-9.
Kendall, G.M., Miles, J.C., Rees, D., Wakeford, R., Bunch, K.J., Vincent, T.J., Little, M.P., 2016a. Variation with socioeconomic status of indoor radon levels in Great Britain: The less affluent have less radon. Journal of environmental radioactivity 164, 84-90.
Kendall, G.M., Wakeford, R., Athanson, M., Vincent, T.J., Carter, E.J., McColl, N.P., Little, M.P., 2016b. Levels of naturally occurring gamma radiation measured in British homes and their prediction in particular residences. Radiat Environ Biophys 55, 103-124.
Kendall, G.M., Wakeford, R., Bunch, K.J., Vincent, T.J., Little, M.P., 2015. Residential mobility and associated factors in relation to the assessment of exposure to naturally occurring radiation in studies of childhood cancer. Journal of radiological protection : official journal of the Society for Radiological Protection 35, 835-868.
Law, G.R., Smith, A.G., Roman, E., 2002. The importance of full participation: lessons from a national case-control study. Br.J.Cancer 86, 350-355.
Liaw, A., 2018. randomForest: Breiman and Cutler's Random Forests for classification and regression. Version 4.6-14., 4.6-14 ed. CRAN - The Comprehensive R Archive Network.
Lin, L.I.K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45, 255-268.
Little, M.P., Rosenberg, P.S., Arsham, A., 2022. Alternative stopping rules to limit tree expansion for random forest models. Sci. Rep. 12, 15113.
Little, M.P., Wakeford, R., Lubin, J.H., Kendall, G.M., 2010. The statistical power of epidemiological studies analyzing the relationship between exposure to ionizing radiation and cancer, with special reference to childhood leukemia and natural background radiation. Radiat. Res. 174, 387-402.
Matérn, B., 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations., Meddelanden fran Statens Skogsforsknings institut. Statens Skogsforsknings institut, Stockholm, pp. 1-144.
Mazzei-Abba, A., Folly, C.L., Kreis, C., Ammann, R.A., Adam, C., Brack, E., Egger, M., Kuehni, C.E., Spycher, B.D., 2021. External background ionizing radiation and childhood cancer: update of a nationwide cohort analysis. J. Environ. Radioact. 238-239, 106734.
McCulloch, W.S., Pitts, W.R., III, 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115-133.
Miles, J.C.H., Appleton, J.D., 2005. Mapping variation in radon potential both between and within geological units. J. Radiol. Prot. 25, 257-276.
National Council on Radiation Protection and Measurements (NCRP), 2018. Implications of recent epidemiologic studies for the linear-nonthreshold model and radiation protection. NCRP Commentary no 27. National Council on Radiation Protection and Measurements (NCRP), Bethesda, MD, USA.
Nikkilä, A., Arvela, H., Mehtonen, J., Raitanen, J., Heinäniemi, M., Lohi, O., Auvinen, A., 2020. Predicting residential radon concentrations in Finland: model development, validation, and application to childhood leukemia. Scand. J. Work Environ. Health 46, 278-292.
Nikkilä, A., Erme, S., Arvela, H., Holmgren, O., Raitanen, J., Lohi, O., Auvinen, A., 2016. Background radiation and childhood leukemia: A nationwide register-based case-control study. Int. J. Cancer 139, 1975-1982.
Office for National Statistics (ONS), 2024. National Statistics Postcode Lookup File 2019 (https://www.ons.gov.uk/methodology/geography/geographicalproducts/namescodesandlookups/lookups). Office for National Statistics, UK.
Petermann, E., Meyer, H., Nussbaum, M., Bossew, P., 2021. Mapping the geogenic radon potential for Germany by machine learning. Science of The Total Environment 754, 142291.
Rybach, L., Bachler, D., Bucher, B., Schwarz, G., 2002. Radiation doses of Swiss population from external sources. Journal of environmental radioactivity 62, 277-286.
Sheather, S.J., Jones, M.C., 1991. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. Journal of the Royal Statistical Society. Series B (Methodological) 53, 683-690.
Smith, A., 2013. Digital Geological Map of Great Britain, information notes, 2013. British Geological Survey Open Report OR/13/007. British Geological Survey, Keyworth, Nottingham, UK, pp. 1-54.
Spix, C., Grosche, B., Bleher, M., Kaatsch, P., Scholz-Kreisel, P., Blettner, M., 2017. Background gamma radiation and childhood cancer in Germany: an ecological study. Radiat Environ Biophys 56, 127-138.
Spycher, B.D., Lupatsch, J.E., Zwahlen, M., Röösli, M., Niggli, F., Grotzer, M.A., Rischewski, J., Egger, M., Kuehni, C.E., Swiss Pediatric Oncology, G., Swiss National Cohort Study, G., 2015. Background ionizing radiation and the risk of childhood cancer: a census-based nationwide cohort study. Environ Health Perspect 123, 622-628.
Stamp, L.D., 1931. The Land Utilization Survey of Britain. The Geographical Journal 78, 40-47.
Stram, D.O., Kopecky, K.J., 2003. Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: some observations. Radiat. Res. 160, 408-417.

Tosteson, T.D., Buzas, J.S., Demidenko, E., Karagas, M., 2003. Power and sample size calculations for generalized regression models with covariate measurement error. Statistics in Medicine 22, 1069-1082.
United Kingdom Childhood Cancer Study Investigators, 2002a. The United Kingdom Childhood Cancer Study of exposure to domestic sources of ionising radiation: 1: radon gas. Br. J. Cancer 86, 1721-1726.
United Kingdom Childhood Cancer Study Investigators, 2002b. The United Kingdom Childhood Cancer Study of exposure to domestic sources of ionising radiation: 2: gamma radiation. Br. J. Cancer 86, 1727-1731.
United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 1977. Sources and effects of ionizing radiation. 1977 report to the General Assembly, with annexes. Annex B. Natural sources of radiation. United Nations, New York, pp. 35-114.
United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2000. Sources and effects of ionizing radiation. UNSCEAR 2000 report to the General Assembly, with scientific annexes. Volume I: Sources. United Nations, New York.
United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2010. UNSCEAR 2008 Report. Annex B. Exposures of the public and workers from various sources of radiation. United Nations, New York.
United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2017. Sources, effects and risks of ionizing radiation. UNSCEAR 2017 report to the General Assembly. Scientific annex B. Epidemiological studies of cancer risk due to low-dose-rate radiation from environmental sources. United Nations, New York.
Warnery, E., Ielsch, G., Lajaunie, C., Cale, E., Wackernagel, H., Debayle, C., Guillevic, J., 2015. Indoor terrestrial gamma dose rate mapping in France: a case study using two different geostatistical models. Journal of environmental radioactivity 139, 140-148.
Wrixon, A.D., Green, B.M.R., Lomas, P.R., Miles, J.C.H., Cliff, K.D., Francis, E.A., Driscoll, C.M.H., James, A.C., O'Riordan, M.C., 1988. Natural radiation exposure in UK dwellings. National Radiological Protection Board, Chilton, Didcot, Oxon OX11 0RQ, UK, pp. i-ii+1-188.
Zhang, Z., Preston, D.L., Sokolnikov, M., Napier, B.A., Degteva, M., Moroz, B., Vostrotin, V., Shiskina, E., Birchall, A., Stram, D.O., 2017. Correction of confidence intervals in excess relative risk models using Monte Carlo dosimetry systems with shared errors. PloS one 12, e0174641.