Indoor radon measurements in south west England explained by topsoil and stream sediment geochemistry, airborne gamma-ray spectroscopy and geology

Antonio Ferreira[1], Z. Daraktchieva[2], D. Beamish[1], C. Kirkwood[1], T.R. Lister[1], M. Cave[1], J. Wragg[1], K. Lee[1]

1 British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK; 2 Public Health England, UK

Corresponding author:

Antonio Ferreira

British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK

Email:antonio@bgs.ac.uk

Tel: +44(0)115 936 3465

KEYWORDS

Indoor radon, Geology, Airborne gamma-ray, stream sediment geochemistry, topsoil geochemistry, analysis of variance, compositional data

HIGHLIGHTS

- Tested datasets (geology, airborne gamma-ray, stream sediment geochemistry and topsoil geochemistry) are useful for radon mapping purposes
- "Complete" topsoil geochemistry is preferable to the use of uranium concentrations only for radon prediction.
- For SW England the indoor radon variability is higher at short distances than at long distances

Submitted to Journal of Environmental Radioactivity in November 2015

Revised version, March 2016

HIGHLIGHTS

- Tested datasets (geology, airborne gamma-ray, stream sediment geochemistry and topsoil geochemistry) are useful for radon mapping purposes
- "Complete" topsoil geochemistry is preferable to the use of uranium concentrations only for radon prediction.
- For SW England the indoor radon variability is higher at short distances than at long distances

Indoor radon measurements in south west England explained by topsoil and stream sediment geochemistry, airborne gamma-ray spectroscopy and geology

Antonio Ferreira[1], Z. Daraktchieva[2], D. Beamish[1], C. Kirkwood[1], T.R. Lister[1], M. Cave[1], J. Wragg[1], K. Lee[1]

1 British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK; 2 Public Health England, UK

Corresponding author:

Antonio Ferreira

British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK

Email: antonio@bgs.ac.uk

Tel: +44(0)115 936 3465

KEYWORDS

Indoor radon, Geology, Airborne gamma-ray, stream sediment geochemistry, topsoil geochemistry, analysis of variance, compositional data

HIGHLIGHTS

- Tested datasets (geology, airborne gamma-ray, stream sediment geochemistry and topsoil geochemistry) are useful for radon mapping purposes
- "Complete" topsoil geochemistry is preferable to the use of uranium concentrations only for radon prediction.
- For SW England the indoor radon variability is higher at short distances than at long distances

Submitted to Journal of Environmental Radioactivity in November 2015

Revised version, March 2016

ABSTRACT

Predictive mapping of indoor radon potential often requires the use of additional datasets. A range of geological, geochemical and geophysical data may be considered, either individually or in combination. The present work is an evaluation of how much of the indoor radon variation in south west England can be explained by four different datasets: a) the geology (G), b) the airborne gamma-ray spectroscopy (AGR), c) the geochemistry of topsoil (TSG) and d) the geochemistry of stream sediments (SSG). The study area was chosen since it provides a large (197,464) indoor radon dataset in association with the above information. Geology provides information on the distribution of the materials that may contribute to radon release while the latter three items provide more direct observations on the distributions of the radionuclide elements uranium (U), thorium (Th) and potassium (K). In addition, (c) and (d) provide multi-element assessments of geochemistry which are also included in this study.

The effectiveness of datasets for predicting the existing indoor radon data is assessed through the level (the higher the better) of explained variation (% of variance or ANOVA) obtained from the tested models. A multiple linear regression using a compositional data (CODA) approach is carried out to obtain the required measure of determination for each analysis.

Results show that, amongst the four tested datasets, the soil geochemistry (TSG, i.e. including all the available 41 elements, 10 major – Al, Ca, Fe, K, Mg, Mn, Na, P, Si, Ti - plus 31 trace) provides the highest explained variation of indoor radon (about 40%); more than double the value provided by U alone (ca. 15%), or the sub composition U, Th, K (ca. 16%) from the same TSG data. The remaining three datasets provide values ranging from about 27% to 32.5%. The enhanced prediction of the AGR model relative to the U, Th, K in soils suggests that the AGR signal captures more than just the U, Th and K content in the soil.

The best result is obtained by including the soil geochemistry with geology and AGR (TSG+G+AGR, ca. 47%). However, adding G and AGR to the TSG model only slightly improves the prediction (ca. +7%), suggesting that the geochemistry of soils already contain most of the information given by geology and airborne datasets together, at least with regard to the explanation of indoor radon.

From the present analysis performed in the SW of England, it may be concluded that each one of the four datasets is likely to be useful for radon mapping purposes, whether alone or in combination with others. The present work also suggest that the complete soil geochemistry dataset (TSG) is more effective for indoor radon modelling than using just the U (+ Th, K) concentration in soil.

1. INTRODUCTION

Radon (^{222}Rn) is a geogenic radioactive gas permanently produced in all rocks and soils by radioactive decay of radium (^{226}Ra), which primarily derives from uranium (^{238}U). However uranium levels tend to vary with rock type, with higher Rn production being observed in uranium (and/or radium)-rich rocks, as expected. High concentrations of U (and/or Ra) can be found in rocks such as, uranium ores, some granites, black shales, some sandstones, phosphate rocks, some limestones and U (and/or Ra)-rich soils (e.g., over some limestones) (Appleton, 2013). In general, locations overlying rocks containing high levels of uranium show higher radon potential than those overlying rocks with low levels of U. Radon gas released from rocks and soils reaches open air, and rapidly disperses in the atmosphere to average levels of 4 $Bq.m^{-3}$ (HPA, 2010). As a consequence, radon typically becomes a health risk only in confined areas, where it tends to accumulate, such as in underground caves and mines, as well as inside houses and other buildings. Radon from the ground can enter houses through gaps and cracks in the floor, walls and pipes mainly due to differential indoor - outdoor – soil air conditions (pressure and temperature) (Appleton, 2013). As radon is a heavy noble gas, it tends to sink and has the potential to concentrate further in poorly ventilated basements and cellars. In the UK, the average indoor radon concentration is 20 $Bq.m^{-3}$ (ranging from less than 10 $Bq.m^{-3}$ to over 17000 $Bq.m^{-3}$) and thus is 10 times lower than the 200 $Bq.m^{-3}$ Action Level established for indoor radon in the UK (NRPB, 1990). Radon concentration in the soil pore space is quite variable, ranging from less than 1 to more than 2500 $kBq.m^{-3}$ (Appleton, 2013), thus it is generally 1 to 3 orders of magnitude higher than in the air above the ground, both indoor and outdoor (Varley and Flowers, 1998). In addition to the rock and soil capacity for radon gas production, other ground characteristics, such as permeability, water content,

77 organic matter or proximity to faults, as well as the weather conditions, temperature, pressure and
78 precipitation, are among the factors to be considered in order to account for the radon concentration in soil.
79 Despite of other relevant factors (atmospheric conditions, construction of the building, lifestyle including
80 heating and ventilation routines) contributing to radon levels inside houses, geology is the most important
81 one, as it is the source. According to Appleton and Miles (2010) approximately 25% of the total variation of
82 indoor radon in England and Wales can be explained by the geology. Hunter et al. (2009) found that
83 geology, at 19.7%, is the most important factor explaining the variance of UK indoor radon, while less than
84 half (8.9%) was attributed to the sum of seven house-related factors, including, house type (3.8%), double-
85 glazing (2.3%), date of building (1.1%), floor level (1.0%), floor type (0.5%), ownership (0.1%) and draught
86 proofing (0.1%).
87
88 ***Radon mapping***
89 The dependence of indoor radon on the geology has led to the development of radon mapping methods that
90 use geological maps together with indoor radon measurements. Other proxies, such as Rn concentration in
91 the soil gas; U (and Ra) content measured in rock, soil or stream sediment; the U content estimated from
92 ground or airborne Gamma-Ray spectroscopy; soil / rock permeability data, or proximity to faults are also
93 used for the assignment of a radon potential classes (Appleton, 2013). Often, the proxies are used according
94 to their availability. Relative to the other mentioned proxies, indoor radon measurements have the
95 advantage of being directly related to the space where people spend most of their time (i.e. home and other
96 buildings). Typically, the indoor radon measurements are spatially very unevenly distributed, clustered in
97 highly populated areas and rare or absent in rural areas. This can be partially overcome by using the geology
98 to extend the trends to areas with sparse or absent indoor data. Nevertheless, it may constitute a drawback
99 for large areas with no human occupation at present, but which may be planned for the near future. In fact,
100 the implementation of radon preventative measures in new buildings on radon prone areas (e.g., in the UK),
101 will, over time, induce an artificial reduction of the estimated radon potential, in case this is based on indoor
102 radon measurements.

In the UK, Public Health England (PHE) and the British Geological Survey (BGS) collaborate in the production of radon potential maps; these information have already been released for England and Wales (Miles et al., 2007), for Scotland (Miles et al., 2011) and for Northern Ireland with the collaboration of the Geological Survey Northern Ireland (Daraktchieva et al., 2015). Digital geological information (DiGMapGB-50k, mostly at 1:50 000 scale, http://www.bgs.ac.uk/products/digitalmaps/DiGMapGB_50.html), is used together with the indoor radon measurements for mapping the radon risk. The bedrock and the superficial 1:50 000 scale units are simplified (by age and lithology, or permeability and genetic type respectively) to produce a new set of codes, which are then combined together in a bedrock-superficial parent material (**BS**, geological combination) polygon code. Finally, the **BS** polygons are intercepted with the British National Grid (BNG) 1-km grid squares, resulting in the **KM1BS** polygons which constitute the basic mapping units. The following steps of the UK radon mapping methodology are performed after allocating each indoor radon measurement to the **KM1BS** polygon underlying it. The radon potential is then computed for each basic polygon (**KM1BS**) using the nearest data and provided that it belongs to same geological combination (**BS**). After mapping each **BS** separately, these are assembled in a final map. This radon potential map provides an estimation of the probability of homes in the UK having radon concentrations above the UK Action Level (200 $Bq.m^{-3}$). Further explanation about the UK radon mapping method and main results can be found in Miles and Appleton (2005), Miles et al. (2007, 2011) and Daraktchieva et al. (2015).

This paper is a contribution to the European Geogenic Radon Potential / Natural Radiation mapping project, led by the Joint Research Centre (JRC) of the European Commission. JRC started to design a European map of the geogenic radon potential exploring different approaches used by different European countries for development of national radon risk maps. A general classification scheme applicable to all countries is under development, which reflects the different experiences and type of information used at national level. Datasets including key variables such as indoor radon, geology, soil gas radon, U (and Ra) concentrations in soil and bedrock, airborne gamma-ray data, terrestrial gamma dose rate and permeability, seem to be the most obvious and appropriate approach.

The main question addressed in the present work is to what extent **geology (G)**, **airborne gamma-ray spectroscopy (AGR)**, **topsoil geochemistry (TSG)** and the **stream sediment geochemistry (SSG)** are able to explain the **indoor radon** variation, and based on the results, to assess the usefulness of these proxies for geogenic radon mapping. The above parameters are tested on the radon prone area of SW England, which is one of the most tested areas for indoor radon in the world.

2. MATERIALS AND METHODS

2.1. Geology and the KM1BS polygons

The geology of SW England, as shown in Figure 1, mainly consists of a range of metasediments deposited in palaeo sedimentary basins during the Devonian and the Carboniferous, to which followed the granite intrusion of the prominent Cornubian batholith, emplaced in the later stages of the Variscan orogeny, Late Carboniferous to Early Permian (see Kirkwood et al., 2016 and references herein). Post-Variscan sediments (Permian and Triassic to Neogene, Figure 1) are also preserved namely in small extensions near the eastern limit of the study area. The Cornubian batholith, which is exposed onshore in five major granite outcrops (labelled in *italic* Figure 1) and other minor ones and extends offshore south-westwards, is U-enriched and one of the most radon-productive formations in the UK (Scheib et al., 2013). The intrusion of the granite caused dissemination of several elements into the intruded Devonian (and Devonian-Carboniferous) sedimentary systems, giving rise to local enrichments in the form of polymetallic and other mineralisations, including in Uranium (see mine locations in Figure 1).

FIGURE 1 HERE

Most of the area is not covered by superficial units (Figure 2, the geological units in greyscale), being alluvium and costal clays, silts and sands the most represented ones by far. Also, peat (1.5 % of the study area, the palest colour of the grayscale in Figure 2) covers some moorlands in upland areas, namely over the

154 *Dartmoor* granite and, in a lesser extent, over Devonian mudstones and sandstones in the northernmost of
155 the study area and the *Bodmin* granite.
156 A simplified geological classification based on **BS** codes, derived from the BGS 1:50 000 scale geology
157 (DiGMapGB-50k), as previously discussed, was developed by the BGS for radon mapping purposes in the UK.
158 For the SW England study area (8,841 km^2) a total of 134 **BS** codes were derived from combinations of 42
159 simplified bedrock (BED) and 8 simplified superficial (SUP) codes (Figure 2).
160 From the interception of the **BS** polygons with the 1 km square grid framework, a total of 27,633 **KM1BS**
161 polygons were derived. The **KM1BS** polygons are the basic mapping units to which the other datasets are
162 allocated. Joining the datasets by the **KM1BS** polygons, ensures that both the spatial location and the
163 geology (**BS**) are taken into account.
164
165 FIGURE 2 HERE
166
167 **2.2. Indoor radon measurements**
168 In addition to geological factors, indoor radon could be affected by changes in atmospheric factors, such as
169 outdoor temperature, pressure and wind direction. To account for this variation, radon tests over a
170 continuous three month period are offered to householders by Public Health England (PHE), with the data
171 being stored in a database, and the associated house locations being accurately recorded (Miles and
172 Appleton, 2005).
173 In the present work the available 197,464 indoor radon measurements (annual radon concentrations
174 calculated as average of "living room" and "bedroom " results applying occupancy factors, and seasonally
175 corrected) carried out by the Public Health England are distributed across 9,144 **KM1BS** of the total of
176 27,633 **KM1BS** polygons, about 22 radon measurements per km^2 (Figure 3). This is one of the densest
177 collections of indoor radon measurements in the world although it inevitably remains unevenly distributed,
178 with the high number of measurements in densely populated areas, such as Plymouth, contrasting with the
179 areas with few or no houses, such as the Dartmoor uplands (Figure 3). The spatial distribution of indoor

radon concentrations show high values over Carboniferous – Permian granites and, in a lesser extent, over Devonian and Devonian – Carboniferous metasedimentary rocks, while the lower values are mainly related to Namurian sandstones and mudstones, Permian sandstones and Meso - Cenozoic units (Figure 4). The values shown in Figure 4 are the geometric mean (GM) of indoor radon measurements in each KM1BS polygon.

In the present study, as indoor radon data is allocated by **KM1BS**, the centroid coordinates of these polygons are attributed to the indoor radon data instead of the original house coordinates. For modelling purposes, the indoor radon concentrations are log-transformed (**lnRn**), as they follow a lognormal distribution (histogram of Figure 4) as expected (Miles, 1998). The histogram of Figure 4 also shows that the shape and the central values of the distribution of the indoor radon measurements are fairly similar before (histogram bars of Figure 4, 197464 indoor radon measurements) and after averaging the indoor radon measurements in each KM1BS (histogram curve of Figure 4, 9144 KM1BS GM indoor radon data polygons). This averaging procedure, however, leads to a slight decreasing in the variability of the data (e.g., compare IQR, interquartile range, in Table 1).

Table 1 – Summary statistics of indoor radon concentrations for the available 197,464 measurements (data points) and for the geometric mean of indoor radon measurements in each KM1BS polygon (data polygons, 9,144 values, data (2) in table 5)

Data	N	Min	Q25	Mdn	Q75	Q95	Q99	Max	IQR	AM	GM	GSD
Data points	197464	1.1	37.0	77.5	162.4	523.2	1224	24343	125.4	154.3	**78.6**	**3.1**
Data polygons	9144	1.1	42.5	75.9	139.8	368.7	735.1	5014	97.2	121.8	**78.4**	**2.5**

N: number of samples; **Min**: minimum value; **Max**: maximum value; **Q##**: ##th quantile; **Mdn**: median; **IQR**: interquartile range; **AM**: arithmetic mean; **GM**: geometric mean; **GSD**: geometric standard deviation. Values are in $Bq.m^{-3}$, except for **N** and **GSD** (dimensionless). The most used indoor radon statistics (GM and GSD) are in bold.

FIGURE 3 HERE

FIGURE 4 HERE

205 **2.3. Airborne Gamma-Ray**

206 Gamma radiation measured by airborne surveys relates to the surface layer, often < 50 cm, possibly deeper
207 in less dense unconsolidated material such as some soils, to a maximum of a few metres in dry peat. Such
208 data provide potassium abundance (%) directly, as gamma-rays are emitted from ^{40}K decay to argon, while U
209 and Th concentrations ($mg.kg^{-1}$) are estimated from the gamma-ray emission peaks associated with ^{214}Bi and
210 ^{208}Tl respectively, and are expressed as equivalent uranium (eU) and equivalent thorium (eTh). Total counts
211 (in counts per second, cps), across a wider energy range, is also frequently reported.

212 The airborne gamma-ray spectrometry data (**AGR**) available for the present study were acquired under the
213 BGS Tellus South West airborne geophysical survey (http://www.tellusgb.ac.uk/home.html), carried out
214 during the second half of 2013. The survey comprised a high resolution magnetic/magnetic gradient survey
215 combined with a multichannel (256 channel) radiometric survey. The survey was carried out using 200 m
216 (north–south) line separations at a mean elevation of 91 m. The survey provided 60,323 line-kilometres of
217 data (Beamish et al., 2014). Calibration and processing of the airborne data was done following standard
218 procedures described in IAEA (IAEA, 2003) and AGSO (Minty et al., 1997) publications.

219 The 684,384 **AGR** measurements covering the study area provide a sampling density of about 77 per km^2.
220 The spatial distribution of equivalent uranium (eU, in $mg.kg^{-1}$) from the **AGR** dataset shows the highest
221 values over Carboniferous – Permian granites, except when these are covered by peat (Figure 5). The eU
222 spatial distribution at KM1BS resolution, that is, after averaging the **AGR** data in each KM1BS (Figure 6,
223 23,573 data polygons), shows the same patterns as Figure 5, although necessarily with less detail. From the
224 27,633 **KM1BS** basic mapping unit polygons 23,573 include **AGR** data. The number of **AGR** data points in
225 each KM1BS (ranging from 1 to 145) is mainly dependent on the size of the polygon, as the distribution of
226 **AGR** points is quite regular. Only **KM1BS** polygons not fully covered by the survey (e.g., near the mouth of
227 River Tamar, close to Plymouth), or reporting a very small area (thus behaving like a sliver), have no airborne
228 data inside.

229 The statistics for all data points (N = 684,384) and after allocating data to the KM1BS polygons (N = 23,573)
230 are quite similar (Table 2, e.g., compare median and GM values for each of the four variables). This indicates

231 that the distribution after the averaging procedure keep the same structure of the original data, despite of a
232 very small decreasing in the variability of the data (Table 2, e.g., compare IQR and GSD).
233
234 FIGURE 5 HERE
235 FIGURE 6 HERE
236
237 Table 2 – Summary statistics for total counts (TC), potassium (K), equivalent uranium (eU) and
238 equivalent thorium (eTh) from the available airborne gamma ray (**AGR**) data points (N = 684,384) and
239 after averaging **AGR** data in each KM1BS polygon (data polygons, N = 23,573 values).

	Data points (N = 684,384, Figure 5)				Data polygons (N = 23,573, Figure 6)			
	TC (cps)	**K (%)**	**eU ($mg.kg^{-1}$)**	**eTh ($mg.kg^{-1}$)**	**TC (cps)**	**K (%)**	**eU ($mg.kg^{-1}$)**	**eTh ($mg.kg^{-1}$)**
N%>0	99.93	99.75	99.83	99.92	99.96	99.89	99.84	99.93
Min	-471	-1.43	-7.10	-3.47	-150	-0.22	-2.03	-0.74
Q25	1215	0.99	1.41	7.20	1208	1.01	1.47	7.11
Mdn	**1428**	**1.28**	**1.73**	**8.45**	**1419**	**1.28**	**1.73**	**8.31**
Q75	1684	1.63	2.11	9.73	1667	1.61	2.08	9.47
Q95	2203	2.14	3.46	12.01	2170	2.05	3.43	11.42
Q99	2870	2.57	4.94	14.86	2724	2.42	4.68	14.12
Max	20157	4.79	84.20	25.76	4550	4.24	12.82	18.72
IQR	469	0.64	0.70	2.53	459	0.60	0.61	2.36
AM	1461	1.32	1.87	8.41	1444	1.31	1.88	8.18
GM*	1386	1.23	1.72	7.98	1370	1.22	1.74	7.77
GSD*	1.45	1.55	1.54	1.46	1.44	1.52	1.51	1.46

240 **N%>0**: number of samples (in %) with positive values; **Min**: minimum value; **Max**: maximum value; **Q##**: ##th quantile;
241 **Mdn**: median; **IQR**: interquartile range; **AM**: arithmetic mean; **GM***: geometric mean (positive values only); **GSD***:
242 geometric standard deviation (positive values only). Units (in brackets) are valid for all parameters except **N%>0** and **GSD**
243 (dimensionless).
244

245 ## 2.4. Topsoil Geochemistry and Stream Sediment Geochemistry

246 Top soil (TS) and stream sediments (SS) both derive mainly from the bedrock beneath. Soil is a complex
247 body developed in the bedrock-atmosphere interface, mainly from weathering of the bedrock, to which is
248 added organic matter, water and air due to interaction with living organisms (often considered a component

249 of the soil too) and with the atmosphere. The active stream sediment is mainly composed by downstream
250 transportation of material comprising of rock, soil and random organic debris. Stream sediment chemistry
251 reflects the original bedrock chemistry from which it is derived, however the sediment is sorted by gravity
252 and also by the energy involved in the downstream dispersion of the parent materials. Soil chemistry should
253 also reflect the original bedrock chemistry, but is modified by bio-physical-chemical processes that are
254 behind its development. Furthermore, soil and close to source stream sediment can be seen, approximately,
255 as two complementary products of the same bedrock; with soil retaining mostly the immobile parts of the
256 elements and minerals whilst the sediment retains (at least in part) the elements and minerals mobilized
257 from the soil. In practice, the content of an element in the soil and its ratio with other elements may be
258 quite different from that in a stream sediment, as well as in the original bedrock, depending on the hosting
259 minerals and the element's mobility at different environmental conditions (pH/Eh).
260 Geochemistry of topsoils (**TSG**) and/or of stream sediments (**SSG**) are among the most commonly available
261 geoscience dataset in many countries of the world. Unfortunately, radium, the direct precursor of radon
262 gas, is not part of the routinely analysed list of elements due to analytical difficulties. On the contrary, U, Th
263 and K, the three main sources of natural geogenic radiation, as well as an additional 30 to 50 elements, allow
264 a very comprehensive understanding of the analysed sample.
265 Data from two systematic geochemical surveys, the UK-wide G-BASE (Geochemical Baseline Survey of the
266 Environment, http://www.bgs.ac.uk/gbase/home.html) project and the England and Wales NSI (National
267 Soils Inventory, http://www.landis.org.uk/data/nsi.cfm), are available for the present work. The 3382
268 stream sediment samples, shown in Figure 7, were collected under the G-BASE project at a sampling density
269 of approximately one site per 2.5 km^2. From the available 987 soil samples, shown in Figure 8, 644 were
270 collected during the 2013 season of the G-BASE project, at a sampling density ranging from about one
271 sample per 10 km^2 to one sample per 30 km^2, depending on the nature and variability of underlying
272 geological units. The remaining 343 soil samples were collected under the NSI at a regular sampling density
273 of 1 sample per 25 km^2. The overall soil sample density is about 1 sample per 9 km^2. All the processes from
274 sampling to analysis, as well as the data, were subjected to quality control to ensure countrywide

consistency and continuity with other existing G-BASE and NSI geochemical data. XRF data for the < 150 µm size fraction of the stream sediment samples and for the < 2 mm size fraction of the soil samples are available for the present work for 41 elements, including 10 major (Al, Ca, Fe, **K**, Mg, Mn, Na, P, Si and Ti) and 31 trace (As, Ba, Br, Ce, Co, Cr, Cs, Cu, Ga, Ge, Hf, La, Mo, Nb, Nd, Ni, Pb, Rb, Sb, Sc, Se, Sm, Sn, Sr, **Th**, **U**, V, W, Y, Zn and Zr).

The most visible feature of the spatial distribution of uranium concentrations (U, in $mg.kg^{-1}$) from both stream sediments (Figure 7) and topsoils (Figure 8) is that high values are generally overlying the Carboniferous – Permian granites.

After the 3,382 stream sediment data points were allocated to the underlying KM1BS, 3,027 data polygons was obtained. The number of stream sediment data points in each of these 3,027 KM1BS is only one in most cases (2,699) and range from two to four in the remaining 328 KM1BS polygons. The 3,027 stream sediment data polygons (i.e., after averaging SS data in each KM1BS polygon) show very similar statistics to the 3,382 data points as can be seen for K, U and Th in Table 3. An extra table including all elements is shown in supplementary material (Extra Table A).

A number of data polygons (n=947) was obtained after allocating the 987 top soil data points to the underlying KM1BS. The number of top soil data points in each of these 947 KM1BS is only one in most cases (907) and two in the remaining 40 KM1BS polygons. The 947 top soil data polygons (i.e., after averaging TS data in each KM1BS polygon) show very similar statistics to the 987 data points as can be seen for K, U and Th in Table 4. An extra table including all elements is shown in supplementary material (Extra Table B).

FIGURE 7 HERE

Table 3 – Summary statistics for potassium (K), uranium (U) and thorium (Th) from stream sediment geochemistry of the SW England (N= 3,382 data points) and after taking the average of the data points in each KM1BS polygon (N = 3,027 data polygons).

Data points (N = 3,382, Figure 7)	**Data polygons (N = 3,027)**

	K_2O (%)	U (mg.kg^{-1})	Th (mg.kg^{-1})	K_2O (%)	U (mg.kg^{-1})	Th (mg.kg^{-1})
Min	0.53	1.00	2.60	0.53	1.00	2.60
Q25	2.33	2.50	9.80	2.33	2.50	9.88
Mdn	**2.75**	**2.90**	**11.30**	**2.75**	**2.90**	**11.30**
Q75	3.18	3.50	12.90	3.19	3.50	13.00
Q95	3.92	24.48	43.38	3.91	23.91	44.20
Q99	4.53	44.44	77.24	4.48	43.87	76.41
Max	6.78	206.9	321.7	6.17	206.9	321.7
IQR	0.85	1.00	3.10	0.86	1.00	3.13
AM	2.78	5.70	15.01	2.78	5.70	15.07
GM	2.69	3.64	12.53	2.70	3.66	12.58
GSD	1.29	2.10	1.64	1.28	2.11	1.64

N: number of samples; **Min**: minimum value; **Max**: maximum value; **Q##**: ##th quantile; **Mdn**: median; **IQR**: interquartile range; **AM**: arithmetic mean; **GM**: geometric mean; **GSD**: geometric standard deviation. Units (in brackets) are valid for all parameters except **GSD** (dimensionless).

FIGURE 8 HERE

Table 4 – Summary statistics for potassium (K), uranium (U) and thorium (Th) from top soil geochemistry of the SW England (N= 987 data points) and after taking the average of the data points in each KM1BS polygon (N = 947 data polygons).

	Data points (N = 987, Figure 8)			**Data polygons (N = 947)**		
	K_2O (%)	U (mg.kg^{-1})	Th (mg.kg^{-1})	K_2O (%)	U (mg.kg^{-1})	Th (mg.kg^{-1})
Min	0.01	0.70	0.30	0.01	0.66	0.30
Q25	1.96	2.50	8.80	1.95	2.50	8.78
Mdn	**2.47**	**2.90**	**10.30**	**2.47**	**2.90**	**10.25**
Q75	2.96	3.40	11.60	2.96	3.43	11.60
Q95	3.43	6.10	13.40	3.43	5.87	13.40
Q99	3.96	11.71	17.21	3.92	11.70	17.25
Max	4.94	45.70	22.20	4.94	35.35	22.20
IQR	1.01	0.90	2.80	1.01	0.93	2.82
AM	2.41	3.34	9.99	2.41	3.31	9.99
GM	2.20	3.01	9.43	2.19	3.00	9.42
GSD	1.84	1.49	1.51	1.86	1.48	1.51

N: number of samples; **Min**: minimum value; **Max**: maximum value; **Q##**: ##th quantile; **Mdn**: median; **IQR**: interquartile range; **AM**: arithmetic mean; **GM**: geometric mean; **GSD**: geometric standard deviation. Units (in brackets) are valid for all parameters except **GSD** (dimensionless).

2.5. Integration of Datasets and models produced

As explained in the previous subsections, the point based datasets (**Indoor Radon**, **AGR**, **TSG**, **SSG**) were allocated to the **KM1BS** polygons, through joining by spatial location (in ArcGIS). **KM1BS**s are polygons corresponding to the interception of a 1km^2 grid with the geology (**BS**) polygons. If two or more data points of a certain dataset are within the same **KM1BS** polygon, the resulting record will correspond to the ***average*** value. After grouping the available 197,464 $\log_e$-transformed indoor radon measurements (data points) according to the **KM1BS**, a dataset of 9,144 records (data polygons) was generated. By the same procedure, the available 684,384 airborne data points were grouped into 23,573 data polygons; the available 3382 stream sediment data points were grouped into 3027 data polygons and the 987 top soil data points were grouped into 947 data polygons.

Finally the information in these new files was integrated in *ACCESS*, according to the data requirements and availability for each specific model. Table 5 reports the final data files used in the present work and the respective number of records. Indoor radon modelled by **AGR** can be carried out by using data (3), (4), (5) or (6), as shown in Table 5, however with a decreasing number of available records. The models including **SSG** can be performed either with 1182 (data (4)) or with 83 (data (6)) records; while models with **TSG** are based either on 600 (data (5)) or on 83 (data (6)) records (Table 5). Indoor radon models exclusively with geology can be carried out by any of the data ((1) to (6)) referred to in Table 5, however the value obtained with data (1) cannot be directly compared with those from the remaining data ((2) to (6)). While these last datasets only account with the variance *between* **KM1BS** polygons, models with data (1) will account with both the variance *within* and *between* **KM1BS** polygons.

The spatial join by **KM1BS** ensures that data is joined over the same geology, contrary to a simple join based exclusively on distance. It should also be noted that **KM1BS** polygons, which may be sized up to a maximum of 1 km^2, constitute the maximum detail shown in UK radon maps, thus, most likely providing enough detail also for the European radon mapping. Further explanation about the variances can be found under section *2.6.2. Variance Explained*.

Table 5: number of records used for **lnRn** modelling according to data required.

Data file used	Number of records
(1) Rn	197464
(2) Rn (averaged by KM1BS)	9144
(3) Rn + AGR	8905
(4) Rn + AGR + SSG	1182
(5) Rn + AGR + TSG	600
(6) Rn + AGR + SSG + TSG	83

Rn: $\log_e$ of indoor radon measurements; **AGR**: airborne gamma-ray data; **SSG**: Stream sediment geochemistry data; **TSG**: Topsoil geochemistry data.

The resulting **lnRn** ($\log_e$ of indoor radon) models produced for the present work are summarised in section **3** (Results).

Models using only the geology (**G**) as predictor variable were the first set to be run. As these can be produced using all the data files (i.e., (1) to (6) of Table 5), and using 3 types of geological classification (**GPE**: simplified geological period classes, **BED**: simplified bedrock classes and **BS**: simplified bedrock/superficial geological combinations), a total of 18 models have been produced (Table 6). Running all these models allows comparison of those obtained with the three different geological classification types (i.e., values along the same line in Table 6) as well as checking for consistency of results as the number of records decreases from data file (2) to data file (6) of table 5 (i.e., along the same column in Table 6).

Models using airborne gamma ray (**AGR**) data as predictor follow those with geology only, as the spatial cover of **AGR** data is much more complete than that of **SSG** and **TSG**. A total of 4x6 = 24 models have been produced (Table 7), from four data files (from (3) to (6) of Table 5) and using six different predictor combinations, namely total counts (**AGR[TC]**); eU concentrations (**AGR[U]**); eU, eTh and K concentrations (**[AGR]**); eU, eTh, and K concentrations and total counts (**[AGR] + AGR[TC]**); eU concentration and geology (**AGR[U]+G**, with **BS** classes) and finally eU, eTh and K concentrations and geology (**[AGR] + G**, with **BS** classes). This set of models provides an understanding of how good the **AGR** data is (i.e., eU, eTh, K) in predicting indoor radon, weighting and comparing uranium relative to the whole **AGR**, concluding about the

usefulness of total counts in radon prediction and measuring the improvement of models as a result of adding **AGR** to **G** in the set of predictor variables.

Finally models including stream sediment geochemistry (**SSG**) or top soil geochemistry (**TSG**) as predictor variables are shown in Table 8 and Table 9 respectively. Data files (4) and (5) of Table 5 are used for models with **SSG** and **TSG** respectively. A total of 3x4 = 12 models including **SSG** data (Table 8) or **TSG** data (Table 9) have been produced. Three possible sets of chemical elements are used, namely uranium concentrations only (**SSG[U]**, **TSG[U]**); uranium, thorium and potassium concentrations (**SSG[UThK]**, **TSG[UThK]**) and the entire set of the 41 available elements (**[SSG]**, **[TSG]**). Four different groups of variables are used, namely the geochemistry only (first column of results in Table 8 and Table 9); geochemistry and geology (second column of results in Table 8 and Table 9); geochemistry and airborne gamma ray (third column of results in Table 8 and Table 9); and geochemistry, airborne gamma ray and geology (last column of Table 8 and Table 9). These two sets of models allow comparison of the performance of the geochemistry of the two media in predicting radon, weighting and comparing the two subsets of elements (uranium and U, Th, K) with each other and relative to the whole composition, and measuring the improvement of models as a result of adding the geochemistry to **AGR** and **G** in the set of predictor variables. Models including both **SSG** and **TSG** (data (6) of Table 5) were not considered after less consistent results obtained from the same data file (6) for models with **AGR** and **G** (Table 7).

Before going into detail on the results a brief introduction to compositional data, compositional regression and the tools used are presented.

2.6. Compositional data

Geochemical datasets, including the **TSG** and **SSG** used in the present work, are typical examples of *compositional (closed) data, CODA,* (Aitchison, 1986), as the sum of all elements (parts) in a sample necessarily sum up to a constant (often 1, 100 % or 1 000 000 $mg.kg^{-1}$). This implies that the concentration of one element does not vary independently from the others, and thus that the interpretation and statistical evaluation of the observed element concentrations is only meaningful if the relationship to the

387 concentration of the remaining elements is taken into account (Aitchison, 1986; Filzmoser et al., 2010;
388 Reimann et al., 2012).
389 Aitchison (1986) developed a geometry on the simplex on which compositional data fully adjust.
390 Compositional data must be transformed into the correct geometry first, after which classical methods can
391 be fully applied. In CODA context the statistics are performed after opening the data through one of three
392 possible log-ratio (logarithm of a ratio) transformations. The basic idea behind these transformations is that
393 the meaningful information for an element is not in the reported concentration, but only achievable if the
394 other elements are taken into account. The ***alr*** - additive log-ratio transformation, proposed by Aitchison
395 (1986), uses one of the elements as divisor, against which each one of the other elements is presented after
396 taking the logarithm; when three elements (D=3, ***x1***, ***x2***, ***x3***) are reported in a compositional dataset ***x***, the
397 respective ***alr*** will be composed by 2 log-ratios (coordinates): alr(***x***) = [ln(***x1***/***x3***; ln(***x2***/***x3***)]. The ***clr*** – centred
398 log-ratio transformation, proposed by Aitchison (1986), use the geometric mean (g(***x***)) of all measured
399 elements as the divisor against which each element is "normalized"; when three elements (D=3, ***x1***, ***x2***, ***x3***)
400 are reported in a compositional dataset ***x***, the respective ***clr*** will contain 3 log-ratios (coordinates): ***clr***(***x***) =
401 [ln(***x1***/g(***x***)); ln(***x2***/g(***x***)); ln(***x3***/g(***x***))], where g(***x***) is the geometric mean of ***x1***, ***x2***, ***x3***. This is perhaps the most
402 popular log-ratio transformation as there is a direct relation to the original element. However working with
403 this transformation requires some care as the covariance and correlation matrices are singular because the
404 ***clr*** coefficients necessarily sum to zero (Buccianti, et al., 2006). The ***ilr*** – isometric log-ratio transformation,
405 proposed by Egozcue et al. (2003), have the best geometrical properties for multivariate data analysis
406 (Buccianti, et al., 2006) and linear regression (van den Boogaart & Tolosana-Delgado, 2013), but the results
407 are difficult to interpret as there is no direct relation with the original elements (parts). The ***ilr***
408 transformation is given by ***ilr(x)***:= ***clr(x).V***, where ***V*** is a matrix with D rows and D-1 columns (with D the
409 number of original parts) which columns form an orthonormal basis of the ***clr***-plane. Thus ***ilr*** relates with the
410 ***clr*** transformation, allowing to reconstruct the names of the original parts through an inverse
411 transformation; when three elements (D=3, ***x1***, ***x2***, ***x3***) are reported in a compositional dataset ***x***, the
412 respective ***ilr*** will contain 2 log-ratios (coordinates): ***ilr***(***x***) = [sqrt(1/2).ln(***x1***/***x2***); sqrt(1/6).ln(***x1.x2***/***x3.x3***)]

(Pawlowsky-Glahn & Egozcue, 2006). Further explanation on how to analyse compositional data can be found in e.g. Buccianti et al. (2006), Pawlowsky-Glahn & Buccianti (2011), Tolosana-Delgado (2012), van den Boogaart & Tolosana-Delgado (2013), among other publications. Further discussion and practical applications of CODA principles to a European geochemical mapping project (GEMAS) can be found in Reimann et al. (2014) and Reimann et al. (2012). The latest discussions about compositional data mapping are found in McKinley et al. (2015), Buccianti et al. (2015), Pawlowsky-Glahn and Egozcue (2015).

2.6.1. Compositional Regression

The indoor radon modelling using the four datasets described above was carried out in R (***compositions*** package), and using a compositional approach. The main reason is rooted in the compositional nature of some of the datasets, namely the stream sediment geochemistry, soil geochemistry, and eventually the airborne gamma-ray dataset.

The ***compositions*** is an **R** package, developed by Gerald van den Boogaart, Raimon Tolosana and Matevz Bren (https://cran.r-project.org/web/packages/compositions/compositions.pdf), providing functions for the consistent analysis of compositional data. Both **R** (programming language) and ***compositions*** are available as free software. **R** (Rx64 3.1.0) was downloaded from http://www.cran.r-project.org,while ***compositions*** package (version 1.40) was downloaded and installed using an **R** friendly interface, the RStudio (version 0.98.1102 of the open source edition), which is available in https://www.rstudio.com/products/rstudio/. More information about ***compositions*** package can be found in K.G. van den Boogaart and R. Tolosana-Delgado (2013). Other open source packages, such as the CoDaPack (http://ima.udg.edu/codapack/), are available for Compositional Data Analysis.

According to van den Boogaart & Tolosana-Delgado (2013), compositional regression is analogous to multiple linear regression, although more complex in the details. For the present case, the *dependent* variable (***Y***) is the indoor radon ($Bq.m^{-3}$) transformed by the natural logarithm (***Y*** = $\log_n$(indoor radon) or simply ***Y*** = ***lnRn***) as the available indoor radon data follows a lognormal distribution as expected (Miles,

1998). Each one of the datasets **G**, **AGR**, **SSG**, **TSG** will be part of the *independent* variables (***X***), with **G** as a *categorical* (*factor*) variable, and **AGR**, **SSG**, **TSG** as *continuous compositional* variables.

According to van den Boogaart & Tolosana-Delgado (2013), the ***ilr*** isometric logratio transformation is the only one that completely fulfils the requirements in linear modelling involving compositions, namely when metric variances are required. This is true for the present work, as **SSG** and **TSG** (and eventually **AGR**) are compositions and the R^2, a measure of determination will be used to compare the different models.

Following procedures found in van den Boogaart & Tolosana-Delgado (2013), the indoor radon prediction was performed using ANOVA and a compositional regression model of the general form $\boldsymbol{Y_i} = \boldsymbol{a} + (\boldsymbol{b}, \boldsymbol{X_i})_A + \boldsymbol{\varepsilon_i}$ (**Equation 1**), where ***Y*** is a real random variable, ***X*** is an independent compositional variable, and the parameter $\boldsymbol{a}$, **b** and ε are, respectively, a real-valued intercept, a regression "slope" composition (belonging to a D-part *Simplex*), and a real-valued zero-mean error (typically modelled as normally distributed).

According to van den Boogaart & Tolosana-Delgado (2013) the estimation of the regression parameters will be a multiple linear regression problem of the type $\boldsymbol{Y_i} = \boldsymbol{a} + \sum_{k=1}^{D-1} \boldsymbol{\beta_k} \boldsymbol{ilr_k}(\boldsymbol{X_i}) + \boldsymbol{\varepsilon_i}$ (**Equation 2**), with β= β_k a vector of slope parameters.

The multiple linear regression are carried out using a simple script line under the ***compositions*** R package, ***(model = lm (Y ~ ilr(X_A)))*** (**Script line 1**), where $\boldsymbol{X_A}$ is a compositional dataset (van den Boogaart & Tolosana-Delgado, 2013). This can be applied to the entire composition (e.g., a dataset including contents of "all" elements in soil samples) as well as to a sub-composition (e.g., a subset including the U, Th, K contents in soil samples). If two compositional datasets ($\boldsymbol{X_A}$, $\boldsymbol{X_B}$) are used together with the geology categories (***G***), the script line will be ***(model = lm (Y ~ ilr(X_A) + ilr(X_B) + G))*** (**Script line 2**).

Models including 3 or more elements of the same dataset (i.e., **[AGR]**, **SSG[UThK]**, **[SSG]**, **TSG[UThK]** and **[TSG]**) were processed including a residual (**R**) in addition to the respective elements, and using a compositional (CODA) approach, according to the following procedures. First, all elements were converted into $mg.kg^{-1}$. Secondly, the data was closed (using the *acomp()* function in ***compositions*** R package) taking

into account the residual part (**R**) with respect to the total (1,000,000 $mg.kg^{-1}$) computed as **R = 1,000,000 – (sum of elements)** (**Equation 3**), where **sum of elements** is K + eU + eTh for **[AGR]**, K + U + Th for both **SSG[UThK]** and **TSG[UThK]**, and the sum of the 41 elements available (see list in section 2.4.) for both **[SSG]** and **[TSG]**, with concentrations in $mg.kg^{-1}$ for all mentioned cases. Thus, in short, each one of the models **[AGR]**, **SSG[UThK]** and **TSG[UThK]** is based in a four-parts composition (D = 4, i.e., 3 elements and the respective residual **R**), while each one of the models **[SSG]** and **[TSG]** is based in a 42-parts composition (D = 42, i.e., 41 elements and the respective residual **R**). Finally, the isometric logratios for each composition are computed directly, using the *ilr()* function in ***compositions*** R package, while running the required model (see script lines 1 or 2 above). It should be emphasized, thus, that the predictor variables for these models are a set of dimensionless **D-1 logratios** (logarithm of a ratio of 2 or more elements), the ***ilr*** coordinates, not a set of elements (see section 2.6. Compositional data for further explanation). Also, the significance for each of the D-1 ***ilr*** coordinates can be obtained from the output of the model, but not the significance for each element. However, the focus of the present work is on the significance of each model as a whole, rather than evaluating the significance of the ***ilr*** coordinates. Models with sub-compositions (e.g., **SSG[UThK]**), or with one element only (e.g., **SSG[U]**) are mainly used to frame the results of the whole relative to a set of elements commonly used in indoor radon modelling.

To note that the use of the residual R as one of the parts in a composition is not compulsory in CODA context. However it was used in the present work as it may contain valid information (Kirkwood et al., submitted). As an example, peat soil samples from SW England always have a high residual R, which is related to the Loss on Ignition (Kirkwood et al., submitted) of the samples.

2.6.2. Variance Explained

Two measures of determination R^2 and ***adjusted*** R^2 (***adj*** R^2) can be used to numerically evaluate the strength of the relationship being modelled and be interpreted as an estimate of the portion of the total variability that is explained by the model. **Adjusted** R^2 will be preferably shown instead of R^2, as it takes into consideration the sample size and the number of predictor variables, contrary to R^2, which thus tend to

increase with increasing number of parameters. These two and other model parameters can be retrieved by calling for a model summary (script line 3: *summary(model)*) and an ANOVA table (script line 4: *anova(model)*) in ***compositions*** R package (van den Boogaart & Tolosana-Delgado, 2013). The ANOVA table provides an overall significance test for each dataset included in the model, thus allowing to check whether the information provided by each dataset as a whole influences or not the indoor radon. The model summary provides a significance test for each of the individual factors of the geology, and for the *individual* coordinates (that is, each one of the log-ratios) of the ***ilr***-transformed composition(s) (van den Boogaart & Tolosana-Delgado, 2013). This allows assessment of the relevant ***ilr*** coordinates, but not the relevant original elements directly, for the indoor radon explanation.

The ***total*** $\log_e$ indoor radon variability ($\boldsymbol{V_t}$), can only be computed from data (1) in Table 5, that is, directly using all indoor radon measurements. In models computed with data (1) the mentioned explained variation ($\boldsymbol{V_e}$, **adj. R^2** in %) is a ***portion*** ($\boldsymbol{V_p}$) of the ***total*** $\log_e$ indoor radon variability $\boldsymbol{V_t}$ ($\boldsymbol{V_e}$=100x$\boldsymbol{V_p}$/$\boldsymbol{V_t}$). This $\boldsymbol{V_t}$ can be decomposed in two main sources of variability: $\boldsymbol{V_t}$ = $\boldsymbol{V_b}$ + $\boldsymbol{V_w}$, where $\boldsymbol{V_b}$ is the variability ***between*** **KM1BS** polygons (or ***lateral*** variability) and $\boldsymbol{V_w}$ is the variability ***within*** **KM1BS** polygons (or ***local*** variability). The $\boldsymbol{V_w}$ can only be computed from data (1) in Table 5, as well. This is because the ***local*** variability $\boldsymbol{V_w}$ is automatically removed after grouping and averaging data per **KM1BS**, as happen with data (2) to data (6) in Table 5. Also, the overall ***lateral*** $\log_e$ indoor radon variability ($\boldsymbol{V_b}$) can be computed from data (1) as the $\log_e$ indoor radon variation explained by **KM1BS**, 32.3% (Table 6). Thus, for models computed from data (2) to data (6) the mentioned explained variation ($\boldsymbol{V_e}$, **adj. R^2** in %) is a ***portion*** ($\boldsymbol{V_p}$) of the overall ***lateral*** $\log_e$ indoor radon variability $\boldsymbol{V_b}$ ($\boldsymbol{V_e}$=100x$\boldsymbol{V_p}$/$\boldsymbol{V_b}$, with $\boldsymbol{V_b}$=32.3%), not of the ***total*** $\log_e$ indoor radon variability ($\boldsymbol{V_t}$). As an example, the **lnRn** variation explained by **BS**s in data (1), 10.1%, is relative to $\boldsymbol{V_t}$, while the value reported for data (2), 29%, is relative to $\boldsymbol{V_b}$ (Table 6). Note that 29% is similar to 10.1% if computed relative to 32.3% (29x32.3/100=9.4%), being the difference between them (10.1%-9.4% = 0.7%) an effect of the highly variable number of indoor radon measurements available per **KM1BS** (from 1 to 933). The effect explained for data

515 (2) also applies to data (3), (4), (5) and (6). Thus, for models with data (2) to data (6), the explained variance
516 (V_e) is a portion (V_p) of the overall ***lateral*** $\log_e$ indoor radon variability V_b (V_e=100xV_p/V_b, with V_b=32.3%).
517
518 The **adjusted R^2** is the statistical parameter chosen to evaluate and compare models, and thus shown in the
519 four tables reporting the results for the 67 models (Table 6 to Table 9). To note that all models presented
520 report a **p-value < 0.001** except when indicated otherwise. The models shown in Figure 9 to Figure 12
521 include information about the number of records, the significance level (p value), the Pearson R, the R^2 and
522 the adj. R^2. All these parameters have been retrieved from the model outputs after following instructions
523 above.
524
525 **3. RESULTS**
526 **3.1. Indoor radon explained by Geology**
527 The $\log_e$ indoor radon (**lnRn**) variance explained (adjusted R^2) by geology after regression analysis is shown in
528 Table 6 for models from three different geological classifications: as Geological Period (**GPE**), as bedrock
529 (**BED**) or as bedrock-superficial (**BS**s).
530 For SW England, the proportion of $\log_e$ transformed indoor radon (**lnRn**) variation explained by geology,
531 whether bedrock (**BED**) or bedrock-superficial (**BS**), is about 10% (Table 6 and Figure 9 A). The *superficial*
532 units only add about 1% of explained variability to that from *bedrock* units alone (10.1-9.2=0.9%). If data is
533 classified according to Geological Period (GPEs), the variation of **lnRn** explained by geology is only 1.4%
534 (Table 6). Thus, not surprisingly, the variation of indoor radon explained by geology increases with the level
535 of detail of the geology. Also, for the present case, a classification based on geological periods is not
536 effective for indoor radon analysis, contrary to a classification based on bedrock. SW England shows a low
537 proportion of **lnRn** variation explained by geology, in comparison to the average value found for England and
538 Wales (24.6%, Appleton and Miles, 2010), Scotland (21%, Scheib et al. 2009) or in a lesser extent, to values
539 found in Belgium (15.4% - 17.7%, Tondeur et al., 2014), and similar to those found in Austria (11.2%, Bossew
540 et al., 2008), Northern Ireland (12.1%, Appleton et al., 2015). The low value is a result of 1) the short

541 geological time range, most of the area belongs to two successive periods (Devonian and Carboniferous),
542 thus constraining the explanation of indoor radon variability given by the geology, and 2) the uranium
543 mineralization (Appleton and Miles, 2010), which are spatially and genetically dependent on granitic rocks,
544 but scattered throughout other surrounding geological units, several Devonian and Devonian-Carboniferous
545 sedimentary basins. This leads to a general increasing of the **lnRn** variation within geological units, thus
546 decreasing contrast between them.
547 The overall **lnRn** variability observed ***between* KM1BS**s (V_b) is only 32.3% of the *total* **lnRn** variability (V_t, last
548 column of Table 6). This implies that, in SW England, the overall **lnRn** variability ***within* KM1BS**s is about two
549 thirds (100-32.3= 67.7%) of the ***total* lnRn** variation (V_t), that is, most of the **lnRn** variability is observed at
550 ***local*** level.
551 To note that the adj R^2 values reported in Table 6 for data (2) to (6) with **BS**s as geology are quite similar,
552 with an average value of 29.1% and a standard deviation of 2.5%. Three from the five values are between
553 29% and 30% (Figure 9 B), while the other 2 are one underestimation (25.1%) and one overestimation
554 (32.1%, Figure 9 C) relative to the average, which can be seen as a reference value. Furthermore, the
555 average coincide with the value for data (2), 29.0%, which is based on the largest number of **KM1BS**
556 polygons.
557 Values for geological units as **BS**s (shaded values in table 6) are those being considered in the remaining of
558 the present work.
559
560 Table 6 – Proportion of the variation (adjusted R^2 in %) of natural log transformed indoor radon (**lnRn**)
561 explained by Geology, based on three simplified geological classifications, geological period (GPEs), bedrock
562 (BEDs) and bedrock-superficial (BSs) geology polygons, derived from the BGS DiGMapGB-50k.

		Geology as **GPEs**		Geology as **BEDs**		Geology as **BSs**		**KM1BS**	
Data file used	**Number of records**	**Number of units**	**Adj. R^2**	**Number of units**	**Adj. R^2**	**Number of units**	**Adj. R^2**	**Number of units**	**Adj. R^2**
(1) Rn	197464	6	1.4	40	9.2	90	10.1	9144	32.3

(2) Rn (joined by KM1BS)	9144	6	0.6	40	27.9	90	29.0	9144	100
(3) Rn + AGR	8905	6	0.6	37	28.5	82	29.8	-	-
(4) Rn + AGR + SSG	1182	4	1.7	20	24.7	41	25.1	-	-
(5) Rn + AGR + TSG	600	4	~~0.4~~	19	30.7	29	32.1	-	-
(6) Rn + AGR + SSG + TSG	83	3	~~0.2~~	12	31.9	15	29.5	-	-

Records in data (2) to (6) are data polygons, that is, average values per KM1BS polygon. The significance for the adjusted R^2 is $p < 0.001$, except for strikethrough values. Columns in grey highlight the geological classification type (**BS**) that will be used in the following models (Table 7 to Table 9).

FIGURE 9 HERE

3.2. Indoor radon explained by AGR

The $\log_e$ indoor radon (**lnRn**) variance explained (adjusted R^2) by airborne data after regression analysis is shown in Table 7 for six models, either using $\log_{10}$(eU) (**AGR[U]**, Figure 10 A), $\log_{10}$ of total counts (**AGR[TC]**), or the set of elements available, eU, eTh, K, in a compositional (CODA) approach (**[AGR]**, Figure 10 B). For the last model, the isometric logratios of the four part composition are used. Two of the models also include geology (G), one of them shown in Figure 10 C.

For SW England, the **lnRn** variation *between* **KM1BS**s explained by airborne gamma-ray using a CODA approach (**[AGR]**) is about **29%** (Table 7), as this is the value reported from the model with data (3), which have the largest number of **KM1BS**s. The value for data (4) is slightly underestimated (-1.8%) while for data (5) is slightly overestimated (+3.5%, Figure 10 B) relative to the reference value (29%), but both acceptable. The value from data (6) is clearly overestimated (about +17%), indicating that its 83 available records are not statistically representative of the 8905 data (3) polygons which include both indoor radon and airborne information. Consequently, models including data (6) will not be shown further and /or subject to further discussion.

The models including **AGR[U]** (e.g., Figure 10 A), explain a larger portion (+3.1% to +6.6%) of **lnRn** variation than the models with **AGR[TC]**, and a similar but slightly lower (-1.9% to -4.3%) portion than that of **[AGR]**

585 (e.g., Figure 10 B). These values can be checked by comparing the respective columns in Table 7. Also, a
586 maximum of 0.1% is added to the explained variance by appending **AGR[TC]** to the **[AGR]** model (Table 7).
587 When geology (**G**, the **BS**s) is added to the model with CODA airborne data (**[AGR]**), the explained **lnRn**
588 *lateral* variation increase from about 5% to 9% (Table 7, Figure 10 C).
589 These results suggest that the inclusion of Total Counts (**AGR[TC]**) does not enhance equivalent U (**AGR[U]**)
590 or CODA eU, eTh, K (**[AGR]**) models, and that the last is (slightly) preferred to that with eU only (compare
591 Figure 10 A and Figure 10 B). Also, the **[AGR]**+**G** (Figure 10 C) provide a better model than geology (Figure 9
592 C) or airborne gamma-ray alone (Figure 10 B), suggesting that a (small) part of the information provided by
593 these two datasets is complementary to each other in respect to indoor radon.
594 Note that the **AGR[U]** model provides a similar to slightly higher level of **lnRn** explained variation to that
595 obtained by Appleton et al. (2008) in an area of the Northern Ireland (R^2=21%) and slightly lower to that
596 obtained by Scheib et al. (2006) for an area in Central England (R^2=31.4%).
597 **[AGR]** and **[AGR]+G**, with **BS**s used as the geological classes (shaded values in table 7) are those considered
598 in the remaining of the present work.
599
600 Table 7 – Proportion of the variation (adjusted R^2 in %) of natural log transformed indoor radon
601 concentrations (**lnRn**) explained by airborne gamma-ray (**AGR**), and by **AGR** combined with Geology (**G**).

			Adjusted R^2					
Data file used	**Number of records**	**N. of BSs**	**AGR[TC]**	**AGR[U]**	**AGR[U] + G**	**[AGR]**	**[AGR] + AGR[TC]**	**[AGR]+ G**
(3) Rn + AGR	8905	82	21.5	25.7	37.8	29.0	29.1	37.9
(4) Rn + AGR + SSG	1182	41	19.8	22.9	32.3	27.2	27.3	32.6
(5) Rn + AGR + TSG	600	29	24.0	30.5	39.0	32.5	32.4	38.8
~~**(6) Rn + AGR + SSG + TSG**~~	~~83~~	~~15~~	~~31.5~~	~~40.6~~	~~42.3~~	~~45.7~~	~~45.5~~	~~49.4~~

602 **[AGR]**: the isometric logratio transformation of the composition eU, eTh, K and the residual R (Equation 3); **AGR[TC]**: $\log_{10}$ of Total
603 Counts; **AGR[U]**: $\log_{10}$ of equivalent **U**. The significance for the adjusted R^2 is $p < 0.001$ for all models. Models from data file (6) are

strikethrough indicating that they are excluded from further discussion. Columns in grey highlight the predictor combinations (**[AGR]** and **[AGR]+G**) that will be used in the following models (Table 8 and Table 9).

FIGURE 10 HERE

3.3. Indoor radon explained by stream sediment geochemistry

Table 8 summarises the $\log_e$ indoor radon (**lnRn**) variance explained (adjusted R^2) by stream sediment geochemistry after regression analysis with examples of 3 models shown in Figure 11. The **lnRn** models have been developed using stream sediment geochemical data from data (4) (referred to in Table 5, Table 6 and Table 7), and are separated in three groups. A first group of simple models (i) for which only $\log_{10}$(U) was used (**SSG[U]** in Table 8, and e.g., Figure 11 A); a second group of models (ii) for which a sub composition including three elements (K, Th, U in $mg.kg^{-1}$) is used together with the associated residual part (R) (**SSG[UThK]** in Table 8); and a third group of models (iii) for which the available composition of 41 elements (in $mg.kg^{-1}$) is used together with the residual part (R) (**[SSG]** in Table 8, e.g., Figure 11 B). For models with (ii) or (iii), the isometric logratio of the four part sub-composition or the 42 part composition are used respectively.

For SW England, the **lnRn** variation *between* **KM1BS**s explained by **[SSG]** is 32.5% (Table 8, Figure 11 B), which is clearly above the value given by **SSG[U]** (17.8%, Figure 11 A) or by **SSG[UThK]** (20%), and at the same level of the explanation given by **[AGR]+G** (32.6%, data (4) in Table 7). Explained variance is enhanced +4.2% by adding **[SSG]** to the model with **[AGR]+G** (32.6% + 4.2% = 36.8%, Figure 11 C), and only +0.6% by adding **SSG[U]** or **SSG[UThK]** to the same **[AGR]**+**G**.

These results show that the stream sediment geochemistry as a whole and using a CODA approach (**[SSG]**) clearly enhance results provided by uranium alone or U, Th, K in stream sediments, and provide a similar level of explanation given by the **[AGR]**+**G** model. The results also suggest that, concerning indoor radon, a (small) part of the information provided by **[SSG]** is complementary to that from **[AGR]**+**G**.

630 Table 8 – Proportion of the variation (adjusted R^2 in %) of natural log transformed indoor radon
631 concentrations (**lnRn**) explained by stream sediment geochemistry (**SSG**), or **SSG** combined with **AGR**
632 and/or Geology.

SSG elements used	**Adjusted R^2**			
	SSG[U]	**SSG[U]+G**	**SSG[U]+[AGR]**	**SSG[U]+[AGR]+G**
(i) Log10 (U)	17.8	29.1	28.3	33.2
	SSG[UThK]	**SSG[UThK]+G**	**SSG[UThK]+[AGR]**	**SSG[UThK]+[AGR]+G**
(ii) U, Th, K (CODA)	20.0	29.0	28.2	33.2
	[SSG]	**[SSG]+G**	**[SSG]+AGR**	**[SSG]+[AGR]+G**
(iii) ALL 41 SSG elements (CODA)	32.5	34.7	34.7	36.8

633 **SSG[U]**: $\log_{10}$ of uranium concentrations in stream sediments; **SSG[UThK]**: the isometric logratio of the sub composition U, Th, K
634 (including the residual R (Equation 3)) concentrations in stream sediments geochemistry; [**SSG]**: the isometric logratios of the
635 available composition with 10 major+31 trace elements (including the residual R (Equation 3)) concentrations in stream
636 sediments geochemistry. Data used for all models: **1182** records of data **(4)** in Table 5, for which 41 **BS** classes are available;
637 **[AGR]**: airborne gamma-ray; **G**: geology (**BS** classes). The significance for the adjusted R^2 is $p < 0.001$.

638

639 FIGURE 11 HERE

640

641 ## 3.4. Indoor radon explained by topsoil geochemistry

642 Table 9 summarises the $\log_e$ indoor radon (**lnRn**) variance explained (adjusted R^2) by topsoil geochemistry
643 after regression analysis with 3 models shown in Figure 12. The **lnRn** models have been developed using
644 topsoil geochemical data from data (5) (referred in Table 5, Table 6 and Table 7), and are separated in three
645 groups. A first group of simple models (i) for which only $\log_{10}$(U) was used (**TSG[U]** in Table 9, e.g., Figure 12
646 A); a second group of models (ii) for which a sub composition including three elements (K, Th, U in $mg.kg^{-1}$)
647 is used together with the associated residual part (R) (**TSG[UThK]** in Table 9); and a third group of models
648 (iii) for which the available composition of 41 elements (in $mg.kg^{-1}$) are used together with the residual part

649 (R) (**[TSG]** in Table 9, e.g., Figure 12 B). For models with (ii) or (iii), the isometric logratio of the four part sub-
650 composition or the 42 part composition are used respectively.
651 For SW England, the **lnRn** variation *between* **KM1BS**s explained by **[TSG]** is 39.7% (Table 9, Figure 12 B),
652 which is clearly above the value given by **TSG[U]** (14.7%, Figure 12 A) or by **TSG[UThK]** (16.3%), and at the
653 same level of the variation explained by **[AGR]**+**G** (38.8%, data (5) in Table 7). Explained variance is
654 enhanced +7.8% by adding **[TSG]** to the model with **[AGR]+G** (38.8% + 7.8% = 46.6%, Figure 12 C), and only
655 +1.2% or +1.5% by adding **TSG[U]** or **TSG[UThK]** respectively.
656 These results show that the topsoil geochemistry as a whole and using a CODA approach (**[TSG]**) clearly
657 enhances results provided by uranium (**TSG[U]**) or U, Th, K (**TSG[UThK]**) in topsoils. The topsoil "whole"
658 composition model **[TSG]** also show a better performance than the stream sediment one **[SSG]**, which does
659 not happen if only uranium or the sub-composition with U, Th, K are used. Also, concerning to indoor radon,
660 a part of the information provided by **[TSG]** is complementary to that from **[AGR]**+**G**. The **[TSG]+[AGR]+G**
661 seems to provide the best model of all those considered in the present work, with more than **46%** of
662 variance explained (Figure 12 C).
663
664 Table 9 – Proportion of the variation (adjusted R^2 in %) of natural log transformed indoor radon
665 concentrations (**lnRn**) explained by topsoil geochemistry (**TSG**), or **TSG** combined with **AGR** and/or geology
666 (**G**).

TSG elements used	**Adjusted R^2**			
	TSG[U]	**TSG[U]+G**	**TSG[U]+[AGR]**	**TSG[U]+[AGR]+G**
(i) Log10 (U)	14.7	37.0	33.1	40.0
	TSG[UThK]	**TSG[UThK]+G**	**TSG[UThK]+[AGR]**	**TSG[UThK]+[AGR]+G**
(ii) U, Th, K (CODA)	16.3	37.2	33.2	40.3
	[TSG]	**[TSG]+G**	**TSG+[AGR]**	**TSG+[AGR]+G**
(iii) ALL 41 TSG elements (CODA)	39.7	44.5	42.8	46.6

667 **TSG[U]**: $\log_{10}$ of uranium concentrations in topsoils; **TSG[UThK]**: the isometric logratio of the sub composition U, Th, K (including the
668 residual R (Equation 3)) concentrations in topsoils geochemistry; **[TSG]**: the isometric logratios of the available composition with 10

669 major+31 trace elements (including the residual R (Equation 3)) concentrations in topsoils geochemistry. Data used for all models:
670 **600** records of data **(5)** in Table 5, for which 29 **BS**s are available; **[AGR]**: airborne gamma-ray; **G**: geology (**BS** classes). The
671 significance for the adjusted R^2 is $p < 0.001$.
672
673 FIGURE 12 HERE
674
675 **4. DISCUSSION**
676 For SW England, the indoor radon variability is higher *within* **KM1BS** polygons than *between* them. Thus, a
677 wide difference can be observed between neighbouring houses built over the same geology, which reflects
678 the different construction (namely the effectiveness of radon isolation from the ground) and ventilation
679 (poorly ventilated usually induces higher radon concentrations) of the buildings, apart from any possible
680 geological difference at *local* level. The **lnRn** variation observed *between* **KM1BS** polygons (32.3% of the
681 *total* variation, last column of Table 6) can be seen as an estimation of the maximum amount of indoor
682 radon variability which can potentially be explained by the geological (s.l.) datasets here tested.
683 For the SW England, the indoor radon model with uranium (U) in stream sediments (**SSG[U]**, Table 8 and
684 Figure 11 A) performs (slightly) better than uranium (U) in topsoils (**TSG[U]**, Table 9 and Figure 12 A). If Th
685 and K are added to U, for both stream sediments (**SSG[UThK]**, Table 8) and soils (**TSG[UThK]**,Table 9) the
686 model only slightly improves, while the model with stream sediments (20.0%) keeps performing better than
687 that with soils (16.3%). However, geology (ca. 25% to 32%, Table 6, Figure 9) or **[AGR]** (ca. 27% to 32.5%,
688 Table 7, Figure 10) alone or combined (**[AGR]**+**G**, ca. 32.5% to 39%, Table 7) provide a better explanation
689 than any of the models above. Moreover, the U, Th, K sub-composition whether in stream sediments (Table
690 8) or in soils (Table 9) do not provide further explanation (<1.5%) to models including airborne and geology
691 (**[AGR]**+**G**, Table 7). There is an enhanced **lnRn** prediction of the **[AGR]** model (32.5%, Table 7, Figure 10 B)
692 relative to the **TSG[UThK]** (16.3%, Table 9), and that of **AGR[U]** (22.9% to 30.5%, Table 7, Figure 10 A)
693 relative to **SSG[U]** (17.8%, Table 8, Figure 11 A) or **TSG[U]** (14.7%, Table 9, Figure 12 A). In short, these
694 results indicate that the airborne gamma-ray data generally provide better models than the U, Th, K

concentrations in topsoil or stream sediment geochemistry, suggesting that the **AGR** signal may capture other characteristics than just the uranium (and thorium and potassium) concentration in soils, characteristics (e.g. permeability, water content, organic matter?) which are favourable for indoor radon explanation. In fact, the **AGR** data is a geophysical signal corresponding to the radiation emitted from the ground, thus depending on the "radiation permeability". A typical case is the ground covered by peat which returns a depressed **AGR** signal (due to high water content) whatever the bedrock underneath. Also, this good performance of **AGR** also complies with Appleton et al. (2011) which, in an exercise carried out for the Northern Ireland, concluded that, among a set of soil geochemistry and airborne gamma-ray variables, the airborne eU (with 24% of variance explained) was the most significant parameter in modelling the radon potential.

For models with only one dataset, **[TSG]** shows the best performance, with an explained variance of about 40% (Table 9, Figure 12 B), thus clearly above the other datasets, namely **[SSG]** (32.5%, Table 8, Figure 11 B), **[AGR]** (ca. 29.5% ± 2.7%, e.g., Figure 10 B) and **G** (ca. 29.0% ± 2.5%, e.g., Figure 9 B). Similar or improved results are obtained from the **[TSG]** model than that using airborne and geology together (**[AGR]**+**G**, 36.4% ± 3.4%, Table 7, e.g., Figure 10 C) or that using **[SSG]**+**[AGR]**+**G** (36.8%, Table 8, Figure 11 C).

The best explanation of all tested models is given by those including the soil geochemistry (**[TSG]**, i.e. including all the available 41 elements, 10 major – Al, Ca, Fe, K, Mg, Mn, Na, P, Si, Ti - plus 31 trace, in a CODA approach), together with geology (**[TSG]+G**, 44.5%) or geology and **[AGR]** (**[TSG]+G+[AGR]**, 46.6%), shown in Figure 12 C. Models including combinations with **[TSG]** and **[SSG]** together were not included as the number of samples (83) is too low.

Adding geology **G** (**BS** geological classes) to the models **[SSG]** or **[TSG]**, only slightly improves the prediction (+2.2%, +4.8% respectively), suggesting that the "near-complete" geochemistry of stream sediments and of soils has already captured most of the bedrock-superficial (geology) variation. The same happens for **[AGR]** (2.2% for **[SSG]** and 3.1% for **[TSG]** models). This suggest that the information given by the **AGR** is mainly contained within the information given by the soil (or the stream sediment) geochemistry, at least with regard to the explanation of indoor radon.

721 The model with U (as well as the model with U, Th, K) concentration in stream sediments provides a slightly
722 improved outcome than that with U (as well as that with U, Th, K) concentration in topsoil. This may reflect
723 (i) the different processes behind the formation of these two materials, including the mobilization of U from
724 soils to the stream sediments; (ii) the existence of organic peat soils over some (U-rich) granitic areas, which
725 may show low levels of uranium, contrary to the stream sediments derived from the bedrock underneath
726 peat; and (iii) the different size fractions analysed (<2 mm for soils and 150 μm for stream sediments), as the
727 fine fraction often concentrate trace elements. Figure 13, providing information about the stability (CSR, the
728 closest to 1 the more stable is the ratio) and of the median (MDN) of the U (or Th) / K_2O plot, illustrates the
729 net difference in U (and Th) concentrations between topsoils (<2 mm) and stream sediments (<150 μm), as
730 well as over granitic rocks or other rock types.
731 However if the "complete" geochemical composition is used, topsoils (**[TSG]**) perform better than stream
732 sediments (**[SSG]**) and doubles, at least, the explanation provided by **SSG[U]** models. This suggests that,
733 from the point of view of the indoor radon, the information provided by the soil chemistry (of <2 mm size
734 fraction) as a whole is more complete than that provided by the stream sediment (<150 μm), perhaps
735 containing information about the structure of the soil and a more consistent signature of the bedrock.
736 Although, the lower performance of **[SSG]** relative to **[TSG]** models may also be an effect of the expected
737 lower accuracy for parent material classification of stream sediments relative to soil samples. In fact, this
738 lower accuracy may be the reason for the lower **lnRn** *lateral* variation explained by geology observed for
739 data (4) which includes the stream sediment data (25.1%, in Table 6), relative to that for data(5) which
740 includes the topsoil data (32.1%, in Table 6).
741
742 FIGURE 13 HERE
743
744
745 Despite obtaining better results from geochemistry datasets, namely that on soils, this type of data has
746 several obvious disadvantages relative to the other tested datasets (geology and airborne data). Airborne

747 data is usually collected in a much denser set of locations, thus constituting a very rich dataset, with high
748 potential for (very) detailed mapping of the spatial radon variation. However, the land coverage of **AGR**
749 datasets is generally relatively small compared with stream sediment or topsoil geochemical datasets, as is
750 the case in the UK. Geology has an obvious advantage as it is composed of polygons, contrary to the other
751 datasets which are composed by a set of points. This characteristic of the geology allows the point-based
752 information to be extrapolated to areas where these datasets do not exist. However, it seems that the
753 indoor radon variation explained by geology is sensitive to the level of detail used.
754 The ***ilr*** transformation assures that all requirements for regression analysis of compositional data are fulfilled
755 as their coordinates are in an orthogonal system. Yet, expressions for the calculation of ***ilr*** coordinates
756 (isometric logratios) are complex and various rules to generate them exist (Egozcue et al., 2003). Using an
757 adequate compositional regression approach, with ***ilr*** coordinates, the relevance of each original chemical
758 element generally cannot be assessed from the model and their associated tests should therefore be ignored
759 (van den Boogaart & Tolosana-Delgado, 2013).
760 At this point, it should be remembered that detailing the results of each ***ilr*** coordinate (i.e., each ***ilr*** logratio),
761 namely in the **[SSG]** and **[TSG]** cases, in attempting to detect which are the most important elements (other
762 than the obvious U, Th and K) for indoor radon explanation, is out of the scope of the present work.
763 A high number of negative values in an airborne dataset, which may occur after the standard calibrations
764 and data processing (see section 2.3.), may disable the use of a CODA approach, as this is only valid for
765 positive values. This was not the case for the present case study as only 3 records out of 8508 have been
766 removed from data (3). Results (not shown) from **lnRn** models with **AGR** data (eU, eTh, K) after a simple
767 log10-transformation, are virtually the same as those reported in table 7 (i.e., after the CODA transformation
768 used). This suggest that the airborne data is not necessarily compositional in its nature, despite of data
769 being reported with units (as % for K and as ppm for eU and eTh) usually associated to compositional
770 datasets. For the present case study we conclude that there is no obvious benefit on a CODA transformation
771 of the **AGR** dataset.

772 The **AGR[U]**, **SSG[U]** and **TSG[U]**, were developed using a simple classical log10 transformation of the U
773 concentrations only (no other elements were taken into account). This is not critical for **AGR[U]** as the **AGR**
774 data may not be compositional in nature. However, this is important for **SSG[U]** and **TSG[U]** as the **SSG** and
775 the **TSG** datasets are clearly compositional. Thus, it may be concluded that the explained variance obtained
776 for the models **SSG[U]** (Table 8) and **TSG[U]** (Table 9) is biased, that is, those values are not exclusively
777 representative of uranium contents in stream sediments and topsoils respectively.
778
779 **5. CONCLUSIONS**
780 Results from an indoor radon modelling exercise carried out in SW England is presented here. Models were
781 developed using geology, airborne data, topsoil geochemistry and stream sediment geochemistry. The main
782 purpose was to understand how much indoor radon variation can be explained by the above earth science
783 datasets, whether alone or combined. This was achieved by measuring the *lateral* variation of indoor radon,
784 with *lateral* variation referring to the variation observed *between* the **KM1BS**s polygons, which size is equal
785 or less than 1 km^2.
786 For SW England, the indoor radon variability is higher *within* **KM1BS** polygons than *between* them.
787 Results show that any of the tested datasets can be useful for radon mapping purposes as each one is able to
788 explain part of the indoor radon variance observed *between* **KM1BS**s. The stream sediment geochemistry,
789 provided that they refer to small watershed areas, can be seen as a valid alternative to topsoil geochemistry,
790 namely for areas where this last type of data is absence.
791 By combining two or more datasets, models are often enhanced but the variance explained never doubles,
792 meaning that a large portion of the "information" provided by each one of the datasets is already contained
793 in the "information" provided by the other datasets. The best model indicates that almost half (ca. 47%) of
794 the **lnRn** variation *between* **KM1BS**s in SW England can be explained by combining soil geochemistry (**[TSG]**),
795 airborne data (**[AGR]**) and geology (**G**).
796 A "complete" topsoil (or stream sediment) geochemistry (in a CODA approach) is preferable to the use of
797 uranium concentrations only. Not only because uranium concentration in a sample is dependent on the

concentrations of the other elements (leading to the CODA approach), but also because the information provided by the other elements may reflect other soil (stream sediment) characteristics (such as grain size, element's associations) which are potentially helpful in explaining indoor radon. This also applies to the U, Th, K sub-composition.

Using "complete" airborne data (eU, eTh, K) is preferable to the use of eU only. There is no obvious benefit on using a CODA transformation of the **AGR** dataset.

The present work was a first step to modelling indoor radon in the SW England using a multi-dataset multivariate CODA approach. The present work was focused on the significance of each dataset's model, as a whole, rather than evaluating the significance of each ***ilr*** coordinate or, in the case of stream sediments and topsoils, searching for elements, other than U, Th, K that may contribute for the explanation of indoor radon. Further work is planned to address this last point. Mapping these models is another point to be addressed in the near future.

Acknowledgments

This paper is published with the permission of the Executive Director, British Geological Survey and with the permission of the Public Health England. The authors thank to the 2 anonymous reviewers for useful comments and suggestions that certainly improved the manuscript.

REFERENCES

Aitchison, J., **1986**. The statistical analysis of compositional data. London: Chapman & Hall; 416 pp.

Appleton, J.D., **2013.** Radon in air and water. In: Selinus, O.; Alloway, B.J.; Smedley, P., (Eds.) Essentials of Medical Geology. Dordrecht, Netherlands, Springer, 239-277.

Appleton J.D., Daraktchieva Z., Young M.E., **2015**. Geological controls on radon potential in Northern Ireland. *Proceedings of the Geologists' Association*, 126, 328-345

Appleton JD, Miles JCH, Green BMR, Larmour R., **2008**. Pilot study of the application of Tellus airborne radiometric and soil geochemical data for radon mapping. *Journal of Environmental Radioactivity*, 99, 1687-1697.

Appleton, J.D., Miles, J.C.H., **2010**. A statistical evaluation of the geogenic controls on indoor radon concentrations and radon risk. *Journal of Environmental Radioactivity*, 101 (10), 799-803.

Appleton, J.D., Miles, J.C.H., Young, M., **2011**. Comparison of Northern Ireland radon maps based on indoor radon measurements and geology with maps derived by predictive modelling of airborne radiometric and ground permeability data. *Science of the Total Environment*, 409, 1572-1583.

Beamish, D., Howard, A., Ward, E.K., White, J., Young, M. E., **2014**. Tellus South West airborne geophysical data. NERC-Environmental Information Data Centre. Doi: 10.5285/73848363-57c1-480a-a64e-c732e15c4b37

van den Boogaart, K.G. and Tolosana-Delgado, R., **2013**. Analyzing Compositional Data with R, Use R!. Springer-Verlag Berlin Heidelberg DOI 10.1007/978-3-642-36809-7

Bossew, P., Dubois, G., Tollefsen, T., **2008**. Investigations on indoor Radon in Austria, part 2: Geological classes as categorical external drift for spatial modelling of the Radon potential. *Journal of Environmental Radioactivity*, 99, 81-97.

847 Buccianti, A., Mateu-Figueras, G. & Pawlowsky-Glahn, V., **2006**. Compositional Data Analysis in the
848 Geosciences: From Theory to Practice. *Geological Society, London, Special Publications*, 264. The Geological
849 Society of London.
850
851 Buccianti A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R. and De Vivo, B., **2015**. Exploring topsoil
852 geochemistry from the CoDA (Compositional Data Analysis) perspective: The multi-element data archive of
853 the Campania Region (Southern Italy). *Journal of Geochemical Exploration*, 159, 302-316
854
855 Daraktchieva Z., Appleton J.D., Rees D.M., Adlam K.A.M., Myers A.H., Hodgson S.A., McColl N.P., Wasson
856 G.R. and Peake L.J., **2015**. Radon in Northern Ireland: Indicative Atlas. PHE-CRCE-017 14 pp. Public Health
857 England, Chilton, Oxfordshire, ISBN: 978-0-85951-764-5
858
859 Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barcel´o-Vidal, C., **2003**. Isometric logratio
860 transformations for compositional data analysis. *Mathematical Geology*, 35(3), 279-300.
861
862 Filzmoser, P, Hron, K., Reimann, C., **2010**. The bivariate statistical analysis of environmental (compositional)
863 data. *Science of the Total Environment*, 409, 4230-4238.
864
865 HPA (Health Protection Agency), **2010**. Limitation of human exposure to radon. HPA RCE-15, 20 pp.
866
867 Hunter, N., Muirhead, C.R., Miles, J.C.H., **2009**. Uncertainties in radon related to house-specific factors and
868 proximity to geological boundaries in England. *Radiation Protection Dosimetry*, 136 (1), 17-22.
869
870 IAEA, **2003**. Guidelines for Radioelement Mapping Using Gamma Ray Spectrometry Data. IAEA-TECDOC-
871 1363. International Atomic Energy Agency, Vienna, Austria.
872

873 Kirkwood, C., Everett, P., Ferreira, A., Lister, T.R., **2016**. Stream sediment geochemistry as a tool for
874 enhancing geological understanding: An overview of new data from south west England. *Journal of*
875 *Geochemical Exploration*, 163, 28–40.
876
877 Kirkwood, C., Cave M., Beamish D., Grebby S., Ferreira A., **submitted**. A machine learning approach to
878 geochemical mapping. Submitted to the *Journal of Geochemical Exploration*.
879
880 Miles, J.C.H., **1998**. Mapping radon prone areas by lognormal modelling of house radon data. *Health Physics*
881 74: 370-378.
882
883 Miles, J.C.H., Appleton, J.D., **2005**. Mapping variation in radon potential both between and within geological
884 units. *Journal of Radiological Protection*, 25, 257-276.
885
886 Miles, J.C.H., Appleton, J.D., Rees, D.M., Green, B.M.R., Adlam, K.A.M., Myers, A.H., **2007**. Indicative Atlas of
887 Radon in England and Wales. HPA-RPD-033 29 pp. Health Protection Agency, Chilton, Oxfordshire,
888
889 Miles, J.C.H., Appleton, J.D., Rees, D.M., Adlam, K.A.M., Scheib, C., Myers, A.H., Green, B.M.R., McColl, N.P.,
890 **2011**. Indicative Atlas of Radon in Scotland. HPA-CRCE-023 HPA, Chilton, UK.
891
892 Minty, B.R.S.; Luyendyk, A.P.J.; Brodie, R.C., **1997**. Calibration and data processing for airborne gamma-ray
893 spectrometry. AGSO *J. Aust. Geol. Geophys.*, 17, 51–62.
894
895 McKinley J.M., Hron K., Grunsky E.C., Reimann C., de Caritat P., Filzmoser P., van den Boogaart K.G. and
896 Tolosana-Delgado R., **2015**. The single component geochemical map: Fact or fiction? *Journal of Geochemical*
897 *Exploration*, 162, 16-28.
898

National Radiological Protection Board, **1990**. Limitation of Human Exposure to Radon in Homes. Documents of the NRPB: Volume 1, No. 1, ISBN: 0-89591-322-X.

Pawlowsky-Glahn, V. & Buccianti, A., **2011.** Compositional Data Analysis: Theory and Applications. Wiley.

Pawlowsky-Glahn, V., Egozcue, J.J., **2006**. Compositional data and their analysis: an introduction. In: Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (eds) Compositional Data Analysis in the Geosciences: From Theory to Practice. *Geological Society, London, Special Publications*, 264, 1-10. The Geological Society of London.

Pawlowsky-Glahn V., Egozcue, J.J., **2015**. Spatial analysis of compositional data: A historical review. *Journal of Geochemical Exploration*.

Reimann C, Filzmoser P, Fabian K, Hron K, Birke M, Demetriades A, Dinelli E, Ladenberger A and The GEMAS Project Team, **2012**. The concept of compositional data analysis in practice — Total major element concentrations in agricultural and grazing land soils of Europe. *Science of the Total Environment*, 426, 196-210

Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (Eds.)., **2014**. Chemistry of Europe's Agricultural Soils. Part A: Methodology and Interpretation of the GEMAS Data Set. Geol. Jb. B 102: 528 pp. 358 figs., 86 Tables, 1 DVD; Hannover.

Scheib, C., Appleton, J.D., Jones, D., Hodgkinson, E., **2006**. Airborne gamma spectrometry, soil geochemistry and permeability index data in support of radon potential mapping in Central England. In: Barnet, I., Neznal, M., Pacherova, P. (Eds.), Proceedings of the 8th International Workshop on the Geological Aspect of Radon

924 Risk Mapping, 26–30 September 2006, Prague, Czech Republic. Czech Geological Survey, RADON Corp.,
925 Prague, 210-219.
926
927 Scheib, C., Appleton, J.D., Miles, J.C.H., Green, B.M.R., Barlow, T.S., Jones, D., **2009**. Geological controls on
928 radon potential in Scotland. *Scott. J. Geol.*, 45 (2), 147–160.
929
930 Scheib, C., Appleton J.D., Miles, J.C.H. Hodgkinson, E., **2013**. Geological controls on radon potential in
931 England. *Proceedings of the Geologists' Association*, 124, 910-928.
932
933 Tolosana-Delgado R., 2012. Uses and misuses of compositional data in sedimentology. *Sedimentary Geology*,
934 280, 60-79.
935
936 Tondeur F., Cinelli G., Dehandschutter B., **2014**. Homogeneity of geological units with respect to the radon
937 risk in the Walloon region of Belgium. *Journal of Environmental Radioactivity*, 136, 140-151.
938
939 Varley, N. and Flowers, A., **1998**. Indoor Radon Prediction from Soil Gas Measurements. *Health Physics*, 74
940 (6), 714-718.
941
942
943 **FIGURE CAPTIONS**
944
945 **Figure 1**. Geological Periods in SW England, with main faults (black lines) and U (large red dots) and other
946 mines (small black dots). The Carboniferous-Permian units (reddish) essentially refer to the U-rich granite
947 outcrops (the major five are labelled in *brown italic*) referred in the text.
948
949 **Figure 2**. The 42 simplified bedrock (BED) and 8 simplified superficial (SUP) units from which derived the 134
950 simplified bedrock-superficial geological combinations (**BS**s) of SW England (based on the BGS DiGMapGB-
951 50k) used in the present study. The major five granite outcrops are labelled in *brown italic*.

952

953 **Figure 3**. The spatial distribution of the 197464 indoor radon measurements throughout 9144 KM1BS
954 polygons (out of the total 27633) in the SW England. The underlying greyscale map refers to the BS
955 geological combinations.

956

957 **Figure 4**. The spatial distribution of indoor radon concentrations (in $Bq.m^{-3}$) in SW England after allocating
958 the each measurement to the underlying KM1BS polygon. Each concentration value refers to the geometric
959 mean (GM) of measurements allocated to a KM1BS polygon. The histogram (*bars*) refers to the 197,464
960 indoor radon measurements, while the histogram density *curve* refers to the 9,144 KM1BS GM values. The
961 underlying greyscale map refers to the BS geological combinations.

962

963 **Figure 5**. Spatial distribution of equivalent U (eU) concentrations ($mg.kg^{-1}$) from airborne gamma-ray
964 spectroscopy (AGR). The total number of data points in the map is 684,384.

965

966 **Figure 6**. Spatial distribution of equivalent U (eU) concentrations ($mg.kg^{-1}$). Values are the arithmetic mean
967 of eU data points (Figure 5) in each KM1BS polygon. The total number of KM1BS polygons with AGR data in
968 the map is 23,573.

969

970 **Figure 7**. U content ($mg.kg^{-1}$) in 3382 stream sediment samples from the SW England. Data from the G-BASE
971 project. The underlying greyscale map refers to the BS geological combinations.

972

973 **Figure 8**. U content ($mg.kg^{-1}$) in 987 topsoil samples from the SW England. Data from the G-BASE project and
974 the NSI inventory. The underlying greyscale map refers to the BS geological combinations.

975

976 **Figure 9**: **lnRn** (Y axis) modelled by Geology (G the BS geological classes derived from the BGS DiGMapGB-
977 50k), using **A) 197484** log indoor radon measurements used (data (1) of Table 5), 10.1% of the *total* **lnRn**
978 variance explained; **B) 8905** records of log indoor radon averages at KM1BS (data (3) of Table 5), 29.8% of
979 the *lateral* **lnRn** variance explained; **C) 600** records of log indoor radon averages at KM1BS (data (5) of Table
980 5), 32.1% of the *lateral* **lnRn** variance explained. The X axis units are not relevant, corresponding to BS
981 geological classes recoded to numbers.

982

983 **Figure 10**: **lnRn** (Y axis) modelled by airborne gamma ray data (AGR) using **600** records (data (5) of Table 5)
984 predicted by **A)** the **$\log_{10}$(eU)** (with eU concentrations in $mg.kg^{-1}$), 30.5% of the *lateral* **lnRn** variance
985 explained; **B)** the **[AGR]**, i.e., the 3 isometric logratios of the 4-part composition (eU, eTh, K and the residual
986 R (Equation 3) with elements concentrations in $mg.kg^{-1}$), 32.5% of the *lateral* **lnRn** variance explained; **C)** the

[AGR]+**G**, with [AGR] as in graph B) and G as the BS geological classes derived from the BGS DiGMapGB-50k, 32.5% of the *lateral* **lnRn** variance explained. The X axis units for B) and C) graphs are dimensionless (the logarithm of the ratio between two or more elements) and not relevant in compositional data analysis.

Figure 11: **lnRn** (Y axis) modelled by stream sediment geochemistry (**SSG**) using **1182** records (data (4) of Table 5) predicted by **A)** the **log_{10}(U)** (with U concentrations in $mg.kg^{-1}$), 17.8% of the *lateral* **lnRn** variance explained; **B) [SSG]**, i.e., the 41 isometric logratios of the 42-part **SSG** composition (41 elements and residual R (Equation 3) with elements concentrations in $mg.kg^{-1}$), 32.5% of the *lateral* **lnRn** variance explained; **C) [SSG]**+**[AGR]+G,** with **[SSG]** as in graph B), [AGR] as the 3 isometric logratios of the 4-part composition (eU, eTh, K and the residual R (Equation 3) with elements concentrations in $mg.kg^{-1}$) and G as the BS geological classes derived from the BGS DiGMapGB-50k, 36.8% of the *lateral* **lnRn** variance explained. The X axis units for B) and C) graphs are dimensionless (the logarithm of the ratio between two or more elements) and not relevant in compositional data analysis.

Figure 12: **lnRn** (Y axis) modelled by topsoil geochemistry (**TSG**) using **600** records (data (5) of Table 5), predicted by **A)** the **log_{10}(U)** (with U concentrations in $mg.kg^{-1}$), 14.7% of the *lateral* **lnRn** variance explained; **B) [TSG]**, i.e., the 41 isometric logratios of the 42-part **TSG** composition (41 elements and the residual R (Equation 3) with elements concentrations in $mg.kg^{-1}$), 39.7% of the *lateral* **lnRn** variance explained; **C) [TSG]**+**[AGR]+G**, with **[TSG]** as in graph B), [AGR] as the 3 isometric logratios of the 4-part composition (eU, eTh, K and the residual R (Equation 3) with elements concentrations in $mg.kg^{-1}$) and G as the BS geological classes derived from the BGS DiGMapGB-50k, 46.6% of the *lateral* **lnRn** variance explained. The X axis units for B) and C) graphs are dimensionless (the logarithm of the ratio between two or more elements) and not relevant in compositional data analysis.

Figure 13: scatterplots of K_2O and Th (A, B) and K_2O and U (C, D) in soils (A, C) and stream sediments (B, D) from SW England. Red dots represent samples over Carboniferous-Permian granitic rocks (CG); black dots represent samples over the remaining bedrock units (OB). CSR: coefficient of stability of the ratio (Filzmoser et al., 2010); MDN: median; N: number of samples. The dashed lines indicate a constant ratio (corresponding to the median of the log-ratio) among the pair, while the solid lines indicate a ratio 2 times higher (upper line) or one half (lower line) the constant ratio for samples over OB (in grey) and CG (in red).

SUPPLEMENTARY MATERIAL

Extra Table A – Summary statistics for 41 elements from **stream sediment geochemistry** of the SW England (N= 3,382 data points) and after taking the average of the data points in each KM1BS polygon (N = 3,027 data polygons)

	N	**min**	**Q25**	**mdn**	**Q75**	**Q95**	**Q99**	**max**	IQR	mean	GM	GSD
Al_2O_3	**3382**	**6.20**	**13.99**	**16.29**	**18.59**	**21.29**	**23.09**	**28.19**	4.60	16.26	**15.93**	1.23
Al_2O_3	**3027**	**6.20**	**13.99**	**16.29**	**18.59**	**21.29**	**23.08**	**28.19**	4.60	16.26	**15.94**	1.23
CaO	**3382**	**0.25**	**0.49**	**0.55**	**0.69**	**1.25**	**2.84**	**7.82**	0.20	0.67	**0.61**	1.47
CaO	**3027**	**0.28**	**0.49**	**0.56**	**0.70**	**1.29**	**2.89**	**7.82**	0.21	0.69	**0.62**	1.48
Fe_2O_3	**3382**	**1.00**	**4.99**	**6.17**	**7.44**	**9.68**	**12.34**	**22.08**	2.45	6.28	**5.93**	1.42
Fe_2O_3	**3027**	**1.00**	**5.01**	**6.17**	**7.45**	**9.77**	**12.37**	**22.08**	2.44	6.30	**5.94**	1.42
K_2O	**3382**	**0.53**	**2.33**	**2.75**	**3.18**	**3.92**	**4.53**	**6.78**	0.85	2.78	**2.69**	1.29
K_2O	**3027**	**0.53**	**2.33**	**2.75**	**3.19**	**3.91**	**4.48**	**6.17**	0.86	2.78	**2.70**	1.28
MgO	**3382**	**0.70**	**1.00**	**1.20**	**1.50**	**2.00**	**3.72**	**15.10**	0.50	1.30	**1.23**	1.36
MgO	**3027**	**0.70**	**1.00**	**1.20**	**1.50**	**2.00**	**3.87**	**15.10**	0.50	1.31	**1.24**	1.37
MnO	**3382**	**0.02**	**0.12**	**0.18**	**0.27**	**0.60**	**1.32**	**8.90**	0.15	0.24	**0.18**	2.06
MnO	**3027**	**0.02**	**0.12**	**0.18**	**0.27**	**0.60**	**1.35**	**6.56**	0.15	0.24	**0.18**	2.04
Na_2O	**3382**	**0.20**	**0.70**	**0.90**	**1.00**	**1.30**	**1.80**	**3.00**	0.30	0.87	**0.83**	1.38
Na_2O	**3027**	**0.20**	**0.70**	**0.90**	**1.00**	**1.30**	**1.80**	**3.00**	0.30	0.87	**0.82**	1.38
P_2O_5	**3382**	**0.03**	**0.16**	**0.20**	**0.25**	**0.39**	**0.53**	**0.85**	0.09	0.21	**0.20**	1.45
P_2O_5	**3027**	**0.03**	**0.16**	**0.20**	**0.25**	**0.39**	**0.53**	**0.85**	0.09	0.22	**0.20**	1.44
SiO_2	**3382**	**12.69**	**54.47**	**59.34**	**63.24**	**69.23**	**74.73**	**85.61**	8.77	58.57	**58.03**	1.15
SiO_2	**3027**	**16.98**	**54.25**	**59.24**	**63.14**	**68.70**	**74.66**	**85.61**	8.89	58.41	**57.89**	1.15
TiO_2	**3382**	**0.24**	**0.80**	**0.86**	**0.92**	**1.14**	**1.73**	**3.37**	0.11	0.87	**0.85**	1.23
TiO_2	**3027**	**0.24**	**0.80**	**0.86**	**0.92**	**1.15**	**1.77**	**3.37**	0.12	0.87	**0.85**	1.23
As	**3382**	**1.40**	**11.10**	**16.90**	**37.10**	**197.8**	**786.7**	**11000**	26.00	65.05	**23.07**	2.97
As	**3027**	**1.40**	**11.20**	**17.15**	**39.20**	**200.5**	**844.7**	**11000**	28.00	67.27	**23.64**	3.00
Ba	**3382**	**101.0**	**377.3**	**444.0**	**492.0**	**569.0**	**642.2**	**5000**	114.8	432.7	**417.4**	1.31
Ba	**3027**	**142.0**	**375.0**	**444.0**	**491.0**	**568.0**	**636.5**	**5000**	116.0	431.5	**416.2**	1.31
Br	**3382**	**0.50**	**9.70**	**15.90**	**29.00**	**67.30**	**139.5**	**427.0**	19.30	24.21	**17.11**	2.22
Br	**3027**	**0.50**	**9.80**	**16.10**	**29.30**	**67.27**	**137.7**	**427.0**	19.50	24.43	**17.37**	2.21
Ce	**3382**	**22.0**	**67.0**	**75.0**	**88.0**	**154.0**	**261.6**	**1940**	21.0	86.4	**80.4**	1.39
Ce	**3027**	**22.0**	**67.0**	**75.0**	**88.8**	**152.7**	**264.7**	**1940**	21.8	86.7	**80.6**	1.39
Co	**3382**	**1.8**	**16.1**	**21.7**	**30.3**	**60.4**	**106.0**	**1926**	14.2	27.0	**21.8**	1.85
Co	**3027**	**2.4**	**16.2**	**21.8**	**30.3**	**60.7**	**106.3**	**984.6**	14.1	26.9	**21.9**	1.85
Cr	**3382**	**2.10**	**73.40**	**86.80**	**102.8**	**127.8**	**274.7**	**4097**	29.40	93.76	**81.20**	1.65
Cr	**3027**	**7.90**	**73.60**	**86.93**	**102.8**	**129.2**	**280.6**	**4097**	29.20	94.52	**81.46**	1.65
Cs	**3382**	**1.00**	**7.00**	**10.00**	**18.00**	**40.00**	**66.00**	**153.00**	11.00	14.20	**10.47**	2.16
Cs	**3027**	**1.00**	**7.00**	**10.00**	**18.00**	**40.00**	**64.74**	**153.00**	11.00	14.36	**10.64**	2.15
Cu	**3382**	**3.70**	**17.80**	**24.10**	**34.00**	**91.77**	**635.8**	**12000**	16.20	59.50	**26.94**	2.15
Cu	**3027**	**3.70**	**17.93**	**24.50**	**34.63**	**95.04**	**780.6**	**8000**	16.70	60.69	**27.54**	2.17
Ga	**3382**	**2.60**	**16.70**	**20.20**	**23.90**	**29.40**	**35.80**	**67.90**	7.20	20.46	**19.65**	1.34
Ga	**3027**	**4.60**	**16.70**	**20.20**	**23.90**	**29.40**	**35.80**	**67.90**	7.20	20.53	**19.74**	1.33
Ge	**3382**	**0.10**	**0.80**	**1.00**	**1.40**	**1.80**	**2.30**	**7.70**	0.60	1.06	**0.94**	1.74
Ge	**3027**	**0.10**	**0.80**	**1.00**	**1.33**	**1.80**	**2.30**	**7.70**	0.53	1.07	**0.95**	1.72
Hf	**3382**	**0.50**	**6.50**	**8.60**	**11.80**	**26.10**	**95.07**	**1014**	5.30	12.73	**9.30**	1.84
Hf	**3027**	**0.50**	**6.50**	**8.60**	**11.70**	**25.08**	**101.8**	**548.7**	5.20	12.50	**9.25**	1.83
La	**3382**	**10.00**	**37.00**	**42.00**	**49.75**	**84.00**	**130.2**	**320.0**	12.75	47.12	**44.37**	1.38
La	**3027**	**10.00**	**37.00**	**42.00**	**50.00**	**84.70**	**131.0**	**318.0**	13.00	47.34	**44.54**	1.38
Mo	**3382**	**0.10**	**0.60**	**0.90**	**1.50**	**3.40**	**6.00**	**11.70**	0.90	1.29	**1.00**	1.96
Mo	**3027**	**0.10**	**0.60**	**0.90**	**1.50**	**3.30**	**6.00**	**11.70**	0.90	1.28	**1.00**	1.95
Nb	**3382**	**8.40**	**16.40**	**17.40**	**19.00**	**32.40**	**50.88**	**350.0**	2.60	19.73	**18.73**	1.31

Nb	**3027**	**8.40**	**16.50**	**17.40**	**19.10**	**32.70**	**50.77**	**350.0**	2.60	19.84	**18.80**	1.32
Nd	**3382**	**3.00**	**31.40**	**35.70**	**41.60**	**71.60**	**118.7**	**313.2**	10.20	39.91	**37.44**	1.39
Nd	**3027**	**3.00**	**31.50**	**35.90**	**41.80**	**72.67**	**122.4**	**313.2**	10.30	40.12	**37.60**	1.40
Ni	**3382**	**3.80**	**32.73**	**44.10**	**57.50**	**96.09**	**173.3**	**1689**	24.78	49.91	**40.23**	1.94
Ni	**3027**	**4.10**	**32.90**	**44.40**	**57.80**	**97.05**	**178.7**	**1041**	24.90	50.13	**40.43**	1.95
Pb	**3382**	**5.00**	**26.43**	**34.10**	**50.58**	**121.0**	**481.1**	**10000**	24.15	68.65	**39.87**	1.95
Pb	**3027**	**5.00**	**26.80**	**34.60**	**51.45**	**125.5**	**508.4**	**10000**	24.65	70.04	**40.54**	1.96
Rb	**3382**	**27.20**	**99.70**	**126.9**	**164.6**	**308.1**	**477.0**	**916.4**	64.88	147.07	**131.26**	1.58
Rb	**3027**	**27.20**	**100.2**	**127.8**	**167.4**	**307.0**	**475.3**	**916.4**	67.20	147.84	**132.12**	1.58
Sb	**3382**	**0.10**	**0.70**	**1.10**	**2.00**	**6.60**	**16.34**	**128.3**	1.30	2.13	**1.19**	2.62
Sb	**3027**	**0.10**	**0.70**	**1.10**	**2.00**	**6.70**	**16.64**	**128.3**	1.30	2.16	**1.21**	2.59
Sc	**3382**	**0.80**	**10.40**	**13.30**	**16.10**	**19.80**	**23.00**	**34.50**	5.70	13.25	**12.51**	1.44
Sc	**3027**	**0.80**	**10.40**	**13.35**	**16.00**	**19.70**	**23.00**	**34.50**	5.60	13.28	**12.53**	1.43
Se	**3382**	**0.10**	**0.50**	**0.80**	**1.40**	**2.81**	**4.83**	**11.63**	0.90	1.09	**0.79**	2.28
Se	**3027**	**0.10**	**0.50**	**0.90**	**1.40**	**2.80**	**4.87**	**11.60**	0.90	1.11	**0.81**	2.27
Sm	**3382**	**0.10**	**4.60**	**6.00**	**8.00**	**14.10**	**22.21**	**72.60**	3.40	6.86	**5.96**	1.74
Sm	**3027**	**0.10**	**4.70**	**6.00**	**7.90**	**14.24**	**22.85**	**72.60**	3.20	6.89	**6.00**	1.73
Sn	**3382**	**1.00**	**3.40**	**5.50**	**62.00**	**1571**	**2000**	**2212**	58.60	190.66	**16.05**	8.08
Sn	**3027**	**1.00**	**3.50**	**5.90**	**75.85**	**1627**	**2000**	**2212**	72.35	199.10	**17.17**	8.24
Sr	**3382**	**18.80**	**66.40**	**80.70**	**97.28**	**129.1**	**164.6**	**334.2**	30.88	83.06	**78.60**	1.41
Sr	**3027**	**18.80**	**66.65**	**80.70**	**97.10**	**129.5**	**165.2**	**334.2**	30.45	83.27	**78.93**	1.40
Th	**3382**	**2.60**	**9.80**	**11.30**	**12.90**	**43.38**	**77.24**	**321.7**	3.10	15.01	**12.53**	1.64
Th	**3027**	**2.60**	**9.88**	**11.30**	**13.00**	**44.20**	**76.41**	**321.7**	3.13	15.07	**12.58**	1.64
U	**3382**	**1.00**	**2.50**	**2.90**	**3.50**	**24.48**	**44.44**	**206.9**	1.00	5.70	**3.64**	2.10
U	**3027**	**1.00**	**2.50**	**2.90**	**3.50**	**23.91**	**43.87**	**206.9**	1.00	5.70	**3.66**	2.11
V	**3382**	**10.30**	**81.60**	**105.1**	**131.0**	**160.0**	**191.7**	**289.1**	49.38	104.7	**96.77**	1.53
V	**3027**	**20.40**	**82.05**	**105.0**	**131.5**	**160.0**	**193.1**	**289.1**	49.45	104.9	**96.98**	1.53
W	**3382**	**0.10**	**2.30**	**3.10**	**5.40**	**53.59**	**153.0**	**2000**	3.10	14.15	**4.61**	3.18
W	**3027**	**0.10**	**2.30**	**3.10**	**6.00**	**55.21**	**157.7**	**1314**	3.70	14.35	**4.74**	3.22
Y	**3382**	**12.60**	**28.30**	**30.30**	**34.20**	**67.79**	**113.2**	**531.8**	5.90	35.43	**33.06**	1.37
Y	**3027**	**12.60**	**28.30**	**30.30**	**34.30**	**67.94**	**114.3**	**531.8**	6.00	35.36	**33.06**	1.37
Zn	**3382**	**20.00**	**98.33**	**127.7**	**175.2**	**427.7**	**1300**	**12000**	76.85	192.4	**138.7**	1.88
Zn	**3027**	**20.00**	**99.30**	**128.7**	**176.9**	**445.3**	**1408**	**7594**	77.60	194.7	**140.5**	1.88
Zr	**3382**	**30.00**	**240.0**	**329.8**	**438.1**	**871.1**	**2000**	**2000**	198.1	402.1	**343.8**	1.66
Zr	**3027**	**68.30**	**239.9**	**327.9**	**433.1**	**852.1**	**2000**	**2000**	193.2	398.3	**341.9**	1.65

N: number of samples; **Min**: minimum value; **Max**: maximum value; **Q##**: ##th quantile; **Mdn**: median; **IQR**: interquartile range; **AM**: arithmetic mean; **GM**: geometric mean; **GSD**: geometric standard deviation. Units (in brackets) are valid for all parameters except **GSD** (dimensionless).

Extra Table B – Summary statistics for 41 elements from **top soil geochemistry** of the SW England (N= 987 data points) and after taking the average of the data points in each KM1BS polygon (N = 947 data polygons)

	N	**min**	**Q25**	**mdn**	**Q75**	**Q95**	**Q99**	**max**	IQR	mean	GM	GSD
Al_2O_3	**987**	**0.30**	**12.50**	**15.00**	**17.29**	**20.30**	**21.90**	**23.00**	4.79	14.45	**13.41**	1.62
Al_2O_3	**947**	**0.30**	**12.50**	**15.00**	**17.29**	**20.30**	**21.80**	**23.00**	4.80	14.44	**13.40**	1.62
CaO	**987**	**0.04**	**0.40**	**0.53**	**0.68**	**1.60**	**4.34**	**26.73**	0.28	0.74	**0.53**	1.96
CaO	**947**	**0.04**	**0.40**	**0.53**	**0.68**	**1.67**	**4.36**	**26.73**	0.29	0.74	**0.53**	1.98
Fe_2O_3	**987**	**0.29**	**3.86**	**5.49**	**6.73**	**8.15**	**10.34**	**12.06**	2.87	5.19	**4.44**	1.95
Fe_2O_3	**947**	**0.29**	**3.87**	**5.49**	**6.73**	**8.15**	**10.32**	**11.80**	2.86	5.20	**4.46**	1.94
K_2O	**987**	**0.01**	**1.96**	**2.47**	**2.96**	**3.43**	**3.96**	**4.94**	1.01	2.41	**2.20**	1.84
K_2O	**947**	**0.01**	**1.95**	**2.47**	**2.96**	**3.43**	**3.92**	**4.94**	1.01	2.41	**2.19**	1.86
MgO	**987**	**0.20**	**0.80**	**1.00**	**1.30**	**1.80**	**3.21**	**6.90**	0.50	1.10	**1.00**	1.56
MgO	**947**	**0.20**	**0.80**	**1.00**	**1.30**	**1.80**	**3.25**	**6.90**	0.50	1.10	**1.00**	1.57

MnO	**987**	**0.00**	**0.05**	**0.12**	**0.18**	**0.30**	**0.58**	**1.75**	0.13	0.14	**0.09**	2.83
MnO	**947**	**0.00**	**0.06**	**0.12**	**0.18**	**0.30**	**0.58**	**1.75**	0.13	0.14	**0.09**	2.83
Na_2O	**987**	**0.10**	**0.50**	**0.70**	**0.90**	**1.20**	**1.50**	**2.70**	0.40	0.73	**0.66**	1.58
Na_2O	**947**	**0.10**	**0.50**	**0.70**	**0.90**	**1.20**	**1.50**	**2.70**	0.40	0.72	**0.66**	1.59
P_2O_5	**987**	**0.02**	**0.21**	**0.28**	**0.35**	**0.50**	**0.66**	**0.82**	0.14	0.29	**0.26**	1.61
P_2O_5	**947**	**0.02**	**0.21**	**0.28**	**0.35**	**0.50**	**0.67**	**0.82**	0.14	0.29	**0.26**	1.61
SiO_2	**987**	**1.70**	**52.30**	**57.80**	**62.50**	**69.57**	**73.03**	**79.70**	10.20	55.96	**53.13**	1.54
SiO_2	**947**	**1.70**	**52.28**	**57.80**	**62.57**	**69.52**	**73.11**	**79.70**	10.30	55.94	**53.07**	1.55
TiO_2	**987**	**0.00**	**0.70**	**0.83**	**0.90**	**1.14**	**1.90**	**2.64**	0.20	0.80	**0.72**	1.75
TiO_2	**947**	**0.00**	**0.70**	**0.83**	**0.90**	**1.12**	**1.87**	**2.64**	0.20	0.80	**0.72**	1.76
As	**987**	**4.70**	**15.70**	**23.50**	**39.70**	**134.4**	**356.6**	**1949.2**	24.00	46.50	**28.09**	2.29
As	**947**	**4.70**	**15.81**	**23.60**	**39.80**	**134.4**	**351.8**	**1949.2**	24.00	46.00	**28.18**	2.27
Ba	**987**	**13.20**	**284.0**	**374.6**	**435.5**	**514.0**	**556.0**	**641.0**	151.5	354.7	**326.5**	1.63
Ba	**947**	**13.17**	**284.0**	**374.2**	**435.9**	**513.7**	**556.0**	**610.0**	152.0	354.1	**325.5**	1.64
Br	**987**	**4.00**	**24.20**	**32.60**	**44.95**	**131.2**	**254.0**	**459.1**	20.75	44.95	**35.10**	1.88
Br	**947**	**4.00**	**24.66**	**32.90**	**44.80**	**130.5**	**256.1**	**459.1**	20.15	45.15	**35.26**	1.88
Ce	**987**	**6.40**	**54.90**	**66.90**	**75.00**	**84.79**	**92.91**	**134.10**	20.10	63.07	**59.57**	1.47
Ce	**947**	**6.35**	**54.78**	**66.90**	**75.05**	**84.77**	**92.96**	**134.09**	20.27	63.07	**59.56**	1.47
Co	**987**	**0.10**	**5.70**	**13.10**	**18.00**	**27.70**	**40.70**	**89.30**	12.30	13.12	**9.23**	2.74
Co	**947**	**0.10**	**5.71**	**13.00**	**18.05**	**27.71**	**40.98**	**89.30**	12.35	13.14	**9.26**	2.74
Cr	**987**	**7.60**	**60.75**	**80.90**	**95.90**	**112.8**	**267.3**	**1577.8**	35.15	83.17	**67.33**	1.90
Cr	**947**	**7.60**	**61.04**	**80.93**	**96.00**	**113.9**	**298.5**	**1577.8**	34.97	83.84	**67.75**	1.90
Cs	**987**	**1.80**	**7.00**	**10.00**	**17.00**	**36.87**	**64.21**	**240.0**	10.00	14.67	**11.16**	2.00
Cs	**947**	**1.79**	**7.00**	**10.00**	**17.00**	**36.00**	**63.08**	**184.5**	10.00	14.42	**11.12**	1.98
Cu	**987**	**0.30**	**16.70**	**25.70**	**37.20**	**85.86**	**189.0**	**690.5**	20.50	34.27	**24.58**	2.21
Cu	**947**	**0.30**	**16.95**	**25.88**	**37.30**	**85.12**	**189.4**	**690.5**	20.35	34.42	**24.76**	2.21
Ga	**987**	**0.20**	**15.40**	**19.10**	**23.60**	**28.17**	**30.10**	**34.60**	8.20	19.06	**17.39**	1.73
Ga	**947**	**0.20**	**15.39**	**19.10**	**23.50**	**27.94**	**29.95**	**32.30**	8.11	19.01	**17.34**	1.73
Ge	**987**	**0.10**	**0.70**	**1.00**	**1.30**	**1.80**	**2.20**	**7.10**	0.60	1.04	**0.90**	1.82
Ge	**947**	**0.10**	**0.70**	**1.00**	**1.30**	**1.80**	**2.20**	**7.10**	0.60	1.04	**0.91**	1.80
Hf	**987**	**0.50**	**5.00**	**6.20**	**7.40**	**9.60**	**11.10**	**14.60**	2.40	6.26	**5.87**	1.48
Hf	**947**	**0.46**	**5.00**	**6.20**	**7.40**	**9.51**	**11.08**	**14.58**	2.40	6.26	**5.87**	1.49
La	**987**	**0.20**	**29.00**	**35.00**	**41.00**	**47.00**	**51.84**	**79.40**	12.00	33.62	**30.46**	1.85
La	**947**	**0.24**	**29.00**	**35.00**	**41.00**	**46.95**	**51.96**	**79.35**	12.00	33.59	**30.39**	1.86
Mo	**987**	**0.10**	**0.80**	**1.10**	**1.40**	**2.47**	**4.31**	**17.50**	0.60	1.28	**1.10**	1.70
Mo	**947**	**0.10**	**0.85**	**1.11**	**1.40**	**2.47**	**4.30**	**17.50**	0.56	1.28	**1.11**	1.70
Nb	**987**	**0.60**	**15.10**	**16.80**	**18.00**	**22.91**	**39.18**	**107.0**	2.90	16.93	**16.00**	1.46
Nb	**947**	**0.64**	**15.10**	**16.72**	**18.00**	**22.69**	**36.96**	**107.0**	2.90	16.87	**15.93**	1.47
Nd	**987**	**0.20**	**21.60**	**28.30**	**32.30**	**37.40**	**44.71**	**86.90**	10.70	26.38	**23.83**	1.74
Nd	**947**	**0.20**	**21.60**	**28.20**	**32.30**	**37.40**	**44.42**	**72.83**	10.70	26.33	**23.82**	1.74
Ni	**987**	**1.70**	**13.15**	**26.00**	**38.95**	**58.57**	**148.9**	**564.0**	25.80	30.64	**21.40**	2.39
Ni	**947**	**1.65**	**13.32**	**26.00**	**38.95**	**59.04**	**154.3**	**564.0**	25.63	30.89	**21.56**	2.39
Pb	**987**	**10.80**	**35.00**	**45.10**	**66.75**	**125.50**	**206.6**	**521.2**	31.75	57.60	**49.59**	1.66
Pb	**947**	**10.80**	**35.40**	**45.95**	**66.88**	**125.95**	**207.0**	**521.2**	31.48	58.09	**49.99**	1.66
Rb	**987**	**3.40**	**104.2**	**140.1**	**179.9**	**286.6**	**389.8**	**551.3**	75.65	150.4	**132.1**	1.79
Rb	**947**	**3.40**	**104.4**	**140.1**	**179.3**	**283.1**	**383.7**	**551.3**	74.90	149.6	**131.4**	1.80
Sb	**987**	**0.10**	**0.70**	**1.20**	**2.30**	**6.80**	**19.71**	**79.70**	1.60	2.29	**1.32**	2.57
Sb	**947**	**0.10**	**0.70**	**1.20**	**2.30**	**6.86**	**20.59**	**79.70**	1.60	2.32	**1.33**	2.58
Sc	**987**	**0.50**	**8.30**	**12.20**	**14.95**	**18.20**	**21.40**	**24.20**	6.65	11.48	**10.08**	1.80
Sc	**947**	**0.50**	**8.40**	**12.20**	**15.00**	**18.20**	**21.35**	**24.15**	6.60	11.51	**10.12**	1.80
Se	**987**	**0.10**	**0.60**	**0.80**	**1.10**	**1.90**	**3.31**	**6.80**	0.50	0.95	**0.84**	1.67
Se	**947**	**0.10**	**0.60**	**0.80**	**1.10**	**1.90**	**3.35**	**4.80**	0.50	0.95	**0.84**	1.66
Sm	**987**	**0.10**	**2.90**	**4.20**	**5.30**	**7.17**	**8.73**	**20.40**	2.40	4.19	**3.63**	1.90
Sm	**947**	**0.10**	**2.98**	**4.20**	**5.30**	**7.16**	**8.69**	**15.72**	2.32	4.19	**3.63**	1.90
Sn	**987**	**2.10**	**5.85**	**12.50**	**41.10**	**230.2**	**494.5**	**2000**	35.25	51.28	**17.64**	3.74
Sn	**947**	**2.10**	**5.90**	**12.47**	**40.15**	**231.2**	**495.8**	**2000**	34.26	51.33	**17.60**	3.72
Sr	**987**	**12.50**	**54.40**	**70.00**	**90.20**	**144.0**	**321.5**	**1305**	35.80	81.06	**69.51**	1.67

Sr	**947**	**12.50**	**54.45**	**70.08**	**90.62**	**142.5**	**351.3**	**1305**	36.17	81.49	**69.73**	1.68
Th	**987**	**0.30**	**8.80**	**10.30**	**11.60**	**13.40**	**17.21**	**22.20**	2.80	9.99	**9.43**	1.51
Th	**947**	**0.30**	**8.78**	**10.25**	**11.60**	**13.40**	**17.25**	**22.20**	2.82	9.99	**9.42**	1.51
U	**987**	**0.70**	**2.50**	**2.90**	**3.40**	**6.10**	**11.71**	**45.70**	0.90	3.34	**3.01**	1.49
U	**947**	**0.66**	**2.50**	**2.90**	**3.43**	**5.87**	**11.70**	**35.35**	0.93	3.31	**3.00**	1.48
V	**987**	**11.60**	**79.95**	**108.9**	**135.1**	**170.2**	**204.3**	**250.8**	55.10	103.9	**89.42**	1.89
V	**947**	**11.60**	**80.40**	**109.4**	**135.1**	**170.3**	**203.5**	**241.7**	54.66	104.1	**89.71**	1.88
W	**987**	**0.10**	**2.50**	**3.30**	**6.35**	**16.17**	**42.70**	**279.60**	3.85	6.44	**4.06**	2.25
W	**947**	**0.10**	**2.50**	**3.30**	**6.30**	**15.88**	**38.79**	**270.35**	3.80	6.14	**4.02**	2.22
Y	**987**	**1.30**	**20.50**	**25.40**	**27.70**	**30.40**	**40.51**	**82.20**	7.20	23.57	**22.01**	1.54
Y	**947**	**1.30**	**20.50**	**25.30**	**27.61**	**30.42**	**39.55**	**76.98**	7.11	23.51	**21.96**	1.54
Zn	**987**	**14.60**	**56.75**	**84.20**	**120.4**	**204.6**	**389.4**	**720.9**	63.60	97.48	**81.70**	1.80
Zn	**947**	**17.25**	**57.23**	**84.74**	**120.4**	**204.6**	**389.6**	**720.9**	63.21	97.85	**82.23**	1.79
Zr	**987**	**0.50**	**180.0**	**223.6**	**268.0**	**334.5**	**385.3**	**551.2**	88.00	223.2	**196.9**	2.18
Zr	**947**	**0.50**	**180.5**	**222.9**	**267.7**	**334.4**	**383.8**	**551.2**	87.15	222.7	**195.9**	2.20

N: number of samples; **Min**: minimum value; **Max**: maximum value; **Q##**: ##th quantile; **Mdn**: median; **IQR**: interquartile range; **AM**: arithmetic mean; **GM**: geometric mean; **GSD**: geometric standard deviation. Units (in brackets) are valid for all parameters except **GSD** (dimensionless).

Figure01

Click here to download high resolution image

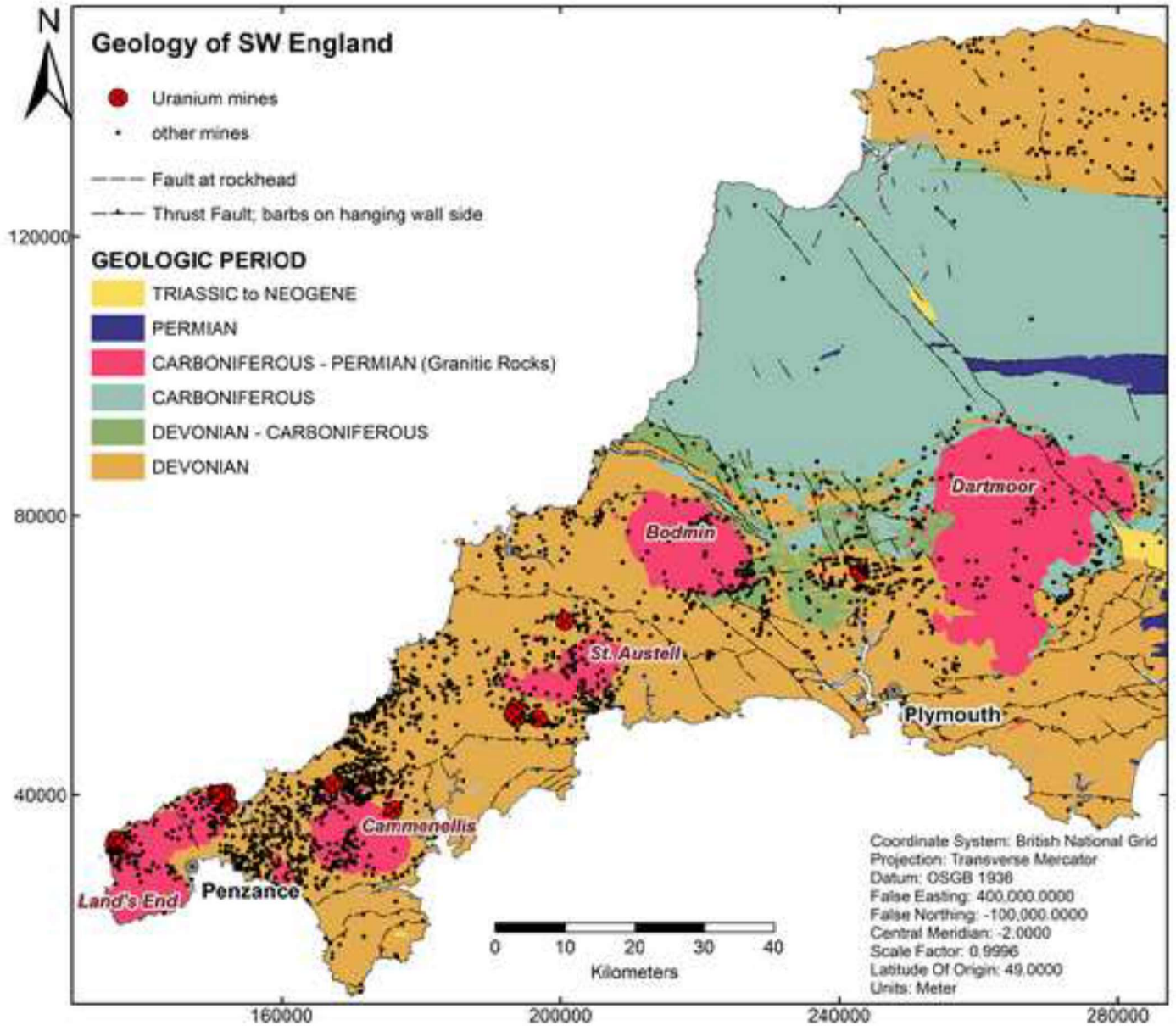

Figure02

Click here to download high resolution image

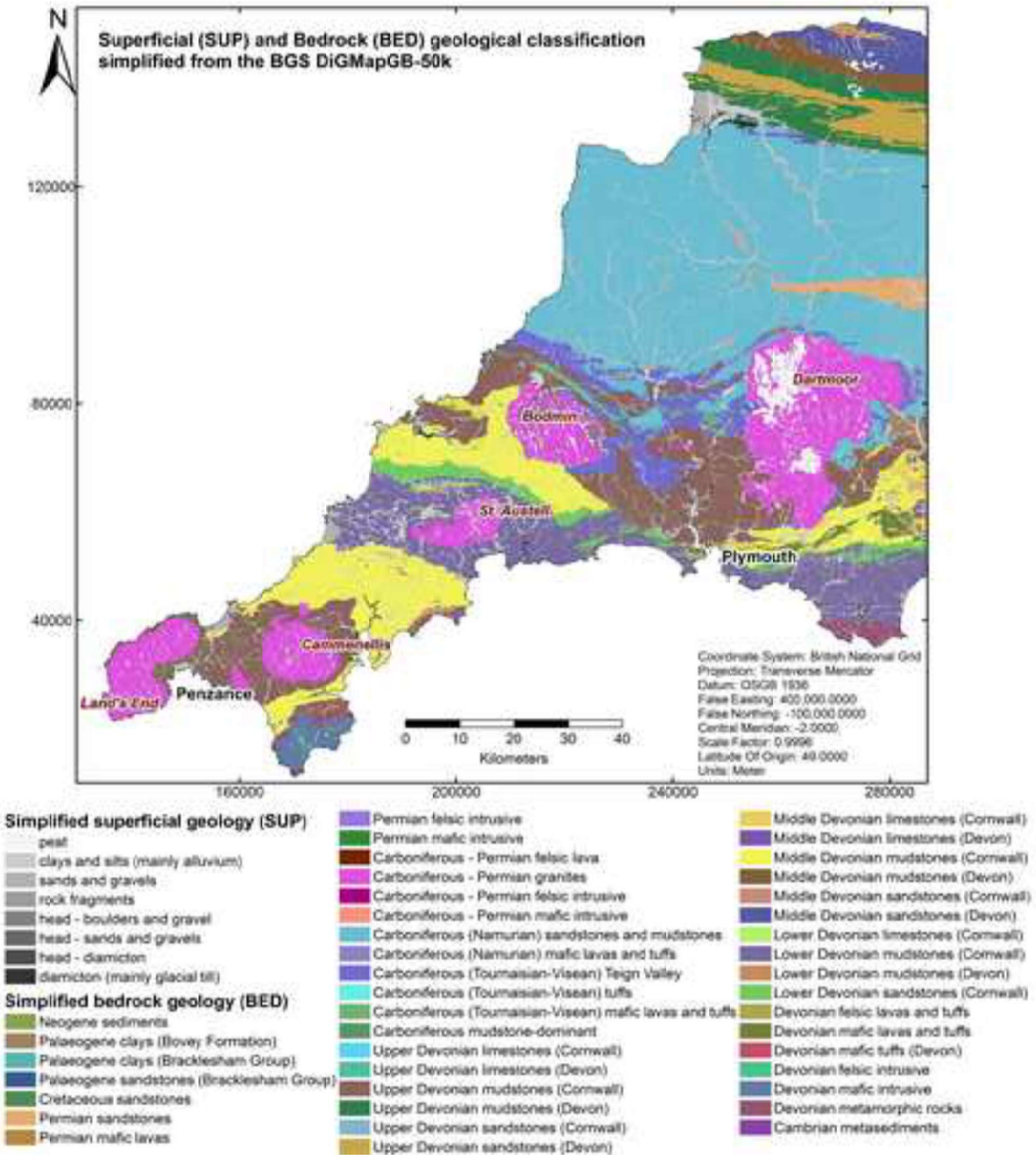

Figure03

Click here to download high resolution image

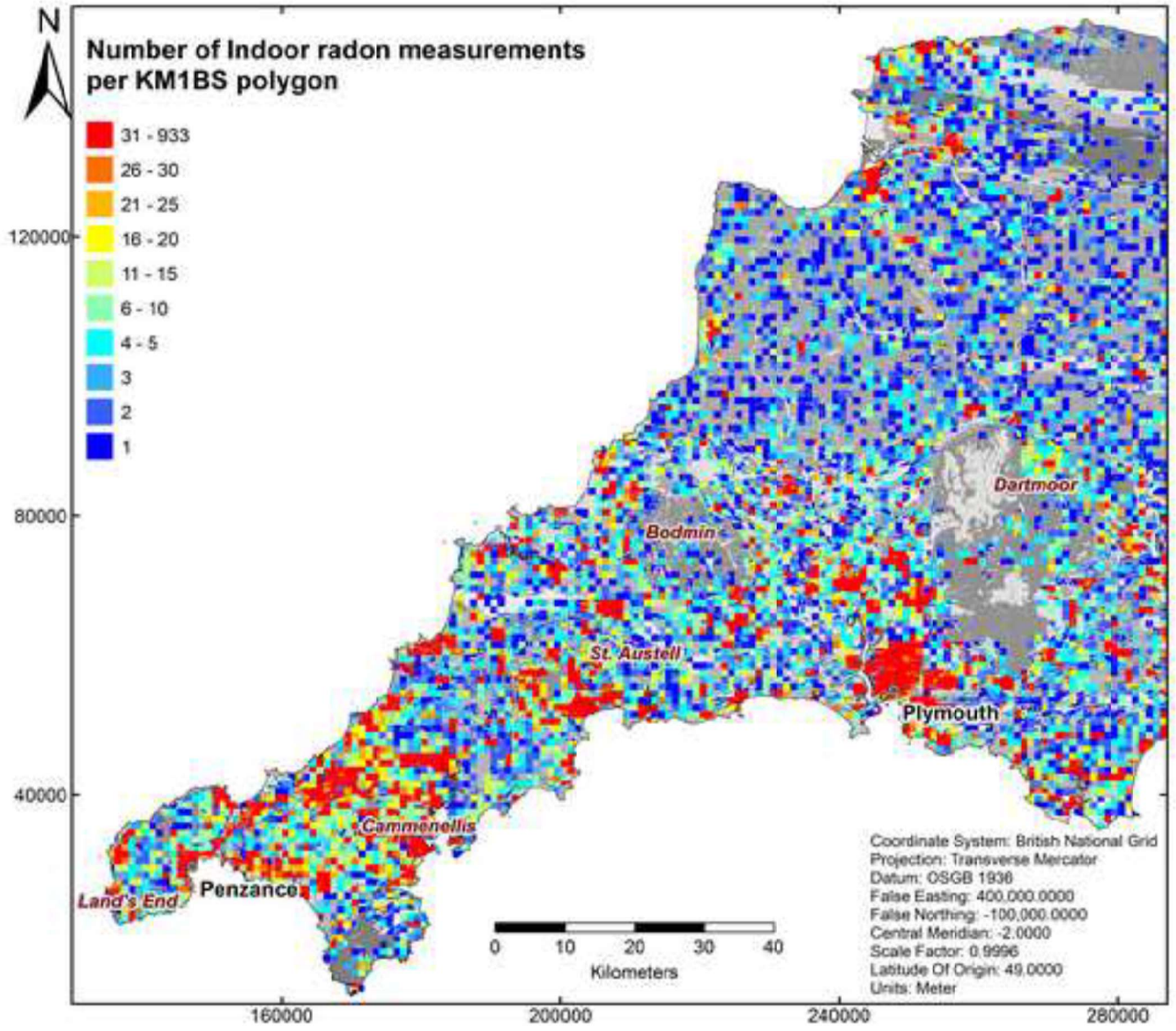

Figure04

Click here to download high resolution image

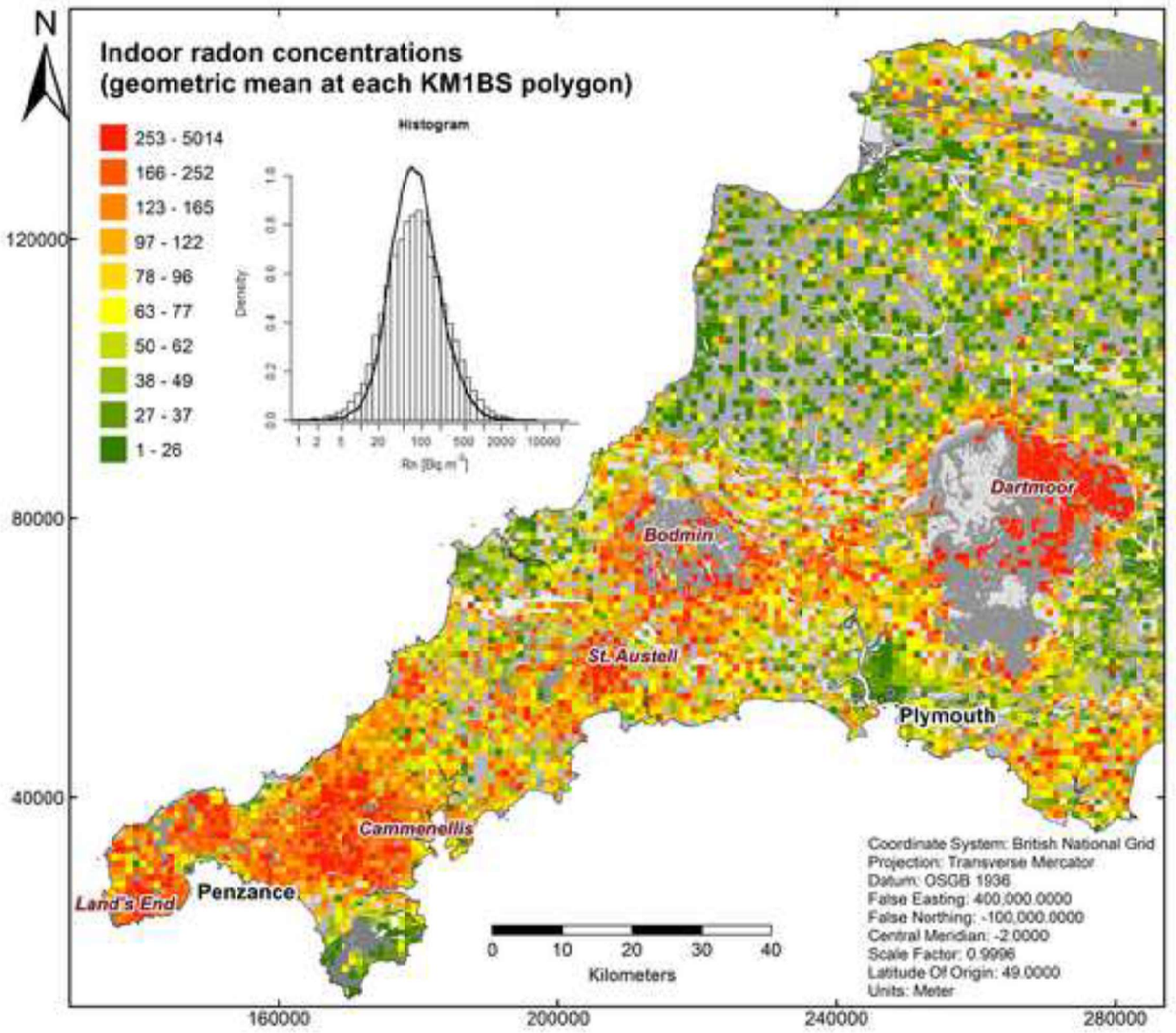

Figure05

Click here to download high resolution image

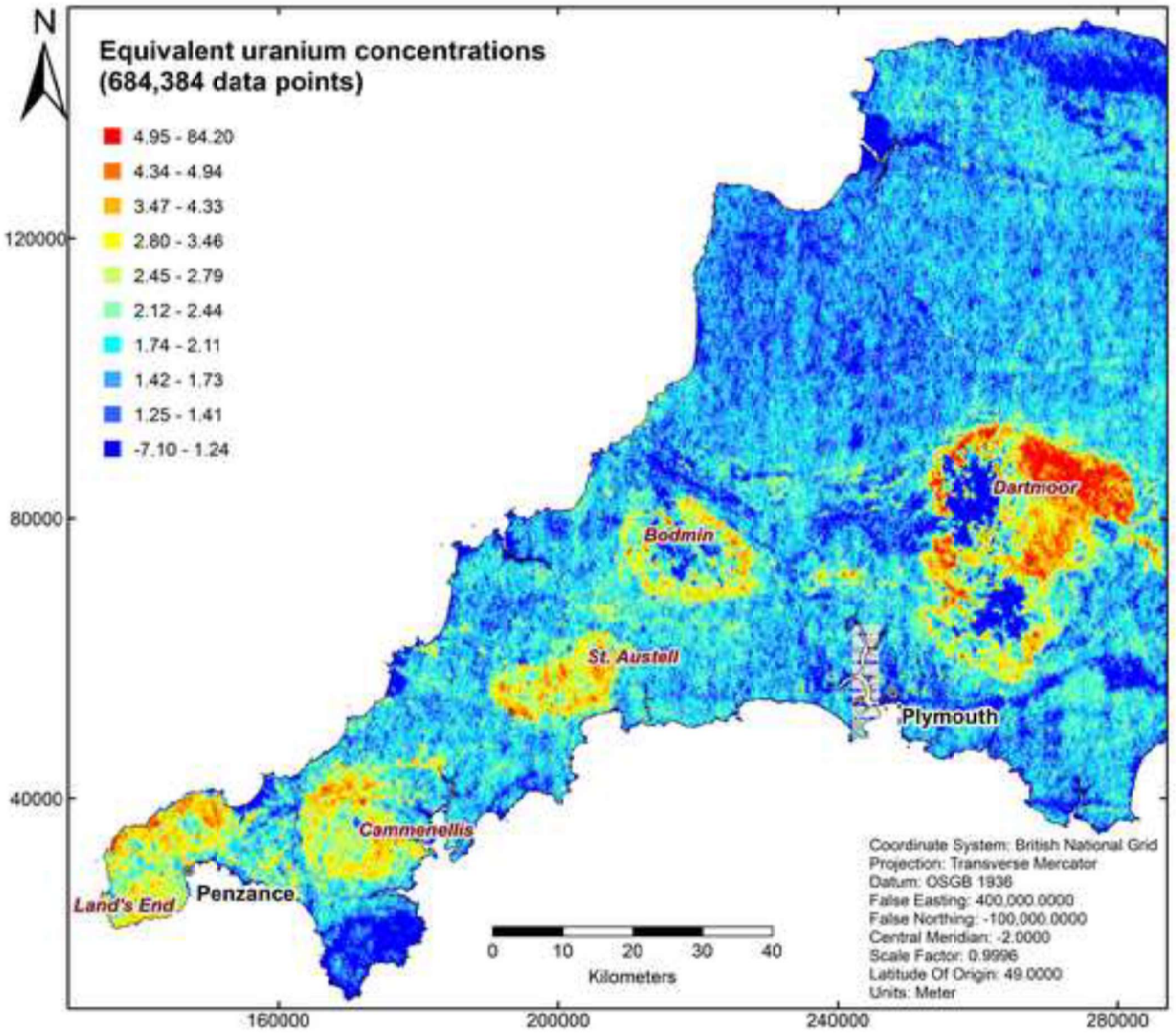

Figure06

Click here to download high resolution image

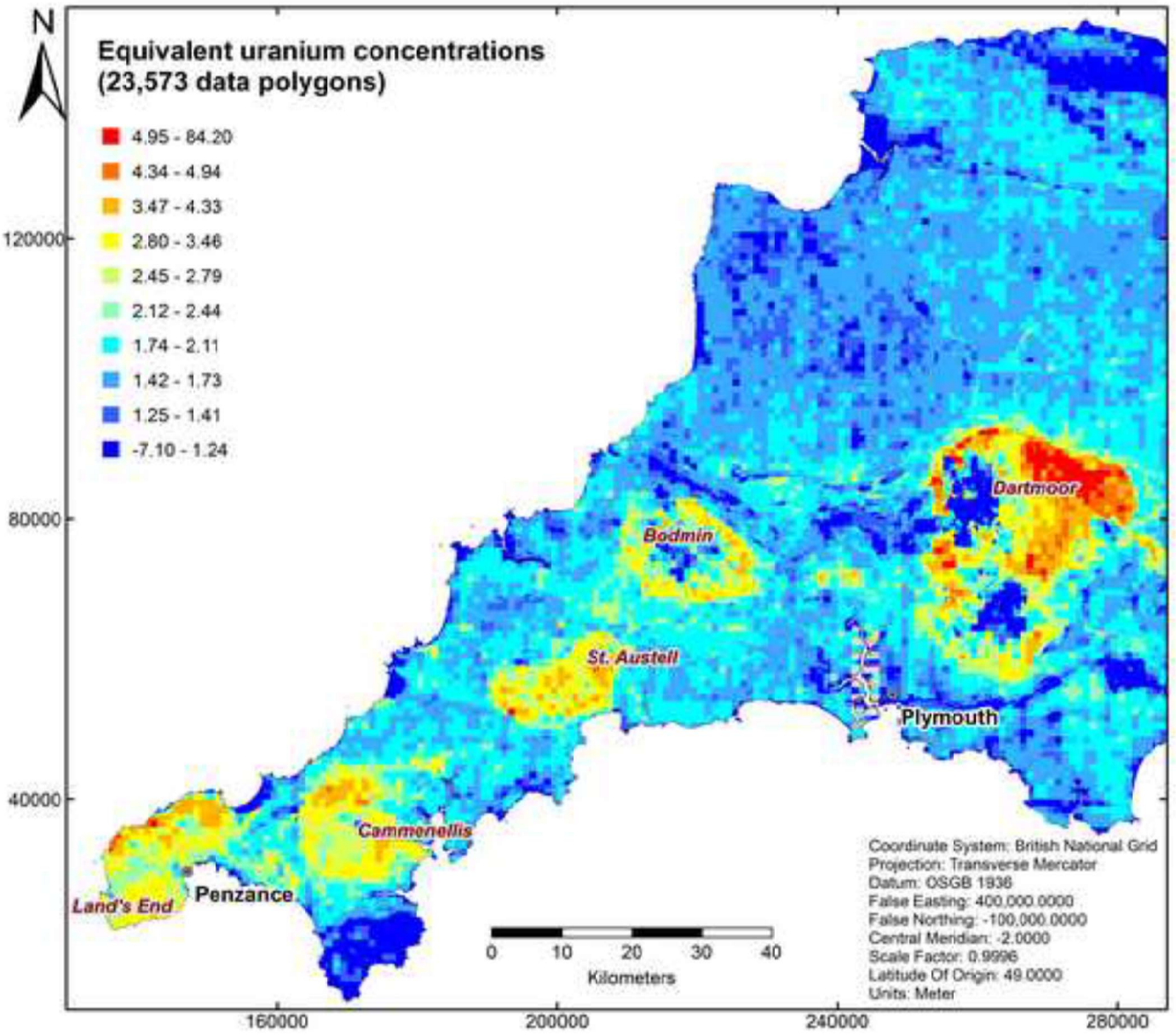

Figure07
Click here to download high resolution image

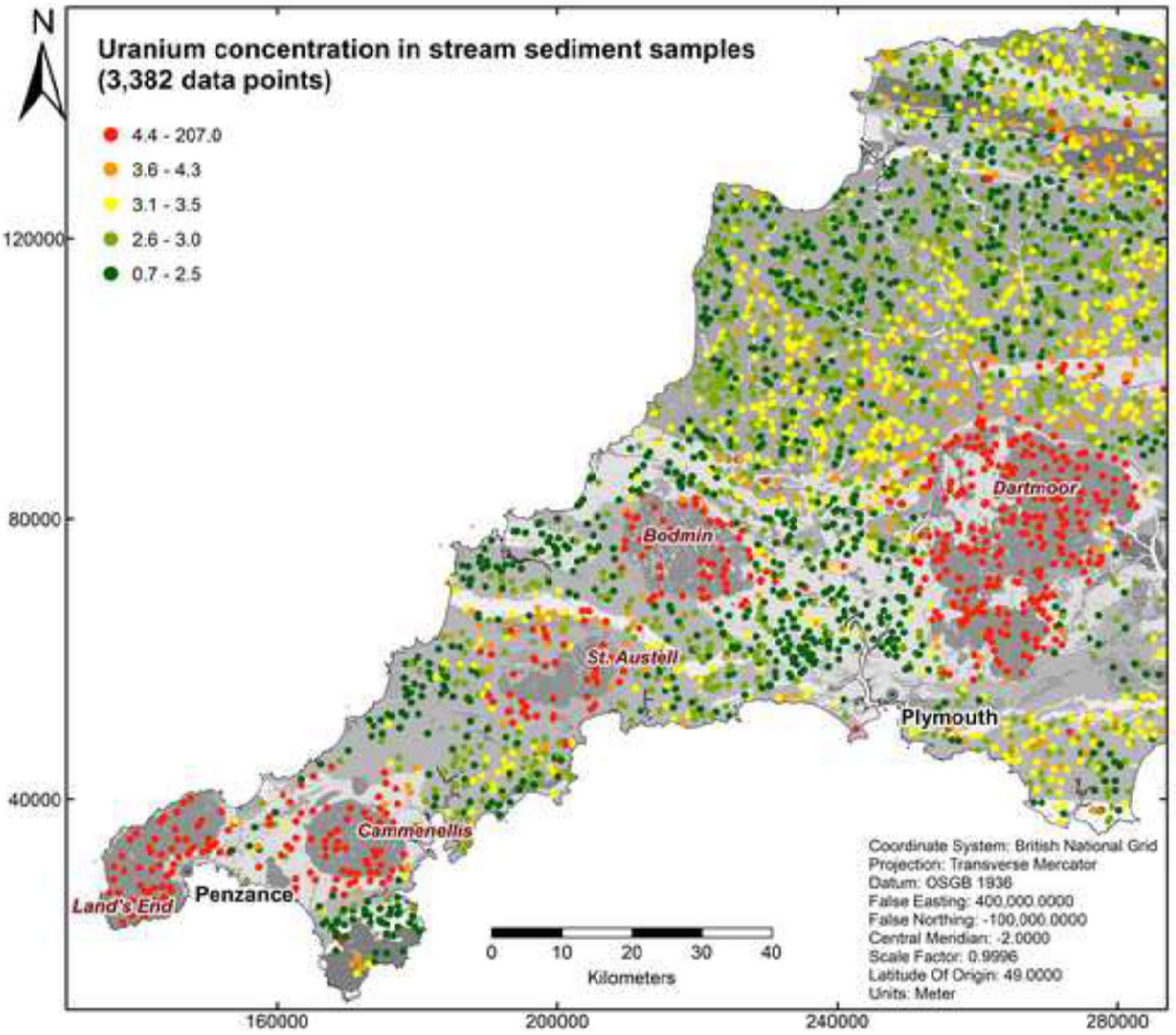

Figure08
Click here to download high resolution image

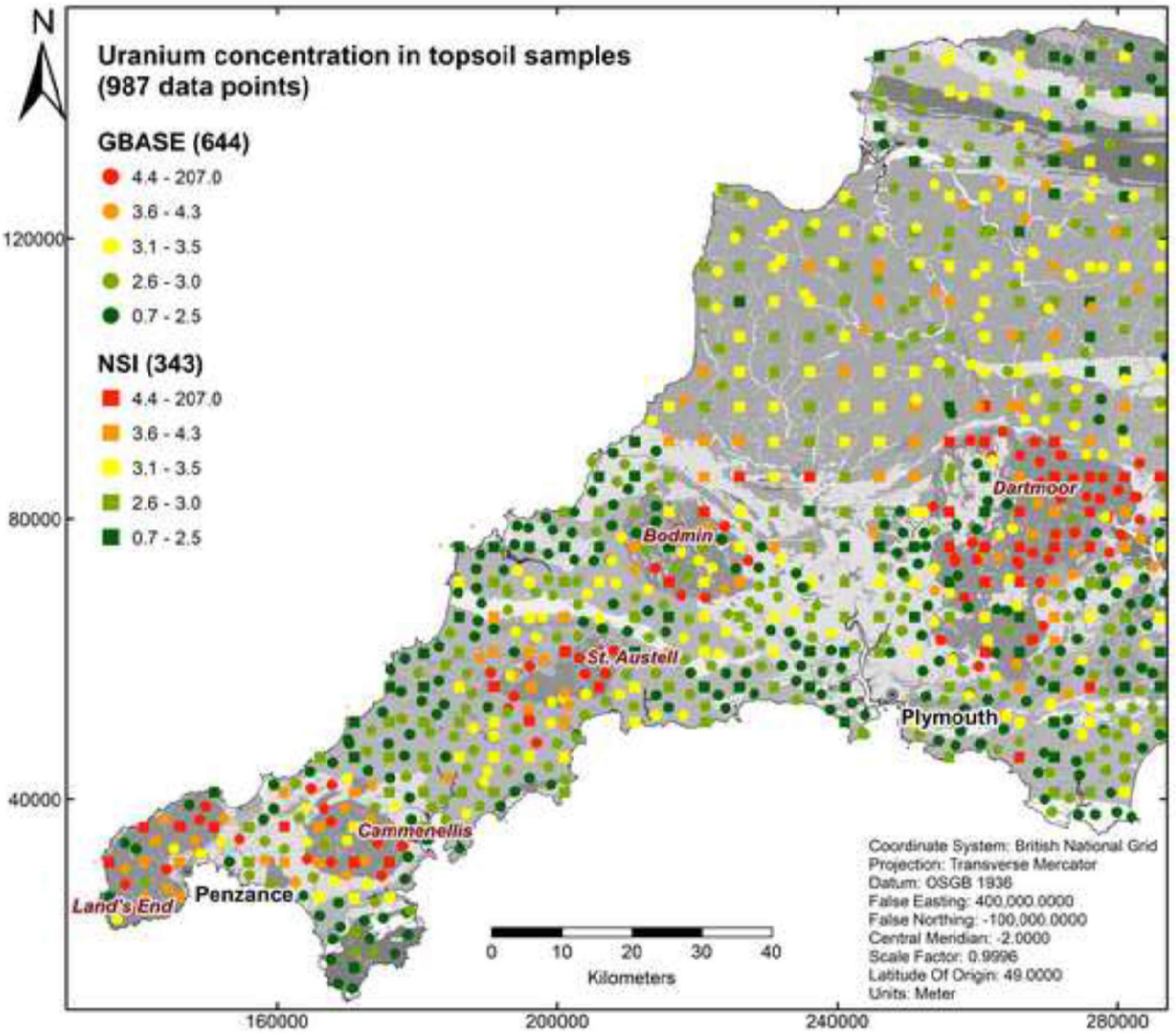

Figure09
Click here to download high resolution image

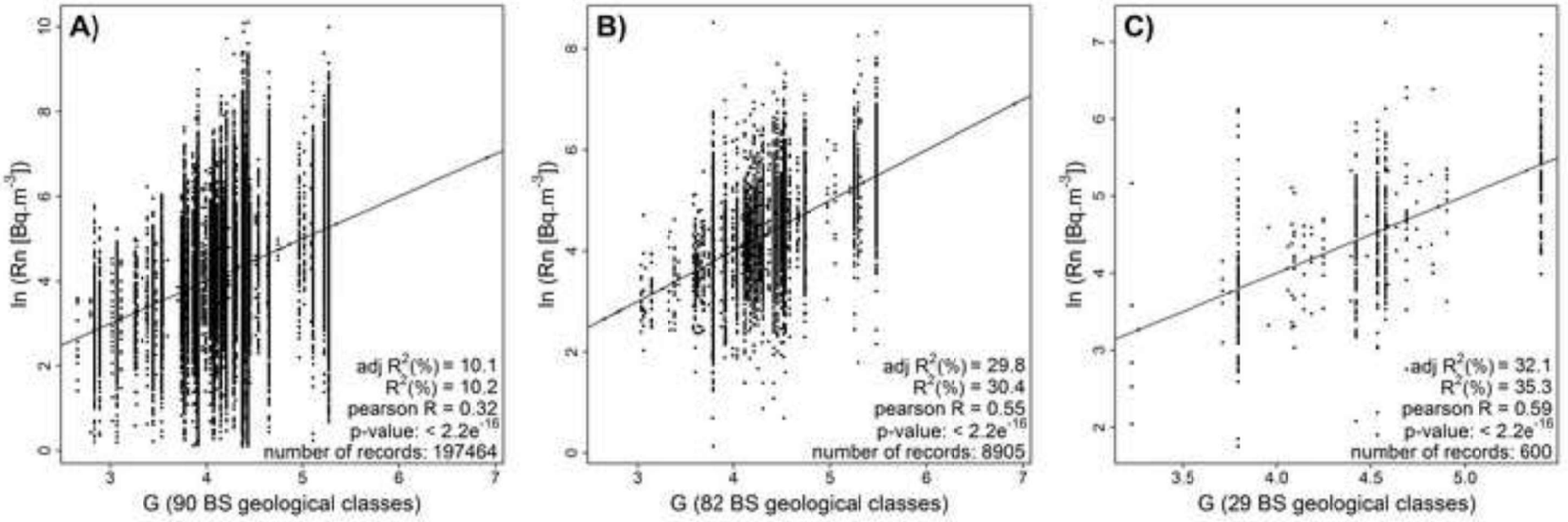

Figure10
Click here to download high resolution image

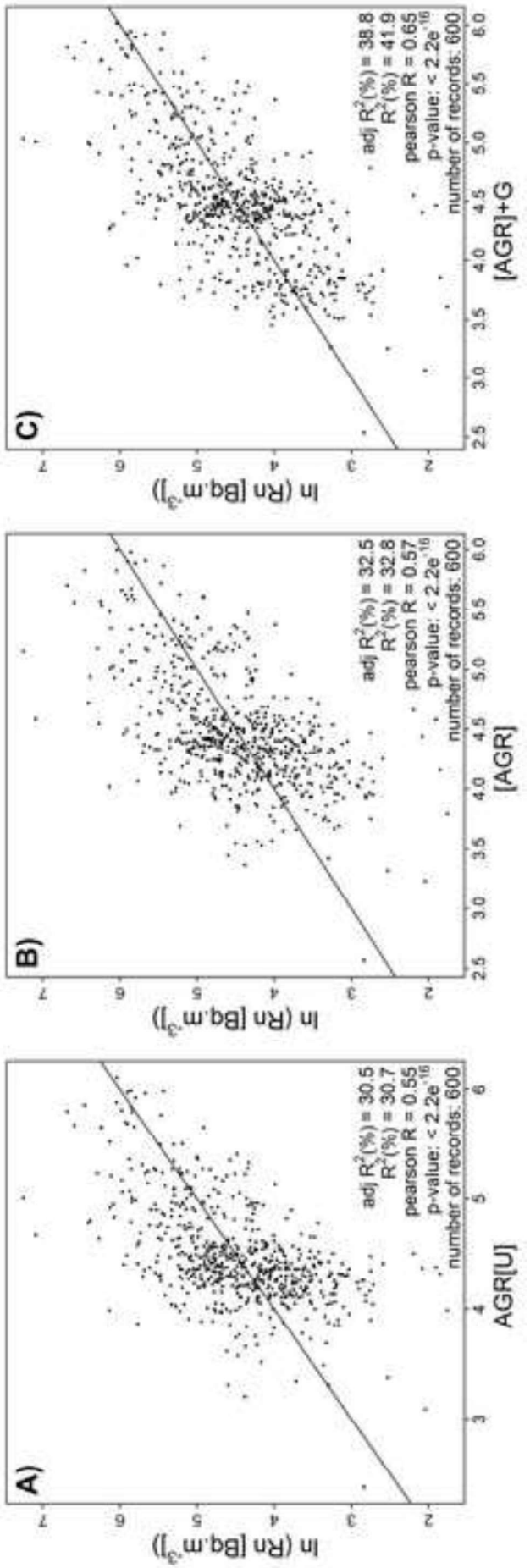

Figure11
Click here to download high resolution image

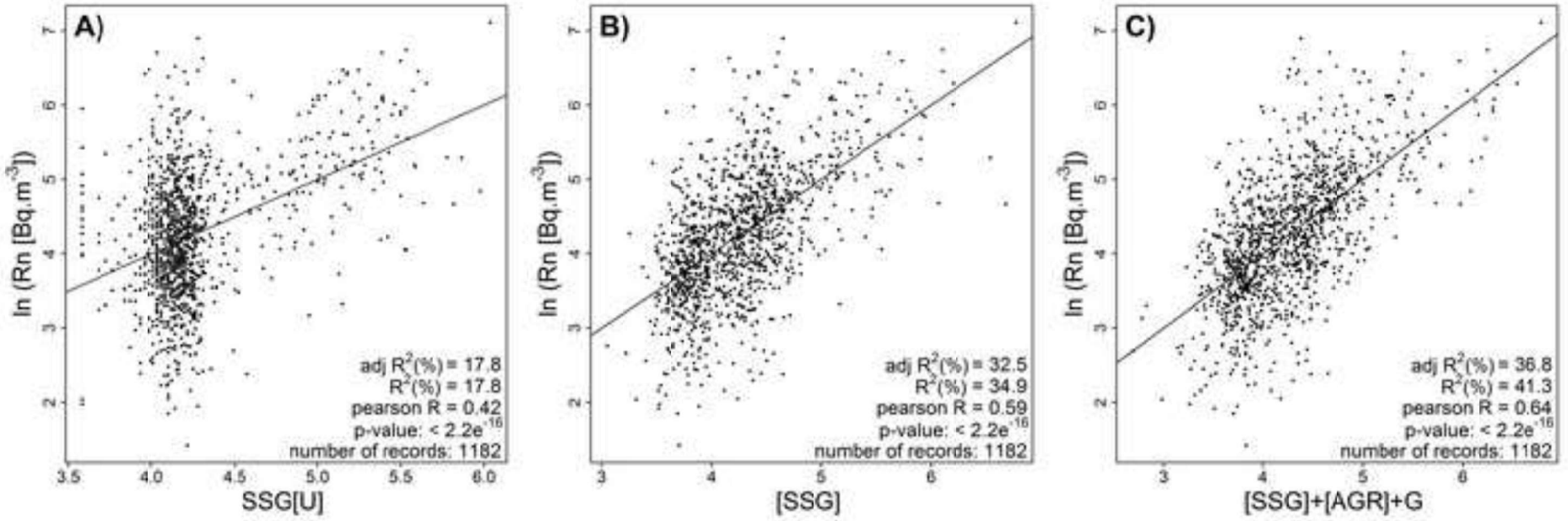

Figure12
Click here to download high resolution image

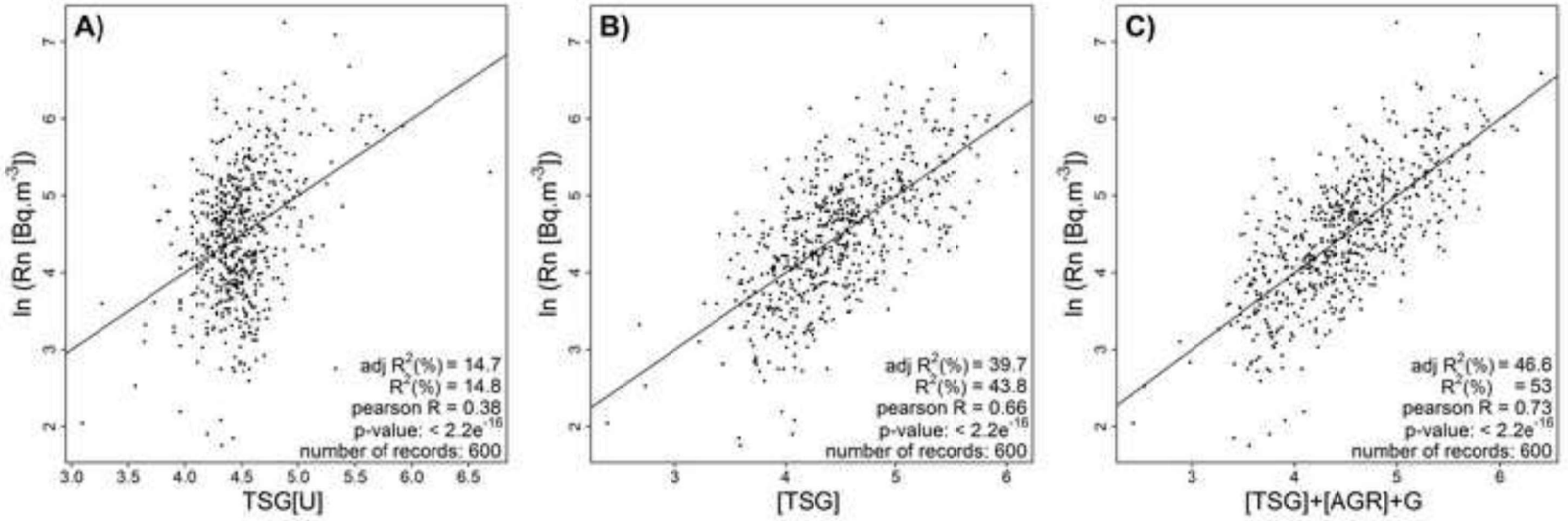

Figure13
Click here to download high resolution image