



16ENV10 MetroRADON

Deliverable 6

Report on the concept and establishment of a Radon Hazard Index (RHI) including an RHI map of Europe showing areas with high geogenic radon potential and conclusions on the relationships and correlation between indoor Rn concentration and quantities related to geogenic Rn.

Lead organisation: European Commission, Joint Research Center (JRC)

Other involved organisations: JRC, BEV-PTP, BFKH, VINS, AGES, BfS, IRSN, SUBG, UC

Due date: March 2020

Submission: November 2020

EMPIR



The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

Table of contents

1	Introduction	4
1.1	Aim	4
1.2	Structure	5
2	RPA estimation and classification uncertainty	7
2.1	Estimation	7
2.2	Uncertainty definition.....	14
2.3	Uncertainty of input quantities and factors	15
2.4	Uncertainty of aggregated quantities.....	23
2.5	Quantifying classification uncertainty	35
2.6	RPAs as random objects – statistical considerations.....	42
2.7	Proposal of methodology for assessing and quantifying RPA uncertainty.....	42
3	Application of retrospective radon measurements to RPA assessment	44
4	RPA classification based on occurrence of extremes	49
4.1	Results: case-studies in France	50
4.2	Results: case studies in Spain	58
4.3	Discussion and perspectives	68
5	Radon hazard index RHI.....	69
5.1	Introduction and rationale of RHI.....	69
5.2	Concept and objective	69
5.3	Review of existing GRHI trials	70
5.4	Statistical background, methodology, challenges	73
6	Summary, conclusions and recommendations.....	81
7	References	83

List of authors (alphabetic order)

Budapest Főváros Kormányhivatala, Hungary – BFKH

Á. Nagy, D. P. Nagy, R. Botos, K. Rózsa N., Szabó, N. Szilágyi, L. Szűcs, Z. Nagyné Szilágyi, D. Párkányi

German Federal Office for Radiation Protection, Berlin, Germany – BfS

P. Bossew, E. Petermann

Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH, Austria – AGES

V. Gruber, S. Baumann, W. Ringer

Universidad De Cantabria, Spain – UC

S. Celaya, A. Fernandez, E. Fernandez, I. Fuente, J. Quindos, L. Quindos, R. Pol, D. Rabago, C. Sainz

European Commission, Joint Research Centre, Ispra, Italy – JRC

G. Cinelli

Institut de Radioprotection et de Surete Nucleaire, France – IRSN

C. Greau, G. Ielsch, J. Guillevic

Physikalisch-Technischer Prüfdienst des Bundesamt für Eich- und Vermessungswesen, Austria – BEV-PTP

F.J. Maringer, H. Wiedner, M. Stietka

Sofiiski Universitet Sveti Kliment Ohridski, Bulgaria - SUBG

I. Dimitrova, S. Georgiev, K. Mitev, D. Pressyanov

Radonova Laboratories AB, Sweden

J. L. Gutiérrez Villanueva

"Vinča" Institute of Nuclear Sciences, Belgrade, Serbia – VINS

I. Čeliković, G. Pantelić

CNR-IGAG, Italy

G. Ciotoli

LNR/Uni Coimbra

A. Pereira

1 Introduction

This document, titled “Report on the concept and establishment of a Radon Hazard Index (RHI) including an RHI map of Europe showing areas with high geogenic radon potential and conclusions on the relationships and correlation between indoor Rn concentration and quantities related to geogenic Rn” represents the deliverable D6 of the MetroRADON project.

It reports the results of the activities developed in **Task 4.3** of Work Package 4 – WP4: Radon priority areas (RPAs) and the development of the concept of a “geogenic radon hazard index” (RHI) of the EURAMET 16ENV10 MetroRADON project.

The aim of Work Package 4 is to analyse and develop methodologies for the identification of radon priority areas (RPA), to investigate the relationships between indoor Rn concentration and quantities related to geogenic Rn, including soil exhalation (see deliverable D3 of MetroRADON project) and to develop the concept of a “geogenic radon hazard index” (RHI) as a tool to help identify radon priority areas.

Work Package 4 is divided into four Tasks:

4.1: Evaluation of the concepts for the definitions of radon priority areas

The aim of this task is to review and evaluate the concepts which have already been proposed to define and to estimate RPAs. The work and results of Task 4.1 are discussed in detail in Annex 1 of the deliverable D5 of MetroRADON project.

4.2: Relationship between indoor radon concentration and geogenic radon

The aim of this task is to estimate relationships between indoor Rn or derived quantities such as the probability of exceeding a reference level within an area, and quantities related to geogenic Rn such as the Rn potential or uranium concentration in the ground, as some concepts for mapping the geogenic Rn potential and RPA crucially depend on such relationships. The work and results of Task 4.2 are discussed in detail in Annex 2 of the deliverable D5 of MetroRADON project.

4.3: New developments in estimation of radon priority areas

The aim of this task is to review and to propose new technical developments related to the RPA estimation, including the development of a methodology for a harmonised “Rn hazard index” (RHI) as a tool to visualise radon priority areas, and to address uncertainty budgets and classification errors which emerge in this context.

4.4: Harmonisation of radon priority areas across borders

The aim of this task is to develop a strategy to harmonise defined RPAs across borders and to incorporate it in a guideline. The work and results of Task 4.4 are discussed in detail in the deliverable D5 of MetroRADON project.

1.1 Aim

The aim of this report is to review and to propose new technical developments related to the RPA estimation, including the development of a methodology for a harmonised “Rn hazard index” (RHI) as a tool to visualise radon priority areas, and to address uncertainty budgets and classification errors which emerge in this context.

In the European Council directive 2013/59/Euratom, Article 103, Paragraph 3 states that Member States should identify areas in which it is expected that annual average indoor radon concentration will exceed

national reference level in significant number of dwellings (EC 2014). These areas are often called “radon priority areas - RPA”. The delineation of these areas will allow to plan and to prioritise measures within the national action plan and has implications in that radon measurements in workplaces located in these areas may be required (EU, 2020). Further to legally binding requirements, such a prioritisation can also be useful for radon prevention for new buildings (for example, through specific building codes), as well as the promotion of actions aimed at reducing exposure to radon (EU, 2020).

RPAs, however defined, are objects that can be interpreted as highly aggregated quantities, which renders uncertainty budgets complicated. As estimated from data and based on models, the uncertainty budgets can be considered “random objects” (an object whose properties are defined only in probabilistic terms, in this case the geometry of RPA, i.e. area, shape, border, topology) that are subject to different types of uncertainty that propagate into the target quantity from its constituents. Experience has shown that establishing uncertainty budgets for (in this case spatially) aggregated Rn related quantities is challenging.

RPAs are sometimes estimated differently in different countries or regions which can lead to inconsistency across borders. This can have a significant impact on the credibility of the prediction of radon hazard areas and renders the data incomparable, which has been subject of Task 4.4 of the project (see Deliverable 5). However, in this report, a major innovation will be the development of a methodology for a harmonised “Rn hazard index” (RHI). The concept of the RHI is to provide a universally applicable tool to quantify the susceptibility of an area to geogenic Rn and hence to quantify the “Rn prone-ness”, independent of available datasets, and applicable irrespective borders, e.g. throughout Europe. The RHI could serve, to an extent to tackle the problem of inconsistency at a European level.

1.2 Structure

The report starts with a review of concepts of classification uncertainty (**section 2. RPA estimation and classification uncertainty**). A methodology to assess RPA classification uncertainty is proposed and a proposal on how to undertake "top-down" harmonisation of existing approaches towards quantifying uncertainty is developed.

Then, precision and applicability of the compact disc method (“CD method”) for retrospective Rn measurements for use for identification of RPAs are evaluated (**section 3. Application of retrospective Rn measurements to RPA assessment**). The uncertainties due to local and temporal variability are validated and assessed in experiments performed in UC’s Saelices el Chico laboratory, by exposing a number of CD/DVD samples over a period of about six months, under variable conditions representing a wide range of Rn concentrations, in parallel with conventional monitors.

Two new techniques, namely liquid scintillation counting of polymers and track-etching of CDs, for measurement of radon exhalation from soil based on radon absorption in plastic samplers have been developed and evaluated.

Several methods have already been developed to map RPAs. A new complementary approach that focuses on identification of areas that could have a significant proportion of dwellings with very high indoor radon concentrations of several thousands of Bq/m³ is developed and tested in **section 4. RPA classification based on occurrence of extremes**. Generally, this will concern locations where a significant proportion of indoor radon concentrations exceed a reference level of a few hundreds of Bq/m³ (maximum 300 Bq/m³ as given by the European BSS). This new approach will provide additional information for existing radon maps.

This method is tested in France and Spain, where such cases occur regionally. It is based on the analysis of available quantities such as the geogenic radon potential, measurements of indoor radon concentration,

dwellings characteristics, and recent results of quantitative radon risk assessment etc., complemented by statistical modelling. Such a method would allow targeting specific prevention and remediation actions in heavily affected regions to reduce the exposure in buildings significantly.

The concept of the RHI (“geogenic radon hazard index”) is to provide a universally applicable tool to quantify the susceptibility of an area to geogenic Rn and hence to quantify the “Rn prone-ness”, independent of available datasets and applicable irrespective borders. The aim is that the geogenic radon hazard index should be a tool to visualise radon priority areas. In section **5.Radon hazard index RHI**, the concept of RHI is introduced and statistical background explained. An RHI map is presented, based on available information from the EU member states, which JRC and BfS have access to, input from the stakeholder interest groups formed and results related to soil radon potential (see deliverable 5). Currently there are attempts, coordinated by the JRC, to design a methodology for defining and estimating a generic multivariate RHI, which could be adapted to different data availability. Initial results that are available from the work previously undertaken by JRC are evaluated and the method developed further. The information collected on countries’ experiences is another important input for the improvement and validation of the methodology (deliverable 5).

2 RPA estimation and classification uncertainty

The text of this section is based on a report by VINS, previous work by BfS, [Bossew \(2018\)](#) and a contribution of the same author to the *European Atlas of Natural Radiation (EC, 2019)*.

There are different definitions of RPA within different countries, yet each country has to develop certain classification scheme in order to delineate RPAs, or different levels of priorities.

By assigning some region to certain class, there is a certain probability to misclassify that area. It is important to make this classification as reliable as possible, given the economic and possibly legal consequences of misclassification. Action required in an RPA can be expensive for various reasons.

- *Measurement campaigns*: Rn measurement itself is not expensive, but the logistic effort behind a representative survey is big.
- *Remediation, mitigation*: It is indirectly dependent on RPA as measurements in RPA are obligatory (in workplaces) or recommended (dwellings), and remediation is done based on measurement results.
- *Prevention*: “Rn proof” construction implies additional costs due to enhanced Rn-tight insulation against the ground and possibly proactive installation of Rn mitigation devices, such as sub-surface ventilation. The costs are usually only in the order of percent of total construction costs, but are still a factor.
- *Secondary consequences*: It has been claimed that declaring a region RPA may have negative economic consequences. As examples, a negative impact on investment and property value has been cited; however, no example is known to us up to now.

Non-compliance with regulation on the one hand, and unnecessary measures due to misclassification of an area on the other, can thus have economic and legal consequences for those responsible of RPA classification; in general these are regional or national administrations and governments.

Understandably, administrations, decision makers and stakeholders altogether want to be on the safe side, in order to avoid possible economic, legal and political cost resulting from uncertainty and error.

It has to be communicated however, that uncertainty of RPA delineation cannot be reduced to zero for reasons inherent to the spatial distribution of the Rn hazard, in terms of which ever quantity. As experience shows, this communication task is not an easy one.

2.1 Estimation

According to the chosen definition, RPAs have to be estimated from data. These are, most commonly, measured indoor Rn data, but predictor or proxy quantities may be required instead or additionally. These may be geology, tectonics, soil properties, Rn concentration in the ground, geochemical concentrations, terrestrial dose rate and others, as physically and statistically related to indoor Rn. For example, the decision about whether a geographical unit shall be assigned RPA or not, or which grade or class should be assigned to that unit (in the case of multinomial definition, i.e. grades of “Rn priorityness”) amounts to a classification problem. If estimated from secondary quantities, one faces the task of conditional and cross-classification. Effectively, the spatial domain (a country) is classified into two or several mutually excluding subsets, labelled according “priorityness”, or RPA / non-RPA in the binomial case.

Indoor Rn concentration is subject to high spatial variability. This is due (a) to the variability of the geogenic Rn potential which in most cases is the main source of indoor Rn, (b), to the variability of the physical properties of buildings and (c) to the one of usage habits (ventilation). Each component can be conceptualized as

consisting of a trend (geologic realm for (a), cultural and climatic factors – themselves partly contingent - for (b) and (c)), a correlated random and an uncorrelated random “noise” component.

Taxonomy of estimation approaches – certainly incomplete – is given in Figure 1. The methods are ordered according to whether spatial autocorrelation of the underlying field is considered or not, and whether uni- or multivariate (i.e. supported by secondary quantities) is performed. The indicator transform or cut-off is defined,

$$I(x) \equiv \Theta(x-T) \equiv 1_T(x) = 1 \text{ if } x>T, 0 \text{ otherwise; } I: \mathfrak{R}^+ \rightarrow \{0,1\}.$$

Soft indicator transform replaces the “hard” step at $x=T$ by a sigmoidal transition, typical tgh- or erf-like. $I(Z^*)$ means, first model Z , then indicator transform; $I^*(z)$, first indicator transform data, then model.

	spatial correlation not considered	spatial correlation considered
univariate	<ul style="list-style-type: none"> • sample stat, $stat(z)$ • enhanced by assuming univar. distribution, e.g. LN, $stat'(z)$ → cut-off, $I(stat(z))$	<ul style="list-style-type: none"> • geostat. model & cut-off, $I(Z^*)$ • indicator kriging (hard/soft), $I^*(z)$
multivariate	<ul style="list-style-type: none"> • ANOVA type • logistic-type regression, $logi(z)=g(y)$ • geographically weighted, local regression etc. → cut-off, $I(f(y))$ <ul style="list-style-type: none"> • full bivariate through copula • cross-classification 	<ul style="list-style-type: none"> • co-kriging et al. & cut-off, $I(Z^*;y)$ • regression kriging & cut-off, $I(f^*(y))$ • indicator regression kriging, $I^*(f(y))$ • indicator co-kriging, $I^*(z;y)$ • machine learning

Figure 1: Taxonomy of estimation approaches. Z, Y – primary and secondary variably; I – indicator transform; * - estimation / interpolation.

a) Methods

Different procedures have been developed to capture this structure. Simpler versions aggregate data within the estimation support (e.g. the map presented in the European Atlas of Natural Radiation, [Cinelli et al. 2019](#)), forming statistics such as empirical averages or exceedance probabilities. Versions that are more elaborate attempt to develop quantitative models of the spatial structure using geostatistical tools. This includes machine learning ([Kropat et al. 2015](#), [Timkova et al. 2017](#)), quantile regression ([Sarraf et al. 2016](#)) and local regression ([Pasculli et al. 2014](#)). A hierarchical regression model where spatial dependence enters via lithology as predictor has been proposed in [Borgoni et al. \(2014\)](#) and further developed as generalised additive mixed model in Austria ([Gruber et al. 2017](#)), not yet fully published). Methods based on the indicator (co-) kriging approach have not been explored yet for RPA estimation, to our knowledge. Literature is growing fast in this field and new techniques pop up in rather fast pace.

Common estimation supports, i.e. the areas over which statistics are computed, are grid cells or municipalities. For example, 10 km × 10 km grids have been chosen for the European indoor Rn map, part of the European Atlas of Natural Radiation (EC, 2019) and for the current version of the German maps of RPA and of the geogenic Rn potential (GRP, see below). In general the municipalities option is preferred for administrative

reasons. However, both ignore natural reality, as the main spatially controlling factor is geology. Also using polygons representing geological units as estimation supports has indeed been considered, e.g. in Spain, [Garcia-Talavera et al. \(2013\)](#).

b) Multivariate estimation

Particular challenges emerge if RPAs shall be estimated from secondary quantities such as the GRP or uranium (U) concentration in soil, e.g. because not sufficient data of the primary quantity (indoor Rn concentration) is available, or to improve estimates by including additional information. The secondary variable(s) act as RPA predictors, “calibrated” by the primary one. Two different types of approaches are conceivable: (1) parametric co-estimation, such as (1a) methods belonging to the co-kriging family. Block estimates over estimation supports are classified according to RPA definition. (1b) As simpler variety, parametric estimation of the primary variable from the secondary by regression. For example, estimation of exceedance probability of indoor Rn concentration, $\text{prob}(C > RL)$, has been demonstrated in [Bossew \(2017\)](#) (predictor: GRP) and in [Elio et al. \(2018\)](#) (several geogenic predictors) using logistic-type regression. (1c) Another option is modelling the bivariate distribution by a copula, from which desired statistics can be derived, [Bossew \(2012, 2013\)](#). Opposed to (1a), methods (1b) and (1c) do not exploit spatial correlation properties.

The second type of approach is justified by the fact that classification does not require full information of the primary variable. The target of RPA delineation is only to determine to which class a certain estimation support belongs. The problem then boils down to finding class limits of the secondary variable(s) which correspond to the ones of the primary variable; i.e. a “regression” between classes instead of numeric variables.

This is usually done by building a “truth table” (in Figure 2 shown for the binomial case, i.e. two classes). “Effect” means that an area is RPA. The threshold of the secondary variable is varied until a given statistic over the table or over ROC space (lower graph), in terms of first and second kind errors, is optimized. The technique has been demonstrated as a classical ROC procedure in [Bossew \(2014\)](#).

		primary variable	
		no effect	effect
secondary variable	estimated no effect	correct estimate true negative (TN)	wrong estimate false negative (FN) second kind error
	estimated effect	wrong estimate false positive (FP) first kind error	correct estimate true positive (TP)

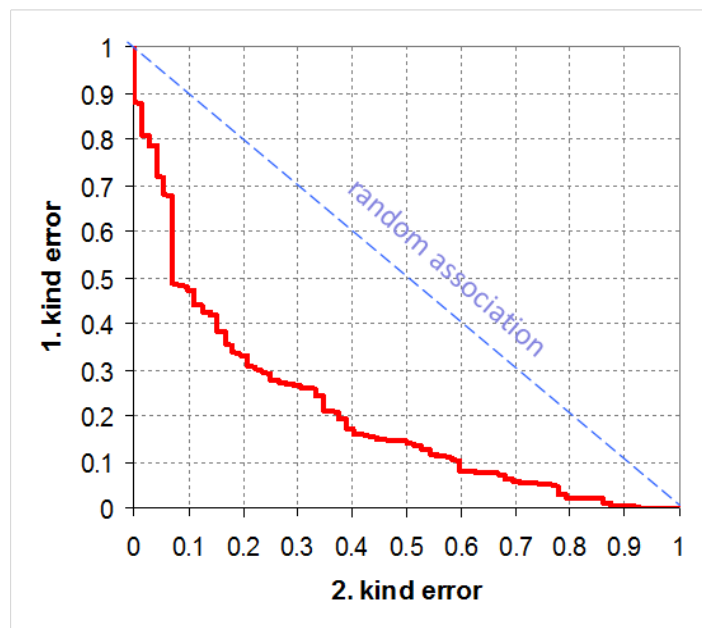


Figure 2: Top: truth table; “effect”: RPA. Primary variable: indoor Rn concentration, secondary variable: e.g. geogenic Rn potential. Bottom: ROC graph, Plot of 1. vs. 2. kind error. The curve is parameterized by the secondary threshold. Upper left corner: low values of this threshold. The higher the deviation of the curve from the diagonal, the stronger the association between the variables, i.e. the further away from random. Data from the example shown in section 4.1.1.5.2.

Requirements of classification reliability in terms of maximum allowable 1. and 2. kind classification error probability can be an additional external constraint to actual RPA delineation. This concept has been implemented in Germany, see the example discussed below. Extending the technique to multinomial classification and inclusion of several secondary variables is technically more demanding. A hybrid approach between (1) and (2) would be a method of indicator co-kriging type, being parametrical, but not carrying the entire information of the variables.

RPA can be estimated from different covariates, as available, such as indoor Rn concentration, the GRP, geochemical concentrations or geological units. If several are available, one would strive to using as much information as available to increase prediction precision. Multivariate RPA estimation involving simultaneously point and areal, as well as numerical and categorical data, which moreover are contingent as being controlled by common underlying “latent” processes, is technically demanding and no sound and accepted procedure seems to exist.

One alternative may be dimensional reduction, i.e. building one quantity out of several predictors. In radon science, this concept has been proposed as geogenic Rn hazard index, GRHI. It shall serve as a tool to quantify the susceptibility of an area to geogenic Rn, applicable independent of which predictor data sets are actually available in a region. The GRHI is subject of section 4.3.4.

Some examples are shown in Figure 3. Database is the indoor Rn dataset underlying the European Indoor Radon Map (EC, 2019). Exceedance probability estimated (p') as in section 2.5. Essentially, the pattern is the same for all methods, but locally, important discrepancies appear. The result of a classification type approach is shown in Figure 20, section 2.5., again looking quite similar. Since it also uses U concentration as predictor, it looks most similar to map C in Figure 3.

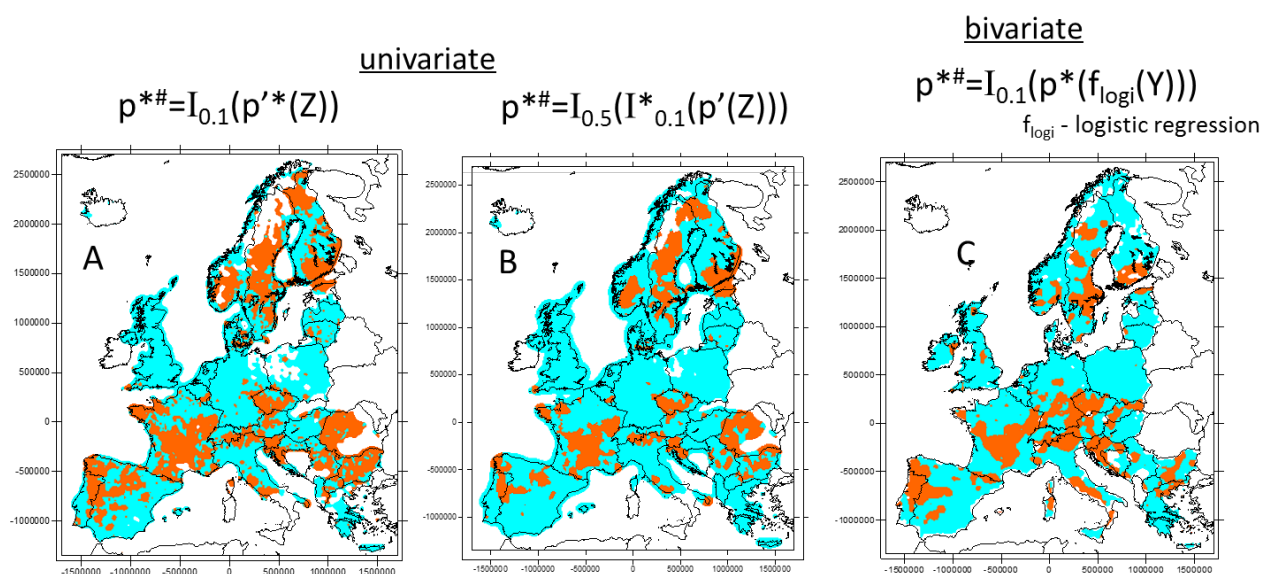


Figure 3: RPA maps of Europe, generated with different methods. Z – indoor Rn concentration; Y in map C – uranium concentration in topsoil. p – probability to exceed RL=300 Bq/m³; nomenclature as in Figure 1. # denotes classification into < and ≥ 0.1.

c) Data

Whatever its definition, estimation of RPAs has to rely on data. A great variety, in terms of quality and quantity, of surveys on indoor Rn and Rn related quantities has been generated in EU Member States for the last decades. Evidently, each country will choose a procedure adapted to the national Rn policy, to the objective Rn situation and to available data.

Indoor Rn

Generation of representative indoor Rn concentration databases with high geographical resolution is administratively demanding, expensive and time consuming. Data protection issues are important (and increasingly prohibitive). Nevertheless, most European countries have undertaken extensive surveys, some

even repeatedly. It can be estimated that Europe wide, considerably over a million indoor Rn data exist. The European Indoor Rn Map, part of the European Atlas of Natural Radiation (EC, 2019) is based on more than 1.1 million individual measurements.

Measurement of indoor Rn entails a number of QA problems on the survey level and on the level of individual measurement. The main challenge on survey level is representativeness (i.e. high accuracy or low bias of derived statistics). Preciseness (low random uncertainty) is a matter mainly of data volume, the more severe, the higher targeted geographical resolution and the higher true spatial variability. On individual measurement level, tasks are reliable calibration and evaluation of monitors and – largely unresolved – measurement protocols that ensure little vulnerability against manipulation.

Covariates, predictors and proxies

Usually a quantity called geogenic Rn potential (GRP) is defined to quantify the availability of geogenic Rn to exhalation into the atmosphere or to infiltration into buildings. Different quantities have been proposed; currently the most used seems to be the so-called Neznal-GRP, Neznal et al. (2004) based on radon in soil gas and soil permeability. Following the physics of Rn generation and transport, the GRP includes source term Rn concentration in the ground or uranium concentration and transport properties, namely soil permeability, emanation factors or soil porosity. The numerical values of the Neznal-GRP are between about 5 and 1000. Ensuring representativeness and logistic is easier for GRP than for indoor surveys.

The main conceptual advantage of GRP (or other geogenic quantity) over indoor Rn based RPA estimation consists in its independence of anthropogenic factors, i.e. building and construction type and usage. In this reasoning, an RPA should not depend on secular (if slow) changes in anthropogenic factors, in analogy to a seismically vulnerable area whose definition is not based on actual buildings but on the geogenic hazard; even if damage observed on buildings is used as one covariate for estimating the hazard potential.

Other predictor or proxy variables are uranium concentration in the ground, terrestrial gamma dose rate, geological units, soil units, hydrogeological features including karstification, and even tectonic properties such as presence of active faults. Dealing with predictors of different type (numerical, categorical), different spatial support (points, areas or lineaments), mutually correlated or contingent (creating collinearity problems in regression-type analysis) can be a statistical challenge.

Uncertainty of input data is traceable to several steps of the observation chain. Most important are errors in the measurement procedure, intrinsic uncertainty due to the stochastic nature of radioactive decay and errors in data attributes, such as location error or wrong assignment of house or room properties.

In practice, striving to use whatever data existent, Rn hazard is sometimes estimated using a mix of both concepts, i.e. including indoor and geogenic data in parallel.

Delineation of radon priority areas is a relatively new field in Rn research. The complication consists in the legal liability that has been introduced by the BSS. Labelling an area RPA or not can make important economical difference, given the possibly high costs of measures (prevention, remediation), which result from assigning an area RPA status. Understandably, stakeholders therefore wish a high degree of quality assurance in RPA delineation.

d) Classification problems

Several problems which occur if areas shall be classified as RPA (or any other criterion, based on measurements) will be described qualitatively here. Figure 4 shows a map of 6 municipalities and Rn measurements. The same are shown in Figure 5 together with a short description of the problems which appear when statistics are to be calculated over the municipalities, in particular if the RPA status shall be assigned.

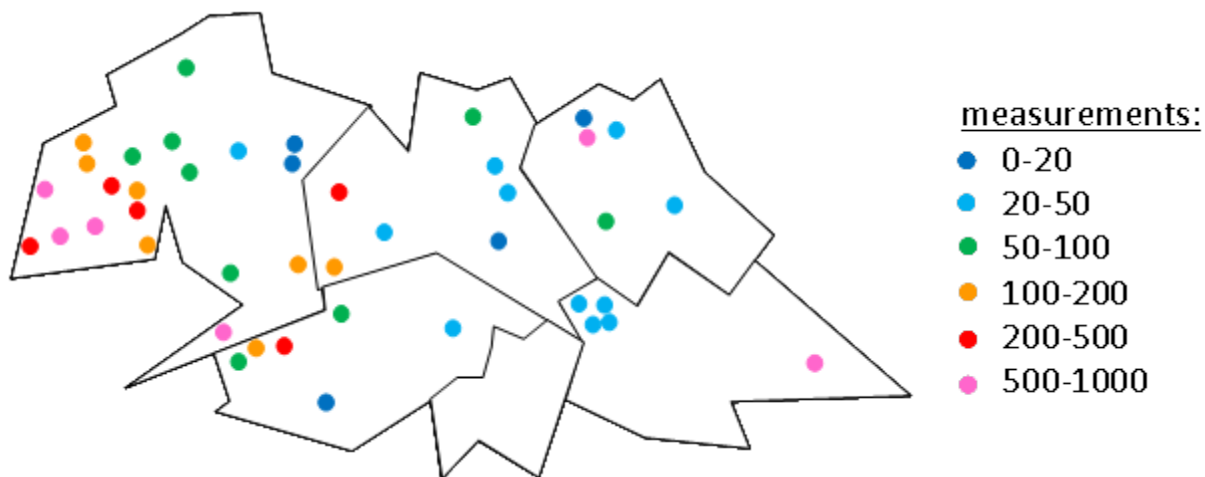


Figure 4: Six municipalities with Rn measurements.

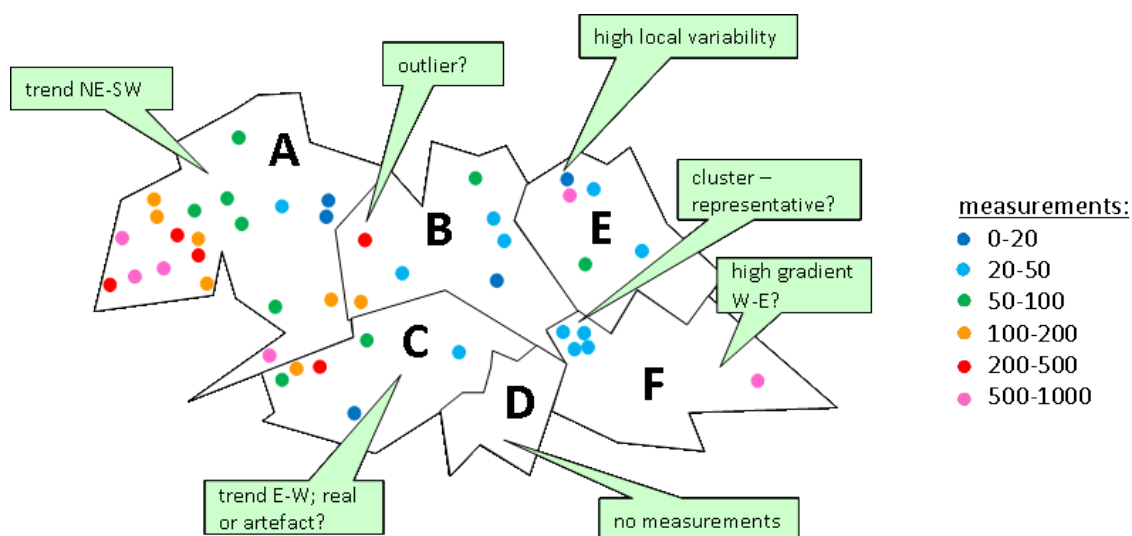


Figure 5: The same municipalities A to F as in Figure 4 with problems arising for classification as RPA.

These problems are:

- A. A trend about NE - SW seems to be present. A mean value will neither be representative for the NE nor for the SW part. The RPA status assigned to the area will be "unjust" for one of the two parts; if labelled RPA, this will not represent the NE part adequately, and v.v.
- B. One measurement may be an outlier in an area with otherwise low values. The outlier may distort the statistics of the area; on the other hand, it may indicate a (small) zone of high values. Optimally, this suggests a closer investigation.
- C. A trend E - W seems to exist. But due to low sampling density in the Eastern part, this impression can be accidental or spurious.
- D. No data available. In such case, inference on RPA status could be made from an interpolated surface or from predictors, typically geogenic controls. For example, if the relation between geological units and Rn values is known from neighbouring municipalities, the same is likely to apply in municipality D if the same geological units are present there.
- E. Two very different values in close vicinity may point to high local variability. However, it may also indicate data error or non-representative samples, i.e. observations made under different conditions or protocols than the other measurements.
- F. There is a cluster of observations in the W corner. On the other hand, the E part is very poorly sampled. The measurements may be located unrepresentatively or a real trend W - E is present. This cannot be decided from the data. If for example the measurements indicate IRC and the E part is low populated while there is a town in the W part, the sample may be indeed representative; but one may ask whether assigning RPA status based on the W cluster is adequate, because it evidently neglects the population in the E part.

In all cases, determination of the RPA status is highly problematic. In real cases, if inference on RPA status is made from few or possibly unrepresentative measurements, this should be addressed explicitly. Assessing uncertainty resulting from these sources is difficult in general, without resorting to extensive local detail studies.

2.2 Uncertainty definition

a) Uncertainty

In *frequentist* or Laplacian reasoning, the uncertainty of the value of a quantity reflects the distribution of outcomes of many observations of that quantity. In *Bayesian* thinking, uncertainty is intrinsic in the sense that it reflects lack of information about the quantity.

Overall uncertainty of seemingly simple quantity such as Rn concentration has different origins, such as stochastic uncertainty due to the final number of counts registered for calculating the concentration and calibration uncertainty. If one is interested in Rn concentration as estimate of the long-term mean concentration in a certain building, measurement design comes into play (detector placement and exposure period). If the measurement shall serve for estimating the mean over an area, representative (according to

objective of the survey) selection of the building and sample size (number of buildings) have to be considered. Also these design related factors are subject to uncertainty.

All factors add to total uncertainty and make up the *uncertainty budget*. Proper QA of a wanted quantity, such as RPA status in the context of this chapter, includes establishing an uncertainty budget.

Realistically, the uncertainty of some contributing factors is very difficult to assess in practice. Therefore, true completeness of uncertainty budgets is rather the exception than the rule. In particular, uncertainty of representativeness is very difficult to quantify.

b) Random quantities and random objects

Being results of estimation, RPAs are random objects in the following senses:

- a) whether a unit is labelled RPA or not is subject to uncertainty;
- b) the topology of a contiguous set of RPA labelled units is subject to uncertainty.

Understood as the realizations of stochastic process, all realizations of a RPA map look differently. In a frequentist sense, this variability constitutes the uncertainty.

Uncertainty has four sources:

1. Uncertainty of input data: target quantity C (long-term indoor R_n concentration), on which any RPA definition is ultimately based qua BSS;
2. Uncertainty of the spatial R_n “measure” Z which is the operational quantity on which the decision is based whether an area is RPA or not, or which is its degree of priority; for example the mean concentration over an area or the probability that in area C exceeds a reference level RL ;
3. Uncertainty of secondary data Y (predictors and proxies) used to estimate C or Z directly, such as GRP, geology etc.
4. Uncertainty of the model which quantifies the relationship between the Y and C or Z ; one has to distinguish between structural or model uncertainty (the analytical form of the relation, if applicable) and the one of the model parameters, which result from a fitting procedure.

These uncertainties propagate into the ones of quantities of concern:

- a) Uncertainty of the assignment of a RPA status or level to a given area;
- b) Uncertainty of the area to which a given status or level is assigned.

a) and b) are further discussed in section 4.3.1.4.c and subsequent. RPAs are results of classification, which is why classification uncertainty is of central concern in this chapter.

2.3 Uncertainty of input quantities and factors

The quantity which serves for decision whether an area is RPA or not, is a quantity aggregated from many more fundamental quantities. So to say, it sits on top of a pyramid of quantities and factors that all may be uncertain. Thus, uncertainties of possibly many sources propagate into classification uncertainty associated to RPA status of an area.

A (non-exhaustive) list of factors that contribute to the uncertainty budget of the quantities on which the decision is based whether an area is RPA or not, is given in the following.

a) Survey design

Since it is obviously impractical and expensive to measure indoor radon in all houses, it is necessary to design a survey carefully to obtain reliable and representative measurements of indoor radon in investigated region. In order to achieve a truly representative survey, it is necessary to have complete list of dwellings, which is seldom available, from which random selection of dwelling should be taken. Any deviation from pure random sampling can cause biases (IAEA, 2013). For example, it was shown that volunteer measurements could be biased due to over-sampling in radon priority areas (Burke and Murphy, 2011).

b) Measured vs. modelled indoor Rn concentration

There are three main sources of radon in dwellings. Usually, the most dominant is soil subjacent to the dwelling, then building material and finally water source. By assuming that radon exhalation from soil and building material is constant, and by neglecting release of radon from water, indoor radon concentration can be derived from simple differential equation, representing a simplified situation:

$$\frac{\partial C}{\partial t} = u_s + \frac{e_B S_B}{V} - (\lambda + \lambda_v)C + \lambda_v C_{van}$$

Where: C is the indoor radon concentration in time t,

u_s is the Rn volume entry rate from soil;

e_B is the surface exhalation rate;

S_B is the surface of the building material;

V volume of the chamber;

λ_v is the ventilation rate;

C_{van} is the outdoor radon concentration;

and λ is the radon decay constant.

Solution of this simplified equation is given by:

$$C(t) = \frac{u_s + \frac{e_B S_B}{V} + \lambda_v C_{van}}{(\lambda + \lambda_v)} \left(1 - e^{-(\lambda + \lambda_v)t} \right)$$

The first term in fraction is contribution from soil, the second one is from building material, while the third one is describing radon entry from outdoor air.

The more indirectly indoor radon is measured/estimated, the higher the uncertainty of indoor radon concentration will be, leading therefore to the higher classification uncertainty.

A schematic representation of measured quantities for estimation of annual average indoor radon concentration together with dominant sources of uncertainties is given in Figure 6.

Thus, minimum uncertainty in determination of average annual indoor radon concentration is by direct measurement of indoor radon concentration in one year, with detectors that have negligible sensitivity to thoron. If measurements are performed only in one period, uncertainty due to seasonal corrections are introduced. If radon measurement devices were not properly designed and they measure thoron and or progenies, results will induce additional uncertainty. For indirect measurement, i.e. by measuring secondary quantities, uncertainty of the estimated indoor radon concentration is becoming larger, and usually very difficult to estimate.

In the following subsections, uncertainties of some components will be discussed.

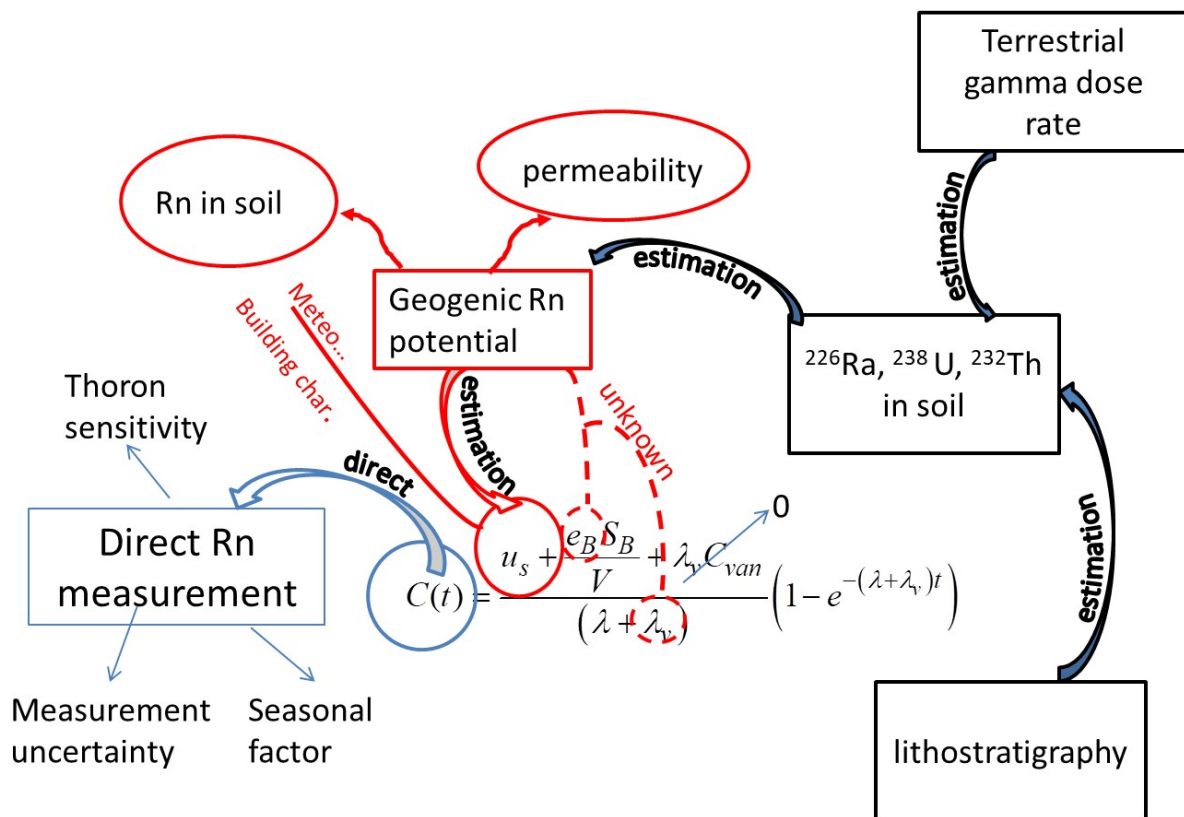


Figure 6: Determination of the indoor Rn concentration directly by measurement or by estimation from its physical constituents and sources of uncertainty.

c) Individual measurement

The smallest uncertainty of the annual average radon concentration is for direct measurement of annual radon concentration. There are numerous radon measuring devices, yet the most used ones are solid state nuclear track detectors (SSNTD), charcoal detectors and active devices. Typical uncertainties of these devices are

- Active devices: ~ 10% (WHO, Rn handbook, 2009)
- Charcoal canisters:
 - 10-35% (WHO, Rn handbook, 2009)
 - 25-35% (Zhukowsky et al., 2010.)
 - typical 10-35%, goes up to 70% (Zivanovic, 2016)

- SSNTD:
 - 10-25% (WHO,2009)
 - up to 30-40% (Zhukowsky et al., 2010.)

Measurement period for charcoal canisters is of the order of days; for active devices of the order of hours (although could be of the order of months if operated over long time), while for the passive devices can be several months, or year.

If a goal is to obtain average annual radon concentration, the most appropriate method would be to use SSNTD. They are also very cheap and easy to handle.

Intrinsic uncertainties of SSNTD are out of the scope of this report, yet we will underline a few sources of errors such as: calibration uncertainty, Poisson error, and not-Poisson errors like: etching process, geometry of the film within the detector, diffusion barrier property, etc. More details on these types of error can be found elsewhere (e.g. Zhukowsky et al., 2010.)

Calibration uncertainty or error enters as systematic error into the reported measurement result. It is treated in Deliverable D1 of MetroRadon project.

d) Uncertainty due to sensitivity to thoron

By design, there are various kinds of radon diffusion chambers. Some of them have significant sensitivity to thoron, of the same order or sensitivity as to radon. This is especially truth for older devices. Relative sensitivity to thoron, assuming that sensitivity to radon is 1, for some typical radon detectors is (Tokonami et al., 2002, 2005):

- Kfk(Germany,1981): 0.78
- RadTrak(USA,1991): 0.68
- NRPB/SSI: 0.05
- Radopot: 0.59 and 0.05 (discriminative Rn/Tn)

Using such diffusion chambers sensitive to thoron will overestimate radon concentration in dwelling. Radon concentration and corresponding uncertainty would be thus be:

$$C_{Rn} = C_{RnMeas} - k \cdot C_{Tn}; \quad \sigma^2_{Rn} = \sigma^2_{Rnmeas} + k^2 \sigma^2_{Tn}$$

Reducing uncertainty:

- place detectors sufficient from the wall,
- use Rn detectors with smallest Tn sensitivity, even better use Rn/Tn detectors
- surveys performed with KfK and RadTrak should be repeated with other detectors

e) Uncertainty due to seasonal factors (Temporal Rn variations)

A strong variation of radon concentrations in time was found. Roughly speaking, one can identify 2 types of variations of indoor radon concentrations: diurnal and seasonal. On daily basis, radon concentration is higher during the night and early morning, while they decrease during the day. Radon concentration is in general higher during the heating season, compared to non-heating season. Therefore, measurements should be long enough to enable averaging these variations. Researchers have also investigated changes of radon concentration in years.

Seasonal corrections should be applied when an Rn measurement covers only one season: 3 to 6 months in order to obtain average annual radon concentration.

Seasonal factors are prone to high uncertainty and could vary for different regions and different dwellings. In a lot of radon surveys performed in Europe measurement took place in winter season, but seasonal corrections were not applied in all surveys (Pantelic et al., 2019). Although there are numerous examples of seasonal factors we will discuss using two case studies in Serbia: Serbian national survey and survey in radon priority area (Niska Banja).

- Case study: Serbian national survey:
 - seasonal factor (year/heat season(6 months)) ≈ 0.83
 - st. dev. ≈ 0.32
 - range: 0.52 - 1.4
 - variation for single house: 60%

- Case study: Radon priority area:(Niska Banja)
 - seasonal factor: (year/summer(3 months)) ≈ 2.7
 - st. dev. ≈ 1.2
 - range: 1.1 - 6.0
 - variation for single house: 2.3 times

Two given examples illustrate large variation of seasonal factors and their misuse could lead to large over or under estimation of annual radon concentration in given region.

Reducing uncertainty:

- preferred 12 month measurement
- "regional" seasonal factors should be applied; dwellings should belong to one distribution

f) Uncertainty due to Year-to-year variation

Radon measurements in consecutive 12 months smooth daily, weekly and seasonal variations, and therefore are recommended for estimation of exposure. Variation of radon concentration in years can have large impact for estimation of risk. Correction of the risk due to exposure uncertainties can increase estimated risk significantly (Bohicchio, 2005). Darby and collaborators have estimated, based on pooled analysis that excess

relative risk for continuous exposure at 100 Bq/m³ was 0.084, while if corrected for uncertainty estimated based on year-to-year data, risk was 0.16, practically doubles (Darby et al., 2005, 2006).

Investigation of year-to-year variation is not only important for reducing uncertainty in risk estimation, but for classification of regions. Namely, large yearly variation would imply that radon survey performed in one year and assigning a class to some region does not necessary mean that selected region will belong to the same class if survey will be performed another year.

The 5 year analysis of the radon exposure in 76 Italian dwellings showed that variation ranges from 3% to 42% with the mean variation of 14% (Bochicchio, 2009). Other authors have reported slightly higher variations: in UK is 50% (Darby et al., 1998) and in China 43% (Lubin et al., 2005).

Variation of annual average indoor radon concentration in 98 dwellings was investigated by Steck (2009). Results of this investigation is presented as another interesting case study:

Case study: 10 years, annual average Rn variations in 98 dwellings (Steck, 2009)

- typical Rn variation: 24- 27%
- variation for single house: 3 - 110%

Reducing uncertainty:

identify dwellings prone to high radon variation:

- renovated houses (HVAC, new windows, doors...)
- houses placed in extreme environment (snowy, windy...)

A few conclusions could be drawn from these findings: Mean year-to-year variation can vary significantly in different countries, which could lead to misclassification of the region. A large variation for the single house (110% in the case of Steck (2009)) could be found when houses are renovated (like HVAC, new windows...). Geological location (radon potential) of the house did not change and in general building characteristics, but results show such a large change. This will be in particular discussed in the next section of uncertainties due to variations in air exchange rates.

g) Uncertainty due to air exchange rate – ventilation rate

Air exchange rate in dwelling mainly depends on climate conditions, building characteristics and living habits. Typical variations of air exchange rate are from 0.2 – 2 h⁻¹ with the average of 0.63 h⁻¹. Higher ventilation rates, with even >10 h⁻¹ are in dwellings with artificial ventilation (ICRP).

Question of air exchange rate in dwellings becomes more important due to trends in the civil engineering to make energy efficient dwellings. A new materials with better thermal insulations are introduced, windows and doors being replaced in old dwelling, having as a consequence decrease of air exchange rate. This consequently leads to an increase of indoor radon concentration (Yarmoshenko 2014).

Air exchange rate is important factor in reduction of indoor radon concentration. There is inverse dependence of air exchange rate on radon concentration.

$$C(t) \propto \frac{1}{\lambda_v}$$

Chao et al. (1997) have investigated ratio of indoor and outdoor radon concentration for different ventilation rates and it was found that for low air exchange rate of 0.2 h⁻¹, ratio goes to 46.5, while for ventilation rates higher than 3 h⁻¹, indoor concentration is close to outdoor. This result indicates that for two dwellings located at the same geological unit, indoor radon concentration can significantly vary due to difference in building characteristics or habit.

Variability of air exchange rates investigated in almost 3000 dwellings in the US is shown in Figure 7.

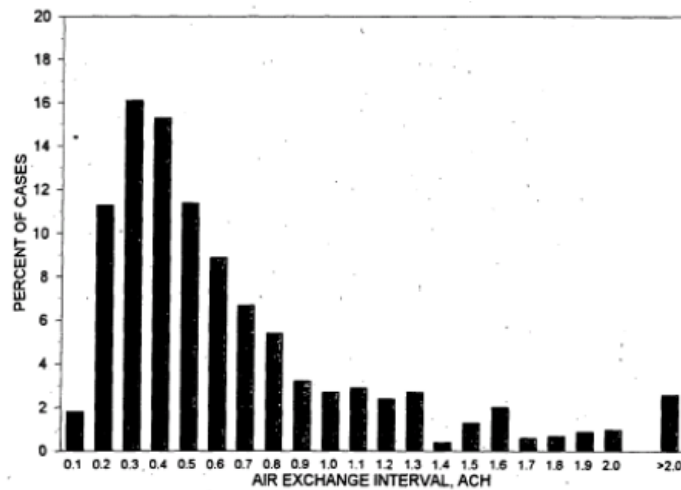


Figure 7: Frequency distribution of estimated residential air exchange rates for all regions combined. (taken from Kennedy, 1995)

h) Uncertainty of predictor and proxy quantities: GRP

Although the main source of indoor radon concentration is soil beneath the dwelling, indoor radon concentration is strongly influenced by the anthropogenic factors such as building characteristics, living habits, climate and meteorological conditions. These variations could be so large that in the Norwegian guidelines for radon measurement is stated that only direct integrated measurements can be used for assessment of indoor radon concentration, while indirect measurements (such as soil gas measurements, external gamma dose rate measurement, geology, etc.,) of neighbouring site cannot be used for decision whether remedial measures are needed or not (Jensen et al., 2004; NRPA, 1996).

In order to have a quantity to assess a radon risk coming from geology, that does not depend on anthropogenic factors, a geogenic radon potential was derived, as a measure what “earth delivers”.

It is derived based on the field measurements of the radon concentration in soil gas and the gas permeability of soils using following formula:

$$RP = \frac{C}{(-_{10} \log k - 10)}$$

Where RP is the radon potential, C is the soil gas radon concentration (kBq m⁻³) and k is the soil permeability (m²).

Additional factors such as bedrock types, the presence of faults, the relief of the terrain, and regional geological units) are also useful for the final determination of radon index (RI) which can take values: low, medium, or high (if RP < 10, then RI is low; if 10 ≤ RP < 35, then RI is medium; if 35 ≤ RP, then RI is high) (Neznal et al., 2004).

Although RP is derived from 2 measured quantities, in literature is uncommon to estimate its measurement uncertainty. Both measurement of radon in soil gas and permeability are instantaneous measurements, therefore reflecting only current meteorological conditions. Nevertheless, geogenic radon potential was investigated throughout whole year and while soil gas radon concentration varies significantly with water content, GRP showed to exhibit much smaller variations. Variations of GRP over a one year period measured at one point in Osijek, Croatia is shown in Figure 8 (Radolic et al., 2017).

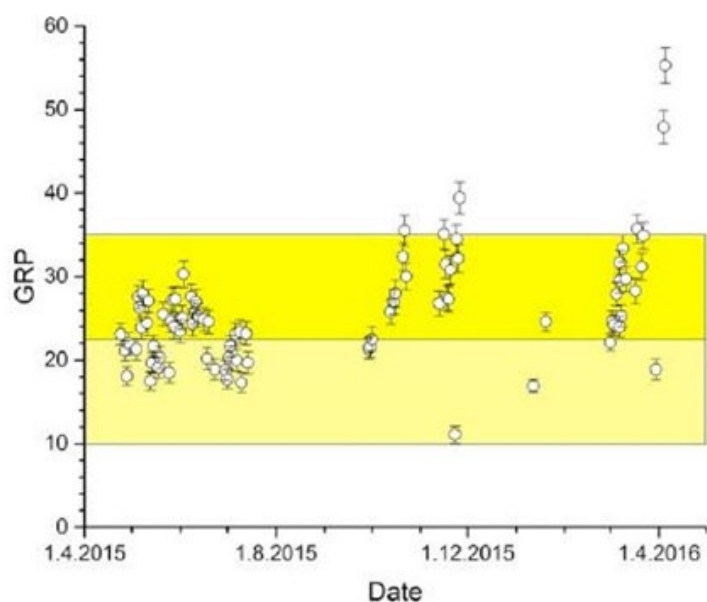


Figure 8: Annual variation of geogenic radon potential measured at one point in Osijek, Croatia (taken from Radolic et al., 2017)

Geogenic radon potential represents secondary quantity for estimation of indoor radon concentration and therefore additional uncertainty is introduced when trying to estimate annual average indoor radon concentration. It only allows estimating radon volume entry from soil, and information on exhalation rate from building material and ventilation rate is missing.

Therefore it is very difficult to estimate the connection between GRP and annual average indoor radon. It is very hard to quantify uncertainties which are mainly of anthropogenic origin. Some of the factors are enlisted as follows:

- building type (rural, urban, house, apartment), building age, floor, basement presence
- quality of building (different in different regions, depending on the development and wealth of the region)
- building materials type, doors/windows tightness (ventilation coefficient)
- the question is should the clusters depend (only) on geology, or on the similarity of the buildings in the cluster

Other secondary (proxy) variables

Indirect estimation of indoor radon concentration i.e. using the secondary (proxy) variables is prone to even larger uncertainties, at least for estimation of average annual indoor radon concentration of a single house. With measurement of geogenic radon potential, only quantity of volume radon entry is estimated, while no information on exhalation rate and ventilation rate are available. When secondary variable is uranium and/or radium content in soil, uncertainty is even larger, since factors that link them to geogenic radon potential such as emanation, porosity, water saturation factor are not known. Even further away from the estimation of indoor radon concentration is lithostratigraphy in which based on lithology where radium and uranium content is assumed.

2.4 Uncertainty of aggregated quantities

a) Regional mean

After estimation of annual average indoor radon concentration in a single house, it is necessary to estimate the annual average over investigated region, since it is not feasible to measure radon in each dwelling. Additional uncertainty arises from the extrapolation of the measured/estimated indoor concentrations in selected dwellings to the whole area.

The area over which radon concentration is averaged is depending on the goal and size of the survey. Many surveys followed approach of JRC and its European Atlas of Natural Radiation in which annual average indoor radon concentrations in ground floor rooms of dwellings is defined within the grid of 10x10 km² cell size. (Dubois et al., 2010). Some surveys apply 5 km x 5 km, 1 km x 1 km, or even 0.5 km x 0.5 km grids. Apart from these artificial grids, other area over which average radon concentration is estimated can be: by municipality or by geological units.

The uncertainty of the mean indoor Rn concentration depends on:

- true spatial variability of Rn concentration
- survey design
 - number of measurement points
 - representativeness
 - variability of building types (rural, urban; house, apartment),

- quality of building (different in different regions, depending on the development and wealth of the region)
- building materials type, doors/windows tightness (ventilation coefficient)
- biased/unbiased sampling
- geological data:
 - variability of bedrocks/soil within the grid
 - geological uncertainties (different maps)
 - Cartographer uncertainty
 - Scale-induced classification (in principle a resolution problem – each “pixel” is considered uniform, although in the reality, it isn’t)
- statistical uncertainty
 - Grid formation: density, clustering of data...
 - Predictors used (Radium concentration in soil, permeability, faults presence, bedrock type...)
 - Interpolation/extrapolation problem (kriging vs spline).

Friedman and collaborators have shown that coefficient of variation scales like $1/\sqrt{n}$, where n is the number of measured dwellings as shown in Figure 9 (Friedman et al., 2017)

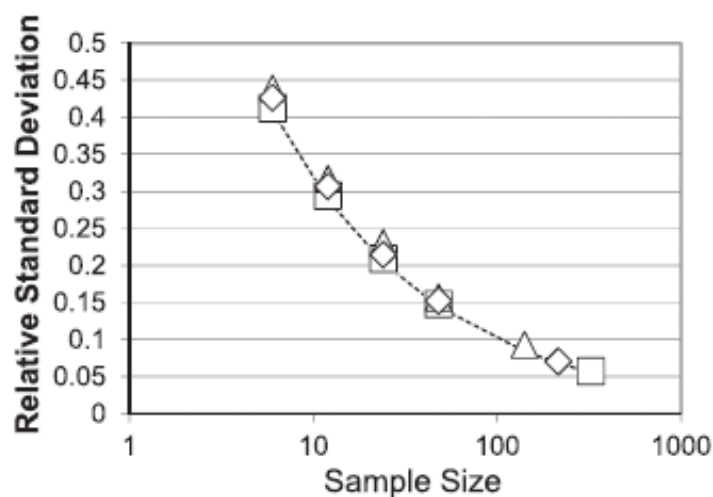


Figure 9: Dependence of the coefficient of variation (relative standard deviation) for the arithmetic mean as a function of the sample size, for 3 different regions (Friedmann, 2017)

On the other hand, it is shown that geometric standard deviation (variance of radon concentration), reduces by “fixing” factors influencing the dispersion of indoor radon concentration.

Based on the data from Sverdlovsk oblast and Niska Banja, it was found that by reducing of heterogeneity, GSD decreases as well for certain factor (Yarmoshenko et al., 2016):

- For geological factors: GSD reduces by factor 1.3

- For geological + building characteristics: GSD reduces by factor 1.7
- For geological + building + living habits: GSD reduces by factor 1.9

Reducing uncertainty:

- To increase number of measurement points per grid cell (making narrower the distribution)
- To reduce heterogeneity of data: based on geological factors; building characteristics (e.g. urban/rural), living habits (e.g. good/bad ventilation)

Therefore, a trade-off between two requirements should be taken into account!

Finally, a numerous investigations appear recently in which spatial distribution of radon, not only by measuring indoor radon concentration, but including all available proxy parameters.

b) Regional exceedance probability

A recurrent task in radon science is estimation of the probability that a reference level (RL) of radon concentration is exceeded within an area. If the observation data are known, an empirical estimate is easily found by counting the observations above the RL and dividing by the number of data or sample size (n). In many cases, the individual data are not available because of data protection, since in particular indoor radon is a sensitive subject.

However, the data may be available in aggregated form. A typical example is the European Atlas of Natural Radiation (Cinelli at al. 2019, EC 2019), whose map of indoor radon concentration (IRC) relies on aggregated data per 10 km × 10 km cells, provided by the participating national authorities. They calculate the statistics per cell, in this case

n - number of measurements;

AM - arithmetic mean;

SD - standard deviation;

AML - arithmetic mean of the natural logarithms;

SDL - standard deviations of the natural logarithms;

Med – median;

Min – lowest value;

Max – highest value.

The individual data remain with the participants. When the European IRC map was designed in the mid-2000s, the procedure including the list of aggregated quantities was chosen such as to allow estimating further parameters. Exceedance probabilities, i.e. $\text{prob}(Z > \text{RL})$, are not provided. Since these are relevant quantities in regulatory practice, they have to be estimated.

Radon priority areas

A common definition of RPAs is of the kind, an area B is RPA if $p = \text{prob}(\text{IRC} > \text{RL}) > p_0$, with RL – the references level and p_0 a set value; for example, $\text{RL} = 300 \text{ Bq/m}^3$ and $p_0=0.1$. Implementation requires estimating p from data, which involves uncertainty.

The straight forward way to determine p in an area is done by counting the number of data $z_i > \text{RL}$ and dividing by the total number of data within B. This estimator is unbiased but very imprecise for low number of data. If the individual data are not available, but only the moments (see above), different estimators have to be chosen. In the following, the performance of three estimators is investigated, regarding uncertainty (precision) and bias (accuracy).

Estimation of exceedance probability

If only moments are given, as in the example of the European IRC map, the task consists in estimating exceedance probabilities from moments. This is commonly done by assuming data to be log-normally (LN) distributed with cell and estimating exceedance probability as tail area above the RL,

$$\text{prob}(Z > \text{RL})(\text{modelled}) = \Phi\left(\frac{\text{AML} - \ln(\text{RL})}{\text{SDL}}\right), \text{ with AML estimated from data and SDL estimated from}$$

data or assumed fixed; the latter is not unreasonable because from experience it is known that IRC with cells of this size typically have $\text{GSD}=2$ and hence $\text{SDL}=\ln(2)$. However, the model rests on the LN assumption. Studies have shown that it is mostly reasonable (among other, [Bossew 2010](#)), but there is no physical reason why the distribution of the population is truly LN. Further, even if it was, the sample is not necessarily representative, so that AML and SDL estimated from data may not reflect the true LN parameters (μ, σ) adequately.

As the next problem, even if the sample was LN, it is not clear whether the estimators of the exceedance probability are unbiased or not. For example, it is known that the common estimator of the standard deviation, $\text{SD}' = \sqrt{\sum_{i=1}^n (z_i - \bar{z})^2 / (n-1)}$, \bar{z} the arithmetic mean, is an unbiased estimator of the true SD of the *sample*, but if the z_i are assumed draws from a normal population $N(\mu, \sigma)$, while $\text{AM}(z)$ is an unbiased estimator of μ , the $\text{SD}'(z)$ is a *biased* estimator of σ .

As a conclusion, there are two tasks:

1. Investigate the performance of estimators of exceedance probability under LN assumption;
2. Investigate the effect from deviation of LN (because either the true distribution is not LN; or the sample is not representative).

(1) will be studied in the following. For (2), it turned out that this goes beyond the Metro Radon project and will therefore left for further studies.

Estimators of the exceedance probability

The following estimators of the exceedance probability $p := \text{prob}(Z > RL)$ are evaluated:

empirical, $p^0 := (\#z > RL)/n$;

$$p^I := \Phi\left(\frac{AML - \ln(RL)}{SDL}\right)$$

$$p^{II} := \Phi\left(\frac{AML - \ln(RL)}{\sigma}\right), \text{ i.e. logarithmic dispersion assumed known}$$

$$p^{III} := t_{n-1}\left(\frac{AML - \ln(RL)}{SDL} \sqrt{\frac{n}{n+1}}\right) \text{ (Bossew et al. 2015 and references there)}$$

p^0 is unbiased but it requires knowing the individual data z , whereas the other rely on the estimated moments.

Sampling statistics

The algorithm is as follows:

(1) Generate n independent LN variates. For normal variates, the common algorithm has been used: $x = \text{sqr}(-2 \cdot \ln(u_1)) \cdot \cos(2\pi \cdot u_2)$; $u_1, u_2 \sim U([0,1])$; $y = \exp(\mu + \sigma x) \sim \text{LN}(\mu, \sigma)$.

n represents the sample from the population. Note that this assumes an infinite population. The algorithm is somewhat more complicated for finite population, such as buildings in reality. But the difference is relevant only for high sampling rates, which one rarely faces (Bossew 2017).

The following sample statistics are computed:

AM, SD, AML:=AM(ln y), SDL:=SD(ln y), CV=SD/AM, GM=exp(AML), GSD=exp(SDL), and the probabilities $p^0, p^I, p^{II}, p^{III}$ as given in section 2.1.

(2) Do this m times ($m=500,000$ chosen) and calculate the AM and SD over the m realizations. The relevant statistics are

- uncertainty, defined as $CV = SD/AM$ over realizations. Evidently, for any parameter Θ , $CV_{\text{real}}(\Theta) = SE(\Theta)/AM(\Theta) = CV(\Theta)/\sqrt{n}$, $CV(\Theta)$ the population statistic. Here, calculation served for verifying the correctness of the algorithm.
- bias, defined as $b = AM/(\text{true parameter}) - 1$.

Although the main interest is in the exceedance probability, also the other statistics were calculated. For SD, SDL and CV, bias corrections were applied (Annex 1). However, for calculation of p^I and p^{III} , the raw SDL were used, because it appeared that their biases increase if the corrected SDL are used. The reason for this unexpected finding should be investigated further.

As true parameters of the LN distribution, $\mu=0$ and $\sigma=\ln(2)$ were chosen corresponding $GSD=2$ which is realistic in spatial dispersion of indoor radon.

From the properties of the LN distribution follows, true AM=1.2715, true SD=0.9986, true CV=0.7854 and true $\text{prob}(Z>RL)=0.0890$ for $RL=k \text{ AM}$ ($k=2$ as an example). The calculation was performed for multiples k between 2 and 3.8 with step 0.2.

Bias of estimated exceedance probabilities

In Figure 10, a population $LN(0, \ln(2))$ is assumed and a number of statistics calculated in dependence of sample size n (x-axis). The reference level is set $RL = 2 \text{ AM}$. Among the statistics are p^0 (labelled p in the legend of the figure) and p^I , p^{II} and p^{III} .

As expected, the estimators AM' and $p^{0'}$ are without bias. The same is true to SDL'_{corr} which is SDL' with bias correction applied (Annex 1). For SD' , bias correction mitigates the bias, but it remains considerable, since bias correction is valid for normal, but not for LN populations. Similarly, bias correction for CV, reduces the bias, but it remains considerable. The raw GSD' has little bias, as known.

Unexpectedly, among estimators of the exceedance probability, p^I has the lowest bias, but still considerable for low sample size. All have positive bias, which means that true probabilities are over-estimated. As a consequence, RPA delineation based on such probability estimators are conservative.

Bias and uncertainty of the exceedance probability estimators p^I , p^{II} and p^{III} in dependence of sample size and for several relative reference levels ($k = RL/AM$) is shown in Figure 11. The estimator p^{III} has the highest bias but the lowest uncertainty. (The strange effect for samples size $n=2$ and 3 is due to that for these, the exact formula of the t distribution was used, and the approximation, Annex 1, for $n \geq 3$. The approximation formula may be inaccurate for low n .)

The lowest bias for small samples has p^I , but still 200% for high RL, whereas it is up to a factor 10 (1000%) for p^{III} . The empirical estimator p^0 has no bias but very high uncertainty.

As a conclusion, it turns out that of the estimators based on moments, p^I has lowest bias, but highest uncertainty. Until estimators with lower bias are found, the safest (least uncertain) choice is p^{III} in spite of its high bias.

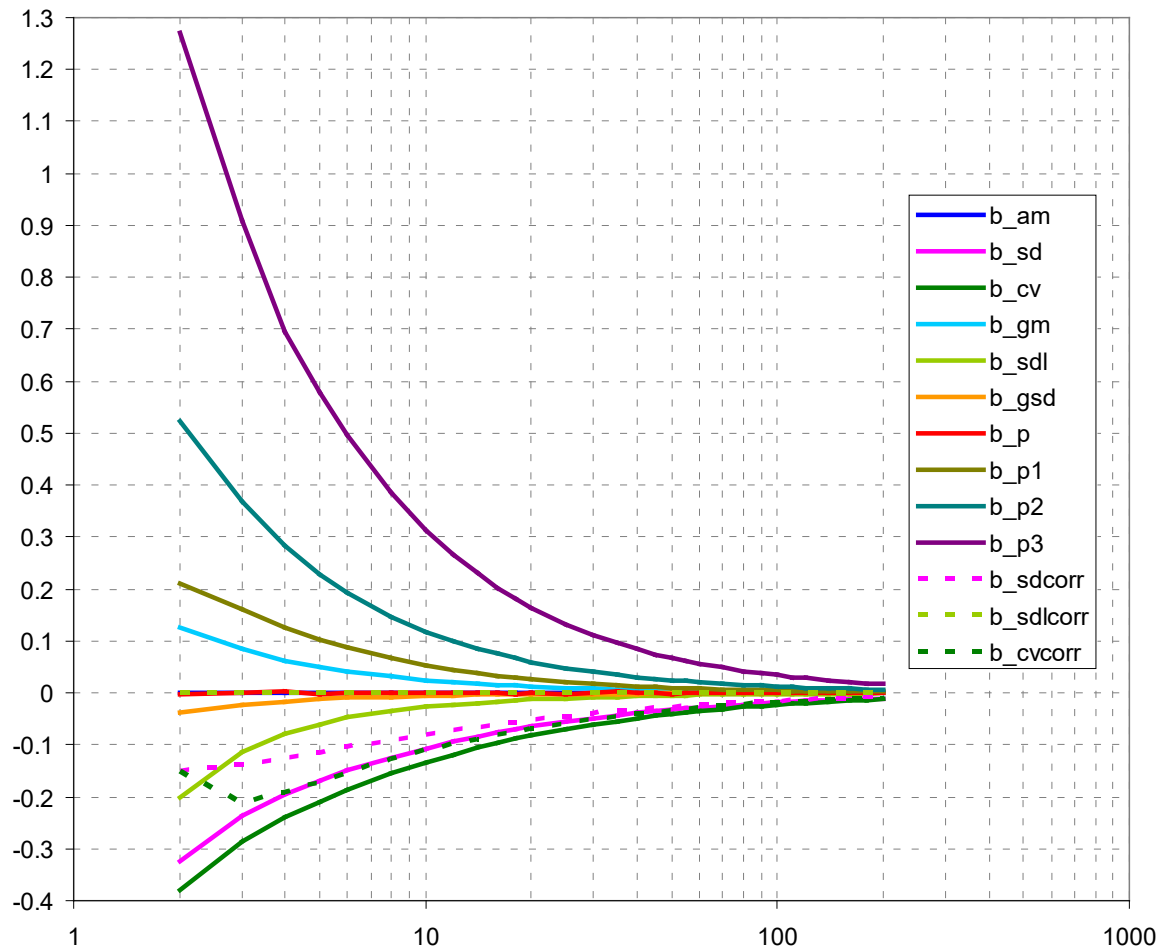


Figure 10: Biases of sample statistics, samples from LN(0,ln(2)); in dependence of sample size. RL=2 AM. x-axis: sample size

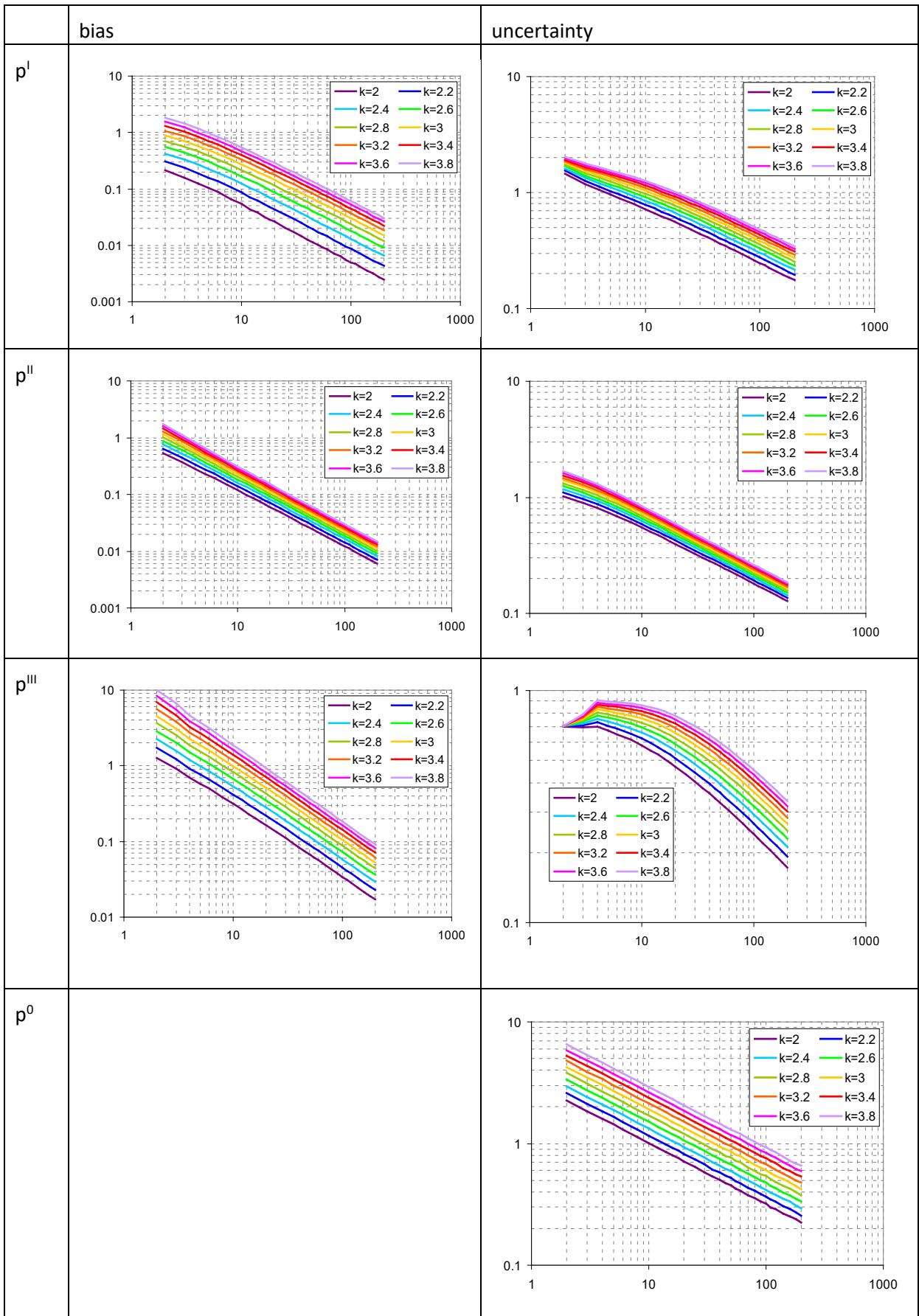


Figure 11: Dependence of the bias (left column) and uncertainty (right column) of different estimators of exceedance probability on sample size, for different multipliers k which define the reference levels RL by $RL=k$ AM. x-axis: sample size. First row: p^I ; second row: p^{II} ; third row: p^{III} ; fourth row: empirical p^0 . Underlying distribution: $LN(0, \ln(2))$.

c) Classified RPA status

The wanted final result of RPA classification is a tiled partition of the domain (e.g. a country) into regions which are assigned certain RPA status, such as “yes / no”, “classes I, II, III, IV”, “low / medium / high”, etc. These assignments are *ordinal* quantities because there is a natural rank between them (but not *interval* quantities, because in general, there is no equal difference between subsequent classes, even if a meaningful distance measure between classes is defined, based on a defined metric on classes).

The RPA status variable is itself a random variable, as it results of an estimation procedure. Therefore, any estimate is affected by uncertainty. But not being an interval-type quantity, it cannot be assigned a standard deviation or confidence interval or the like.

In this spatial context, two aspects of classification uncertainty must be considered:

- Given an area, such as a grid cell or a municipality, its RPA status is uncertain, in general, as result of an estimation procedure. This aspect is further treated in section 4.3.1.5.
- Conversely, given a RPA status, the union of areas to which this status (or class level) is assigned, is uncertain. The areas, to which a certain RPA status is assigned, are *random objects*. Uncertainty is expressed as uncertain area, location, shape and topology. This aspect is further treated in section 4.3.1.6.

The two aspects of uncertainty result from the uncertainties of quantities, through which classes are defined (area mean or exceedance probability, 4.3.1.4a and b) and further from the quantities, from which mean and exceedance probability are calculated, namely indoor Rn concentration or any predictors and proxies..

The aggregation chain and corresponding chain of uncertainty propagation are shown in Figure 12.

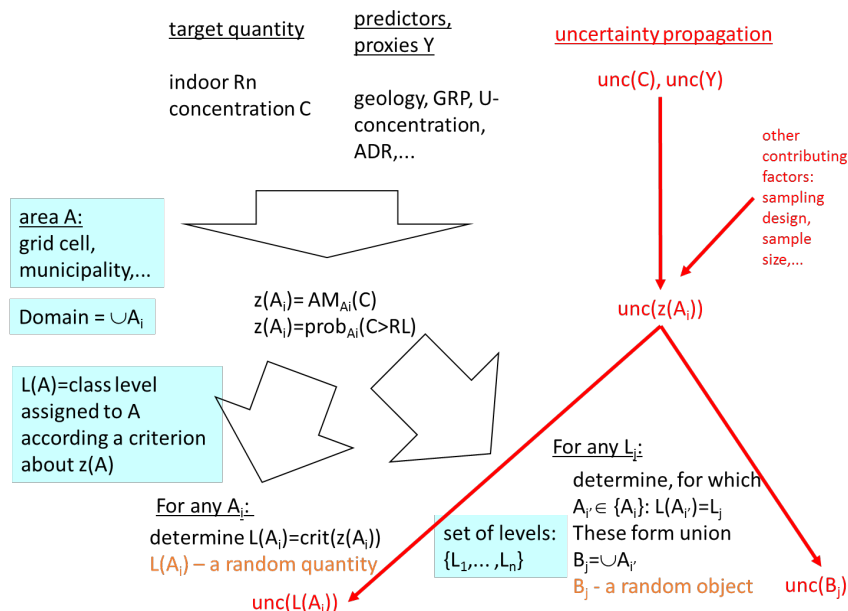


Figure12: Concept: Random objects and their uncertainties

d) Finite and theoretically infinite population

The number of houses in an area is necessarily finite, while the number of theoretical measurement locations of dose rate or of taking soil samples is infinite. Most statistical considerations are performed implicitly anticipating infinite populations. For truly finite populations this is acceptable as long as the sampling rate (sample size / population size) is low. Calculations may become cumbersome and almost intractable for high sampling rates.

For example, assuming infinite population, the standard deviation of the mean of a sample decreases as $1/\sqrt{n}$ with sample size n (more strictly, assuming statistical independence of the observations). This is different for finite true population, because the SD of the mean equals 0, once the sampling rate =1, i.e., the entire population tested.

The sample size effect is summarized in Figure 13, taken from [Bossew 2018d](#).

In a cell U: population N , true “successes” K (i.e. $z > 200$).
True success probability: $p = K/N$.

Sample:

size n (sampling rate n/N), successes k .

Hypergeometric distribution:

$$\text{prob}(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

CDF, for estimating whether success rate $(k/n) = p' > 0.1$ (e.g.):

$$\text{prob}(X \leq k) = 1 - \frac{\binom{n}{k+1} \binom{N-n}{K-k-1}}{\binom{N}{K}} {}_3F_2 \left[\begin{matrix} 1, k+1-K, k+1-n \\ k+2, N+k+2-K-n \end{matrix}; 1 \right]$$

generalized hypergeometric function

Unpleasant! Little chance for practical use!!

If $N, K \gg n$, i.e. low sampling rate, but n still “large”, and $p \gg 0, < 1$:

$$\text{prob}(X \leq k) \approx \Phi \left(\frac{k - np}{\sqrt{np(1-p)}} \right)$$

- conditions often not fulfilled;
- true p not known \rightarrow replace by p'
 \rightarrow replace $\Phi \rightarrow t_{n-1}$??

Figure 13: Mathematical background of the sample size effect for finite population

Since increasingly, we face surveys with high sampling rates, more attention should be directed towards sampling statistics of finite true populations.

e) Uncertainty budget

Many kinds and sources of uncertainty contribute to resulting overall uncertainty (Figure 14, taken from [Bossew 2018c](#)). Intrinsic data uncertainty relates to classical metrological uncertainty. Data as samples from a

population can lead to uncertainty if the sample is not representative (Figure 15). If a modelling step is needed, the model may itself contribute to uncertainty: it may be ill-chosen (e.g. assuming normal instead of log-normal distribution); it may be unduly simplified (e.g. ignoring deviations of log-normality; removing predictors to simplify computation and to avoid over-fitting, etc.); finally parameterization of a model may be imprecise, as this is performed based on limited and again uncertain observations and other data. Regarding choice of model and parameterization, one should also not forget that different validation criteria may lead to preference of different models. (Accuracy, precision, 1st / 2nd kind error rates, cross-validation correlation, RMSE, other metrics?)

Summary: Sources of uncertainty


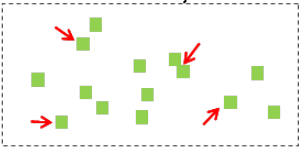
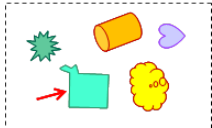
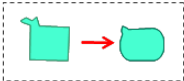
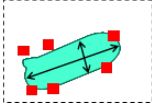
- **Data uncertainty**
 - Intrinsic: data as observations; 
 - Data as samples 
- **Model uncertainty**
 - Structural uncertainty: choice of model 
 - Simplification uncertainty 
 - Parameter estimation uncertainty 

Figure 14: Sources of uncertainty which enter the uncertainty budget.

Data

Data are always

dirty, noisy, incorrect, erroneous,
incomplete, ill-defined, uncertain.



Data as observations:

- measurement uncertainty (not only counting uncertainty! Sampling and measurement procedures include uncertainty, sometimes this is the most important part, but difficult to quantify)
- “semantic” uncertainty
(Value reported ground-floor measurement, in fact first floor,...)
- wrong (e.g., geology wrongly classified)
- sloppiness errors (manual copying of data, wrong insertion into table, Excel misreads decimal point, x and y coordinates confused,...)

Data as samples from a population:

- not representative
(relevant if the target is a statement about the population!)
- finite / limited sample size \Rightarrow estimation uncertainty

Figure 15: Sources of data uncertainty

Complete uncertainty budgets are usually very difficult to establish for highly aggregated data, such as the RPA status of an area.

An example of how one kind of uncertainty propagates into an aggregate is the following (taken from [Bossew 2018d](#)).

Let the observed R_n concentration: Y , true: Z

pdf of Y : $h(y) = \int g(y|z) f(z) dz \dots$ compounded distribution, g =error distribution due to measurement uncertainty, f =true natural distribution.

$g(y|z)$ propagates into exceedance probability $\text{prob}(Y > y_0) = 1 - H_Y(y_0)$

Assume uncertainty unc of observed values proportional the true value, $\text{unc} \sim Z$. Then, $\text{Var } Y = \text{Var } Z + \text{AM}(Z * \text{unc})^2$

For the exhaustive data from the toy example, section 4.3.1.5c), and assuming $Y \sim N(Z, Z * \text{unc})$, numerical calculation leads to the graph, Figure 16. It can be noticed that measurement uncertainty leads to considerable overestimation of exceedance probability.

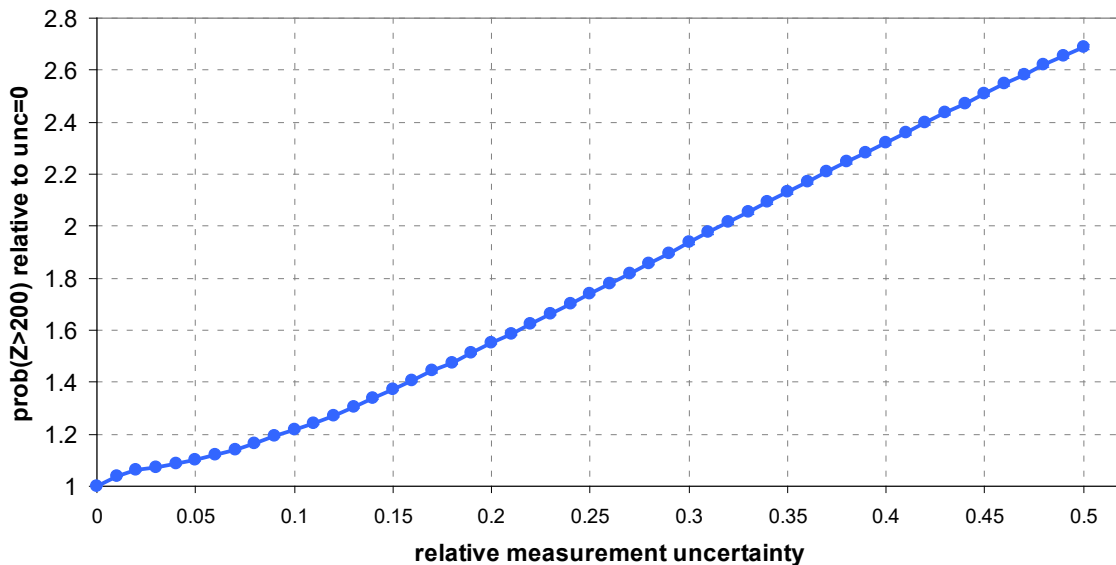


Figure 16: Inflation of the exceedance probability $\text{prob}(Y>200)$ due to measurement uncertainty.

A more realistic error model for IRC measurements of the form $\text{unc} \sim \exp(-\alpha Z)$, calibrated such that $\text{unc} = 30\%$ and 10% for $Z=\text{IRC}=10$ and 1000 Bq/m^3 , respectively, leads to $\text{prob}(Y>200 \text{ Bq/m}^3) = (8.1 \pm 0.6)\%$, i.e. almost 1.5 times the true value (5.6%).

As a conclusion, the RPA status seems to be systematically overestimated (=false positives) due to dispersion inflation caused by measurement uncertainty.

The question, how to “de-compound” or invert, to retrieve true exceedance probability, appears unresolved so far.

2.5 Quantifying classification uncertainty

a) General considerations

Classification means mapping a quantity into set of mutually exclusive classes defined by given criteria. Categorical quantities can be classified by properties or attributes (e.g., geological units into magmatic / metamorphic / sedimentary rocks). Continuous quantities are commonly classified according their values, whether they exceed a threshold which defines the limit between two consecutive classes, or not. The discussion in this report deals with the latter case; the continuous quantity is long-term indoor Rn concentration.

As said, estimation of RPAs is a classification problem, in that a geographical domain (country) is divided into two or more classes according to the RPA definition. Whichever estimation technique applied, assignment of a location or an area to an RPA class will always be affected by uncertainty. Its sources are multiple, from true variability of the mapped quantities on spatial scale below estimation support to data uncertainty to model structural and parameterization uncertainty resulting from estimation. Establishing exhaustive uncertainty

budgets is difficult. Whereas the uncertainties of the estimated actual levels of the Rn measure are commonly quantified by confidence intervals, the ones of classes are given by first and second kind classification error rates. *First kind errors* or false positives or false alarms, denote that an area is falsely labelled RPA although it is not; *Second kind errors* or false negatives or false non-alarms mean that an area is falsely labelled non-RPA, although it is in reality. (The logic can be extended for multinomial classification schemes, i.e. several RPA class levels.)

High classification error chance must be expected in particular for geographical units whose Rn measure is close to the class limits.

Note that in addition to uncertainty of correct assignment of an area to a RPA class, an individual location within an area that corresponds to a class can deviate from the class, again due to the true variability within an area. Apart from classification uncertainty which is inevitable since the respective areas result from a statistical estimation process, and which relates to the geographical units on which the RPA definition is based, it must be expected that individual houses do not conform to the RPA definition. For example, a house located in an area labelled non-RPA, or a cluster or sub-area within the non-RPA, can still have Rn concentration exceeding the reference level. The obvious reason is the high spatial variability of Rn concentrations, resulting in a possibly long "right tail" of the frequency distribution of Rn concentrations.

The physical reason for such phenomenon may be the presence of geographically "small" features which generate high Rn concentrations, well within an otherwise low-radon area. Such features can be tectonic faults, local uranium mineralization, or highly permeable rock formations. Being small in extension, these features contribute little to the mean, but may still pose a radiation problem for that small area. Speaking statistically, the problem emerges because the RPA definition relies on one statistic of the Rn distribution only (e.g. the mean), while occurrence of extremes is measured by other statistics such as high quantiles or dispersion measures. One may therefore think on integrating such additional measures into the criterion which defines RPA or non-RPA, i.e. honouring "small" phenomena although they contribute little to the overall picture. The subject is further discussed in section 4.3.3.

A task, largely untackled so far, to our knowledge, is validation of estimates through additional data, statistical procedures or application of alternative models. Summing up, the statistical aspects of RPA estimation is a relatively new field of environmental science that entails a number of research tasks which can be expected to keep Rn research busy for a while.

Interpretation of classification error probability clearly depends on what shall be optimized. From a radiation protection point of view, second kind errors (i.e. underestimating risk) have more weight than first kind errors (exaggerating risk). On the other hand, Rn protection costs money, so that from an economical point of view one strives to avoid unnecessary action, i.e. minimize first kind error probability. In the end, the problem is one of stakeholder interests.

b) A mental constructed example

To exemplify classification uncertainty, a mental fictive case study was created (from [Bossew 2017, 2018c](#)). In the lovely pre-alpine municipality Gigrizpatschen am Wappelbach, all 1004 houses were measured. The results are shown in Figure 17 together with basic statistics.

As RPA definition, we chose: An area B is RPA if $\text{prob}(z > 200) > 0.1$ in B. The areas B are defined as 5×5 quadratic subareas of the municipality. Their true status, based on the exhaustive data, is shown in Figure 18.

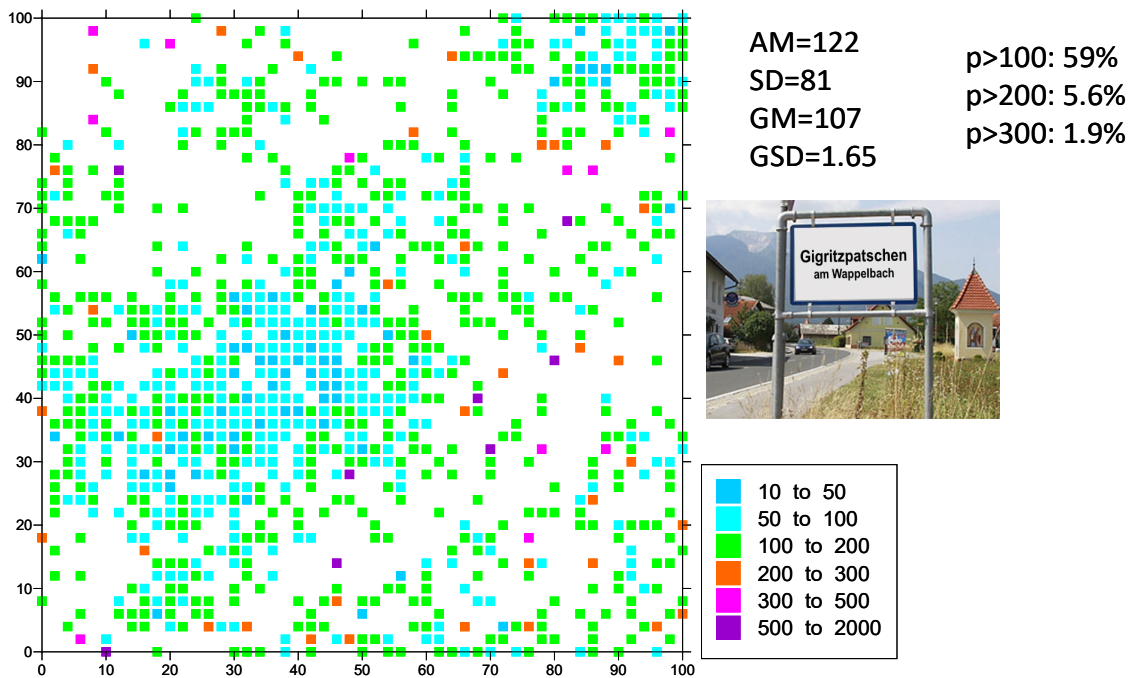


Figure 17: Exhaustive Rn measurements in virtual municipality.

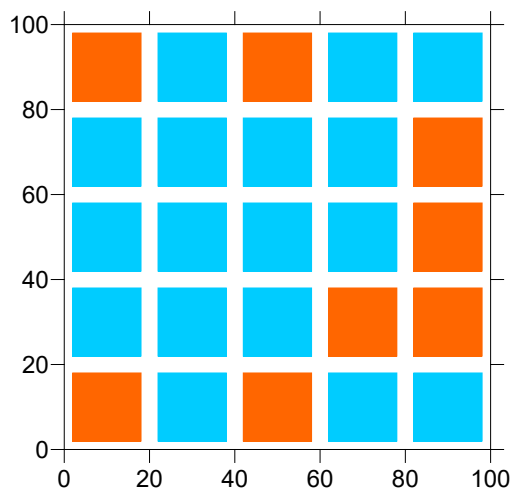


Figure 18: True RPA status of the 5×5 fractions of Gigrizpatschen. Orange: RPA, blue: non-RPA.

To assess the uncertainty in attributing RPA status was assessed by performing virtual sampling campaigns (2000 - 5000 realizations). The false positive (FP) and false negative (FN) rates were calculated for truly non-RPA and truly RPA areas, respectively. The rates are defined as fraction of wrongly classified areas.

The results are shown in dependence of sampling rate in Figure 19. It can be seen that low sampling rate can lead to a high chance of false negative, i.e. erroneously labelling non-RPA a true RPA. On the other hand, false positives, i.e. erroneously labelling a non-RPA as RPA, can also occur with high probability.

High error rate occurs if

- the sample size (sampling rate) is low;
- if the mean in an area is close to the cut-off (or reference) value;
- if variability is high (because then the probability is high that a limited sample catches only extreme values).

Note that variability is a natural, irreducible phenomenon, while the uncertainty of an estimate (e.g. the mean) depends on variability and sample size. The sampling rate refers to the entire population (1004 cases); therefore, the exercise is not entirely realistic, because for high sampling rate but well below 1, certain sub-areas have probably been sampled exhaustively several times.

As said in section 4.3.1.4d), the question of sampling statistics of finite population should be studied in more depth in the future.

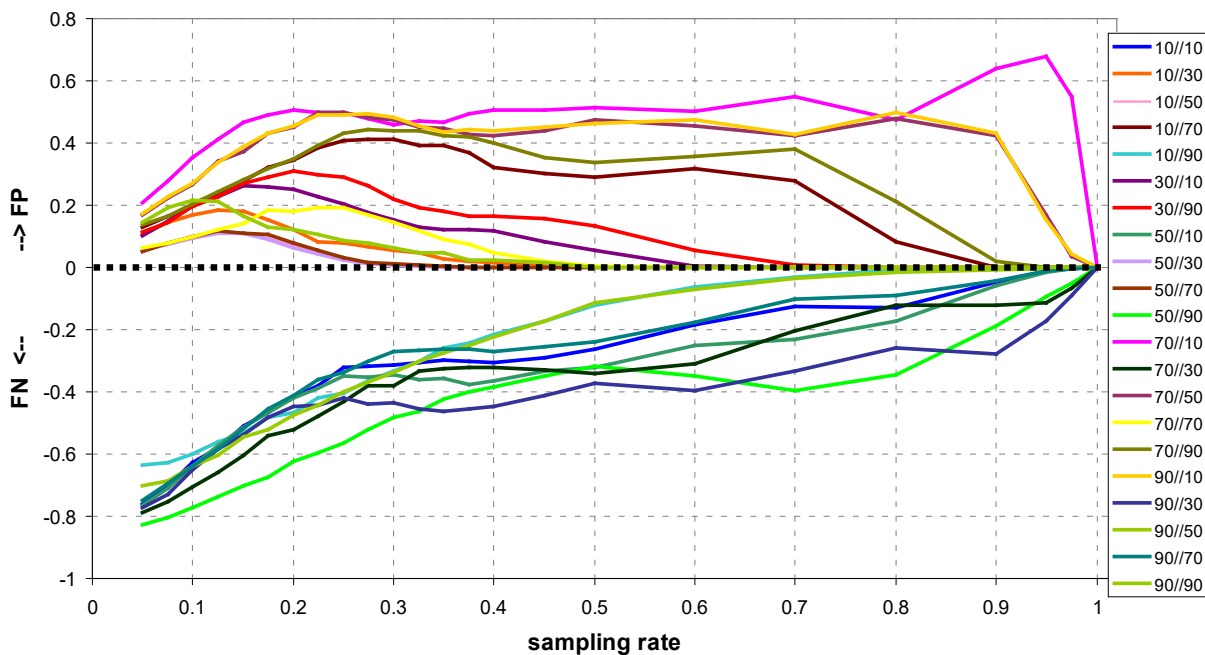


Figure 19: Mis-classification rates (FP, FN = false positives, negatives) in dependence of the sampling rate.

c) Example: RPA, Europe

The example has been taken from [Bossew \(2018b,c\)](#).

RPA criterion: RPA if $\text{prob}(C > 300 \text{ Bq/m}^3) > 10\%$ with $1-\beta$ confidence. Non-RPA, if $\text{prob}(C > 300 \text{ Bq/m}^3) < 10\%$ with $1-\alpha$ confidence.

Input data: European Atlas of Natural Radiation (EC, 2019). Exceedance probability in area B (10 km× 10 km grid cells) estimated as

$$p' := \text{prob}_B(Z > z) = t_{n-1}\left(\zeta \sqrt{\frac{n}{n+1}}\right); \quad \zeta := \frac{\ln z - AM_B(\ln Z)}{SD_B(\ln Z)}$$

Z = Rn concentration C, z = RL = 300 Bq/m³, n – number of data per grid cell B.

RPA status has estimated by cross-classification with uranium concentration (FOREGS / GEMAS, see EC, 2019). The relatively small deviation of the ROC graph from the diagonal indicates that association between U and C is not very strong, in general, Figure 20.

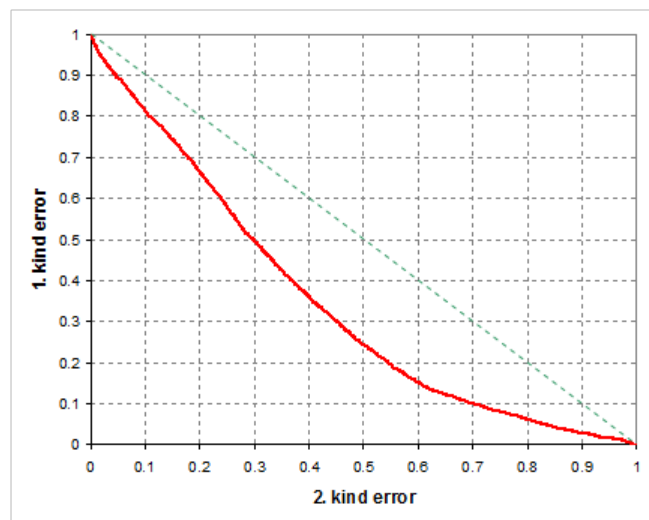


Figure 20: ROC graph of p' vs. U (see text).

The resulting RPA maps for $\alpha=\beta=10\%$ and $\alpha=\beta=20\%$ are shown in Figure 21. The limits of U concentration which define RPA / non-RPA status are given in the legend. 90% confidence intervals of the limits were determined by bootstrap.

The large grey areas which denote that RPA status according the given criterion is undecided, are a consequence of the weak association between C and U. Reversely, defining the U thresholds e.g. by optimizing in the ROC graph, e.g. by choosing the point of the red curve with highest distance from the diagonal (Y-statistic), or which has lowest distance from the lower-left corner (d01 statistic; Bossew 2014), would lead to a truly binary RPA map, but allowing very high first and second kind error probabilities. In the example shown here these are limited by α and β , consequently leaving a part of the domain (the grey area) unclassified.

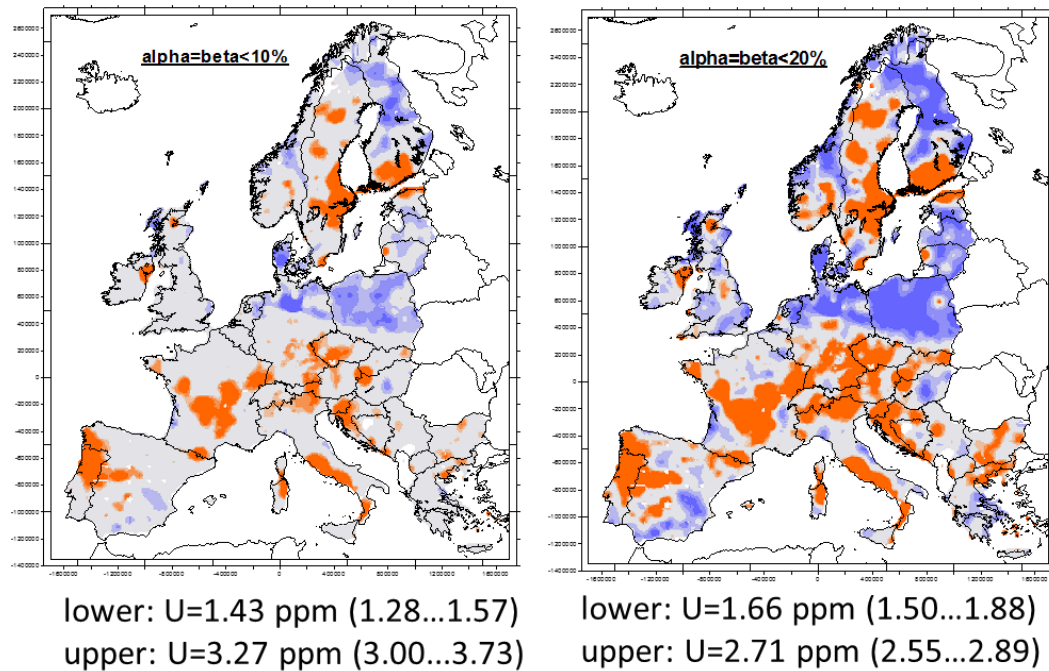


Figure 21: RPA maps of Europe derived from uranium concentration in topsoil. Red: $p > 10\%$ with confidence $1 - \beta$, blue: $p < 10\%$ with confidence $1 - \alpha$. Grey: undecided. Lower / upper in the legend: Definition of red (upper) and blue (lower) areas. In brackets 90% C.I.

d) Example: Germany

The same method applied to Germany, and performing 20,000 bootstraps, leads to an uncertainty map shown in Figure 22.

The uncertainty refers only to model uncertainty, but does not include uncertainty of input variables (IRC exceedance probability and GRP).

For the class limits of RPAs defined through GRP, one finds:

- The reddish hues indicating RPA. $Cl_{90} = (38.2, 52.8)$,
- The greenish hues, non-RPA with $Cl_{90} = (13.1, 26.4)$.

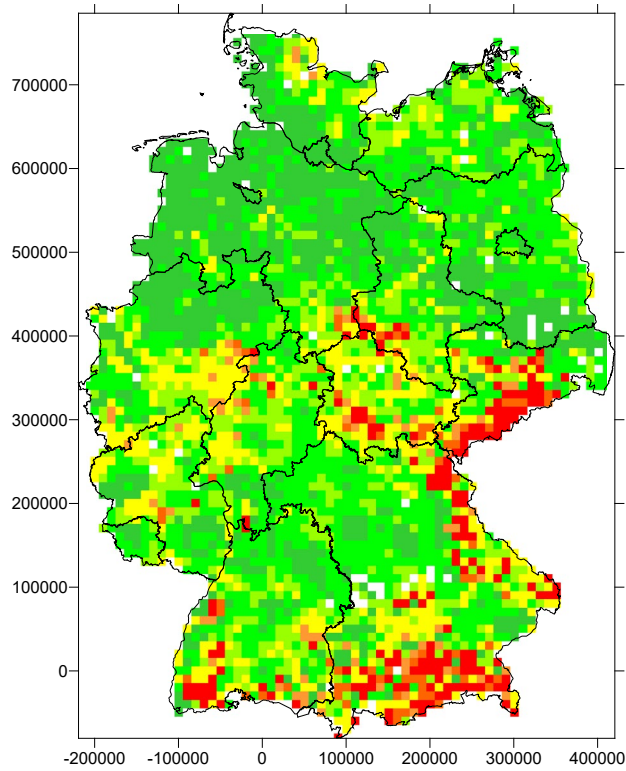


Figure 22: RPA map of Germany with colour tones indicating degree of uncertainty.

The histograms of fractions of the territory of Germany, covered by RPAs and non-RPAs, based on the bootstraps, are shown in Figure 23. While the histogram of RPA fractions appears reasonable, the one of non-RPAs comes rather unexpected.

The fractions which correspond to the central estimates are:

RPA: $GRP > 44.5$, 12.0% of territory

non-RPA: $GRP < 20.2$, 49.8% of territory.

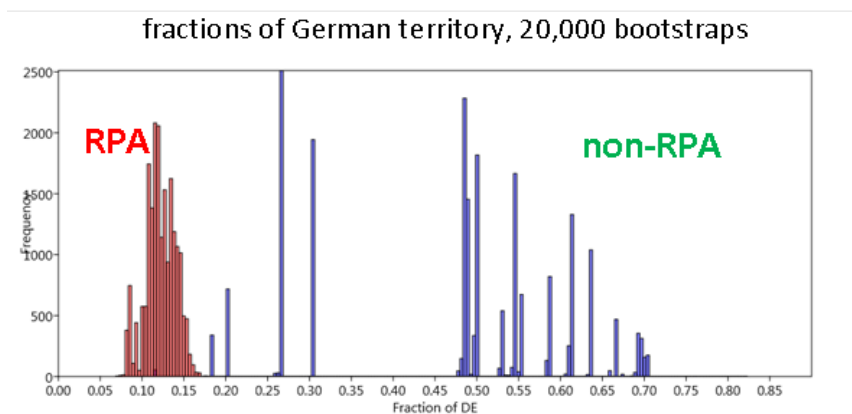


Figure 23: Histograms of the fractions of the territory of Germany covered by RPAs and non-RPAs.

2.6 RPAs as random objects – statistical considerations

As explained in section 4.3.1.5a, two aspects of classification uncertainty must be considered: 1) the uncertainty of RPA status of a given area; 2) given data, estimates the area in which the RPA criterion is fulfilled. Aspect 1) has been discussed above; usually, this is the more relevant one, because in practice, one wants to assign RPA status to given areas.

Therefore, aspect 2) shall be treated only shortly. Data or observations are “draws” from the unknown true fields, which shall be estimated on a given scale (grid size) in order to determine zones in which the RPA criterion is fulfilled (more generally: criterion of anomaly). Since the sample size is finite and hence an imperfect representation of reality, any quantity estimated from data is subject to uncertainty with respect to the true pattern of the field.

The uncertainty can be estimated by I) estimating all (in practice: many) areas that fulfil the RPA criterion, which are compatible with the data. This is done by conditional simulation which honours the values of the data and their structure, typically represented by the variogram (but also multi-point statistics are possible). II) From the ensemble of possible areas, one can estimate local uncertainty e.g. by computing the number of times a given grid cell belongs to the anomalous area. One can also calculate statistics of the topology of the area, e.g. its area size, circumference, topological generation (number of holes), etc.

For an example, see [Bossew \(2010b, 2013\)](#) (in the latter, about Fukushima contamination of Europe).

2.7 Proposal of methodology for assessing and quantifying RPA uncertainty

The only viable method of quantifying RPA uncertainty, if understood as the uncertainty of the RPA status of a given area (first bullet in 4.3.1.4c, aspect 1 in 4.3.1.6), seems to be the estimated misclassification rates of 1. and 2. kind.

If the RPA definition is based on the concept of exceedance probability,

Area B is RPA $\Leftrightarrow p = \text{prob}(Z > z_0) > p_0$ in B,

the uncertainty of the estimated p must be estimated. If the sample size (number of data in an area) is sufficiently high, p is distributed approximately

$p \sim N(p', \text{SD}(p))$, p' the point estimate, $\text{SD} \approx \sqrt{[p'(1-p')/n]}$. More accurate formulae for the quantiles / percentiles of the distribution of p are available.

Then,

1. kind error: $p' > p_0$ but in reality $p < p_0$;

estimated as: if $p' > p_0$: 1. kind error = $\text{prob}(p > p_0) \approx 1 - \Phi[(p' - p_0)/\text{SD}(p)]$;

2. kind error: $p' < p_0$ but in reality $p > p_0$;

estimated as: if $p' < p_0$: 2. kind error = $\text{prob}(p < p_0) \approx \Phi[(p' - p_0)/\text{SD}(p)]$

In case of estimation of RPA through a predictor (section 4.3.1.1c) seems to be most easily done via ROC analysis, as done in Germany, see sections 4.3.1.1c and 4.3.1.5c of this report and report "Activity of 4.1.1 / 4.1.2", chapter 8, p.28: "case study Germany".

This simple method is not viable for more than one predictor. In this case, one would perform dimensional reduction to one summary predictor and use ROC cross-classification again. It seems that so far this has not been tried. A candidate would be the radon hazard index as discussed in section 4.3.4.

To our knowledge, quantification of uncertainty of RPA estimates has not yet been attempted, except in Germany (Annex 1 of deliverable D5), where that uncertainty in terms of 1. and 2. kind errors is a by-product of the estimation method.

3 Application of retrospective radon measurements to RPA assessment

SUBG and UC evaluated the precision and applicability of the compact disc method (“CD method”) for retrospective Rn measurements for use for identification of RPAs. The uncertainties due to local and temporal variability has been validated and assessed by UC and SUBG at UC’s Saelices el Chico laboratory, by exposing a number of CD/DVD samples over a long period, under variable conditions representing a wide range of Rn concentrations, in parallel with conventional monitors.

Two new techniques, namely liquid scintillation counting of polymers and track-etching of CDs, for measurement of radon exhalation from soil based on radon absorption in plastic samplers have been developed and evaluated by SUBG and UC.

Since 2014 the CD/DVD method has been used to identify and study RPAs (Pressyanov et al., 2019a). Within MetroRADON project CDs were exposed at Saelices and Chico laboratory under highly variable ^{222}Rn activity concentrations and temperature. The results were published in International Journal of Environmental Research and Public Health (Pressyanov et al., 2019b) and the article is annexed to this report (Annex 2). During exposure the radon activity concentration and some major environmental parameters were followed continuously (every 10 min) by a reference radon monitor AlphaGUARD PQ2000 PRO (Saphymo/Bertin Instruments) traceable to another AphaGUARD unit calibrated in the Physikalisch-Technische Bundesanstalt (PTB). The experiment was organized as a blind comparison of radon measurements by CDs and continuous radon monitors. The results by the CDs were obtained using two calibration factors ($CF = \text{net track density}/^{222}\text{Rn exposure}$): one corresponding to the average temperature during exposure and another adjusted for the temperature profile during exposure. The results are shown in Table 1. As seen, the difference from the reference exposure in the first case is 3.7% for CDs etched at depth of 80 μm and 13.5% at depth of 120 μm . In the second case the difference is 0.4% at 80 μm and 9.8% at 120 μm . The results justify the conclusion that the CD method provides reliable estimate of the integrated ^{222}Rn activity concentration even under extremely variable ^{222}Rn concentrations and temperature. More details about the organization of the experiment and data processing may be found in the annexed paper (Pressyanov et al., 2019b, Annex 2).

Table 1. Integrated ^{222}Rn activity concentration assessed by CDs with tracks analyzed at 80 μm and 120 μm beneath the front surface. The reference exposure was assessed through continuous measurements by a reference instrument AlphaGUARD PQ2000 Pro.

Scenario	$^{222}\text{Rn exposure (kBq h m}^{-3}\text{)}$		Reference
	At 80 μm	At 120 μm	
With CF at 12.6 $^{\circ}\text{C}$	118000 \pm 12000	106000 \pm 12000	122500 \pm 6100
With CF adjusted for the temperature profile during the real exposure	122000 \pm 12000	110500 \pm 12000	

In previous studies the „traditional version” of the CD/DVD method (see e.g. Pressyanov et al., 2019a) performed well when applied for measurement of ^{222}Rn in soil gas (Mitev et al., 2018). For measurements in soil-gas usually the exposure time is shorter (e.g. one week or even less) than exposure times of months or years - typical for prospective or retrospective measurements of radon indoors. The minimum detectable ^{222}Rn activity concentration by the “traditional CD method” after one-week exposure is about 1500 Bq m^{-3} (Pressyanov, 2010). Such sensitivity may be enough for most of the measurements of radon in soil gas made at depth 60 cm or more. However, for the assessment of radon exhalation from ground by the gradient method for radon flux measurements (Maier and Schack-Kirchner, 2014) quantitative measurements at smaller depths,

close to the surface, are needed. Radon activity concentrations close to the ground surface (e.g. within the top 10-15 cm soil layer) might be much lower than the concentrations at depth of 60 cm or more. On the other hand, close to the surface the temperature variations during exposure might be large. This raised two challenges in front of this direction of potential applications of the CD/DVD method: to develop a version of the method with sufficiently increased sensitivity and to reduce the temperature influence on the results obtained. Within the MetroRADON project a dedicated research was done to address these challenges.

The sensitivity may be sufficiently increased by using DVDs of low background as large area track detectors, and the use of radon absorbing foils as radiators to amplify the signal. Constructively, DVDs consist of two halves stuck together, as shown in Figure 24. The front half is made of polycarbonate material that has radon absorption and track-etch properties. After mechanical splitting of DVDs, the internal surface of the polycarbonate half of the DVD is used as the detection surface. Additionally, it was covered by two foils of Makrofol N. Because of the unique radon absorption ability of Makrofol N (the radon concentration in it is 112 times higher than that in the ambient air, at room temperature (Mitev et al., 2016)) it serves as absorber/radiator that sufficiently amplifies the signal (net track density, after the DVD surface is etched, the tracks are counted and the background track density is subtracted). One version designed especially for measurements in soil is shown in Fig. 25. One benefit of these new detectors is that they may be prepared as coupled detector elements, shown in Fig. 26 with total detection area up to 200 cm² (Pressyanov, 2019). Moreover, using „stacks“ of many detector elements like this one on Fig. 26 the total detector area can be greatly increased and very low detection limits may be achieved (Pressyanov et al, 2018). In this way, the novel detector design show potential for wide scale of applications, not only for radon exhalation measurements.



Figure 24- The structure of a DVD: it consists of two halves stuck together. The front half is made of polycarbonate which can be used as alpha-track detector. The background of the internal polycarbonate surface is very low and can be additionally reduced by thermal annealing.

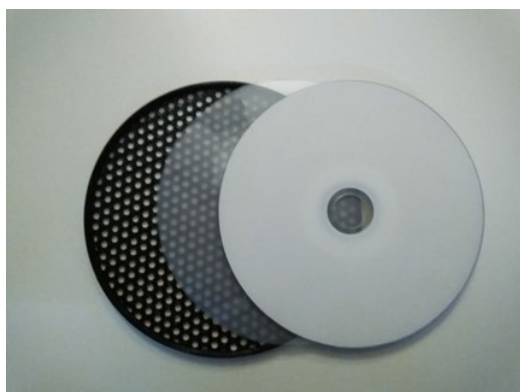


Figure 25- The design scheduled for measurements in soil-gas. The DVD surface is covered by two foils of 43 μm Makrofol N and plastic screen for mechanical protection. The total detection area is 100 cm².

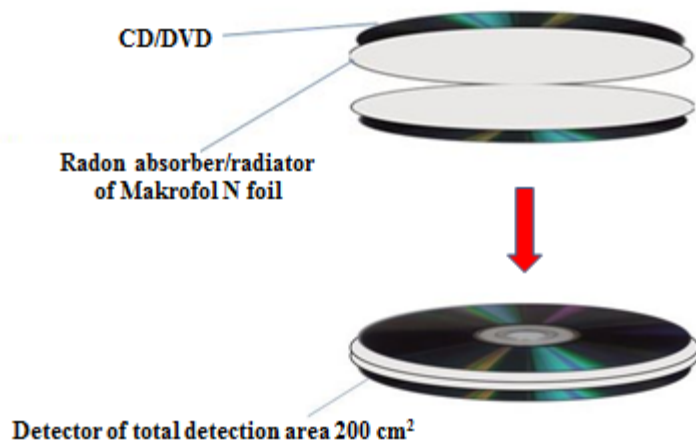


Figure 26- Scheme of a detector element of total detection area 200 cm².

With this new version of the CD/DVD method SUBG participated in 2017/2018 Public Health England (PHE) international radon comparison and the method performed very well. The results are shown in Fig. 27.

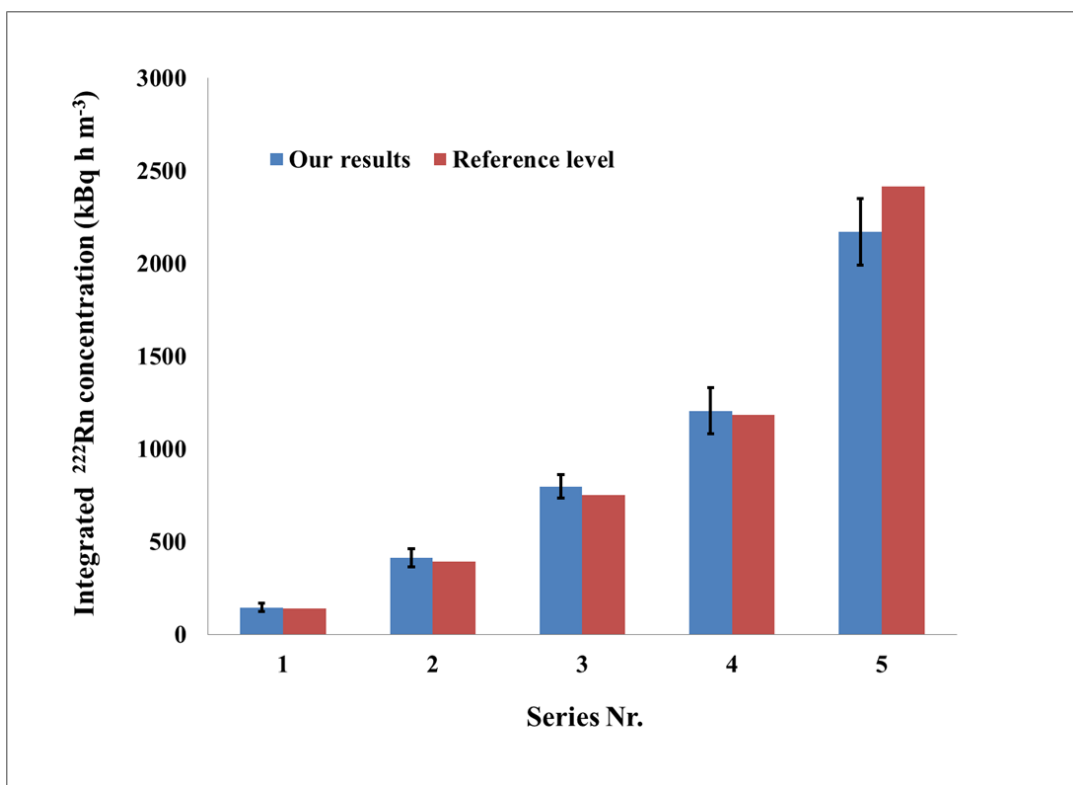


Figure 27- Quality Assurance of the method: performance at the Public Health England (PHE-UK) 2017/2018 radon inter-comparison.

However, the method demonstrates high temperature dependence. It was experimentally obtained that from 5 °C to 35 °C the *CF* decreases more than two times. As under real conditions variable temperatures may be expected, this problem has to be overcome before the method might be recommended for practical use. This problem was resolved within a research related to WP2 - while studying the properties of polymer foils used as anti-thoron barrier.

As explained in more detail in Deliverable 2 of MetroRADON project, when the detector volume is packed/covered by polymer foil to reduce the thoron interference, substantial temperature bias may be introduced. Surprisingly, for the described here version of the CD/DVD method, as well as many other widely used detectors it turned out that the temperature dependence of the detector's response is reciprocal to the temperature bias introduced by the polymer foil. This led to a novel technical concept (Pressyanov, 2019a) with a potential to overcome the temperature dependence problem of many types of detectors. It is possible to design a "compensation module" in which the detector is placed, that ensures reduction or elimination of the temperature dependence of the detector (see Figure 28). To achieve this, the foil material, its thickness, the surface area and the volume of the module should be properly selected, taking into account the temperature dependence of the detector's response, in order to achieve the best temperature compensation. In the same time such module provides an efficient protection against the thoron interference and humidity. This novel concept led to a patent application submitted within MetroRADON project.

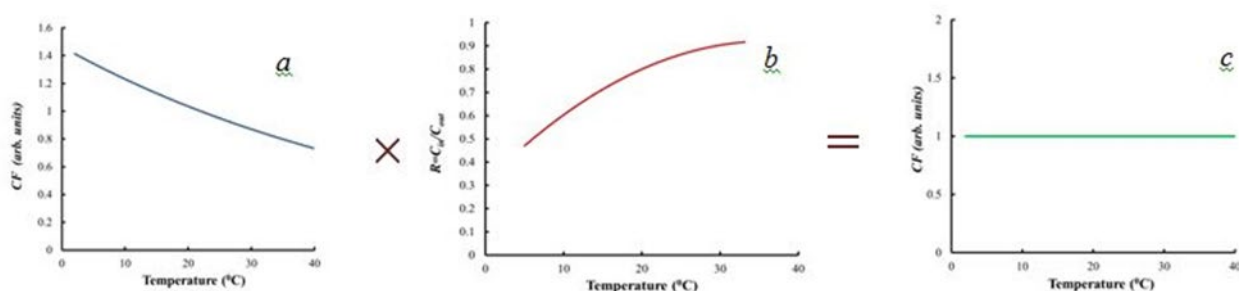


Figure 28 - The concept of the compensation module design (Pressyanov, 2019c): The temperature dependence of many radon detectors (a) and that introduced by polymer anti-thoron barriers (b) are reciprocal. This can be used to reduce/eliminate the temperature dependence (c), the thoron influence and also the humidity influence.

The minimum detectable activity concentration (MDAC) of ²²²Rn after one week of exposure and with a compensation module used is shown in Fig. 29. As seen, in this case even activity concentrations below 100 Bq m⁻³ can be quantitatively measured.

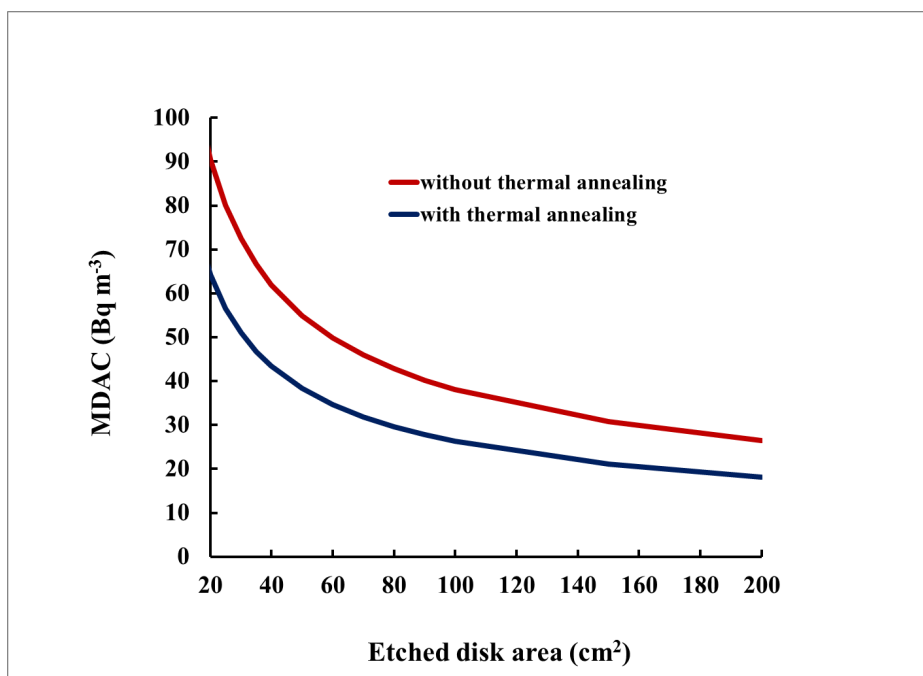


Figure 29- Minimum detectable average ^{222}Rn activity concentrations after one week of exposure. The detectors are packed in a hermetic envelope of low density polyethylene, designed as „compensation module“ that greatly reduces the influence of the temperature on the detector response (Pressyanov, 2019b).

The method based on liquid scintillation counting of polymers is proposed by (Mitev et al., 2019). The method uses the high radon absorption ability of Makrofol N foils to perform unperturbed measurement of the radon profile in the soil, which allows simultaneous determination of the radon concentration in the soil-gas as well as the radon exhalation rate from the soil surface. The soil-gas radon profile is estimated by liquid scintillation counting of Makrofol N foils exposed at different depths in the soil. The method is cumulative in its nature and provides average quantities for the measurement period, typically 1-3 days. It was applied successfully in the Intercomparison of indoor radon and geogenic radon measurements under field conditions, performed in LNR, Saelices el Chico, Spain, which was organized in the framework of MetroRadon. The results obtained in the framework of the intercomparison are described by (Rabago et al., 2019) and show that there is a strong effect of the weather conditions on the radon exhalation rate, especially in highly variable weather conditions. More details and results from the radon in soil-gas and radon exhalation measurements obtained in the intercomparison at LNR Saelices el Chico, Spain are presented in Annex 3.

4 RPA classification based on occurrence of extremes

A new complementary approach that focusses on identification of areas that could have a significant proportion of dwellings with very high indoor radon concentrations of several thousands of Bq/m³ has been developed and tested by IRSN and UC.

Test and development of a new approach to identify areas concerned by a high percentage of very high indoor radon levels (several thousands of Bq/m³): several methods have already been developed to map RPAs. Generally, this concerns areas with a significant proportion of indoor radon concentrations exceeding a reference level of a few hundreds of Bq/m³ (maximum 300 Bq/m³ as given by the EU-BSS).

The aim of this action is to test and develop a complementary approach that could specify existing radon maps further by focusing on the identification of areas that could be concerned by a significant proportion of dwellings with very high indoor radon concentrations of several thousands of Bq/m³. This method has been tested in France and Spain, where such cases occur regionally. It will be based on the analysis of available quantities such as the geogenic radon potential, measurements of indoor radon concentration, dwellings characteristics, and recent results of quantitative radon risk assessment etc., complemented by statistical modelling. Such a method would allow targeting specific prevention and remediation actions in heavily affected regions to significantly reduce the exposure in buildings.

Data available:

- Geogenic radon potential maps;
- data from national and/or local radon surveys in dwellings/buildings (radon measurements, characteristics of buildings) ;
- Geochemical data bases ;
- National and local geological map (1 : 1 000 000 and 1 : 50 000);
- mining prospects and gamma measurements.

Methodology:

1) Based on radon potential maps and all the data available in dwellings (and if possible in public buildings), **calculate the probability of exceeding 300 Bq/m³ in the different categories of radon potential**. The same exercise can then be performed with the level of 1000 Bq/m³.

2) Within the highest categories of potential, **smaller sub-sectors could be delineated** (for example, 10 km x10 km cells). From the radon data available indoor, statistics could be produced on each of these sub-sectors: mean, median, probability of exceeding thresholds (300 Bq/m³, 1000 Bq/m³...). The purpose is to **identify the sub-sectors with the highest concentrations (= RPA)**.

3) For each RPA identified, a **detailed study of the local geology** could be carried out based on geological maps, geochemical data bases, mining prospects and gamma measurements. If available, some building characteristics will also be considered (year of building, interface between soil and first level of the building).

4) Isolate all values **above 1000 Bq/m³**: for each value above the threshold but located outside one of the sub-sectors identified in Part 2, a check to see if it is associated with a particular geology and/or building characteristics will be made.

5) Based on those different observations, a **list of geological features and if possible building properties characterizing a RPA** can be realized. The presence of these geological features and/or buildings characteristics on a territory, even if few radon measurements are available indoor, could allow identifying the areas presenting a risk of high radon exposure. Such knowledge could guide specific prevention and remediation actions.

Terminology:

Qualitatively speaking, an anomaly within a metric dataset is an instance which is significantly different from its neighbours. (This is a dataset in which the data have locations, such as positions in time series or geographical locations.) A metric is required to define what a neighbour or a neighbourhood is; in spatial settings, this is usually (but not necessarily) the Euclidean distance. The concept of anomaly cannot be separated from the one of background (BG), represented by the “normal” neighbourhood of the anomaly. If one succeeds to model the BG, an anomaly may be defined as a statistically significant residual from the model.

Outliers are values which seem not to belong to a population. Evidently, an assumption about the population is required to decide this. Reasons of outliers can be: observation error or uncertainty; an accidentally isolated extreme of the BG population; an instance which belongs to a different population. Importantly, a multivariate outlier is not necessarily an outlier of any individual univariate (marginal) distributions involved.

An *extreme* is simply the highest or lowest value of a set. It does not say anything about its nature.

The term *hot spot* seems to be mostly used for points, or cluster of points, or small regions, where the variable takes anomalously high values. Reasons can be a region in which the background process takes high levels or a region which is the domain of a separate process.

Anomaly and hot spot: mostly seem to denote “true” effects, i.e. not related to observation; “outlier” seems to be neutral in this respect, i.e. can also denote observation effects. A summary on outlier detection and problems involved can be found in [Ben-Gal \(2005\)](#).

4.1 Results: case-studies in France

a) Indoor radon concentrations

For this study, 31,915 measurements spread over the French territory are used. These data were collected in different contexts:

- 12,940 measurements were collected for the national database (IRSN-French Health Ministry) in dwellings over the period 1982-2002 ([Demoury et al. 2013](#));
- Obligations for measuring radon activity levels exist for some public buildings in the French regions most affected by radon risk. 8,253 measurements acquired between 2014 and 2018 were also used (French Health Ministry database);
- 10,722 measurements acquired for local measurement campaigns in dwellings were also collected for this study (local Health Authorities databases).

Table 2 lists the statistics calculated from all these data and the statistics from the national campaign in dwellings. The global dataset overestimates indoor radon concentrations because regulatory measurements in public buildings and local measurements campaigns are rather carried out in high radon potential areas.

All these data cannot be geolocated precisely (coordinates). For half of the data, the indication of the measurement municipality was only available.

Table 2: Statistics calculated from indoor radon concentration data available (global dataset and national campaign in dwellings). AM: arithmetic mean; SD: standard deviation.

	Number of data	Radon concentration (Bq.m ⁻³)						
		Min	Max	AM	SD	1 st quartile	Median	3 rd quartile
Total dataset	31,915	1	28,553	238	586	44	97	229
National campaign in dwellings	12,940	1	4,382	89	159	28	49	93

The measurement results are highly heterogeneous on the French territory. In addition, the sampling density varies widely by region. Figure shows, for each “*département*” (French district), the arithmetic means of indoor radon volume activities and the number of measurements recovered. Areas with the highest arithmetic means are generally those in which the most measurements have been made.

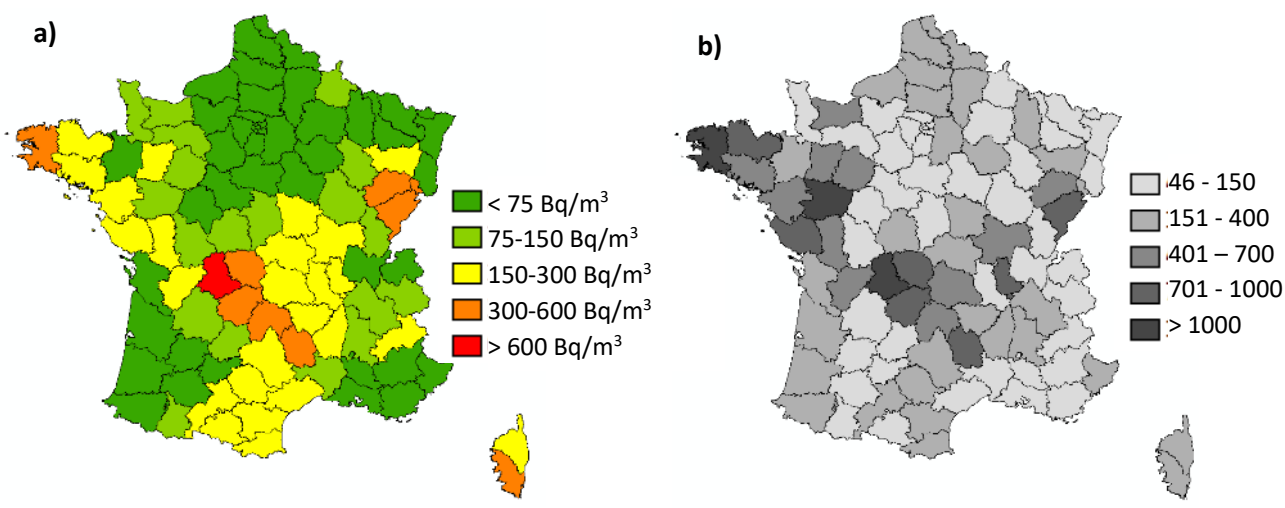


Figure 30: By “*département*”, arithmetic mean of indoor radon activity (a) and number of data (b)

A map of the radon potential of the geological formations has been established by the IRSN (Ielsch et al. 2010, 2017) in order to characterize the capacity of the underlying rocks to generate radon at the surface on the French territory. This mapping is only based on the characteristics of the geological formations (indoor radon measurement results are not taken into account). The two main parameters taken into account are the uranium contents of the underlying rocks and the presence of factors that can facilitate the transport of radon towards the surface (faults, boreholes, mining works ...). This map is based on data from the geological map of France at the scale of 1: 1 000 000.

Based on French geogenic radon potential map, municipalities are classified in three categories Figure 31:

- Category 1: municipalities located entirely on geological formations with low uranium contents and with no factors that may facilitate the transfer of radon to the surface
- Category 2: municipalities also located on geological formations with low uranium contents, but a part of their surface is concerned by geological factors that can facilitate the transfer of radon to the surface

- Category 3: municipalities which present geological formations with higher uranium contents compared to the other formations, on at least a part of their surface. For this category, the presence of radon at high concentrations in buildings is most likely.

This classification of municipalities into three categories is currently used in the French regulation with an obligation of measurement in certain public buildings in the municipalities of category 3.

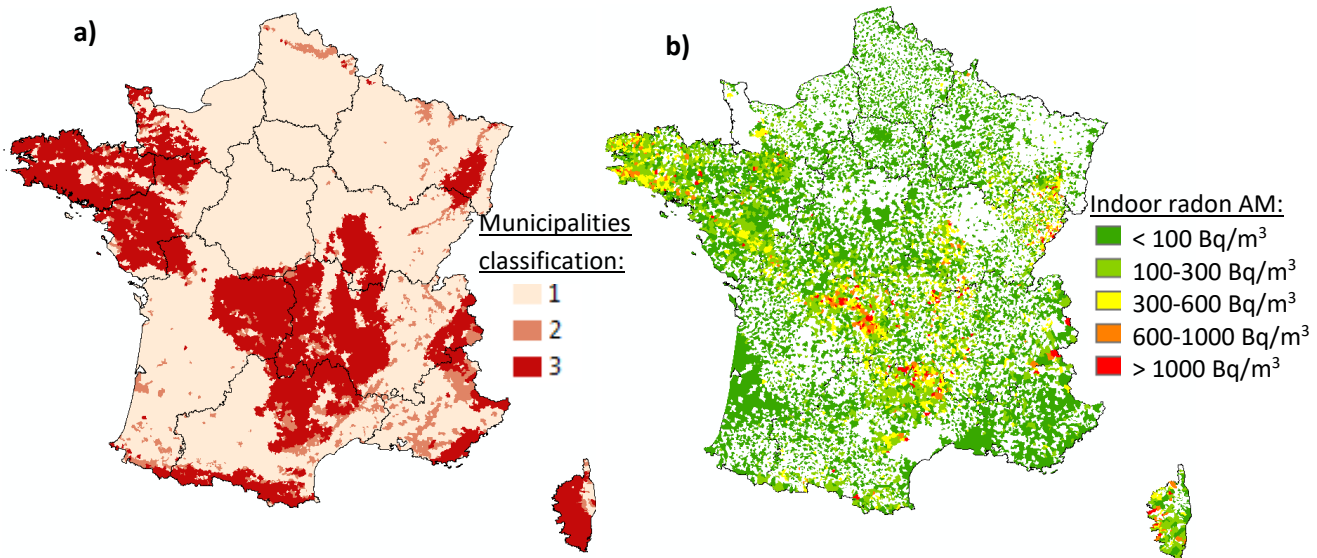


Figure 12: Comparison between the French municipalities' radon classification (a) according to the geogenic radon potential and the arithmetic mean (AM) by municipalities based on 31,915 indoor measurements (b)

The data analysis Table 3 confirms that the municipalities in category 3 are overrepresented in the sampling of this study: whereas they represent only 28.8% of the territory, 60% of the measurements were realized in these areas.

A good correlation is observed between the three categories of municipalities and the results of indoor radon measurement: the AM, the proportion of exceeding the level of 300 Bq/m³ and the proportion of exceeding the level of 1000 Bq/m³ increase between categories 1 and 2 and between categories 2 and 3.

Table 3: Statistics of indoor radon concentration data compare to the municipalities classification. AM: arithmetic mean; SD: standard deviation.

Category	Surface		Data Number		AM	SD	> 300 Bq/m ³ (%)	> 1000 Bq/m ³ (%)
	(km ²)	(%)	-	(%)				
Category 1	349,037	64.7	11,055	35	108	231	6.9	1.0
Category 2	35,187	6.5	1,743	5	144	232	12.1	1.5
Category 3	155,402	28.8	19,117	60	321	721	26.8	5.6
TOTAL	539,627	100	31,915	100	238	586	19.1	3.8

b) Hot spots identification

The term hot spot refers to small regions where the indoor radon concentrations show anomalously high values.

In order to identify these hot spots, the French territory has been subdivided into regular sub-sections using a grid. Different cell sizes were tested (kilometre units): 50x50, 25x25, 20x20 and 10x10 (Table 4). A cell size of 20 km x 20 km was finally selected because it represents the best compromise between the precision of the grid and a sufficient number of cells with a number of data greater than 10, in order to allow a statistical analysis. Figure 18 shows the map of the indoor radon AM per cell, as well as the map of the associated measurement frequency.

Table 4: Comparison of the different sizes of cells tested:

Cell size (km ²)	Total number of cells	Number of cells with more than 10 data
50 x 50	266	249
25 x 25	953	750
20 x 20	1,455	761
10 x 10	5,004	591

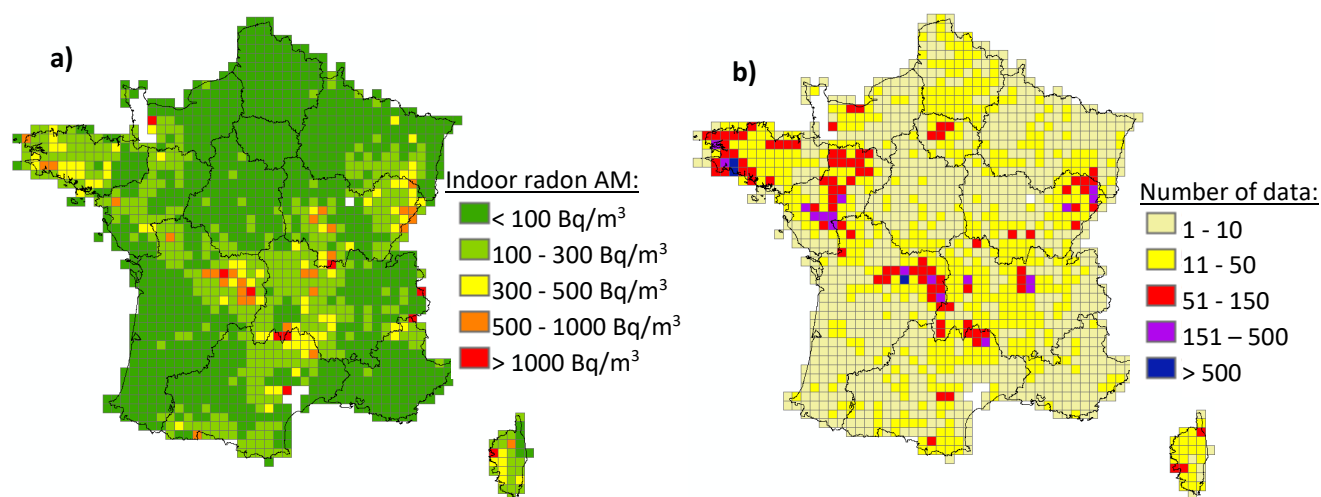


Figure 32: For each cell of a 20x20 grid, AM of indoor radon concentration (a) and number of data (b)

From this grid, only cells with more than 10 measurements were used. The percentages of exceeding value of 300 Bq/m³ and 1000 Bq/m³ were calculated for each cell. Figure 33 and Figure 34 show the results obtained.

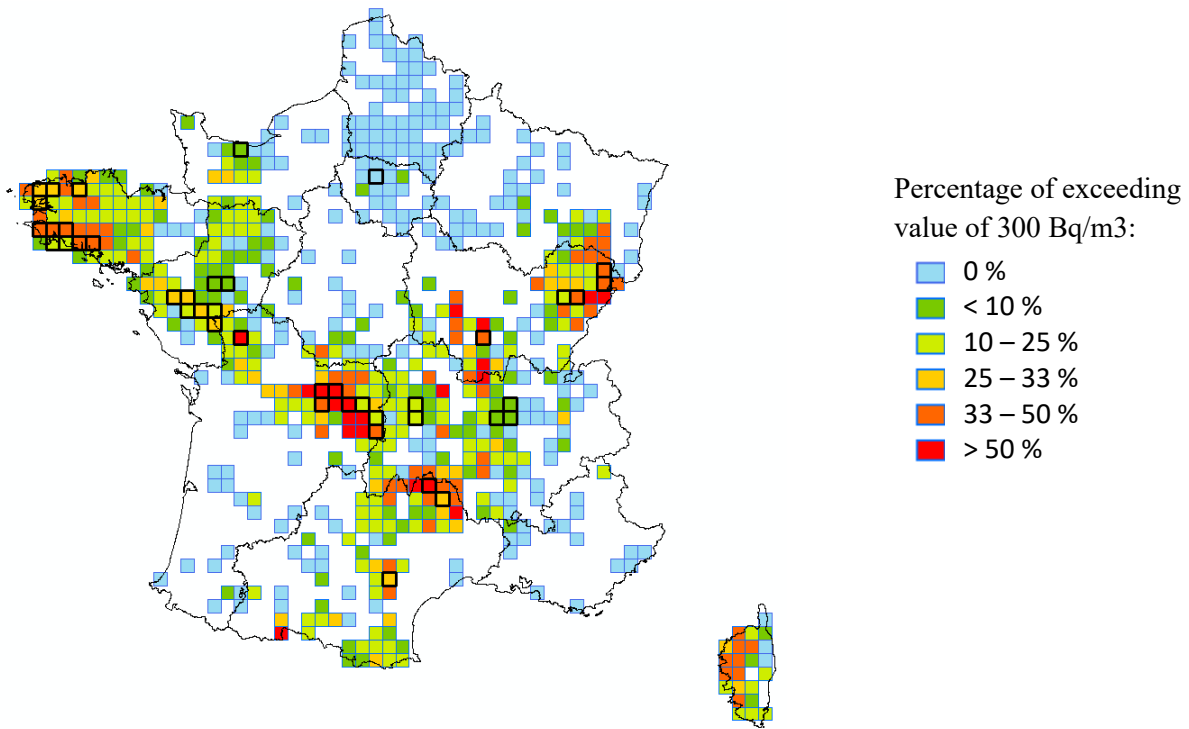


Figure 33: Percentage exceeding the value of 300 Bq/m³ with a 20x20 grid (cells surrounded in bold: cells with more than 100 data)

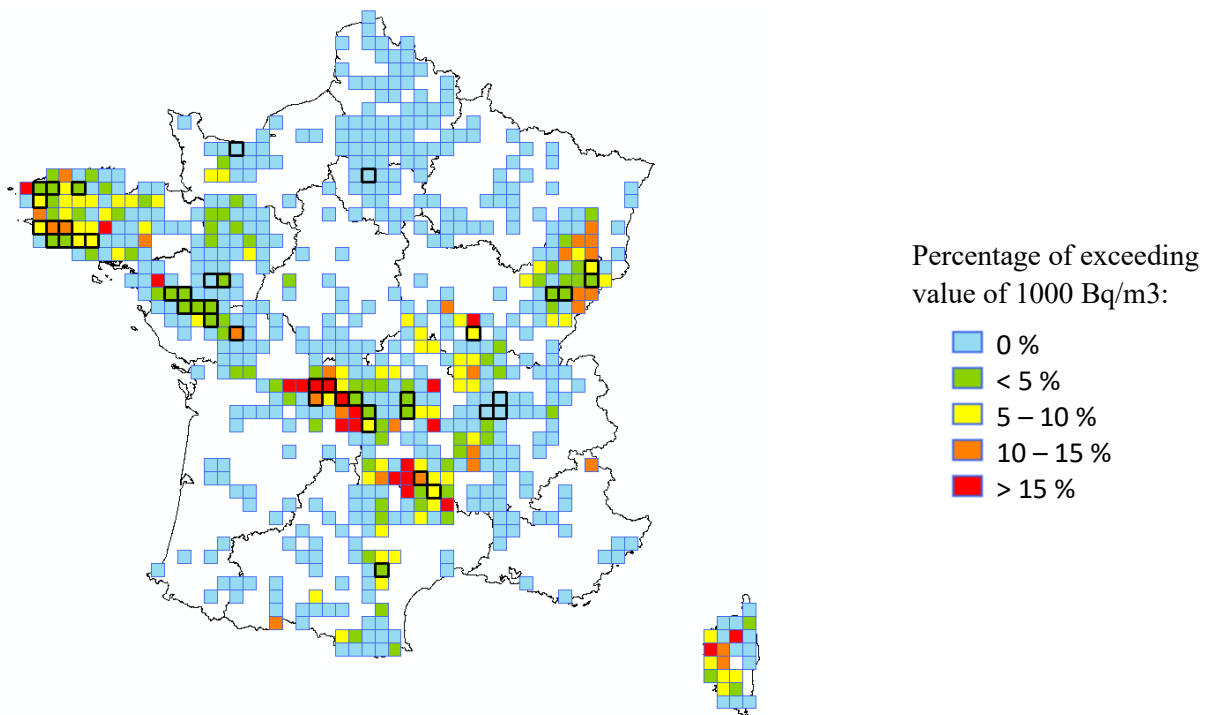


Figure 34: Percentage exceeding the value of 1000 Bq/m³ with a 20x20 grid (cells surrounded in bold: cells with more than 100 data)

From these results, “hot spots” are defined as the cells having at least one of the following characteristics:

- Percentage exceeding 300 Bq/m³ above 50%;
- Percentage exceeding 1000 Bq/m³ above 15%;
- Percentage exceeding 300 Bq/m³ above 33% with more than 100 data in the cell.

According to these criteria, 42 cells can be considered as “hot spots”. Figure 35 locates these 42 cells on the 1: 1 000 000 geological map of France. The distribution of identified hot spots is therefore as follows: 9 hot spots in the Armorican Massif, 25 in the Massif Central, 1 in the Pyrenees, at the border with Spain, 2 in Corsica, and 5 in the Jura Mountains, at the border with Switzerland.

Table A in Annex 4 lists the main geological features associated with each of the 42 identified cells. With the exception of the hot spot identified in the Pyrenees and of the five cells located in the Jura Mountains, all hot spots are associated with some peculiar granite with a clear dominance of monzogranites and peraluminous leucogranites. Thus, according to the available indoor data, this lithology appears to be the most penalizing for the radon issue in France.

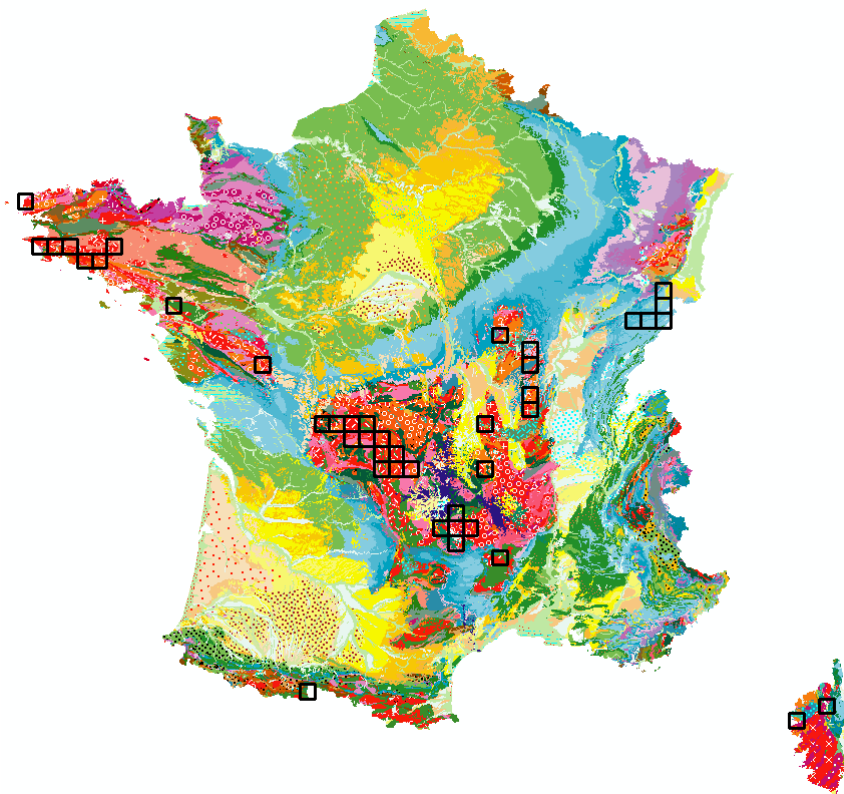


Figure 35: Geological map of France at 1: 1 000 000 and location of identified hot spots

c) Outliers identification

Outliers are values which seem not to belong to a population. For radon, outliers are the high values observed outside areas considered as high radon potential. We focus therefore on the values measured in the municipalities classified in category 1 and 2. Figure 36 shows the map of the AM per cell of 20x20 considering only the indoor radon concentrations measured in municipalities in category 1 and 2.

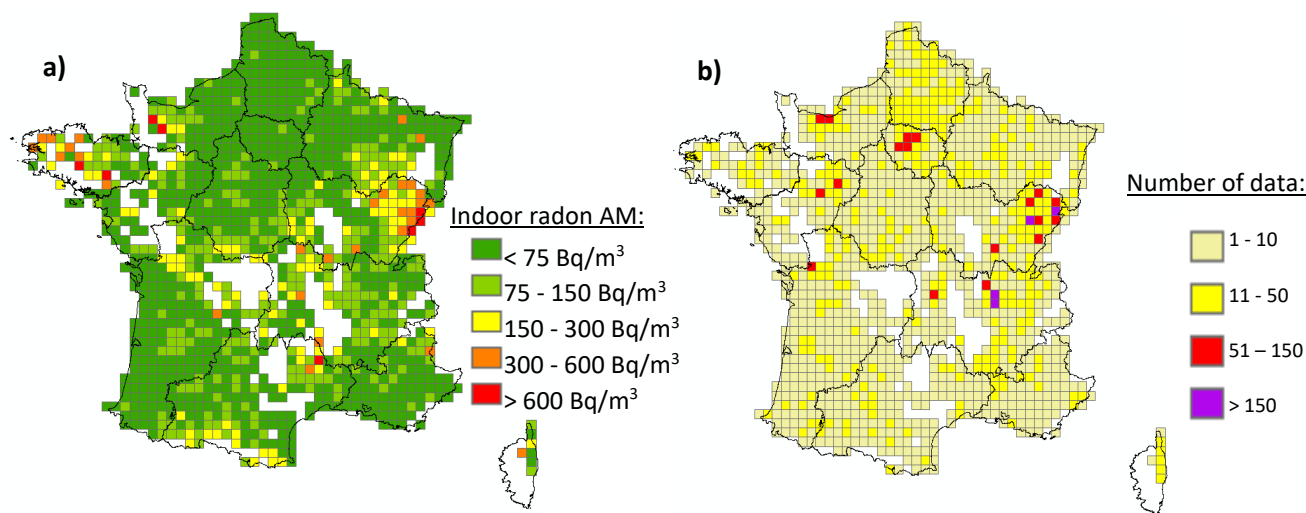


Figure 3613: From a 20x20 grid and considering only the data collected in category 1 and 2 municipalities, AM of the indoor radon concentrations (a) and number of data (b)

From this grid, only cells with more than 10 measurements were used. The percentage of exceeding value of 300 Bq/m³ was calculated for each cell. Figure 37 shows the result obtained. An area located in the East of France stands out clearly. This area corresponds to the Jura karstic area, at the border with Switzerland. Several cells in this sector were also defined as “hot spots” in the previous paragraph. Karstic systems are very complex and their impact on radon potential is not very well known. Indeed, the uranium content of karstic rocks (limestone) is very low but karsts are very permeable geological environments that can facilitate the radon accumulation and/or then the radon transport to the surface in their underground caves, fractures and other typical structures. Last years, IRSN performed a study to enhance knowledge on the influence of karstic structures on the radon production and migration at a regional scale, in a karstic area located in the French Jura Mountains (Gréau et al. 2017, Mansouri et al. 2018). This study confirmed that karstic environments could be the source of locally high radon contents in soils. The data analysis and the modelling show that the average levels of radon activity in soils are essentially the result of radium-226 emanation from the soil. Indeed, on the study area, a relative enrichment of radium-226 was observed in soils due to the important dissolution of limestone in the past (karst formation) and the soil radium-226 contents was quite similar to those observed in some granitic regions. However, the study is still ongoing in other karstic regions in France before to be able to transpose these conclusions to the French geogenic radon potential map.

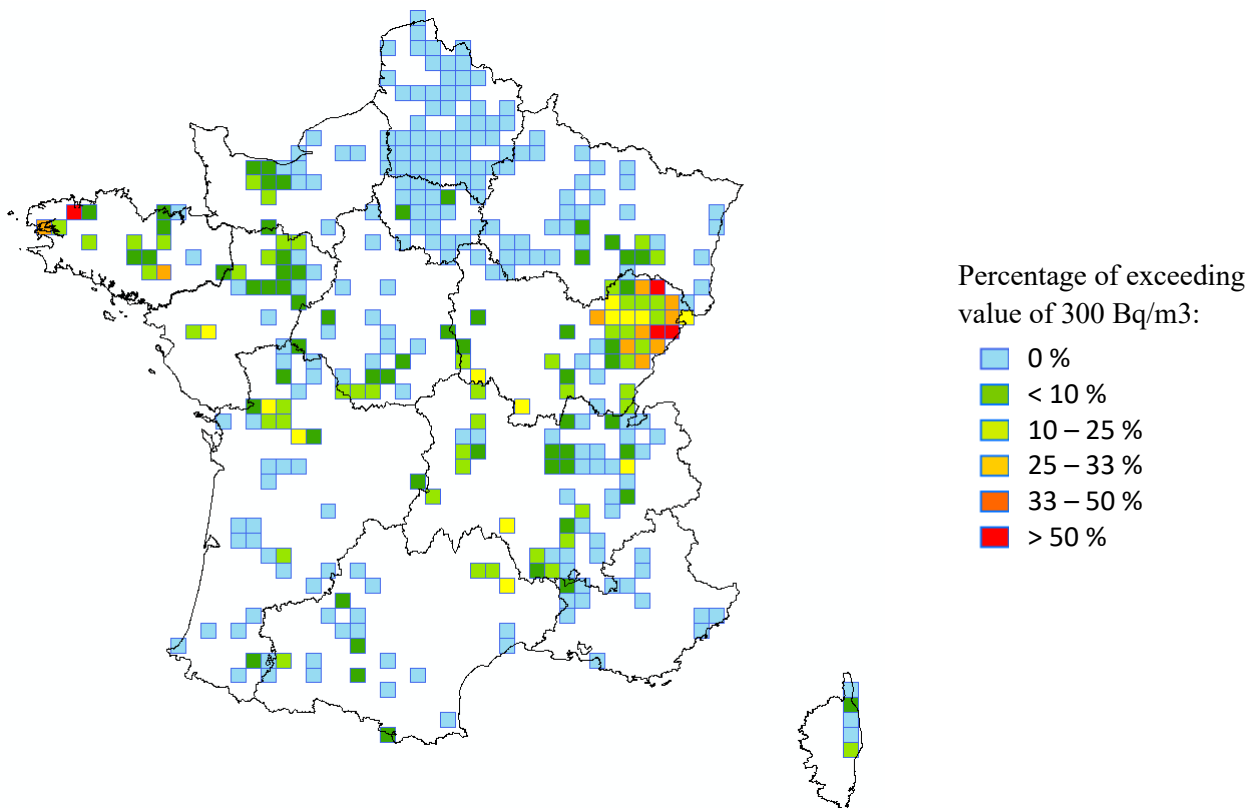


Figure 3714: Percentage of exceeding the value of 300 Bq/m³ with a 20x20 grid calculated with data in category 1 and 2 municipalities and cells with more than 10 data

In order to more specifically identify outliers, all the exceedances of the value of 1000 Bq/m³ measured in categories 1 and 2 municipalities were selected. This corresponds to 132 measurement results (out of 31,915 initial measurements, i.e. 0.4%). Figure 38 shows their location on the map of France. Most outliers thus identified are located in the vicinity of category 3 municipalities (high geogenic radon potential), in the Massif Central and in the Armorican Massif. Numerous outliers are also identified in the Jura karstic area, already mentioned above. Finally, some outliers appear isolated in areas with low geogenic radon potential.

Table B in Annex 4 lists the geological characteristics associated with outliers grouped by French “*départements*”.

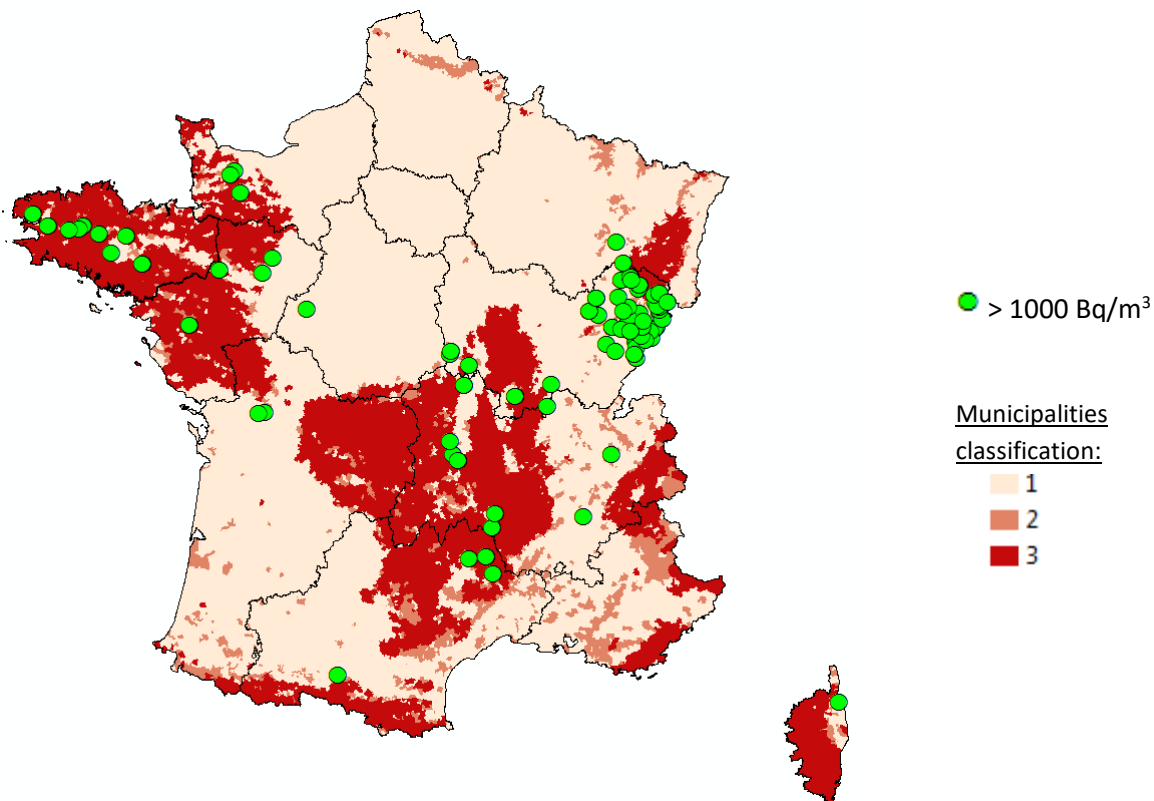


Figure 38: Location of indoor radon measurements above 1000 Bq/m³ in category 1 and 2 municipalities (outliers)

4.2 Results: case studies in Spain

Radon concentration data in dwellings

a) Data

In the Nuclear Safety Council (CSN) database, 9,211 Spanish data are available.

The national radon database in dwellings by the Nuclear Safety Council (CSN) has 12,000 measurements made in different sampling campaigns. The University of Cantabria (UC) contributed until 2014 with 9,211 measurements (statistics in table 5).

One fact to keep in mind is that sampling in Spain was not random. The CSN defined a sampling criteria that intensified the number of measurements in an area previously selected (Sainz et al, 2014; Sainz et al, 2017). This sampling was adjusted to a 10 km x 10 km cells system that will be explained in the following parts.

b) Statistics of radon concentration data in dwellings

Table 5. Statistics of radon concentration data. AM: arithmetic mean; SD: standard deviation.

No. of data	Radon concentration (Bq/m ³)									Skewm	Kurto
	Min	Max	AM	SD	GM	GSD	1st Quart	Median	3rd Quart		
9,211	10	15,4	95	270.57	55	2.61	28	54	103	33.1	1,539

Central tendency measured such as mean, median and mode show that the data do not follow a normal distribution because the arithmetic mean is 95 Bq/m³, the geometric mean 55 Bq/m³ and the median reduces to 54 Bq/m³. The national arithmetic mean is close to the level recommended by the WHO of 100 Bq/m³ (WHO, 2009) to initiate action plans.

In addition, the standard deviation of the arithmetic mean is 270.57 Bq/m³, shows a high data dispersion.

When analyzing the sample distribution shape, a high kurtosis coefficient is observed (K = 1.539) indicating a leptokurtic distribution, while the asymmetry coefficient (CS = 33.1) indicates a positive asymmetry: The measurements distribution has a log-normal distribution.

The use of the arithmetic mean is suggested by the Joint Research Center-European Commission in the European Indoor Radon Map (Tollefsen et al., 2014; JRC-EC, 2019; Dubois et al., 2010; Bossew et al., 2015) because it is the most appropriate in the representation of this variable due to the great variability of measurements. In addition, the use of the arithmetic mean is necessary when there is not a sufficient density of data, and because with this parameter all the radon concentration data obtained are taken into account.

c) Total data number in each category 90th percentile (P90)

The polygons of 90th percentile (P90) are represented in 5 categories according to the radon levels, and the percentage of points made in each of them is analysed.

As it is shown in Table 6, 23% of the measurements were made in areas identified with radon concentrations above 400 Bq/m³ (category 1) and 3% in areas between 301-400 Bq/m³ (category 2). In the areas defined as category 3 (201-300 Bq/m³), 23% of the data was taken. The 50% of the measurements correspond with areas between 101 and 200 Bq/m³ (category 4), while only 1% of the data are included in areas with less than 100 Bq/m³ (category 5).

Table 6. Total data number and surface in each category 90th percentile (P90)

Category	Surface (km ²)	Surface (%)	Data Number	Data Number (%)
Category 1 (>400 Bq/m ³)	71,481	14	2,088	23
Category 2 (301-400 Bq/m ³)	10,411	2	274	3
Category 3 (201-300 Bq/m ³)	118,265	23	2,113	23
Category 4 (101-200 Bq/m ³)	299,778	59	4,607	50
Category 5 (<100 Bq/m ³)	5,774	1	129	1
TOTAL	505,709	100	9,211	100

c1) Probability to exceeding 300 Bq/m³ in each P90 categories

The radon concentration data with more than 300 Bq/m³ are selected. In total there are 397 measurements that represent 4.3% of the measurements made in Spain.

The data analysis (table 7) in each of the five P90 categories reflects that: within category 1, most of the analysed data (59%) exceeds 300 Bq/m³, representing a presence probability of 11.3 %. In the case of category 2 (between 301 and 400 Bq/m³) the presence probability is 8.8%. This probability is reduced in the following categories, since in category 3 its probability is 3.2%, in category 4 it is 1.5% and in category 5 it is 1.6%.

Table 7. Probability to exceeding 300 Bq/m³ in each P90 categories

Category	Data number with > 300 Bq/m ³	%	PROBABILITY (%) Data number in each category with > 300 Bq/m ³ between total data number in each P90
Category 1 (>400 Bq/m ³)	236	59	11.3
Category 2 (301-400 Bq/m ³)	24	6	8.8
Category 3 (201-300 Bq/m ³)	68	17	3.2
Category 4 (101-200 Bq/m ³)	67	17	1.5
Category 5 (<100 Bq/m ³)	2	1	1.6
TOTAL	397	100	4.3

c2) Probability to exceeding 1000 Bq/m³ in each P90 categories

In the same way, the radon concentration data of more than 1000 Bq/m³ are selected. In total there are 34 measurements that represent 0.4% of the measurements made in Spain.

The data analysis (Table 8) shows that within category 1 the majority of these (68%) exceeds 1000 Bq/m³, representing a presence probability of 1.1%. This presence probability of concentrations higher than 1000 Bq/m³ in the area defined as category 2 (between 301 and 400 Bq/m³) is reduced to 0.4%. In the same way as before, the probability is reduced in the following categories: for category 3 its probability is 0.2%, in category 4 it is 0.1% and in category 5 it is 0%.

Table 8. Probability to exceeding 1000 Bq/m³ in each P90 categories.

Category	Data number with > 1000 Bq/m ³	%	PROBABILITY (%) Data number in each category with > 1000 Bq/m ³ between total data number in each P90
Category 1 (>400 Bq/m ³)	23	68	1.1
Category 2 (301-400 Bq/m ³)	1	3	0.4
Category 3 (201-300 Bq/m ³)	4	12	0.2
Category 4 (101-200 Bq/m ³)	6	18	0.1
Category 5 (<100 Bq/m ³)	0	0	0
TOTAL	34	100	0.4

c3) Probability to exceeding 10 times the national average value

The national average value of radon concentration is 95 Bq/m³. When analysing the probability of exceeding 10 times the national average value (950 Bq/m³), data similar to those described in Table 8 appear. The best probabilities are given in category 1 (1.1%) and category 2 (0.7%), reducing these in the other categories (see Table 9).

Table 9. Probability to exceeding 10 times the national average value in each P90 categories

Category	Points Number	PROBABILITY (%) Data number in each category with > 95 Bq/m ³ between total data number in each P90
Category 1 (>400 Bq/m ³)	23	1.1
Category 2 (301-400 Bq/m ³)	2	0.7
Category 3 (201-300 Bq/m ³)	4	0.2
Category 4 (101-200 Bq/m ³)	6	0.1
Category 5 (<100 Bq/m ³)	0	0

Transposing the Spanish radon potential map to the 10 km x 10 km cell system

a) Transposition

The CSN allows the visualization of the Spanish Radon Potential Map on its web page (CSN, 2017a; 2017b; 2019) and download the pdf format (Figure 39). It does not allow the download of polygons in shp or dxf format.

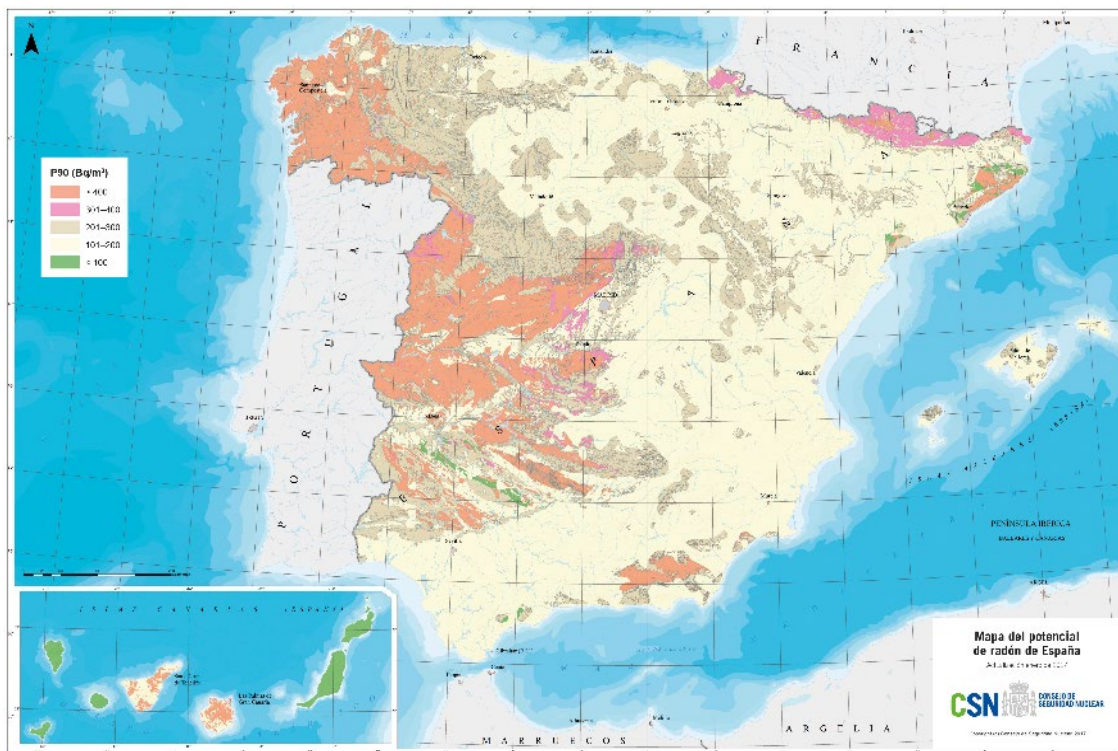


Figure 3915. Spanish Radon Potential Map (CSN, 2017a; 2017b).

The working methodology to obtain the polygons was as follow: Images were captured at an approximately scale 1: 5,000. These images later were georeferenced and from them, the polygons corresponding to each P90 category were digitized with as much detail as possible (between 1: 3,000 and 1: 5,000).

These polygons were intersecting with the 10 km x 10 km cells system, and finally the value of the category P90 with more area in each cell was assigned (Figure 40).

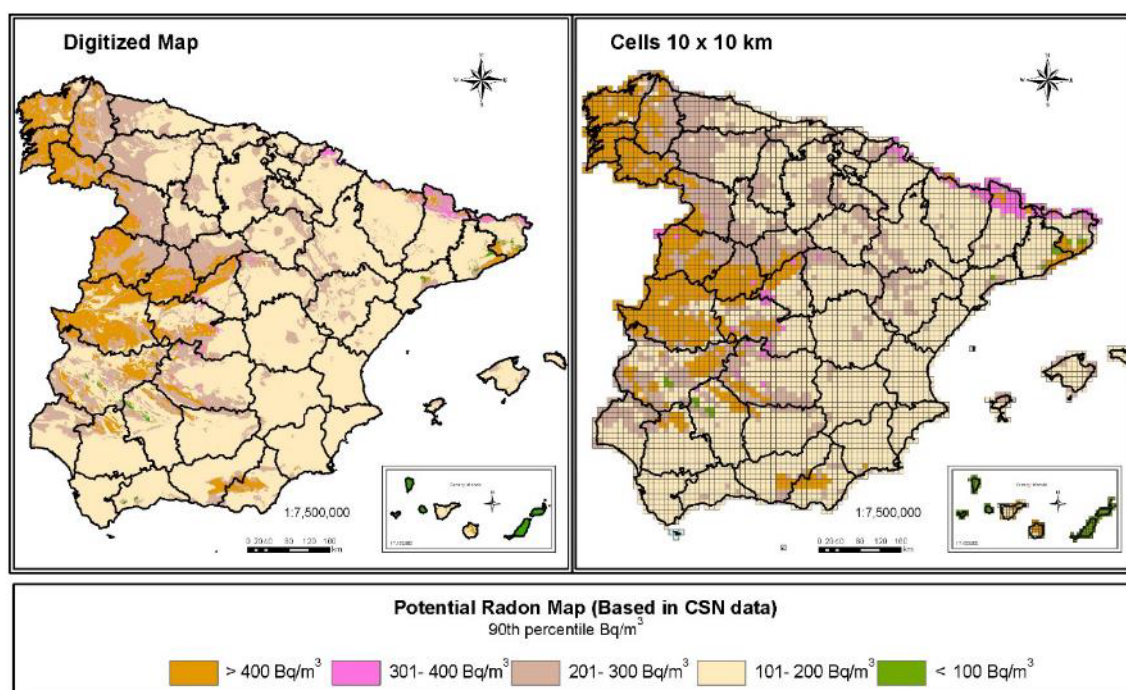


Figure 40. Digitization and transfer to the 10 km x10 km cell system from the CSN radon potential map. (Own development).

The result (Table 10) is that within category 1, 858 cells were generated (16% of Spain) and 115 cells in category 2 (2%). These data reflect that 18% of the territory can present radon levels higher than 300 Bq/m³. Category 3 represents 21% of the surface with 1,158 cells, category four 59% with 3,229 cells, and finally the category five is the least numerous representing 2% of the territory and 106 cells 10 km x 10 km.

Table 10. Number of cells 10x10 km in each category P90.

Cells Category	Number of cells	Number of cells (%)
Category 1 (>400 Bq/m ³)	858	16
Category 2 (301-400 Bq/m ³)	115	2
Category 3 (201-300 Bq/m ³)	1158	21
Category 4 (101-200 Bq/m ³)	3229	59
Category 5 (<100 Bq/m ³)	106	2
TOTAL	5466	100

b) Statistic of 10 km x 10 km cells system regarding to radon concentrations

Previously it was explained that Spain sampling was not random, the number of measures to be taken was intensified in an areas previously selected. The decision of the number of measures to perform in each 10 km x 10 km cell was taken by the CSN taking into account the general objectives established in the European Radon Map, considering a surface criterion, a population criterion and a lithostratigraphic criterion, and according to the terrestrial gamma radiation exposure (Sainz et al, 2014; Sainz et al, 2017).

The statistic of 10 km x 10 km cells can be seen in Table 11. The 26% of measurements were made in cells that were later were defined as category 1. It is observed that the average radon concentration (176 Bq/m³) in these cells is notably higher than the national average (95 Bq/m³). 2% of the measurements were taken in cells defined as category 2, presenting an average radon concentration of 90 Bq/m³. This same average appears in category 3 cells (18% of the data). As for the rest of the categories, its arithmetic average is reduced below 60 Bq/m³.

Table 11. Cells Statistics in each category P90. AM: arithmetic mean; SD: standard deviation.

Cells Category	Number of data in cells	Data (%)	Bq/m ³						
			Min	Max	AM	SD	1st Quartile	Median	3rd Quartile
Category 1	2394	26	10	15403	176	503	51	94	181
Category 2	145	2	10	717	90	88.9	38	67	109
Category 3	1674	18	10	1290	90	102.5	34	61	110
Category 4	4821	52	10	1972	59	74.9	21	40	71
Category 5	175	2	10	481	56	61.4	21	40	63

c) Probability to exceeding 300 Bq/m³ and 1000 Bq/m³ in cells category 1 (> 400 Bq m³) and category 2 (301-400 Bq/m³).

As in the case of point data, the radon concentration data of more than 300 Bq/m³ and 1000 Bq/m³ included within the cells are selected. In category 1, 11% of data exceed 300 Bq/m³ and 1% the 1000 Bq/m³. In category 2, 3% of data exceeds 300 Bq/m³, while no data exceeds 1000 Bq/m³ (see table 12)

Table 12. Cells Statistics in category 1 and 2 with data highest than 300 Bq/m³ and 1000 Bq/m³.

Cells Category	Data with >300 Bq/m ³		Data with > 1000 Bq/m ³	
	Nº	%	Nº	%
Category 1	271	11	29	1
Category 2	5	3	0	0

Geological study about radon potential areas with category 1 (> 400 Bq/m³) and category 2 (301-400 Bq/m³).

This section discusses a geological study of the Radon Potential Areas defined by the CSN as category 1 and 2 by analysing the digitized areas and the 10 km x 10 km cell system. This study will be making from the radon concentrations obtained, taking into account different parameters such as lithostratigraphy analysis 1: 200,000 (IGME, 2009) to continue with the established criteria by the CSN in the cartography definition of the Spanish potential radon map. In addition, to attempt the harmonization of the geological data with the rest of the European countries, the 1: 1000 000 cartography of the One Geology project (IGME, 2019) is used.

a) Digitized categories

a1) Lithostratigraphies 1: 200000

The geologies associated with the areas defined as category 1 (<400 Bq / m³) in Peninsula and Canary Islands, just as the geologies in category 2 (between 301 and 400 Bq / m³) were analysed from the number of measurements made within these areas (Table C, Annex 4).

a1.1) Category 1

In Peninsula, the highest average of radon concentration (541 Bq/m³) is associated with 43 samples taken on "sandstones, silt, and ochre conglomerates" lands. They are lithostratigraphies of detrital type linked to medium permeability. Of these 43 points, 26 % exceeds 300 Bq/m³, and 12% the 1000 Bq/m³.

Above 200 Bq/m³ there are three lithostratigraphs associated with a meta-detrital origin and low permeability: The first is "arcosic sandstones, schists, shales, conglomerates and volcanic rocks" (260 Bq/m³). In this lands, were taken 119 samples of which 15 % exceeded 300 Bq/m³ and 6 % the 1000 Bq/m³. The second corresponds to "schists, phyllites, quartzites, ampelites and lidites" (227 Bq/m³), in these geologies, 49 samples were taken, of which 9 % exceeded 300 Bq/m³ and 2% the 1000 Bq/m³. The third geology corresponds to "schists, paragneis, quartzites, and mica-schists" (219 Bq/m³), where of the 79 samples analyzed, 18 % exceeds 300 Bq/m³.

Above 100 Bq/m³ there are 5 lithostratigraphies mainly associated with meta-detrital and igneous origin and low permeability: they are "slates, schists, meta-areniscs and amphibole gneisses" (182 Bq/m³). They are also Hercynian (= Variscan) plutonic acid rock such as "granites, granodiorites, quartz-diorites" (163 Bq/m³); "Slates, grauwacas and sporadic carbonated levels" (162 Bq/m³); "Metamorphized acid rocks" (133 Bq/m³), and "Mica-schists, quartzites and gneisses" (119 Bq/m³).

With regard to the Canary Islands, a lithostratigraphy appears that approaches 200 Bq/m³, but only one measurement was made in it. The geology with a sufficient sampling density that can reflect accurate data, is the one corresponding to the 76 data taken on "subordinate wind sands", where 14 % of the data exceeds 300 Bq/m³ and 4 % the 1000 Bq/m³

a1.2) Category 2

This category only appears in Peninsula. 4 lithostratigraphies present more than 100 Bq/m³: they are "Schists and paragneiss"; "Limestones, sandstones and shales"; "Arkoses with boulders, conglomerates and clays"; and "Grauwackes and slates".

a2) One Geology Lithologies 1: 1,500,000

In the Iberian Peninsula, there are 7 geologies associated with areas of more than 400 Bq/m³. The highest average radon concentration (165 Bq/m³) corresponds to 1,376 samples taken on "biotite granitoids", which 12 % exceed 300 Bq/m³ and 1 % the 1000 Bq/m³. The following lithologies with high radon concentrations correspond to "Slates, sandstones, quartzites, limestones or vulcanoclastic rocks" (146 Bq/m³) and "acid rocks metamorphosed as peraluminous granitoids" (143 Bq/m³). Other formations that exceed 100 Bq/m³ are the "Serpentinites, metabasites and meta-vulcanites acids"; "Graffitous micachists with garnet"; and "Intermediate and basic igneous rocks". In all exceeds 8% of the data above 300 Bq/m³.

In the case of the Canary Islands, a unique geology appears that exceeded 100 Bq/m³, corresponds to "Calc-alkaline volcanic rocks", where of the 148 sampled data, 9 % exceed 300 Bq/m³, and 2 % 1000 Bq/m³. (Table D, Annex 4).

b) Categories in the 10 km x 10 km cells system.

b1) Lithostratigraphies 1: 200000

The geologies associated with the cells defined as category 1 ($> 400 \text{ Bq/m}^3$) in Peninsula and Canary Islands, just as those in category 2 (between 301 and 400 Bq/m^3) were analysed in base on the number of measurements made within the each cell (Table E, Annex 4)

b1.1) Category 1

In Peninsular Spain there are 3 lithostratigraphies in cells that exceed 200 Bq/m^3 of radon average concentration: "Sandstones, silt, and ochre conglomerates" have a high average concentration (985 Bq/m^3) with 52 % of the data above 300 Bq/m^3 , and 22 % above the 1000 Bq/m^3 . A fact to be taken into account is the sampling shortage in this geology, since only represents 1 % of the cells and 23 points were sampled in them.

The other two lithostratigraphies that exceed 200 Bq/m^3 correspond to "Schists, paragneis, quartzites, and mica-schists"; and "Slates, grauvacas and sporadic carbonated levels". Both geologies are of meta-detrital type and low permeability. They show a high percentage of data above 300 Bq/m^3 (23 % and 11 % respectively) and above 1000 Bq/m^3 (2 % and 3 % respectively).

Between 100 and 200 Bq/m^3 of radon average concentration, 13 lithostratigraphies appear, in many cases the sampling density is insufficient. Therefore, two geologies are identified that present a suitable representation in terms of cell's number and data number sampled within them: The first corresponds to Hercynian plutonic acid rock such as "granites, granodiorites, quartz-diorites" (172 Bq/m^3) where 12 % of the data exceeds 300 Bq/m^3 , and "Metamorphized acid rocks" (132 Bq/m^3) where 9 % of the data exceeds 300 Bq/m^3 .

In Canary Islands there is a single representative lithostratigraphy within category 1. It corresponds to "Wind sands subordinate" (186 Bq/m^3) in which 9 % of the data exceeds 300 Bq/m^3 .

b1.2) Category 2

Within the category 2 cells ($301\text{-}400 \text{ Bq/m}^3$), only one lithostratigraphy can be indicated. They are "Quartzites and slates" (130 Bq/m^3) of meta-detrital type and low permeability, where 13% of the data exceeds 300 Bq/m^3 .

b2) One Geology Lithologies 1: 1500000

b2.1) Category 1

In Peninsula and Canary Islands, there are 12 geologies associated with cells with more than 400 Bq/m^3 (Table F, Annex 4). The highest average radon concentration (485 Bq/m^3) corresponds to 20 samples taken in "Sandstones, shales, quartzites or limestones and conglomerates". The sampling density is insufficient to ensure that this geology can be associated with high radon concentrations.

The most representative lithologies in this case are the "Biotitic Granitoids" (192 Bq/m^3) with 1,404 samples taken in these cells, where 12 % of the data exceeds 300 Bq/m^3 and 1 % the 1000 Bq/m^3 ; the "Serpentinities, metabasites and meta-vulcanites acids" (184 Bq/m^3) where 15 % of the data exceeds 300 Bq/m^3 and 1% the 1000 Bq/m^3 ; the "acid rocks metamorphosed as peraluminous granitoids" (151 Bq/m^3) where 12% of the data exceeds 300 Bq/m^3 ; and the "Slates, sandstones, quartzites, limestones or vulcanoclastic rocks" (146 Bq/m^3) where 8 % of the data exceeds 300 Bq/m^3 .

b2.2) Category 2

Within the category 2 cells (301-400 Bq/m³), only a single lithostratigraphy with representativeness can be indicated. They are the "Slates, schists, sandstones, limestones, ampelites and lidites" (125 Bq/m³) where 12 % of the data exceed 300 Bq/m³.

c) Geological identification of radon concentration data above 1000 Bq/m³.

The radon concentration data measurements higher than 1000 Bq/m³ are identified, and different geological parameters are analysed:

- Lithostratigraphies 1:200,000 (IGME, 2009): typology, permeability and lithostratigraphy origin.
- One Geology Lithologies 1: 1,500,000 (IGME, 2019)
- Karst Lithologies 1: 1 000 000 (IGME, 1986): typology
- Faults 1: 1 000 000 (IGME, 2019): proximity and typology
- Natural gamma radiation exposure rate (CSN, 2001).

The following Figure 41 shows the 34 point data of radon measurements in dwellings in which the 1000 Bq/m³ of annual average concentration is exceeded.

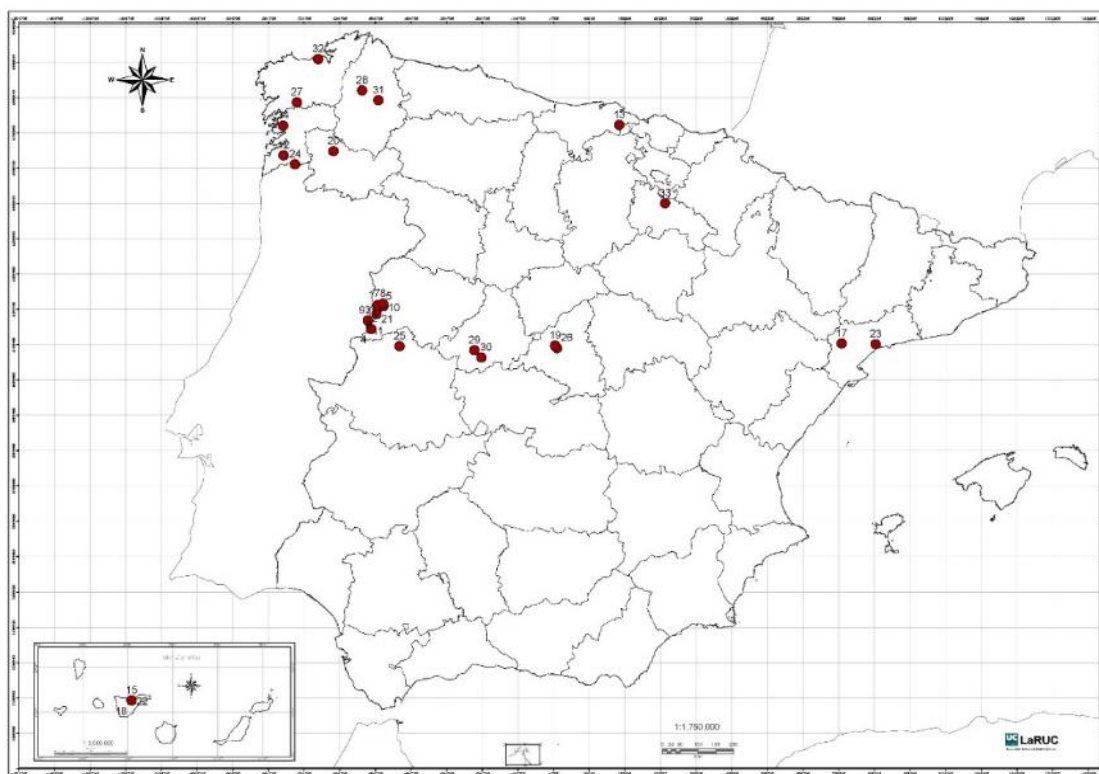


Figure 41. Radon concentration data > 1000 Bq/m³.

Table G in Annex 4 analyzes the geological parameters of these points: Regarding the lithostratigraphies 1: 200,000 it is verified that 26 % of the data were taken on Hercynian plutonic acid rock such as "granites, granodiorites, quartz-diorites"; 18% on "Arkosic sandstones, schists, shales, conglomerates and volcanics

rocks. Inf. and Sup. detrital series"; and 15% on "Arenis., lim.y congl.ocres (Ar.Toro-Corrales,Limos de Geroma,Congl.Villalazán,G.Sup-Inf Paleóg.)".

Mainly these high concentrations are associated with low permeability (59% of the data) or medium permeability (18% of the data), although they also appear on lithostratigraphies with high permeabilities (18%) associated with "Gravels, sands, silts and clays (medium and high terraces deposits)" and "Wind sands subordinates".

The meta-detrital and igneous origin is the most representative of these high concentrations (29% and 26% respectively), but also shows that 21% and 15% of the measurements were made on detrital and quaternary detritic type rocks.

If these points are analysed from the 1: 1 million lithological classification of the One Geology project, it can be seen that 26 % of the samples were taken on "Biotitic Granitoids", 26 % on "Slates, sandstones, quartzites, limestones or vulcanoclastic rocks "and 24 % on" Sandstones, slates, quartzites or limestones and conglomerates "

Referring to the terrestrial gamma radiation rate exposure, it is taken into account that the national fund corresponds to 44 nGy/h (approximately 5 µR/h) (García et al., 2013), it is observed that the 61 % of the data were performed in areas classified as medium exposures (between 44 and 122 nGy/h) and the remaining 39 % in areas of high exposures (> 122 nGy/h)

There are 5 data on karstic formations: Three samples were taken in Canary Islands on "wind sands subordinates" (calcoalkaline volcanic rocks) and the other two in moderately karstified carbonate formations on "sandstones, slate, quartzite or limestone and conglomerates".

Finally, the faults proximity analysis shows that 12 of the sampled points of more than 1000 Bq/m³ are less than 2 km from these structures.

4.3 Discussion and perspectives

For France and Spain, both arithmetic means of indoor radon measurements are quite close and around 90 Bq/m³.

In France, the highest indoor radon concentrations (exceeding 300 Bq/m³ or 1000 Bq/m³) are located in the Armorican Massif, the Massif Central, and the Pyrenees at the border with Spain, in Corsica and in the Jura Mountains at the border with Switzerland. The main high values are associated with some peculiar granite with a clear dominance of monzogranites, peraluminous leucogranites or peralkaline granites. The high values identified in the Jura Mountains are located in karstic areas.

In Spain, the highest indoor radon concentrations are associated with "Biotitic Granitoids", "Slates, sandstones, quartzites, limestones or vulcanoclastic rocks, Sandstones, slates, quartzites or limestones and conglomerates". There are 5 data on karstic formations: Three samples were taken in Canary Islands on "wind sands subordinates" (calcoalkaline volcanic rocks) and the other two in moderately karstified carbonate formations on "sandstones, slate, quartzite or limestone and conglomerates". Finally, the faults proximity analysis shows that 12 of the sampled points of more than 1000 Bq/m³ are less than 2 km from these structures.

The results provide first elements to target areas where more precise studies are needed to acquire more indoor radon data precisely located and the characteristics of buildings associated with the measurements. An analysis of both geological features and building characteristics (mainly the interface between the soil and the building, the building materials, ventilation systems etc.) need to be realized to identify the best indicators of highest indoor radon values.

5 Radon hazard index RHI

The complete discussion can be found in [Bossew et al. \(2020\)](#), “Development of a Geogenic Radon Hazard Index—Concept, History, Experiences”, Annex 3 of the deliverable D5 of MetroRADON Project.

In the following, only introduction and objective of the GRHI are shortly repeated, and some material presented which is not included in the article.

5.1 Introduction and rationale of RHI

The GRHI has been conceived as a possible alternative or complement to the GRP. It shall quantify the hazard originating from geogenic Rn on a deliberate scale, for example from 0 to 1 or from 0% to 100%, etc.. The idea behind is that in most European countries, quantities have been surveyed, or are available as databases, which are physically and statistically related to the GRP (section 4.3.4.5). The idea is visualized in Figure 42.

Discussion about a definition of the GRHI, which on the one hand adequately implements its idea, but is practically manageable, on the other, has been under way for some years. Once a passable definition has been found, the GRHI could be a possible database for a Europe wide map of radon priority areas. In this sense, the GRHI could be a harmonized measure of the *radon priorityness* of an area or of a location.

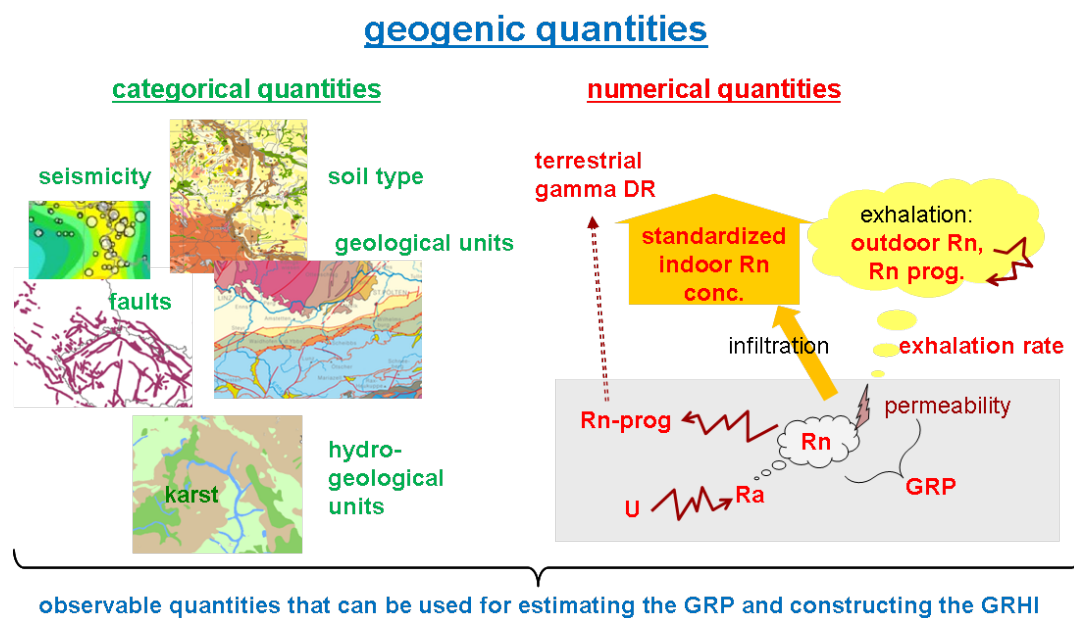


Figure 42: Candidate quantities for constructing the GRHI

5.2 Concept and objective

The GRHI, as a measure of geogenic Rn hazard, shall be comparable across Europe, irrespective of the geogenic databases from which it has been calculated. This can be achieved,

- (1) either, by using the same geogenic database everywhere;
- (2) or using regional databases, but ensure that the resulting value of the GRHI does not depend on the database used.

Such GRHI would be the base of a European map of geogenic Rn and a European wide determination of Rn priority areas.

Remind that C is also – sometimes dominantly - controlled by anthropogenic factors. These can partly be factorized out, e.g. by restricting to *standard situations* (e.g. only ground floor rooms) or to *standardized cases* (Friedmann RP approach). For Europe, the indoor Rn dataset which represents the standard situation is the one underlying the European Indoor Rn Map, i.e. AM or AML within 10 km × 10 km cells.

Whichever standardization of C used, it must be expected that there is a geographical trend in the anthropogenic factors, due to climatic and cultural differences which influence building styles and living habits. Therefore, it must be expected that residuals around C predicted by the GRHI will also show a geographical trend, hence violating the regression requirement of randomness and independence of residuals.

So far, the spatial-statistical properties of the anthropogenic factor(s) have not been paid much attention, in contrast to the geogenic factors.

The objective is to generate a quantity whose regional variability represents as much as possible the variability of the geogenic controls of Rn hazard. In other words, these factors shall be squeezed appropriately into one quantity “GRHI”;

The GRHI can be conceptualized in different terms:

- a quantity which measures the contribution of geogenic factors to the potential risk that exposure to indoor Rn causes;
- a quantity which measures the availability of geogenic Rn at surface level;
- a measure of susceptibility of a location or of an area to increased indoor radon concentration for geogenic reasons;
- a measure of “Rn proneness” or “Rn priorityness” (in the logic of the BSS) of an area due to geogenic factors; i.e., a tool to decide whether an area is RPA.

Desired properties of the GRHI are:

- (I) consistency, across borders between regions, characterized by different databases used for the estimation; this implies independence of the actual database used,
- (II) exhaustiveness, which should reflect as much as possible the available geogenic information;
- (III) simplicity, which should be simple to calculate;
- (IV) predictor of the IRC, which should be a valid predictor of the geogenic contribution of indoor Rn concentration. This is motivated by its very concept.

These properties can be fulfilled only partly to different degrees by different concepts and are even partly contradictory.

5.3 Review of existing GRHI trials

Existing GRHI trials have been reviewed in Bossew et al. (2020). Here only the approaches by [Friedmann \(2011\)](#) and [Bossew \(2016a\)](#) are presented in detail.

a) Friedmann’s top-down radon potential

The following text is adapted from [Long Way \(2011\)](#), section 5.4.3, written by H. Friedmann (Univ. Vienna).

In a working document to the EGRM expert group, Friedmann proposed a purely analytical approach to defining a “radon hazard index” (RH), out of controlling input, or primary quantities, like Rn concentration in soil air, standardized concentration in indoor air, soil permeability, or geochemical quantities, etc., as available. Including proxies which are only partly correlated to Rn, like external dose rate, yields additional uncertainty of the RH, as does inclusion of class variables (also of nominal type, like geological classes) via values which are deemed representative for the class (e.g. a typical value of Rn concentration in soil gas, for a particular geological or lithological unit). Conceptually similar approaches have been proposed by [Smethurst et al. \(2008\)](#) and [Dehandschutter et al. \(2008\)](#).

The starting point of the proposal is that from physics, RH should be a function of soil gas concentration and permeability, similar to the approaches of other authors (e.g. the Neznal-GRP). These primary input variables are either available at a location, or are estimated from proxies, or default values are set.

However, the influence of the permeability must be discussed a little further. The RH is limited by the amount of soil gas radon. This means that above certain permeability it will not increase the RH any more. The same is true for the emanation process from the rock material which depends primarily from the grain size which is primarily responsible for the soil permeability (together with water content). Therefore, the permeability will not linearly influence the RH but it will have more influence at lower values than at higher values. This can firstly simulated by a logarithmic influence of the permeability and the factor permeability in the calculation of the soil-gas concentration and the RH has to be changed into a factor which holds the logarithm of the permeability.

Because the factors have units, we have to normalize them. The normalization must be done in a way that the result is well balanced between the different factors. It seems beneficial for a real RH to use the decimal logarithm of the relevant parameters which means just to add/subtract “scoring points” for the calculation of the RH. A first formula can be

$$RH = 2.5 \cdot \log\left(1 + \frac{Rn}{50 \text{ kBq} / \text{m}^3}\right) + 0.33 \cdot \log\left(1 + \frac{P}{10^{-15} \text{ m}^2}\right)$$

with Rn the soil gas Radon concentration (‘mean definition’) in kBq/m³ and P the permeability in m² (see Table 19).

If these data are not available either mean values from the geological units can be used or the soil-gas Radon concentration must be deduced from uranium concentration, maybe in the form

$$Rn[\text{kBq} / \text{m}^3] = 4 \cdot U[\text{ppm}] \cdot \left(1 + 0.25 \cdot \log\left(1 + \frac{P}{10^{-15} \text{ m}^2}\right)\right)$$

The soil-gas Rn concentration can also be estimated from dose rate measurements. Again a crude formula (assuming common contributions of ⁴⁰K and the Th-decay chain) can be given:

$$Rn[\text{kBq} / \text{m}^3] = 0.2 \cdot D[\text{nGy} / \text{h}] \cdot \left(1 + 0.25 \cdot \log\left(1 + \frac{P}{10^{-15} \text{ m}^2}\right)\right)$$

These formulas can be considered “pragmatic” implementations of the transfer concept.

Additional influences like fault zones, special geological circumstances, uranium ores, mining activities, tailings etc. must be regarded separately because they can only be judged in connection with the local situation.

Table 20: RH computed according to formula (1). The bold lines separate areas with $RH \leq 1, 2, \dots$

Rn in kBq/m ³	P=10 ⁻¹⁴ m ²	P=10 ⁻¹³ m ²	P=10 ⁻¹² m ²	P=10 ⁻¹¹ m ²	P=10 ⁻¹⁰ m ²
20	0,71	1,03	1,36	1,69	2,02
40	1,00	1,34	1,69	2,04	2,39
60	1,22	1,56	1,91	2,26	2,61
80	1,40	1,74	2,09	2,44	2,79
100	1,56	1,89	2,24	2,59	2,94
120	1,69	2,03	2,38	2,73	3,08
140	1,81	2,15	2,50	2,85	3,20
160	1,92	2,26	2,61	2,96	3,31
180	2,02	2,36	2,71	3,06	3,41
200	2,11	2,45	2,80	3,15	3,50

Geology can be included as predictor by assigning default values of the Rn concentration in soil gas, taken from observations, if no measurements are available at a point. From German (Kemski et al. 2001, 2009) and Czech sources (Barnet et al. 2008) values shown in Table 21 are taken.

Table 1: Geological units and estimated mean soil gas radon concentrations in kBq/m³ according to the different definitions of soil gas radon concentration. The data are a combination of data from Germany and from the Czech Republic

Geological Units	Mean definition	Maximum definition
Diorites and Gabbros	25	46
Mesozoic sediments	20	37
Moldanubicum group	30	56
Orthogneis, granulites and migmatites	34	63
Paleozoic folded metamorphosed	40	74
Paleozoic folded unmetamorphosed	37	69
Permocarboniferous sediments	30	56
Proterozoic-palaeozoic volcanites	23	43
Quaternary sediments	20	37
Tertiary sediments	22	41
Tertiary volcanites	33	61
Variscan granites, granodiorites and tonalites	59	109

A similar table could be established for permeability; the controlling categories would be different, referring to surface geology or soil types.

b) An exercise using German data

In an exercise carried out for Germany (Bossew et al. 2016), the GRP, ADER and U in topsoil were chosen as predictors (estimates on $10 \text{ km} \times 10 \text{ km}$ cells).

Several examples were given based on weighted means of correlations between predictors and indoor Rn, or with the GRP. In the latter case the weight given to the GRP was set to 1. For normalizing, the weights were divided by the sum of all. Predictors were transformed into their distributions, $y \rightarrow F_Y(y)$. The former case leads to

$$\text{GRHI} = 0.34 F_{\text{GRP}} + 0.31 F_U + 0.37 F_{\text{ADR}},$$

shown as a map in Figure 43 (left). The right graph shows the association of the GRHI with indoor Rn exceedance probability, $\text{prob}(C > 100 \text{ Bq/m}^3)$.

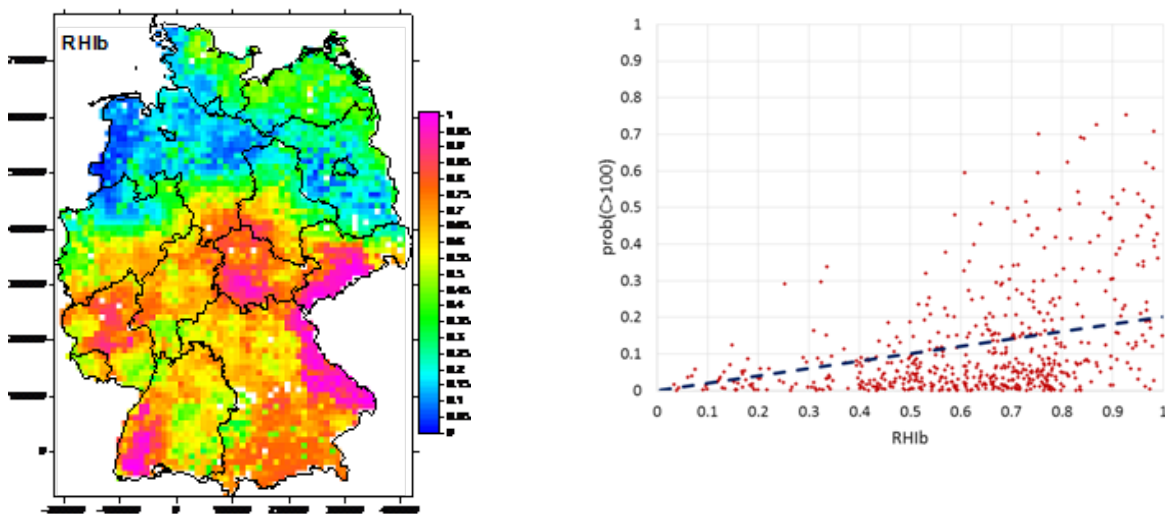


Figure 43: Left: GRHI map of Germany, constructed from GRP, uranium ion topsoil and ambient dose rate. Right: Scatter plot GRHI with indoor Rn exceedance probability. (Dashed line: linear regression through 0, only for orientation.)

5.4 Statistical background, methodology, challenges

a) Type of variable

The GRHI is an ordinal quantity which can be (Figure 44):

- a continuous index (real number), e.g. $\in [0,1]$ or $(-\infty, \infty)$ etc.;
- a discrete index or score, e.g. $\in \{I, II, III, IV\}$ or $\{\text{low, medium, high}\}$ etc.

A quantity of the first type can be transformed into the second type, but not v.v.

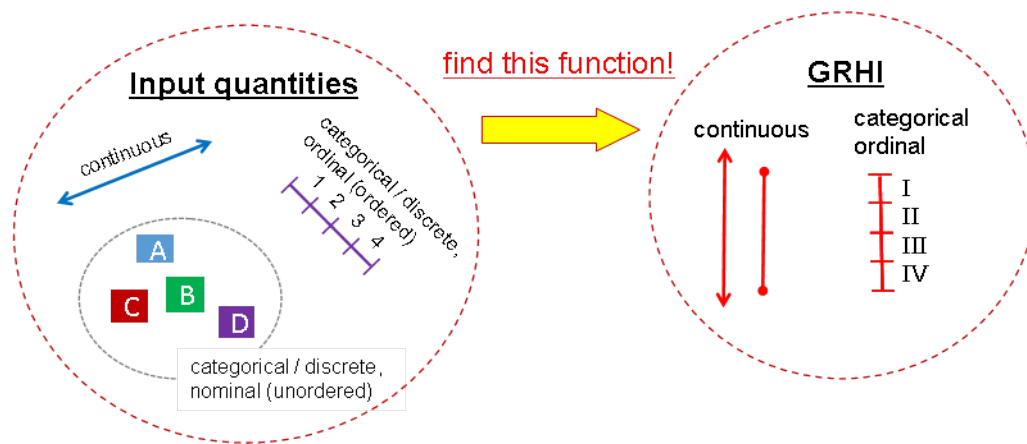


Figure 44: Construction of different types of GRHI-quantity from different types of input quantities.

Continuous quantities can be transformed to [0,1] by sigmoidal function g , for example:

- logit (tanh) type: $g(y)=1/(1+a \exp(-by))^c$
- probit type: $g(y)=(1+\text{erf}(b(y-c)))/2$
- Gompertz type: $g(y)=a \exp(-b \exp(-cy))$
- distribution: $g(y)=F_V(y)$

A quantity confined $[a,b]$ can be transformed into $[0,1]$ by

$$g(y) = (f(y)-f(a)) / (f(b)-f(a)), \text{ in the simplest case } f=1, g(y)=(y-a)/(b-a).$$

b) Areal coverage of predictors

Not only the numerical types of variables are different (real number, ordinal category, nominal unordered category), this is also true for their spatial properties. Geological maps divide a domain into tiles exhaustingly, but dependent on map resolution (A in Figure 45. Remote sensing, e.g. by airborne gamma spectrometry, generates positive real numbers as output, which represent a weighted mean (with complicated weighing function) of an area (detector horizon or "footprint", B in Figure 45). Depending on the mode of operation of the carrier (airplane, helicopter, UAV), remote sensing can yield exhaustive coverage of region. Point samples (C in the same figure), typical for geochemical concentrations in soil samples, soil gas samples, or ambient dose rate measurement, are scattered draws from a "population", as which the theoretical, infinite ensemble, or "field" of the quantity is considered. However, also point samples have in fact a finite, i.e. non-zero footprint or "sampling support", which may be small as for soil samples, but amounts to some 10^3 to 10^4 m² for ADR measured 1 m above ground.

Particular types of point samples are indoor Rn measurements. Their sampling support is the area occupied by a building, some 100 m². A vertical dimension is defined by floor levels. Furthermore, the number of houses is necessarily final, in contrast to the number of theoretically possible ADR or soil measurements. (In fact, also only a final number of soil samples can be taken from an area, but given the small sampling support, some 100 cm², the theoretically possible number is very large.)

For houses, this means that exhaustive sampling is possible, still not covering a domain, because not everywhere stand houses. In indoor Rn maps, and derived maps (RPA etc.), one estimates Rn concentration on locations of hypothetical houses, as if one would be built at the estimation point. Therefore, often Rn maps show values at locations where for physical reasons there cannot be houses, e.g. in mountains, on top of rivers or swamps.

Only to mention, sampling from finite populations has also statistical consequences, mostly more difficult to treat than infinite populations. However, this is not relevant in the context of this chapter and will therefore not be further discussed here.

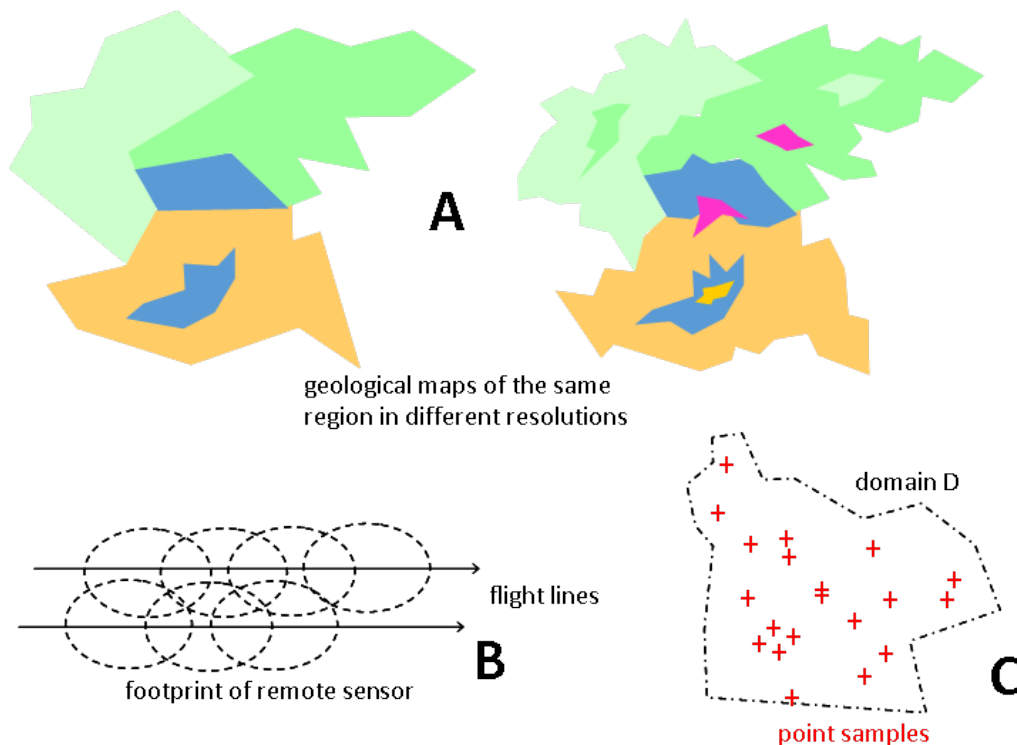


Figure 45: Areal coverage of different types of observations

An important issue is assignment of a value of categorical quantity, typically geology, to an estimation point x^* . The geological unit on which x^* lies on the map, or which is assigned to x^* , depends on the resolution and the classification depth (legend) of the map; the "true" geology is usually not available, which would have to be assessed in situ during sampling.

The type of uncertainty resulting from dependence on geological (hydrological, soil,..) map used, is very difficult to quantify and may add considerably to the "noise" of models in which geology etc. are covariates.

For point-type covariates, the problem consists in that their value has to be estimated at x^* from the ones of sampling locations. Usually this depends on the viability of geostatistical estimation or interpolation, which in turn relies on a covariance structure of that quantity. Simultaneous co-estimation of a number of covariates is theoretically possible, but more often than not, prohibitively complicated in practice. It seems that machine learning (ML) can to some extent overcome the problem.

c) Predictors, proxies, latent variables

Trivial wisdom has it that statistical correlation does not necessarily prove physical causation. Still, correlation (unless by chance) indicates an underlying physical structure which generates it, Figure 46. Correlated variables Z_1 and Z_2 may have a common physical cause, Z_0 , which is the reason for their statistical correlation.

Z_2 shall be the quantity of interest, to be estimated from predictors. The evident physical predictor is Z_0 , but data (observations) of Z_0 may not be available. On the other hand, one may have data of Z_1 . This may serve as substitute or proxy or surrogate for predicting Z_2 .

An example is $Z_2 =$ indoor Rn concentration, $Z_0 =$ GRP. If not available, one may choose $Z_1 =$ ADR (external ambient dose rate), which evidently does not cause indoor Rn, but still may serve as (quite imperfect) proxy-type predictor.

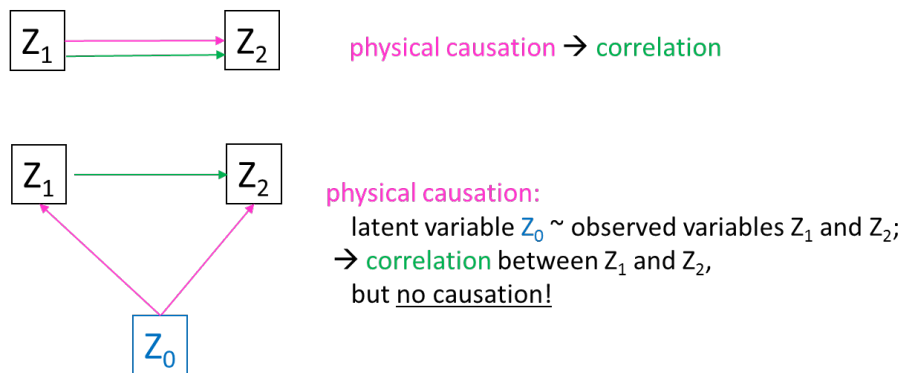


Figure 46: Causation, correlation and latent variables

In reality, the physical dependence structure is much more complicated and convoluted - see the “rock-to-risk” graph. For example, $Z_2 =$ indoor Rn also depends on building characteristics, say Z_0' , which do not generate ADR. On the other hand, $Z_1 =$ ADR also depends on cosmic ray intensity (Z_0'') and ^{137}Cs fallout (Z_0'''), which have no physically causative relationship with the GRP. The controls Z_0' , Z_0'' and Z_0''' blur or confound the correlation of Z_1 and Z_2 (Figure 47 left).

A particular trap appears if Z_2 is caused by Z_0 and Z_0' , but the latter are both partly caused by another quantity, say Z_b . As an example, $Z_2 =$ again indoor Rn, $Z_0 =$ U concentration in the ground and $Z_0' =$ permeability. Both are controlled by $Z_a =$ geology, together with other controls Z_b (e.g. soil type), Z_c (humidity), etc. (Figure 47 right). Z_0 and Z_0' can be expected to be correlated to some extent, which invalidates their function as independent predictors of Z_2 (i.e. absence of collinearity) in a regression model.

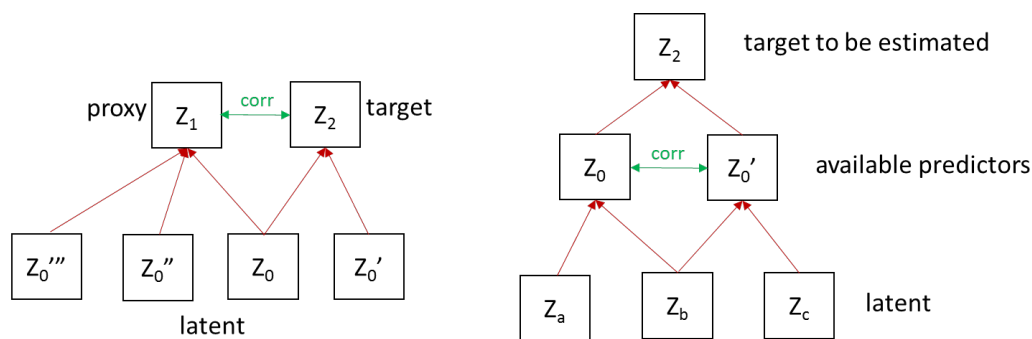


Figure 47: Left: Z_0' , Z_0'' , Z_0''' blur the correlation between Z_1 and Z_2 ; Right: Correlated predictors Z_0 and Z_0'

Obviously, the ability of a quantity to serve as predictor – physically causative or as proxy – has to be explored in each case.

d) Exploratory and confirmatory analysis

(Text mainly taken from Ciotoli et al. 2017c)

Analysis of high-dimensional (dimension = number of covariates), possibly convoluted and nested situations usually proceeds in two steps (Figure 48). As first, *exploratory* step, one tries to reduce complexity by investigating dependence between presumptive covariates. The following confirmatory step serves to establish the dependence between predictors (or their appropriate transforms), and the dependent quantity. The objective is being able to predict the latter from the former.

Practically, one is dealing with a (generalized) regression problem. The difference from common regression is that one has a number of predictors, but is not sure how important each is in the presence of the others. Therefore, one wants to tailor the predictor space such that regression actually makes sense and that redundancy (covariates which carry the same information) as well as non-information (covariates that do not contribute to explaining the dependent variable) is removed.

The confirmatory part proceeds along (more or less) conventional lines.

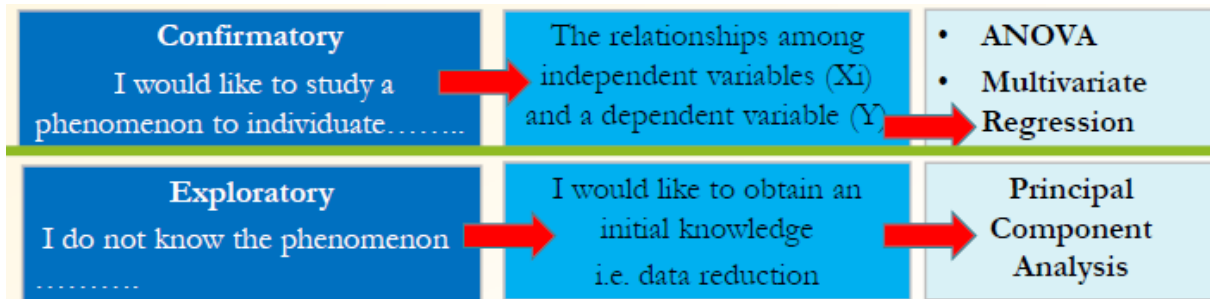


Figure 48: Exploratory and confirmatory analysis. (From Ciotoli et al. 2017c)

In a multi-dimensional setting, such as for R_n prediction from possibly many potentially predicting quantities, one would first attempt to identify the amount of information that the set of covariates actually contains; many of the predictors tend to be correlated between themselves, hence carrying redundancy.

Principal Component Analysis (PCA) is a common method whose main objective is to reduce the data complexity with minimum loss of information and to create m' new non-correlated variables (factors) linearly linked to the m ($\geq m'$) original variables (Figure 49). The factor loadings represent the correlation between the factors and the original variables. PCA is performed in absence of a designated response (or dependent) variable, i.e. just serves to "rearrange" the predicting covariates more efficiently.

The main idea is that the original m variables include some information redundancy caused by some correlation among them. The total information is redistributed in a most efficient way, and some of the original m variables can be excluded with a loss of a minimum percentage of the total information.

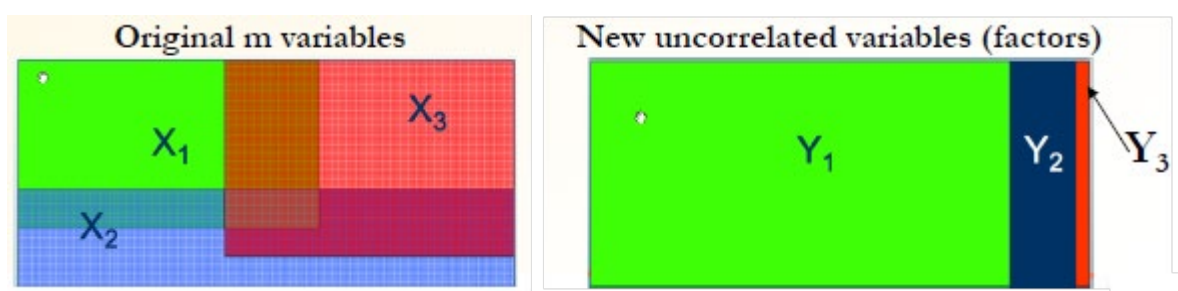


Figure 49: Idea of PCA. (From Ciotoli et al. 2017c)

A newer class of methods consists in machine learning. ML may combine exploratory and confirmatory analysis, seeking appropriate functional combinations of the predictors so that as much as predicting information is retained. The step beyond classical regression (with preceding PCA or not) consists in no necessity to specify a regression structure (or only very generally), as required for the former (e.g. linear regression). Although one would think that the dependence structure is given by the physics of the investigated system, it turns out that sometimes a complex *statistical* reality is better reflected by some intricate structure; this is not the say that it reflects the *physical* structure. In other words, the structure may not have physical meaning. The difference between the approaches is visualized in Figure 50.

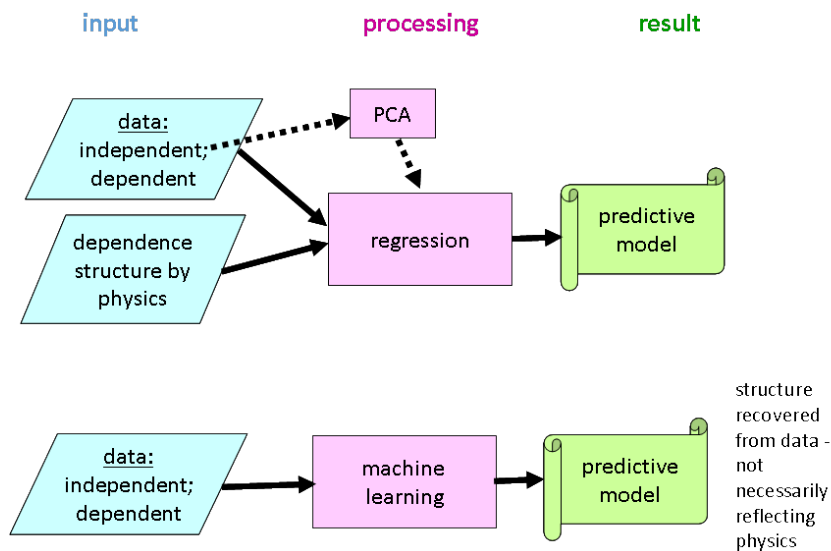


Figure 50: Difference between regression and machine learning

A particular class of regression includes location as independent variable, in order to account for possibly locally variable dependence structure. This can be implemented through geostatistics, which in the simplest form models a response variable as function of location. In more complicated versions (co-kriging family), additional predictors are added. With several possibly correlated and nested predictors the method becomes practically intractable.

Another approach is Geographical Weighted Regression (GWR) which provides a local model of the variable or process that one attempts to predict by fitting a regression equation to every feature in the dataset. GWR constructs these separate equations by incorporating the dependent and explanatory variables of features falling within the spatial bandwidth of each target feature.

e) Back-transform

A notorious problem is back-transform of estimates of transformed into ones of quantities in original space. The origin of the problem is that $E(f(Z)) \neq f(E(Z))$, unless f is a linear function. A well-known case is lognormal kriging: More often than not, environmental quantities are strongly right-skew distributed (often indeed approximately lognormal) and variograms are more easily (and theoretically more satisfyingly) estimated from log-transformed quantities. Back transform of local estimates $(\ln(Z))^{*}(x^{*})^1$ is possible but affords quite exact estimate of the kriging SD which enters as square and whose uncertainty can thus introduce large error. The method has been expanded to transgaussian kriging; for a recent Rn related application see e.g. [Elió et al. \(2019\)](#) (slide 19). An alternative is sequential simulation in transformed space, back transform of individual realizations, followed by computing the wanted statistics (for Rn: e.g. [Bossew 2015](#)). It seems that back transform is particularly complicated for regularized compositional data (CoDa) as response variable. Predictor CoDa in the context of Rn have been treated by [Ferreira et al. \(2018\)](#) and [Elió et al. \(2018\)](#).

A proposal for bias correction in back transforming ML generated estimates of a transformed variable by empirically regularizing histograms has been made by [Petermann et al. \(2019\)](#).

¹ x^{*} = the target location, on which quantity Z is to be estimated; Z^{*} = the estimate
16ENV10 MetroRADON Deliverable No.6

f) Consistency

The value of the GRHI at a location should be independent (up to tolerance) of which quantities it has been estimated from. For example, GRHI calculated from U concentration in soil should have approximately the same value as if calculated from dose rate or GRP, etc..

Given input quantities (for example U, DR, geol. class). Then should be:

$$\text{GRHI}(U, \dots) \cong \text{GRHI}(\cdot, \text{DR}, \cdot) \cong \text{GRHI}(U, \cdot, \text{Geo}) \cong \text{GRHI}(U, \text{DR}, \text{Geo}) \text{ etc.}$$

where \cong means “up to deviations which are due to the imperfect correlation (Figure 30, Figure) between geogenic quantities and statistical uncertainty”

This can also expressed as $E[\text{GRHI}(A) - \text{GRHI}(B)] = 0$, where A and B denote different input databases.

This is equivalent with the requirement of consistency across borders between regions A and B in which different input quantities $Y(A)$ and $Y(B)$ are available, and distance $\rightarrow 0$, see Figure 51.

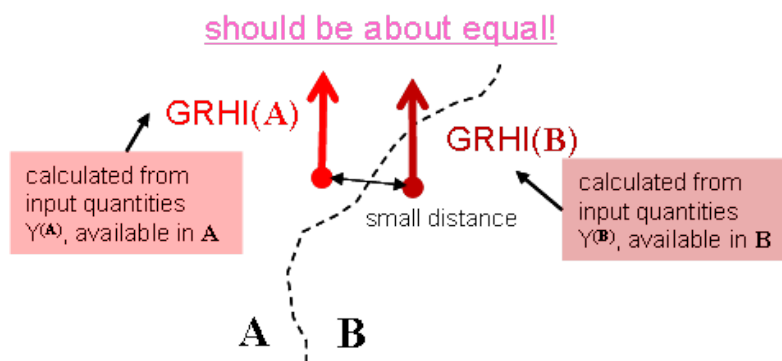


Figure 51: Consistency of the GRHI

If the GRHI is of variant (1) (see taxonomy in [Bossew et al. 2020](#)), i.e. calculated from the same database everywhere, the consistency requirement is trivially fulfilled. For variant (2), it seems to be the most complicated task in defining and calculating a GRHI. The problem could be relaxed by adequate classification of the GRHI.

The reason of inconsistency lies in the imperfect statistical association (correlation) of data from different databases, e.g. between U in topsoil and the GRP. The evident reason is that, in the same example, the GRP is controlled also by permeability.

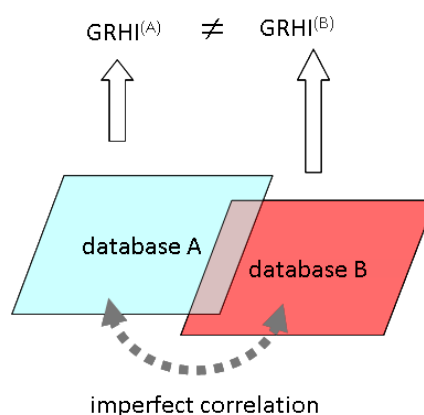


Figure 52: Reason for the lack of consistency between GRHI calculated from different databases.

A simple example is shown in Figure 53 Potential predictors are the GRP and U in topsoil (means of the In in 10 km × 10 km cells, data from Germany). From the leftmost graph it is evident that while the quantities are correlated (as they must be, because U is the primary physical of soil Rn), more factors control the dependence. This is however less evident from the distributions of the quantities and their nscore transforms (centre and right). ($\text{nscore } z = \Phi^{-1}(F_z(z))$.)

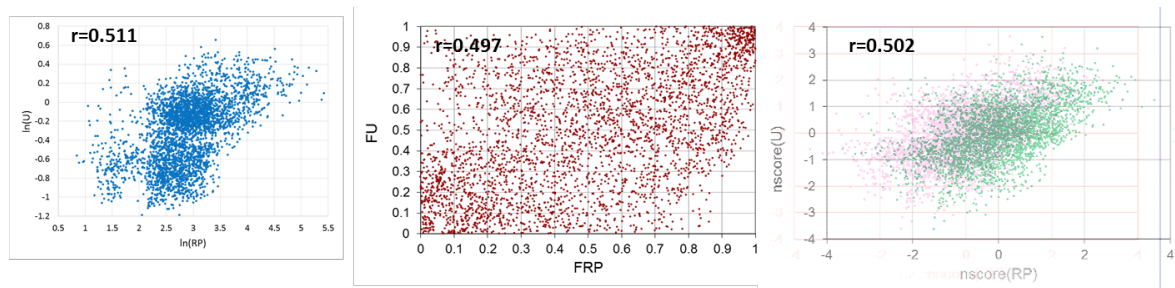


Figure 53: Association between uranium in soil and the GRP. Data points represent 10 km × 10 km cell means, F the distributions of the quantities. r - Pearson correlation. (From [Bossew et al. 2016a](#))

6 Summary, conclusions and recommendations

Action 4.3.1 – Estimation of RPA and classification uncertainty

For the last more than ten years, a variety of operational RPA definitions (i.e., transposition of the generic definition of the BSS into a “formula” to work with in regulatory practice) and estimation methods has been developed. The methods were reviewed and a systematic attempted. Both definitions and estimation models depend on available data: which quantity (IRC, GRP, geology, ambient dose rate, etc.), how well is the domain covered whose RPA status should be assessed, which steps of modelling are required, given the available data, etc.

The RPA status of an area is the outcome of a classification procedure. In most cases known to us, binary classification is chosen (RPA status = yes / no), but also multinomial classification is not precluded by the BSS (e.g., “low / medium / high”). In any case, the RPA status is an non-continuous, ordered random variable which by its nature has uncertainty. However, different from continuous variables, uncertainty cannot be expressed by standard deviation or confidence interval. Instead, it is quantified as misclassification rate, which has to be assessed by particular statistical procedures. These are introduced and discussed in this action. Misclassification means in practice, that an area declared RPA can, with some probability, be no RPA in reality, and inversely, an area declared non-RPA, can in reality be one. This uncertainty cannot be avoided by nature of statistics, but it should be assessed as part of quality assurance. It may be that due the statistical abstractness of the matter, classification QA has been neglected in the past. Therefore, the main recommendation of this action is that if RPA delineation is issued, classification uncertainty should be given attention.

Sources of RPA uncertainty are, apart from the classification procedure, various types of uncertainty of input data, related to the survey design, observation (measurement) uncertainty and model uncertainty, if applicable. To elucidate the subject, practical examples and simulation exercises are presented.

Future research:

Different approaches of RPA definition and estimation are used across Europe, but the legal process laying down a certain definition is not yet finished. For some, uncertainty assessment is not yet clear. A final assessment is therefore not yet possible. Implementation will remain on the agenda.

Action 4.3.2 – Application of retrospective Rn measurements to RPA assessment

Since 2015 the CD/DVD method was used to identify and study RPAs (Pressyanov et al., 2019a). Within MetroRADON project CDs were exposed at Saelices and Chico laboratory under highly variable conditions and the results were published (Pressyanov et al., 2019b) and the paper is annexed to this report (Annex 2). A novel DVDs-based version of the CD/DVD method was developed, with increased sensitivity and compensated temperature influence, suitable for wide range of applications, incl. for radon in soil-gas. Overall, the new results provide strong support to conclude that the CD/DVD method provide reliable results even at extreme conditions and can be used for identification of RPAs.

Action 4.3.3 – RPA classification based on extremes

Several methods have already been developed to map RPAs. Generally, this concerns areas with a significant proportion of indoor radon concentrations exceeding a reference level of a few hundreds of Bq/m³ (maximum

300 Bq/m³ as given by the EU-BSS). A complementary approach was tested to focus on the identification of areas that could be concerned by a significant proportion of dwellings with very high indoor radon concentrations of several thousands of Bq/m³. This method was tested in France and Spain, where such cases occur regionally. It was based on the analysis of available quantities such as the geogenic radon potential, measurements of indoor radon concentration, dwellings characteristics etc., complemented by statistical modelling. The results provide first useful elements to target areas where more precise studies are needed, i.e. to acquire more indoor radon data and the characteristics of buildings associated with the measurements. An analysis of both geological features and building characteristics (mainly the interface between the soil and the building, the building materials, ventilation systems etc.) need to be realized to identify the best indicators of highest indoor radon values. Such a method would allow developing specific prevention (communication and construction rules for new buildings) and remediation actions in heavily affected regions to significantly reduce the exposure in buildings.

Action 4.3.4 – Geogenic Radon Hazard Index (GRHI)

The GRHI can be understood as a generalized complement and extension to the geogenic Radon Potential (GRP) to characterize susceptibility of a location to geogenic radon, as one important control of indoor Rn. The GRHI is more flexible and can deal with data reality which usual GRP definitions cannot handle. Its main application is thought to be large-scale mapping, i.e. on European scale, in contrast to small-scale characterization e.g. of building sites or medium-scale national maps, whose objective is supporting legislative and administrative implementation of the tasks posed by the European Basic Safety Standards (BSS).

Previously existing GRHI attempts were evaluated and “taxonomies” of concepts and estimation methods established. Different concepts correspond to different objectives, while methodology is adapted to availability of data and technical complication which shall be allowed. In course of the project, new methods were conceived and tried. Proposals of Europe wide GRHI maps are presented.

The step to a European map of geogenic radon priority areas (GRPA) is technically easy, but we refrained from doing it within Metro Radon. The reason is that currently the subject is politically delicate. Many countries are still busy in implementing the BSS, which requires extensive legislative and administrative work. Being confronted with a European-scale RPA map may be misinterpreted as being overrun or substituted by a European project, which Metro Radon indeed is.

Once the situation has settled and legislation inclusive sublegislation (into which concrete RPA definitions are confined in many cases) is in place, work on a European RPA map will be resumed.

Future research:

Different approaches will have to be refined and evaluated comparatively. Uncertainty budgeting of highly aggregated quantities like the GRHI is difficult and is yet to be tackled.

7 References

- Barnet, I., Pacherová, P., Neznal, M., Neznal, M. (2008): Radon in geological environment - Czech experience. CGS Special Papers 19. 72 p. Czech Geological Survey. Praha. ISBN 978-80-7075-707-9.
- Ben-Gal I. (2005): Outlier detection, In: Maimon O. and Rockach L. (Editors.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, Kluwer Academic Publishers. <http://www.eng.tau.ac.il/~bengal/outlier.pdf> (6 Oct 2018)
- Bochicchio, F., (2005). Radon epidemiology and nuclear track detectors: methods, results and perspectives. *Radiat. Meas.* 40, 177–190.
- Bochicchio, F., (2009) Results of the first 5 years of a study on year-to-year variations of radon concentration in Italian dwellings, *Radiation Measurements* 44 1064–1068.
- Borgoni R., De Francesco D., De Bartolo D., Tzavidis N., Hierarchical modeling of indoor radon concentration: how much do geology and building factors matter? *J. Environmental Radioactivity* 178 (2014), pp. 227 – 237. <http://dx.doi.org/10.1016/j.jenvrad.2014.08.022>
- Bossey P. (2010): Radon: Exploring the Log-normal Mystery. *JER* 101 (10), 826 - 834,. <http://dx.doi.org/10.1016/j.jenvrad.2010.05.005>
- Bossey P. (2010b): Uncertainty in determining the extension of anomalous zones of spatial environmental variables. Presentation & article, *Accuracy 2010*, Leicester, UK, 20 – 23 July 2010. In N. J. Tate and P. F. Fisher (eds.): *Accuracy 2010, Proc. of the 9th Intl. Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. University of Leicester / Int. Spatial Accuracy Research Association; pp. 281 – 284.
- Bossey P. (2017): Determination of radon priority areas – a classification problem. Presentation, *IWEANR 2017, 2nd International Workshop on the European Atlas of Natural Radiation*. Verbania, Italy, 6 – 9 Nov 2017
- Bossey P., Kuča, Jan Helebrant (2020): Mean ambient dose rate in various cities, inferred from Safecast data. Subm., *J. Environm. Radioactivity*.
- Bossey P., Stochastic dependence of Rn-related quantities. *FERAS 2012*, 2-5. Sept. 2012, Cluj-Napoca, Romania; *Romanian J. Physics* 68 (2013), suppl., S44; www.nipne.ro/rjp/2013_58_Suppl.html (acc. 13 May 2018)
- Bossey P. (2013): Hot spots as random objects. *Radiation Emergency Medicine* 2 (1), 35 - 42.
- Bossey P., A radon risk map of Germany based on the geogenic radon potential. In: E. Pardo-Igúzquiza et al. (eds.): *Mathematics of Planet Earth, Lecture Notes in Earth System Sciences* (2013), pp. 527 - 531, doi:10.1007/978-3-642-32408-6_115, Springer.
- Bossey P., Determination of radon prone areas by optimized binary classification. *J. Environmental Radioactivity* 129 (2014), pp. 121 – 132. <http://dx.doi.org/10.1016/j.jenvrad.2013.12.015>

Bossew, P., Tollefsen, T., Cinelli, G., Gruber, V., and De Cort, M. Status of the European Atlas of Natural Radiation. *Radiation Protection Dosimetry*. Volume 167, Issue 1-3, 1 November 2015, Pages 29–36. <https://doi.org/10.1093/rpd/ncv216>

Bossew P. (2015): Mapping the Geogenic Radon Potential and Estimation of Radon Prone Areas in Germany. *Radiation Emergency Medicine* 4, 2, pp. 13 - 20. http://crss.hirosaki-u.ac.jp/rem_archive/rem4-2 (acc. 13 May 2018)

Bossew P., Local probability of indoor radon concentration to exceed a threshold, estimated from the geogenic radon potential. *Nuclear Technology & Radiation Protection* 32 (2017), 1, pp. 70 – 76; http://ntrp.vinca.rs/2017_1/Contents2017_1.htm (acc. 13 May 2018)

Bossew P. (2017): Determination of radon priority areas – a classification problem. Presentation, IWEANR 2017, 2nd International Workshop on the European Atlas of Natural Radiation. Verbania, Italy, 6 – 9 Nov 2017

Bossew P. (2018a): Radon priority areas – definition, estimation and uncertainty. *Nuclear Technology & Radiation Protection* 33 (3), 286 - 292; <http://doi.org/10.2298/NTRP180515011B>

Bossew P. (2018b): Radon priority areas – definition, estimation and uncertainty. Pres., geoENV-12, Belfast 3-6 July 2018.

Bossew P. (2018c): Estimation of Radon Priority Areas – sources of error and uncertainty. Workshop, GARRM; Geological Aspects of Radon Risk Mapping, Prague, Czech Republic, 18 - 20 September 2018

Bossew P. (2018d): Radon priority areas as random objects. Pres., IAMG 2018, 2 - 8 September 2018, Olomouc, Czech Republic

Bossew P., Cinelli G., Tollefsen T., De Cort M. (2016a): Towards a multivariate geogenic radon hazard index. Presentation, V. Terrestrial Radionuclides in Environment International Conference on Environmental Protection / VIII. Hungarian Radon Forum and Radon In Environment, Veszprém, Hungary, 17 – 20 May 2016.

Bossew P., Cinelli G., Tollefsen T., De Cort M. (2016b): Towards a multivariate geogenic radon hazard index. Presentation, 8th Conference on Protection against Radon at Home and at Work (8th Radon conference) & 13th International Workshop on the Geological Aspects of Radon Risk Mapping (GARRM 13th), Prague, Czech Republic, 12-16th of September 2016.

Bossew P., Cinelli G., Ciotoli G., Crowley Q.G., De Cort M., Elío Medina J., Gruber V., Petermann E., Tollefsen T. (2020): Development of a Geogenic Radon Hazard Index—Concept, History, Experiences. *Int. J. Environ. Res. Public Health* 2020, 17(11), 4134; <https://doi.org/10.3390/ijerph17114134>

Burke, Ó., Murphy, P., (2011). The use of volunteer radon measurements for radon mapping purposes: an examination of sampling bias issues. *J. Radiol. Prot.* 31, 319–328

Chao, C., Tung, T. and Burnett J., (1997) Influence of Ventilation on Indoor Radon Level, *Building and Environment* 32,6, 527-534.

Cinelli, G., Tollefsen T., Bossew P., Gruber V., Bogucarskis K., De Felice L., De Cort M. (2019a): Digital version of the European Atlas of natural radiation, *Journal of Environmental Radioactivity* 196 (2019), 240 - 252, <https://doi.org/10.1016/j.jenvrad.2018.02.008> ; Online version: <https://remon.jrc.ec.europa.eu/About/Atlas-of-Natural-Radiation> (acc. 13 May 2018)

Cinelli G. et al.(2019b): European Commission, Joint Research Centre – Cinelli, G., De Cort, M. & Tollefsen, T. (Eds.): *European Atlas of Natural Radiation*, Publication Office of the European Union, Luxembourg, 2019. ISBN 978-92-76-08259-0, doi:10.2760/520053, Catalogue number KJ-02-19-425-EN-C, EUR 19425 EN. Printed by Bietlot in Belgium 2019 – 190 pp.

Ciotoli GC., Bossew P., Finoia M.G. (2017c): A preliminary exercise to derive the map of potential radon release at European scale. IWEANR 2017, 2nd International Workshop on the European Atlas of Natural Radiation - Verbania, Italy, 6 – 9 November 2017.

CSN (2017a). FDE-02.17 Cartografía del potencial de radón de España [Cartography of radon potential in Spain]. www.csn.es/documents/10182/914801/FDE-02.17%20Cartograf%C3%ADa%20del%20potencial%20de%20rad%C3%B3n%20de%20Espa%C3%B1a

CSN (2017b). Mapa del potencial de radón en España [Spanish radon potential map]. www.csn.es/mapa-del-potencial-de-radon-en-espana

CSN (2019). INT-04.41 Cartografía del potencial de radón de España [Cartography of radon potential in Spain]

CSN (2019). Colección Informes Técnicos 51.2019 Cartografía del potencial de radón de España [Cartography of radon potential in Spain]. <https://www.csn.es/-/el-csn-publica-la-cartografia-del-potencial-de-radon-de-espan-1>

Darby, S., Whitley, E., Silcocks, P., Thakrar, B., Green, M., Lomas, P., Miles, J., Reeves, G., Fearn, T., Doll, R., (1998). Risk of lung cancer associated with residential radon exposure in south-west England: a case–control study. *Br. J. Cancer* 78, 394–408.

Darby, S., Hill, D., Auvinen, A., Barros-Dios, J.M., Baysson, H., Bochicchio, F., Deo, H., Falk, R., Forastiere, F., Hakama, M., Heid, I., Kreienbrock, L., Kreuzer, M., Lagarde, F., Mäkeläinen, I., Muirhead, C., Oberaigner, W., Pershagen, G., Ruano- Ravina, A., Ruosteenoja, E., Schaffrath Rosario, A., Tirmarche, M., Tomasek, L., Whitley, E., Wichmann, H.E., Doll, R., (2005). Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *Br. Med. J.* 330, 218–223.

Darby, S., Hill, D., Deo, H., Auvinen, A., Barros-Dios, J.M., Baysson, H., Bochicchio, F., Doll, R., Falk, R., Farchi, S., Figueiras, A., Hakama, M., Heid, I., Hunter, N., Kreienbrock, L., Kreuzer, M., Lagarde, F., Mäkeläinen, I., Muirhead, C., Oberaigner, W., Pershagen, G., Ruosteenoja, E., Schaffrath Rosario, A., Tirmarche, M., Tomasek, L., Whitley, E., Wichmann, H.E., (2006). Residential radon and lung cancer: detailed results of a collaborative analysis of individual data on 7148 subjects with lung cancer and 14208 subjects without lung cancer from 13 epidemiological studies in Europe. *Scand. J. Work Environ. Health* 32 (Suppl.1), 1–80.

Dehandschutter B., Poffijn A., Ciotoli G. and Klerkx J. (2008): Radon Mapping using indoor, soil gas and geological data. Presentation, IGC33, Oslo, August 2008.

Demoury C., Ielsch G., Hemon D., Laurent O., Laurier D., Clavel J., Guillevic J. (2013). A statistical evaluation of the influence of housing characteristics and geogenic radon potential on indoor radon concentrations in France. *Journal of Environmental Radioactivity* 126 (2013) 216–225.

Dubois, G., Bossew, P., Tollefsen, T., De Cort, M., 2010. First steps towards a European atlas of natural radiation: Status of the European indoor radon map. *J. Environ. Radioact.* 101, 786–798

EC (2014): European Council: Council Directive 2013/59/Euratom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation. *Official Journal of the European Union.* 57(L13), 1 – 73 (2014) <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2014:013:FULL&from=EN> (accessed 25 June 2017)

EC (2019), European Commission, Joint Research Centre – Cinelli, G., De Cort, M. & Tollefsen, T. (Eds.), *European Atlas of Natural Radiation*, Publication Office of the European Union, Luxembourg, 2019. ISBN 978-92-76-08259-0, doi:10.2760/520053.

EC (2020), *Radiation Protection n.193, Radon in workplaces, Implementing the requirements in Council Directive 2013/59/Euratom.* doi:10.2833/552398

Elío J., Crowley Q., Scanlon R., Hodgson J., Long S. (2018b): Logistic regression model for detecting radon prone areas in Ireland. *Science of the Total Environment* 599–600 (2018), pp. 1317–1329; <http://dx.doi.org/10.1016/j.scitotenv.2017.05.071>

Elío J., Bossew P., Cinelli G., De Cort M., Garcia-Talavera N., Gruber V., Gutiérrez-Villanueva J.L., Iurlaro G., Tollefsen T., Udovicic V. (2019): *Advances in the European Indoor Radon Map, towards a European Indoor Radon Dose Map.* Pres., EGU 2019, Vienna 7-12 April 2019.

Ferreira A., Daraktchieva Z., Beamish D., Kirkwood C., Lister T.R., Cave M., Wragg J., Lee K. (2018): Indoor radon measurements in south west England explained by topsoil and stream sediment geochemistry, airborne gamma-ray spectroscopy and geology. *Journal of Environmental Radioactivity* 181. 152 – 171; <https://doi.org/10.1016/j.jenvrad.2016.05.007>

Friedmann H., Baumgartner A., Gruber V., Kaineder H., Maringer F.J., Ringer W., Seidel C. (2017): The uncertainty in the radon hazard classification of areas as a function of the number of measurements. *Journal of Environmental Radioactivity* 173. 6 – 10; <https://doi.org/10.1016/j.jenvrad.2016.08.011>

García-Talavera M., García-Perez A., Rey C., Ramos I., Mapping radon-prone areas using γ -radiation dose rate and geological information, *J. Radiol. Prot* 33 (2013), 3, pp. 605 – 620. doi:10.1088/0952-4746/33/3/605

Gréau C., Ielsch G., Saâdi Z., Mansouri N., Bertrand C., 2017. Influence of karsts on the radon production and migration (Fourbanne site, French Jura Mountains): analysis of experimental data. *Third East-European Radon Symposium*, May 15-19 2017, Sofia, Bulgaria.

Gruber V., Baumann S., Ringer W., Alber O., Kuchling S., Laubichler C., Schleicher C., An extensive indoor radon measurement campaign to define radon priority areas in Austria. Presentation, 2nd International Workshop on the European Atlas of Natural Radiation (IWEANR), Verbania (Italy), 6 – 9 Nov 2017, <https://1drv.ms/f/s!AsdbTr8KIDo1gXecYWJt0RmGCtqu> (acc. 7 June 2018)

IAEA, (2013). National and Regional Surveys of Radon Concentration in Dwellings, Review of Methodology and Measurement Techniques, Analytical Quality in Nuclear Application IAEA/AQ/33; https://www-pub.iaea.org/MTCD/Publications/PDF/IAEA-AQ-33_web.pdf

IGME (2019) <http://info.igme.es/cartografia/oneGeology.asp?mapa=oneGeologyEuropa>

Jensen, C.L., Strand, T., Ramberg, G.B., Ruden, L., Ånestad, K., (2004). The Norwegian Radon Mapping and Remediation Program. 11th Int. Congr. Int. Radiat. Prot. Assoc. 1–10.

Kemski J., Siehl A., Stegemann R. and Valdivia-Manchego M. (2001): Mapping the geogenic radon potential in Germany. *The Science of the Total Environment* 272 (1 – 3), 217 - 230.

Kemski, J., Klingel, R., Siehl, A., Valdivia-Manchego, M. (2009): From radon hazard to risk prediction - based on geological maps, soil gas and indoor measurements in Germany. *Environmental Geology* 56, 1269 – 1279.

Kropat G., Bochud F., Jaboyedoff M, Laedermann JP, Murith C., Palacios (Gruson) M., Baechler S., Improved predictive mapping of indoor radon concentrations using ensemble regression trees based on automatic clustering of geological units. *Journal of Environmental Radioactivity* 147 (2015), pp. 51 – 62; <http://dx.doi.org/10.1016/j.jenvrad.2015.05.006>

Long Way (2011): The European Geogenic Radon Map; working document, version 1. JRC.

Lubin, J.H., Wang, Z.Y., Wang, L.D., Boice Jr., J.D., Cui, H.X., Zhang, S.R., Conrath, S., Xia, Y., Shang, B., Cao, J.S., Kleinerman, R.A., (2005). Adjusting lung cancer risk for temporal and spatial variations in radon concentration in dwellings in Gansu province, China. *Radiat. Res.* 163, 571–579.

Mansouri N., Gréau C., Ielsch G., Saâdi Z., Bertrand C., 2018. Radon production and migration in karstic environment: experimental data and numerical modelling (Fourbanne site, French Jura Mountains). 14th INTERNATIONAL WORKSHOP GARRM (on the GEOLOGICAL ASPECTS OF RADON RISK MAPPING), September 18-20 2018, Prague, Czech Republic.

Mitev K., Georgiev S., Dimitrova I., Pressyanov D. Radon-222 in soil-gas measurements by compact discs. Comparison to diffusion chamber measurements. *Radiat. Prot. Dosim.* 181 (2018) 38-41.

Maier, M., Schack-Kirchner, H. Using the gradient method to determine soil gas flux: a review. *Agric. For. Meteorol.* 192–193 (2014) 78–95.

Mitev K., Cassette P., Georgiev S., Dimitrova I., Sabot B., Boshkova, T., Tartès I., Pressyanov D. Determination of ²²²Rn absorption properties of polycarbonate foils by liquid scintillation counting. Application to ²²²Rn measurements. *Appl. Radiat. Isotopes* 109 (2016) 270–275.

Mitev K., Dutsov Ch., Georgiev S., Boshkova T., Pressyanov D., “Unperturbed, high spatial resolution measurement of Radon-222 in soil-gas depth profile”, *Journal of environmental radioactivity*, 196 (2019) 253-258.

Neznan M., Neznan M., Matolin M., Barnet I., Miksova, J. (2004). The new method for assessing the radon risk of building sites. *Czech Geol. Survey Special Papers*, 16, Czech Geol. Survey, Prague, Section 2.5.1; <http://www.radon-vos.cz/pdf/metodika.pdf> (acc. 13 May 2018)

NRPA (Norwegian Radiation Protection Authority (1996).), *Methods for measuring radon in indoor air and construction sites*, NRPA Radiation Protection series no. 3, Österaas.

Pantelić G. et al., *Qualitative overview of indoor radon surveys in Europe*, *Journal of Environmental Radioactivity* 204, (2019), 163-174, doi: 10.1016/j.jenvrad.2019.04.010

Pasculli A., Palemi S., Sarra A., Piacentini T., Miccadei E., *A modelling methodology for the analysis of radon potential based on environmental geology and geographically weighted regression. Environmental Modelling & Software* 54 (2014), pp. 165 – 181; <http://dx.doi.org/10.1016/j.envsoft.2014.01.006>

Petermann E., Bossew P. (2019): *Mapping the geogenic radon potential considering multiple environmental covariates by applying multivariate adaptive regression splines (MARS)*. EGU, Vienna, 8 - 12 April 2019.

Pressyanov D., Dimitrova I., Mitev K., Georgiev S., Dimitrov D. *Identifying radon priority areas and dwellings with radon exceedances in Bulgaria using stored CD/DVDs*. *J. Env. Radioact.* 196 (2019a) 274-280.

Pressyanov D., Poncela L. S., Georgiev S., Dimitrova I., Mitev K., Sainz C., Fuente I., Rabago D. *Testing and Calibration of CDs as Radon Detectors at Highly Variable Radon Concentrations and Temperatures*. *Int. J. Env. Res. Public Health* 16 (2019b) 3038.

Pressyanov D. *Nuclear tracks in polycarbonates with high radon absorption ability: Opportunities for measuring 222Rn*. Chapter 4 IN: *Nuclear Track Detectors: Design, Methods and Applications*. Nova Science Publishers, Inc., 2010, New York, ISBN: 978-1-60876-826-4.

Pressyanov D., Dimitrova I., Dimitrov D. *High sensitivity passive radon detector for measuring radon in low-background underground nuclear/particle physics laboratories*. *IEEE-NSS/MIC Conference Records* (2018).

Pressyanov, D. (2019a), *Bulg. Pat. Appl. Reg. Nr. 112897, priority: 19.03.2019 (assignee: Sofia University “St. Kliment Ohridski”)*, *WIPO Appl. Reg. Nr. PCT/BG2020/000003*.

Pressyanov D. (2019b), *Highly sensitive passive detectors for short-term pre- and post- mitigation measurements*. *Proc. 33rd AARST International Radon Symposium, Denver, CO, 9-11 September 2019*.

Rabago D., Quindos L.S., Quindos J., Fernandez E., Fernandez A., Quindos L., Celaya S., Fuente I., Sainz C., Intercomparison of indoor radon and geogenic radon measurements under field conditions, MetroRadon report, WP.3.3.3., February, 2019.

Radolić, V., Poje Sovilj, M., Stanić, D., Miklavčić, I., (2017) Radon in soil gas and constructed geogenic radon potential in Croatia, Proceedings of XXIX SYMPOSIUM DZZSCG Srebrno jezero, 192-199, 27- 29. September 2017

Sainz-Fernandez, C., Fernandez-Villar, A., Fuente-Merino, I., Gutierrez-Villanueva, J. L., Martin-Matarranz, J. L., Garcia-Talavera, M., ... & Quindós-Poncela, L. S. (2014). The Spanish indoor radon mapping strategy. Radiation protection dosimetry, 162(1-2), 58-62. <https://doi.org/10.1093/rpd/ncu218>

Sainz, C., Quindos, L. Q., Fernández, A., Fuente, I. , Gutierrez-Villanueva, J. L., Celaya, S. , ... & García M. (2017). Spanish experience on the design of radon surveys based on the use of geogenic information. Journal of environmental radioactivity, 166, 390-397. <https://doi.org/10.1016/j.jenvrad.2016.07.007>

Sarra A., Fontanella L., Valentini Pasquale, Palermi S., Quantile regression and Bayesian cluster detection to identify radon prone areas. Journal of Environmental Radioactivity 164 (2016), pp. 354 – 364; <https://doi.org/10.1016/j.jenvrad.2016.06.014>

Smethurst M. A., Strand T., Sundal A. V. and Rudjord A. L. (2008): A novel approach to large scale radon hazard evaluation utilizing indoor radon concentrations, airborne gamma ray spectrometry and geological mapping. Science of the Total Environment 407 (1), 379 – 393.

Steck, D.J. (2009) Annual average indoor radon variation in two decades, Health Phys. 96(1): 37– 47.

Timkova J., Fojtikova I., Pacherova P., Bagged neural network model for prediction of the mean indoor radon concentration in the municipalities in Czech Republic. Journal of Environmental Radioactivity 166 (2017), 2, pp. 398 – 402; <http://dx.doi.org/10.1016/j.jenvrad.2016.07.008>

Tokonami, S., Takahashi, H., Kobayashi, Y., Zhuo, W., Hulber, E., (2005). Up-to-date radon-thoron discriminative detector for a large scale survey, Rev. Sci. Instrum. 76 11113505–113509.

Tollefsen, T., Cinelli, G., Bossew, P., Gruber, V., & De Cort, M. (2014). From the European indoor radon map towards an atlas of natural radiation. Radiation protection dosimetry, 162(1-2), 129-134. <https://doi.org/10.1093/rpd/ncu244>

WHO (2009). Zeeb, Hajo, Shannoun, Ferid & World Health Organization. (2009). WHO handbook on indoor radon: a public health perspective. World Health Organization. <http://www.who.int/iris/handle/10665/44149>

Yarmoshenko, I., (2016) Variance of indoor radon concentration: Major influencing factors, Science of the Total Environment 541, 155-160

Yarmoshenko, I., Vasilyev, A., Onishchenko, A., Kiselev, S., Zhukovsky, M., (2014). Indoor radon problem in energy efficient multi-storey buildings. Radiat. Prot. Dosim. 160, 1-3, 53-56.

Zhukovski, M. et al., (2010). Radon measurements—discussion of error estimates for selected methods, *Applied Radiation and Isotopes* 68, 816–820

Zivanovic, M. (2016). *Optimisation of Indoor Radon Concentration Measurements by Means of Charcoal Canisters*, University of Belgrade, Doctoral thesis.

Annex A: Approximation formulae and bias correction

A.1 Approximations of the cumulative normal and t-distributions

For practical approximation, cumulative normal and t-distributions have to be approximated. Several approximations are known for $\Phi(x)$. Some newer ones are given in [Vazquez Leal et al. \(2012\)](#) and [Sosanzo and Epure \(2014\)](#). From the latter we chose because of its low remaining error,

$$\Phi(x) \approx \frac{1}{2} + \frac{1}{2} \sqrt{1 - \exp\left(\frac{-1.2735457x^2 - 0.0743968x^4}{2 + 0.1480931x^2 + 0.0002580x^4}\right)}$$

for $x \geq 0$ and $\Phi(-x) = 1 - \Phi(x)$.

A simple formula for the percentiles of the t-distribution are given in [Li and De Moor \(1999\)](#):

$$F_t(x; v) \approx \Phi(\lambda x), \quad \lambda := (4v + x^2 - 1) / (4v + 2x^2)$$

v - degrees of freedom.

A.2 Bias correction of the standard deviation

The common estimator of the SD with Bessel correction, $SD' = (SX^2 - SX^2/n) / (n-1)$, is biased. This can be corrected for *normal populations*, e.g. [Andersen et al. \(2001\)](#) (eq.4):

$SD \rightarrow SD \cdot SD_{corr}(n)$,

$$SD_{corr}(n) = \sqrt{\frac{n-1}{2} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}}$$

By virtue of the property of the gamma function, $\Gamma(x+1) = x \Gamma(x)$, one finds

$$SD_{corr}(n+1) = \sqrt{\frac{n}{n-1}} \frac{1}{SD_{corr}(n)},$$
 which allows calculating the correction factor as recurrence

without need to evaluate the gamma function, except the initial $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(1) = 1$.

In particular for low n , this correction performs much better than the well-known, in which the Bessel factor $(n-1)$ is replaced by $(n-1.5)$, which is however equally useful for larger n .

A.3 Bias correction of the coefficient of variation

The estimator of the CV, $CV' = SD'/AM'$ is biased. For *normal populations*, according [Beigy \(2019\)](#), an approximate bias correction is

$CV \rightarrow CV \cdot CV_{corr}(n)$,

$$CV_{corr}(n) = 1 - \frac{1}{4(n-1)} + \frac{CV^2}{n} + \frac{1}{2(n-1)^2}$$

to be compared with the well-known approximation, $CV_{corr}(n) = 1 + 1/(4n)$ (e.g. [Haldane 1955](#)). For non-normal populations, as in our case, both perform poorly, but better than without any correction.

Also for non-normal cases, corrections exist, but are very complicated and require higher moments of the distributions, not available here.



Article

Testing and Calibration of CDs as Radon Detectors at Highly Variable Radon Concentrations and Temperatures

Dobromir Pressyanov ^{1,*}, Luis Santiago Quindos Poncela ², Strahil Georgiev ¹,
Ivelina Dimitrova ¹, Krasimir Mitev ¹, Carlos Sainz ², Ismael Fuente ² and Daniel Rabago ²

¹ Faculty of Physics, “St. Kliment Ohridski”, Sofia University, 1164 Sofia, Bulgaria

² Radon Group, University of Cantabria, 39005 Santander, Cantabria, Spain

* Correspondence: pressyan@phys.uni-sofia.bg

Received: 23 July 2019; Accepted: 20 August 2019; Published: 22 August 2019



Abstract: The application of the compact disk (CD) method for radon measurements at mines, caves and other workplaces needs testing under highly variable exposure conditions. We present the results from a blind comparison of CDs exposed in the Laboratory of Natural Radiation (Saelices el Chico, Spain). During the exposure the temperature varied from 6.5 to 24.9 °C (average 12.6 °C) and the ²²²Rn activity concentrations varied from <10 Bq m⁻³ to 147 kBq m⁻³. Good correspondence was observed between the integrated ²²²Rn activity concentration determined by the reference instruments in the laboratory (122,500 ± 6100 kBq h m⁻³) and that assessed by analysis of the CDs at a depth 80 μm beneath the front surface (118,000 ± 12,000 kBq h m⁻³) and at a depth of 120 μm (106,000 ± 12,000 kBq h m⁻³). The theoretical modeling of the CD response under variable temperature and radon concentration suggested that the small bias is probably due to the time variation of the calibration factor because of the time variations of the temperature.

Keywords: radon; CD-method; blind comparison; extremely variable concentrations; unstable temperature

1. Introduction

The compact disk (CD) method for radon measurements was proposed in 2001 [1], initially as a method for retrospective measurements. It is based on radon absorption in the polycarbonate material of which CDs and digital versatile disks (DVDs) are made and analysis of alpha tracks at a certain depth beneath the disk surface (higher than 76 μm, usually about 80 μm) as described elsewhere [1,2]. Since 2001 the method has been thoroughly studied in the laboratory and in indoor radon surveys [2]. The temperature is the only identified environmental factor to have an effect on the results, and it can be corrected for a posteriori [2]. Past comparisons made indoors showed good correspondence between the CD method and conventional measurements [2]. However, new applications of this method (e.g., for measurements in mines [3] or caves) require tests of the method under more extreme conditions than those typically found indoors. The comparison of results obtained by CDs under extremely high variations in the radon activity concentration and variable temperature with parallel measurements by reference radon monitors can test the potential of the method for applications at peculiar working places or environmental conditions. Here we describe the results of a blind comparison of radon measurements by CDs and continuous radon monitors, which was carried out in the Laboratory of Natural Radiation (LNR) located in Saelices el Chico (Salamanca, Spain). This is a unique laboratory facility where radon activity concentration can vary by orders of magnitude and in which continuous follow-up of radon activity concentrations and environmental parameters (temperature, humidity, pressure) is made by reference instruments [4].

2. Materials and Methods

The LNR was set up and handled by the University of Cantabria (Figure 1a). It is located inside the former uranium mine of Saelices el Chico (Salamanca, Spain) managed by the Spanish National Uranium Company ENUSA, currently under reclamation process. It has been used for calibration and testing of instruments and detectors for the measurement of natural radiation under environmental conditions. The ground floor has two spaces designed as radon chambers (Room 1 and Room 2) with approximately 45 m³ volume each. Room 1 has no direct connection to the exterior while Room 2 has an artificial ventilation system installed but switched off during the experiment. The radon source is the uranium mine underground soil which has a high radium content.



Figure 1. (a) Photos of the Laboratory of Natural Radiation (b) and the place in Room 1 where the experimental exposure was carried out.

During the blind test a set of 10 CDs (verbatim, recordable) were exposed in Room 1 (Figure 1b) for 171 days from 29 September 2017 to 19 March 2018. The disks were exposed in their “jewel cases” (the protective boxes in which CDs or DVDs are usually stored). The jewel cases are not hermetic, and radon penetrates freely inside them. It has been experimentally proved that CDs exposed to ²²²Rn bare and in their jewel cases give statistically identical results [1,5]. The radon activity concentration and some major environmental parameters were followed continuously (every 10 min) by a reference instrument AlphaGUARD PQ2000 PRO (Saphymo/Bertin Instruments, Frankfurt am Main, Germany) traceable to another AlphaGUARD unit calibrated in the Physikalisch-Technische Bundesanstalt (PTB). The reference instrument was verified at the LaRUC’s radon chamber (Laboratory of Environmental Radioactivity, University of Cantabria) [6].

The average temperature during the exposure was 12.6 °C (range 6.5–24.9 °C, Figure 2a), the average pressure was 944 hPa (903.6–960.2 hPa) and the average relative humidity as 64.4% (27.5%–97.4%). The radon activity concentration varied by orders of magnitude: from <10 to 147,000 Bq m⁻³ (Figure 2b). The variations in radon concentration levels were irregular, while those in the temperature showed a systematic pattern modified by irregular fluctuations. There was a weak negative correlation between the temperature and ²²²Rn activity concentration (Figure 3). However, at any temperature ²²²Rn levels can vary in a wide range, therefore the temperature variations are not considered as the primary cause for ²²²Rn variations. The ²²²Rn activity concentrations measured by the reference monitor in the LNR were exchanged with the Sofia University team once the final results were obtained (after the CDs calibration, etching and analysis). The temperature variations were shared previously as they were needed to calibrate the CDs at the mean temperature.

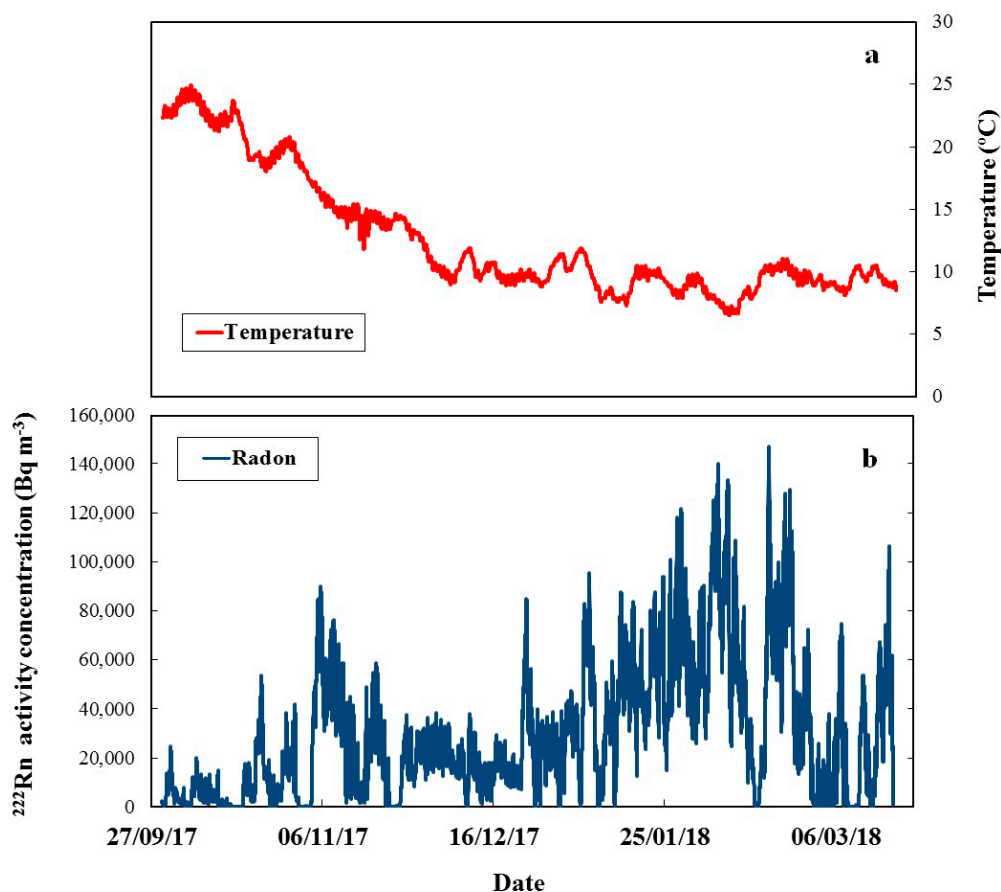


Figure 2. (a) Temperature during the exposure. The average temperature was 12.6 °C; (b) ^{222}Rn activity concentration during the exposure. The concentration varied from <10 to 147,000 Bq m $^{-3}$.

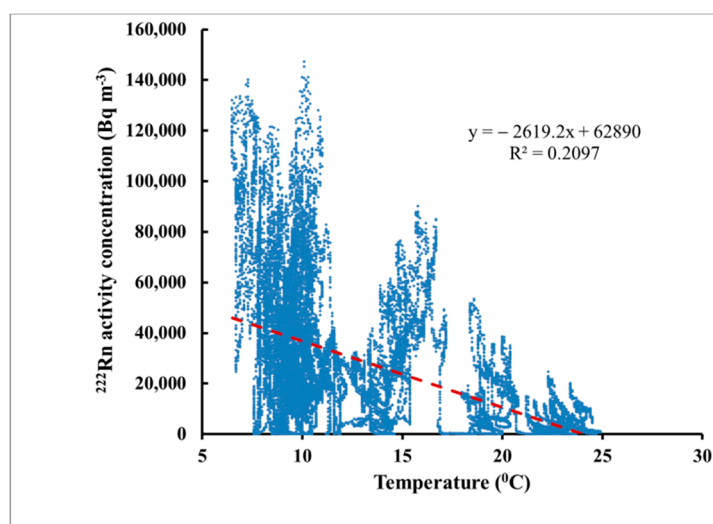


Figure 3. The correlation between the temperature and the ^{222}Rn activity concentration. The statistical analysis made by PAST statistical package [7] showed a statistically significant (at 95% level of confidence) negative correlation.

After exposure the disks were processed at Sofia University, Bulgaria. The processing starts with chemical pre-etching, in order to reach the desired depths (in this case 80 μm and 120 μm) by chemical removal of the surface layer. After that, electrochemical etching is applied and the tracks are counted automatically. The etching procedure is described in detail in [2] and the automatic track counting

by a computer scanner in [8]. The analyzed signal is the net track density (the track density after the background is subtracted). The background of unexposed CDs of the kind used in the experiments was $3.8 \pm 1.3 \text{ cm}^{-2}$.

The calibration of the CDs was carried out at Sofia University, Bulgaria by exposure of identical unexposed disks at reference radon concentrations at the average temperature of the exposure in the LNR (12.6 °C). The calibration exposure was done using the calibration facility described in [9] (Figure 4). The reference concentration was measured by the reference monitor AlphaGUARD PQ2000 PRO (Saphymo/Bertin Instruments, Frankfurt am Main, Germany). The calibration factor ($CF = \text{net track density/radon exposure}$) was determined for two depths beneath the disk surface: 80 μm and 120 μm . The CF values at the average temperature were as follows:

$$CF (80 \mu\text{m}) = 0.00946 \pm 0.00054 \text{ cm}^{-2}/\text{kBq h m}^{-3}$$

$$CF (120 \mu\text{m}) = 0.00286 \pm 0.00024 \text{ cm}^{-2}/\text{kBq h m}^{-3}$$



Figure 4. A photo of the exposure facility [9]. The detectors for calibration are placed in the 50 L exposure box (1) that is placed in the programmable thermostat (2).

Since the track density decreases in depth [8], the CF at temperature 12.6 °C at depth 80 μm is 3.3 times greater than the CF at 120 μm . Analysis at depths greater than 80 μm can be useful when the signal at 80 μm is high and approaches the saturation level. While it is hard to analyze the tracks in a saturated track detector, the CDs give the opportunity to analyze them at a greater depth at which the tracks are less and to ensure quantitative measurements.

3. Results and Discussion

The integrated ^{222}Rn activity concentration (^{222}Rn exposure) was determined by numerical integration of the values of the ^{222}Rn activity concentration measured by the reference continuous monitor. Its value for this experiment was $I = 122,500 \pm 6100 \text{ kBq h m}^{-3}$. The ^{222}Rn exposure by CDs was determined by the net track-density at two depths beneath the CD surface, 80 μm and 120 μm , considering the obtained calibration factors. The results of the blind comparison are illustrated in Figure 5. The individual results for the ^{222}Rn exposure by the single CDs analyzed at 80 μm and 120 μm are shown in Figure 6.

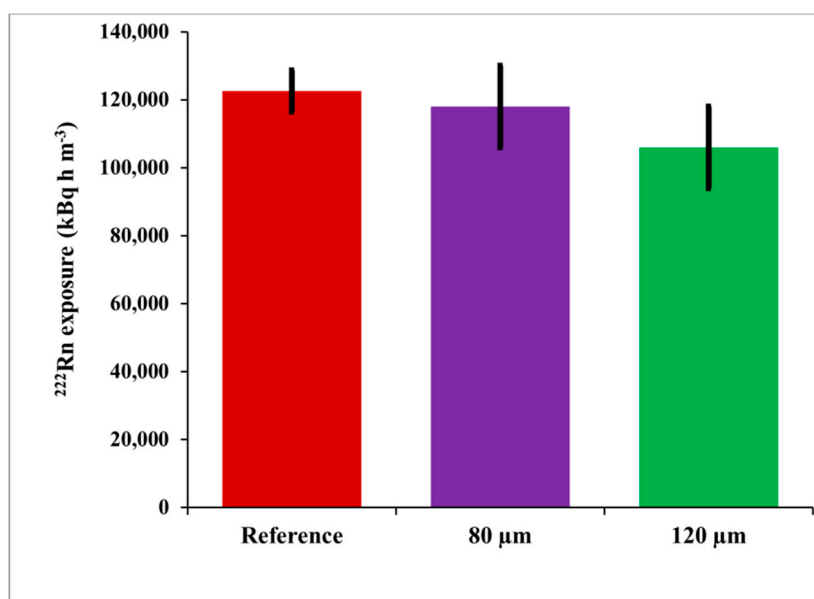


Figure 5. ^{222}Rn exposure assessed by reference measurements, compact disks (CDs) analyzed at 80 μm and 120 μm deep.

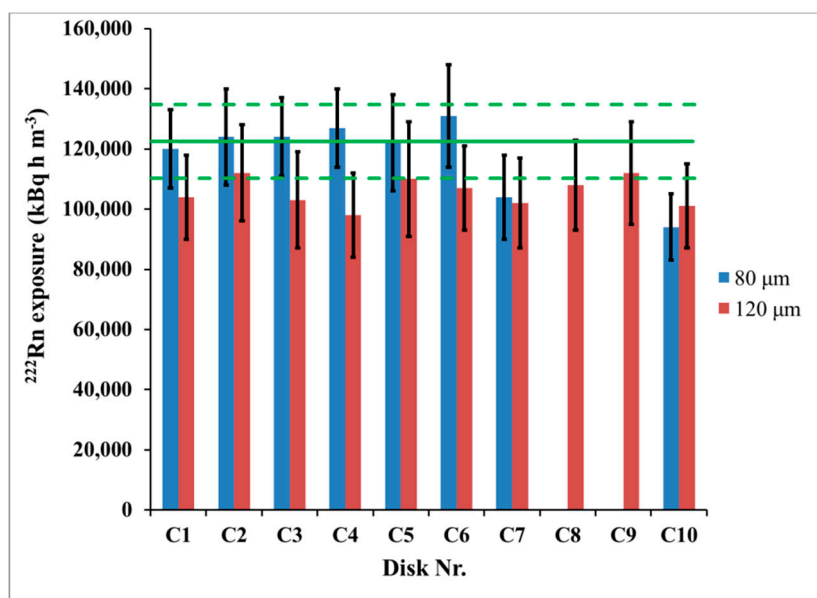


Figure 6. Variation of individual results between disks from one set at depths 80 μm and 120 μm . Disks C8 and C9 were analyzed only at 120 μm . The error bars correspond to the “one sigma” combined uncertainty (counting uncertainty and calibration uncertainty). The horizontal line represents the reference ^{222}Rn exposure and the dashed lines show its 95% confidence interval (“two-sigma” interval).

The differences between the reference activity concentration and that assessed by CDs were 3.7% at 80 μm and 13.5% at 120 μm (Figure 5). The *t*-test [7,10] showed that they are not statistically significant at 95% level of confidence. However, a small and systematic bias was observed at both depths analyzed. Therefore, after the results from the blind comparison became available, we explored potential reasons for such bias. The CD calibration factor depends on the temperature, and the time variations of the temperature may incur bias in the results obtained by using the *CF* value estimated during the calibration exposure at “the average” temperature. To study this bias, theoretical modeling which follows the model described in [11] was employed. In the theoretical model [11] the dependence of the *CF* is modeled analytically and numerically as a function of the temperature within the temperature

interval 5–38 °C. The model [11] considers the radon absorption and the track-etch properties of the polycarbonate material of which the commercial CDs/DVDs are made. The temperature dependence of the $CF(T)$, modeled for the studied temperature interval according to [11] is illustrated in Figure 7 for the two depths at which the signal is analyzed: 80 μm and 120 μm .

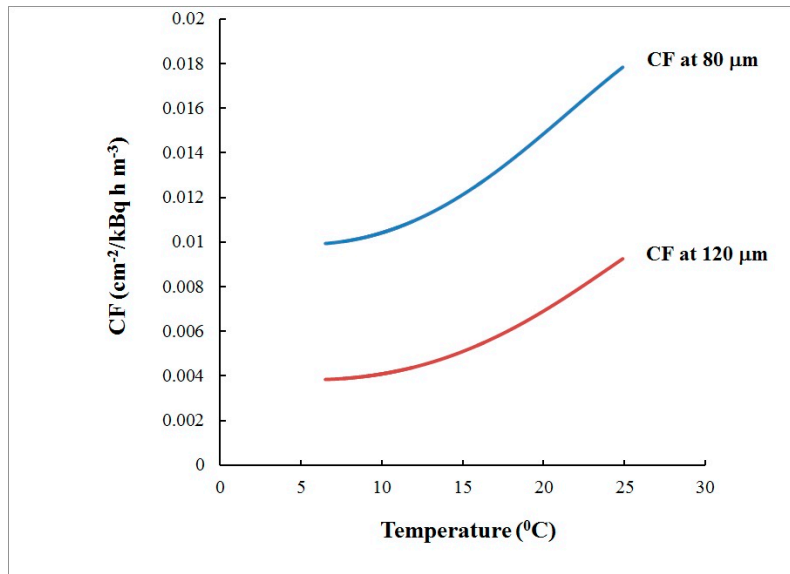


Figure 7. Dependence of the modeled calibration factor [11] on the temperature within 6.5–24.6 °C at depths of 80 μm and 120 μm .

In the real exposure the calibration factor depends on the temperature T , which depends on the time t . By combining the temperature dependence of the $CF(T)$ with the time dependence $T(t)$ of the temperature (Figure 2a), the time dependence of the calibration factor $CF(T(t))$ can be determined (Figure 8). On the other hand, the ^{222}Rn activity concentration $C_A(t)$ also depends on the time (Figure 2b). The “true” calibration factor \overline{CF} is the ratio between the signal n and the ^{222}Rn exposure (I) at the specific exposure conditions (i.e., $n = \overline{CF}I$). Any small time interval dt at which $CF(t)$ and $C_A(t)$ can be considered practically constant contributes to the signal by $dn = CF(t)C_A(t)dt$. Therefore, for the signal one obtains the following expression, used in the modeling below:

$$n = \overline{CF} \cdot I = \overline{CF} \int_0^{t_{exp}} C_A(t) dt = \int_0^{t_{exp}} dn = \int_0^{t_{exp}} CF(t) C_A(t) dt \tag{1}$$

where t_{exp} is the exposure time. The “true” calibration factor \overline{CF} depends on the exposure scenario and may differ from the calibration factor $CF(\overline{T})$ at the average temperature \overline{T} , where:

$$\overline{T} = \frac{1}{t_{exo}} \int_0^{t_{exp}} T(t) dt \tag{2}$$

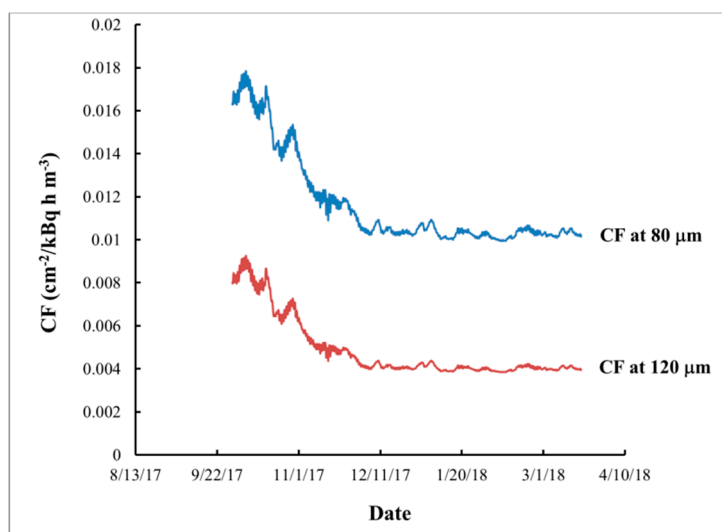


Figure 8. Dependence of the modeled calibration factor [11] on time at depths of 80 μm and 120 μm.

To study the effect of the eventual difference between the “true” calibration factor and that used in the blind comparison (determined in the laboratory and corresponding to the average temperature) a model approach was used. The “true calibration factors” were calculated for the known exposure conditions, by adjusting the calibration factors at the average temperature for the real exposure profile. The obtained results are:

$$\overline{CF} (80 \mu\text{m, true exposure profile}) = 0.966 \cdot CF(80 \mu\text{m, } 12.6 \text{ }^\circ\text{C})$$

$$\overline{CF} (120 \mu\text{m, true exposure profile}) = 0.958 \cdot CF(120 \mu\text{m, } 12.6 \text{ }^\circ\text{C})$$

The results of the integrated ²²²Rn activity concentrations without and with such adjustment are shown in Table 1.

Table 1. Integrated ²²²Rn activity concentration assessed by CDs with tracks analyzed at 80 μm and 120 μm beneath the front surface. The reference exposure was assessed by continuous measurements by a reference instrument AlphaGUARD PQ2000 Pro. CF = net track density/radon exposure.

Scenario	²²² Rn Exposure (kBq h m ⁻³)		
	At 80 μm	At 120 μm	Reference
With CF at 12.6 °C	118,000 ± 12,000	106,000 ± 12,000	122,500 ± 6100
With CF adjusted for the real exposure	122,000 ± 12,000	110,500 ± 12,000	

As seen, adjustment for the real exposure temperature improves the correspondence between the results, making it almost perfect for CDs etched at a depth of 80 μm (deviation reduced from 3.7% to 0.4%). For CDs etched at 120 μm, the deviation between the results and the reference value is reduced from 13.5% to 9.8%. The theoretical modeling revealed that the influence of the temperature variability is greater at a depth of 120 μm and therefore greater temperature bias can be expected. However, there are situations in which the analysis at a greater depth may be preferred. At a depth of 80 μm the “upper limit” of the method (corresponding to track density saturation) is at an integrated ²²²Rn activity concentration about 260,000 kBq h m⁻³ [5]. However, the upper limit can be increased significantly by etching at a greater depth and/or by modifying the etching regime [12]. This adds the possibility to make the upper limit of this method quite greater than that of the conventional radon

detectors. Thereby, the method is applicable for the measurement of very large radon exposures, either for a long exposition time or at very high radon activity concentration.

According to the results from the experimental comparison and theoretical modeling, a possible reason for the bias between the reference value and the CD results in the blind comparison could be the great time variations of the temperature and ^{222}Rn activity concentration. However, this bias appears to be small even under these extreme variations.

4. Conclusions

In this work, a blind test of the CD method for radon measurement under extreme conditions is presented. There is a very good correspondence between the results obtained by CDs and the reference value despite the large variations in the activity concentration of radon and the temperature and the high integrated radon activity concentration. The observed small systematic bias of 3.7% at 80 μm and 13.5% at 120 μm is explained by the significant variability of the temperature and ^{222}Rn concentrations during exposure. In conclusion, when an appropriate temperature correction is applied, the CD method provides a reliable estimate of the integrated radon concentration even under extremely variable conditions. This might be important for the public health at least in two directions: (1) The CD method is usable for retrospective measurements, which are directly related to the radon risk as it is due to the exposure received in the past; (2) since there is a new legislation requiring measurements of radon in workplaces, one can find situations with very high ^{222}Rn levels at which the standard detectors become saturated. By using this new technique, we minimize the probability for this, because the upper limit of the CD method is substantially higher than that of the widely used commercial detectors. Further investigations will focus on the effect of variable temperature at different exposure scenarios.

Author Contributions: Conceptualization, D.P. and L.S.Q.P.; data curation, S.G., I.D. and D.R.; formal analysis, S.G., I.D., I.F. and D.R.; funding acquisition, D.P., L.S.Q.P. and K.M.; investigation, D.P., L.S.Q.P., S.G., K.M., C.S. and D.R.; methodology, I.D., I.F. and D.R.; project administration, D.P.; supervision, D.P., L.S.Q.P. and C.S.; validation, D.P. and L.S.Q.P.; writing—original draft, D.P.; writing—review and editing, L.S.Q.P., S.G., I.D., C.S., I.F. and D.R.

Funding: This research was funded by the European Metrology Programme for Innovation and Research (EMPIR), JRP-Contract 16ENV10 MetroRADON (<http://www.euramet.org>). The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Pressyanov, D.; Buysse, J.; Van Deynse, A.; Poffijn, A.; Meesen, G. Indoor radon detected by compact discs. *Nucl. Instrum. Meth.* **2001**, *457*, 665–666. [[CrossRef](#)]
2. Pressyanov, D.; Mitev, K.; Georgiev, S.; Dimitrova, I. Radon mapping by retrospective measurements—An approach based on CDs/DVDs. *J. Environ. Radioact.* **2010**, *101*, 821–825. [[CrossRef](#)] [[PubMed](#)]
3. Dimitrov, D.; Pressyanov, D. The CD/DVD method as a tool for the health physics service and ventilation diagnostics in underground mines. *Radiat. Prot. Dosim.* **2018**, *180*, 30–33. [[CrossRef](#)] [[PubMed](#)]
4. Gutierrez-Villaneuva, J.L.; Sainz Fernández, C.; Fuente Merino, I.; Quindós López, L.; Quindós López, J.; Fernández Villar, A.; Casal Ordas, S.E.; López Abascal, D.; Arteché Laso, D.; Fernández López, E.; et al. *International Intercomparison Exercise on Natural Radiation Measurements under Field Conditions*; Editorial Universidad de Cantabria: Cantabria, Spain, 2012.
5. Pressyanov, D.; Buysse, J.; Poffijn, A.; Meesen, G.; Van Deynse, A. The compact disk as radon detector—A laboratory study of the method. *Health Phys.* **2003**, *84*, 642–651. [[CrossRef](#)] [[PubMed](#)]
6. Fuente, M.; Rabago, D.; Herrera, S.; Quindos, L.; Fuente, I.; Foley, M.; Sainz, C. Performance of radon monitors in a purpose-built radon chamber. *J. Radiol. Prot.* **2018**, *38*, 1111–1127. [[CrossRef](#)] [[PubMed](#)]
7. Hammer, Ø.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Paleont. Electron.* **2001**, *4*, 9.

8. Mitev, K.; Madzhunkov, Y.; Gerganov, G.; Dimitrova, I.; Georgiev, S.; Pressyanov, D. Automatic counting of electrochemically etched tracks in compact discs. Application to retrospective measurements of Rn-222. *IEEE Trans. Nucl. Sci.* **2010**, *57*, 300–308. [[CrossRef](#)]
9. Pressyanov, D.; Mitev, K.; Georgiev, S.; Dimitrova, I.; Kolev, J. Laboratory facility to create reference radon + thoron atmosphere under dynamic exposure conditions. *J. Environ. Radioact.* **2017**, *166*, 181–187. [[CrossRef](#)] [[PubMed](#)]
10. Gardner, M.; Altman, D. *Statistics with Confidence*; British Medical Journal: London, UK, 1992; pp. 20–27.
11. Pressyanov, D. Modeling a ^{222}Rn measurement technique based on absorption in polycarbonates and track-etch counting. *Health Phys.* **2009**, *97*, 604–612. [[CrossRef](#)] [[PubMed](#)]
12. Pressyanov, D.; Mitev, K.; Georgiev, S.; Dimitrova, I. Optimization of etching conditions of CDs/DVDs used as detectors for ^{222}Rn . *Radiat. Meas.* **2015**, *83*, 36–40. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



16ENV10 MetroRADON

Report on Activity A.4.3.2

SUBG Results for Radon in Air, Radon in Soil-gas and Radon exhalation from soil, obtained in the frames of “METRORADON: Intercomparison on indoor radon at LNR Saelices el Chico (Salamanca, Spain)”, organized by LaRUC

Strahil Georgiev, Ivelina Dimitrova,
Krasimir Mitev, Chavdar Dutsov

Sofia University St. Kliment Ohridski, Bulgaria (SUBG)

SUBG Results for Radon in Air, Radon in Soil-gas and Radon exhalation from soil
obtained in the frames of “METRORADON: Intercomparison on indoor radon at LNR Saelices el Chico (Salamanca, Spain)”, organized by LaRUC

contributors: Strahil Georgiev, Ivelina Dimitrova, Krasimir Mitev, Chavdar Dutsov
Sofia University St. Kliment Ohridski, Bulgaria (SUBG)

1. Measurements of radon in air

A novel type of detector is used - a DVD (used as a solid state track detector) and two thin Makrofol N foils (used as radon absorber) facing the DVD surface. Another DVD half is used to keep the foils close to the disk. The detector is designed in such way that radon progeny in the air doesn't have an influence on the signal. The background exposure of the detectors was not estimated based on the transit detectors, but with an alternative approach! That is because the transit detectors showed very high exposure – 79 kBq.h/m³. It is possible that significant part of this exposure was accumulated in the LNR lab while the other sets of detectors were exposed and it shouldn't be taken into account.

The detectors were calibrated by exposure of an identical set of detectors to radon atmosphere at the calibration facility at Sofia University. The exposure was carried out at the same temperature as the exposure in the intercomparison at LNR. The obtained results are shown in the Table below.

	Overall Results (kBq h m ⁻³)	
	Value	Uncertainty
1st Exposure	317	32
2nd Exposure	796	62

Table 1. Results from the exposures of radon in air carried out in the frames of the intercomparison at LNR in November 2018. The presented uncertainties are at the level of 1 standard deviation and include the calibration uncertainty and the standard deviation of the results of the 10 detectors in each group.

2. Measurements of radon in soil gas and radon exhalation

The method based on liquid scintillation counting of polymers is proposed in [1]. For the measurements of radon in soil-gas and radon exhalation from soil the metal rod (stainless steel rod with holes along its length with Marofol N foils packed in thin kitchen polyethylene sheet inside) was used. The foils were placed 5 cm apart and the last one reached a depth of 75 cm. The rod with the foils was hammered (see Fig. 1) at 15:30 on 05 Nov. 2018 and pulled-out at 10:30 on 07 Nov. 2018. The temperatures measured at the beginning and at the end of the exposure at about 20 cm below the ground were in the range 10 – 11°C.



Figure 1. Hammering the rod in the “Green”

The foils were placed in glass LS-vials with THM-cocktail shortly after the exposure and measured at the “Triathler” LS-counter provided by the LaRUC-team. For the activity estimation, the signal in the alpha-channel obtained more than 6 hours after the end of exposure was used. The beta-channel signal was not used because of its high and variable background counting rate. Based on the observed count rates in the alpha- and beta-channel it seems that about 10% of the alphas are counted as betas. Despite of the loss of alphas, this separation seems good as no betas are counted in the alpha-channel (Fig.2 (middle)).

The activity concentration in the soil was determined as:

$$C_A = \frac{n_0 k_{des}}{\varepsilon_c \varepsilon_s V}$$

where n_0 is the net counting rate of the foil corrected to the moment of its placement in the vial, k_{des} is a correction factor for the radon desorbing from the foils in the minutes before the foil is placed in the vial, ε_c is the counting efficiency in the alpha-channel of the Triathler, ε_s is the sampling efficiency for the exposure conditions and V is the volume of the foil. The estimation of each of the above values is briefly described below.

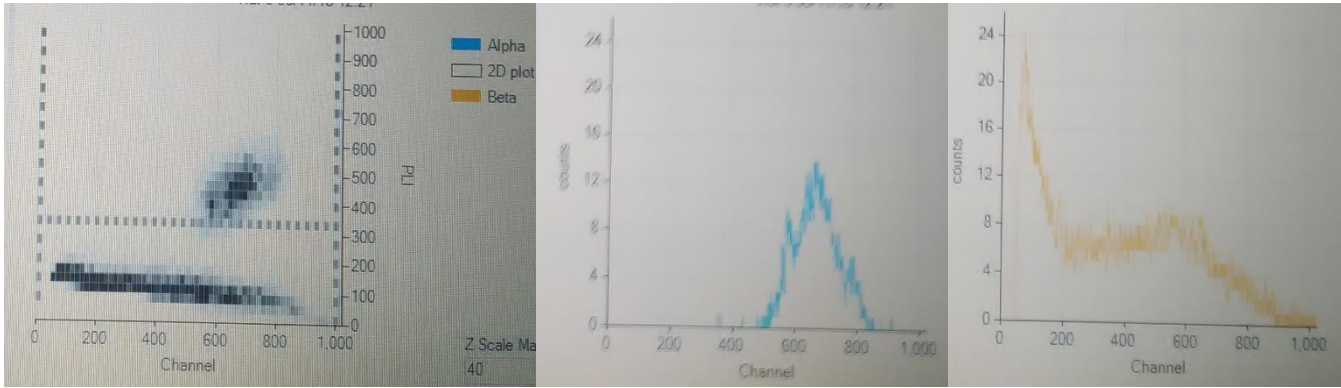


Figure 2. (left) 2D-plot at $Z=40$ – visualization of the alpha/beta separation of the Triathler and the PSA-threshold. (middle) Alpha- and (right) beta-spectra obtained with this PSA.

In order to calibrate the Triathler two of the samples (with the foils buried at the two highest depths) measured at the Triathler were brought in Sofia and measured at the RackBeta. The foils were followed at the RackBeta for a few days and the net signal was extrapolated to the moment of the preparation of the samples (see Fig. 3). The net count-rates of in the alpha-channel of the Triathler were also decay-corrected to that moment. The ratio between the count-rates of the Triathler and the RackBeta and the known counting efficiency of the RackBeta in the two samples were used to estimate the counting efficiency of the Triathler in the alpha-channel.

The obtained estimate is $\varepsilon_c = 2.615(39)$ for the PSA-threshold shown in Fig. 2, which is close to the value used in the preliminary estimation $\varepsilon_c = 2.73(27)$. Additionally, we worked on the 2D plots of the samples in order to improve the alpha/beta separation by moving the PSA-threshold and adding Energy-threshold (see Fig.4). That resulted in about 10% increase in the counting rate in the alpha-channel and the counting efficiency for this PSA-value, estimated using equation (2), is $\varepsilon_c = 2.863(42)$.

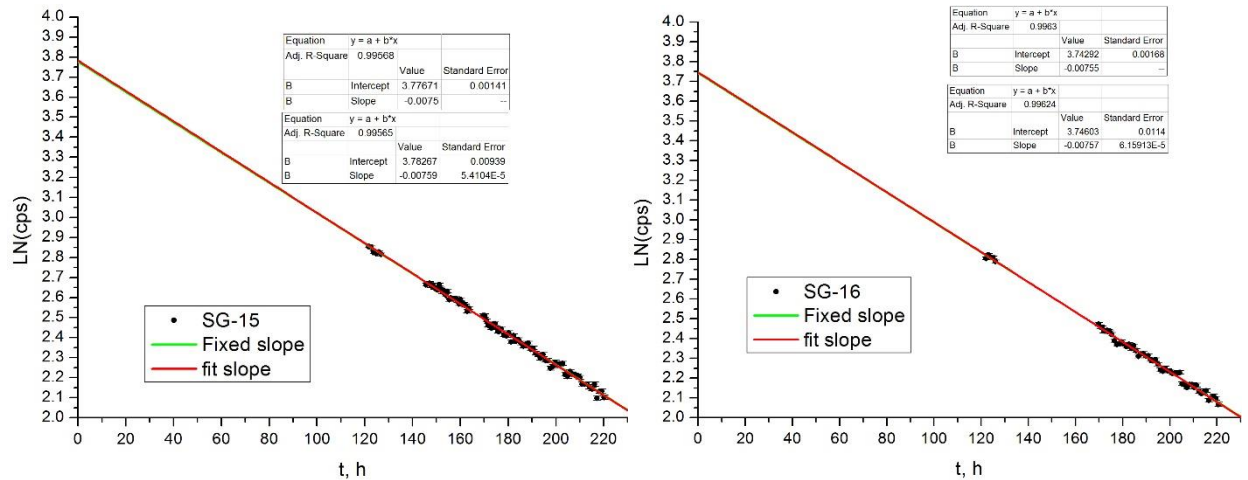


Figure 3. Logarithm of the net signal as a function of time for the two samples. Two fits are applied – one with fixed slope-parameter equal to the Rn-decay constant and the other with free slope-parameter. As it is seen, for each sample the fit-parameters of the two fits coincide within the uncertainties.

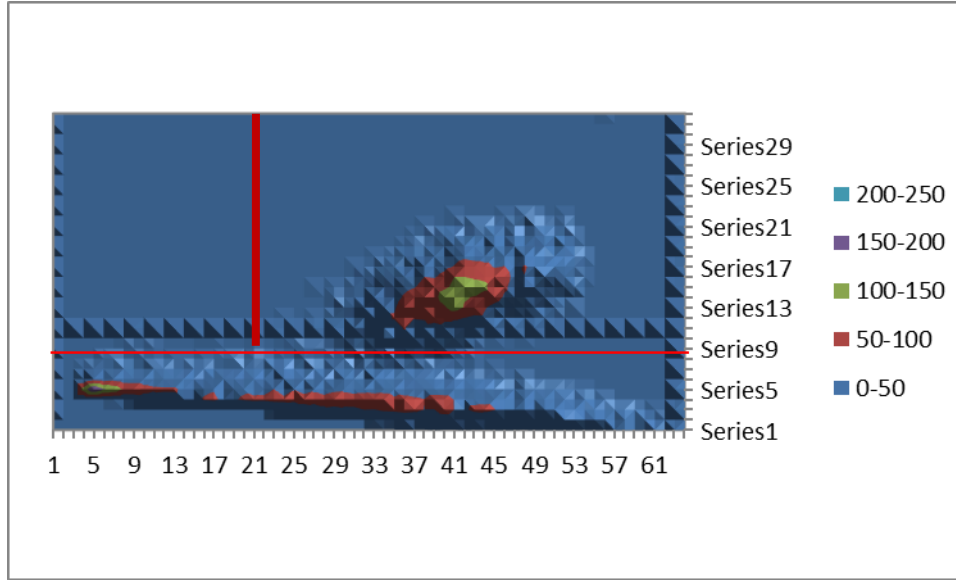


Figure 4. 2D-plot at $Z=0$ – visualization of the alpha/beta separation of the Triathler – the black triangles mark the original PSA-threshold, the horizontal red line marks the new PSA-threshold and the vertical red line the Energy threshold.

In order to determine the sampling efficiency (ε_s) we conducted an experiment to determine the partition coefficient K and L_D of Makrofol N at $t=10^\circ\text{C}$ using an approach presented in [2]. The obtained values are $K=183(12)$ and $L_D=23.9(10)$ μm . A sorption/desorption model [3] was used to estimate the sampling efficiency of the foils for the given exposure conditions assuming constant activity concentration and temperature in the soil. The estimated value for the sampling efficiency multiplied by the volume of the foil is $\varepsilon_s V= 1.49 \times 10^{-5}(14)$ m^3 (actually, that product is needed to estimate the ^{222}Rn concentration). As the foils desorbed from the moment the rod was pulled-out to the moment of their placement in the LSC-vials, the desorption correction for each foil C_{des} was also estimated.

To estimate the activity concentration of ^{222}Rn at infinity ($C_{A,\infty}$) by the gradient method, the depth profile of $C_A(d)$ should be fitted with the function:

$$C_A(d) = C_{A,\infty} \left(1 - e^{-\frac{d}{L_{D,\text{soil}}}} \right)$$

where $L_{D,\text{soil}}$ is the diffusion length of radon in the soil and d is the depth. The exhalation rate J_0 can then be determined with the parameters determined by the fit:

$$J_0 = \lambda L_{D,\text{soil}} C_{A,\infty}$$

As the uncertainty of $C_A(d)$ is dominated by systematic contribution mainly due to the estimate of the ε_s , the fit can be directly applied to the counting rate (after the decay and desorption correction is applied) then $C_{A,\infty}$ can be calculated. The fit was applied to the decay and desorption corrected count rates obtained directly by the Triathler and that obtained with the modified PSA threshold(see

Fig. 5). No qualitative difference between the two fits is found. Moreover, the difference between the obtained values for C_A , and J_0 by the two approaches are smaller than their respective uncertainties. The values of the fit with the higher R -squared value were chosen for the estimates. For visualization purposes the fit curve and the count rates are converted in activity concentration and plotted in Fig. 6.

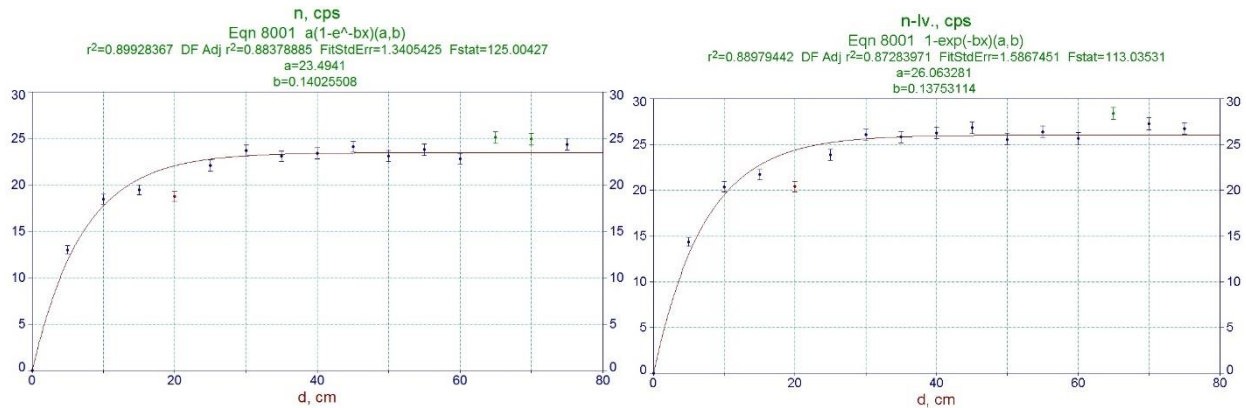


Figure 5. Applying the gradient method (left) on the count rates of the Triathler and (right) on the count rates with the modified PSA-threshold.

The exhalation rate was additionally estimated through the “missing” activity near the surface with the following equation:

$$J_0 = \lambda \sum_i \frac{C_{A,\infty} + C_{A,i}}{2} \Delta d ,$$

where $C_{A,i}$ is the the i -th data point and Δd is the distance between each measurement point. This estimate is 10% greater than the one obtained by the fit due to the “drop” of the point at $d = 20$ cm. There could be a physical reason for that drop. Therefore, this was the estimate of J_0 reported as a final results.

Results:

Location	Radon in soil concentration (kBq m ⁻³)	
	Value	Uncertainty
"Green Ballesteros" (next to the LNR)	602	57
Offices site		

Location	Radon Exhalation (Bq m ⁻² h ⁻¹)	
	Value	Uncertainty
"Green Ballesteros" (next to the LNR)	361	33

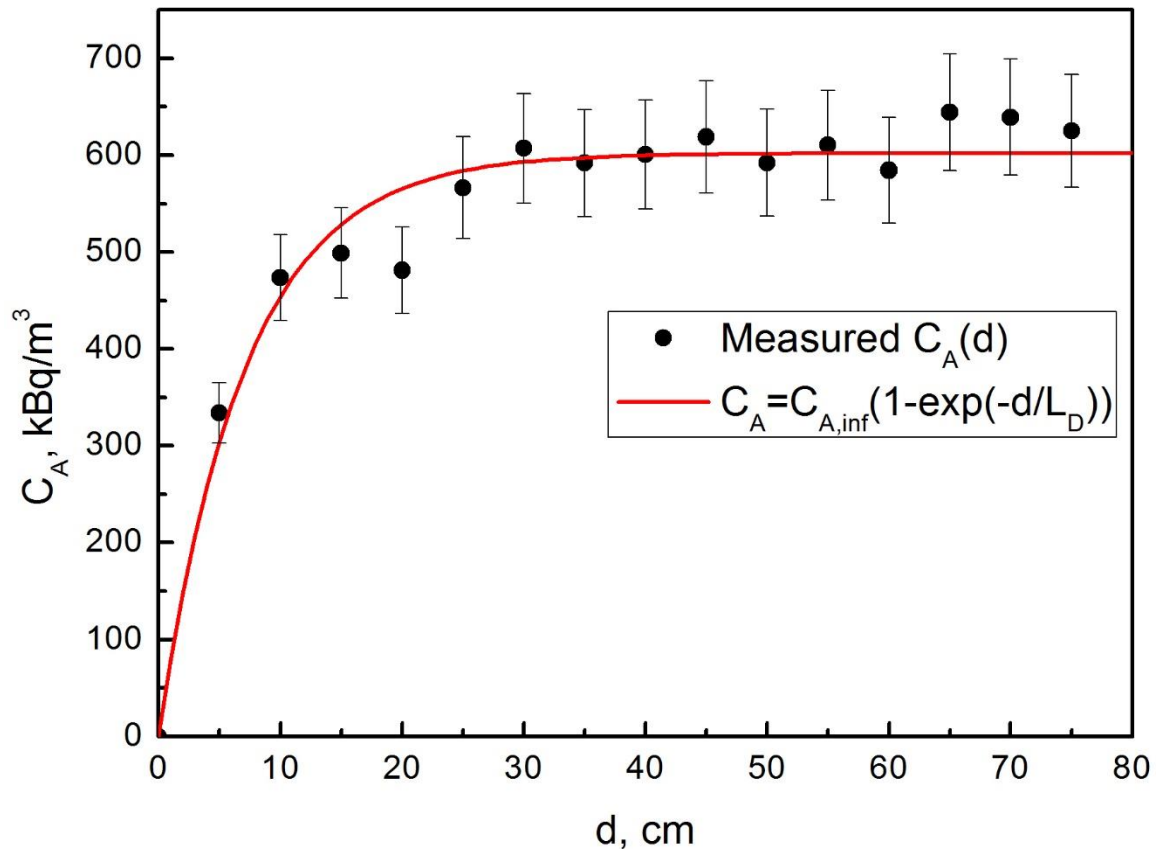


Figure 6. Depth profile of the radon concentration.

3. Comment on the exhalation measurement results and comparison of the results to the other participants

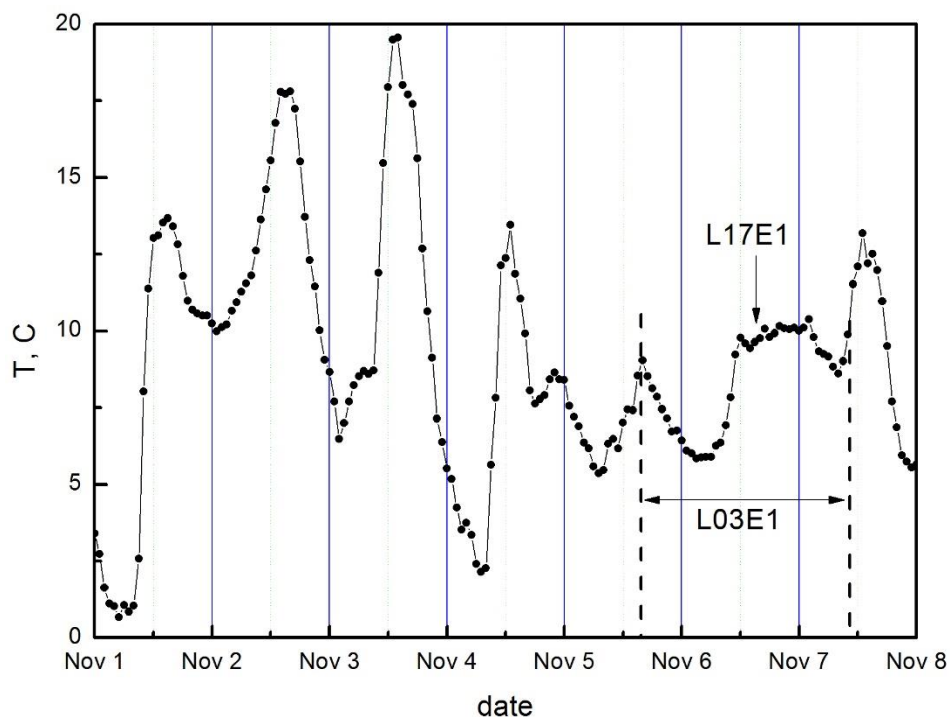
The large differences in the results of the radon exhalation measurements require some explanation. Although the measurements are very few to draw definitive conclusions, still some consideration can be made. As could be seen in Fig. 7 (Table 10 from the Metroradon report, WP.3.3.3.), measurements L03E1 and L17E1 overlap in time, but they differ in two orders of magnitude. However, the first method is cumulative, while the second is discrete. During the period of these measurements, the weather was mostly rainy and windy and the soil was soaked, which would impede the exhalation. That is seen in the radon depth profile shown in Fig.6 (the data is obtained by L03 and used for the gradient method). The diffusion length of radon in soil corresponding to that profile is $L_D=7.1(7)$ cm. Due to the windy weather, there were some cloudless and sunny time windows and the discrete measurement L17E1 was carried out in such a window (see the weather data in Fig. 8). The sun would dry the soil and lead to increase in the radon exhalation, which could be a possible explanation for the observed difference between the results of the two measurements.

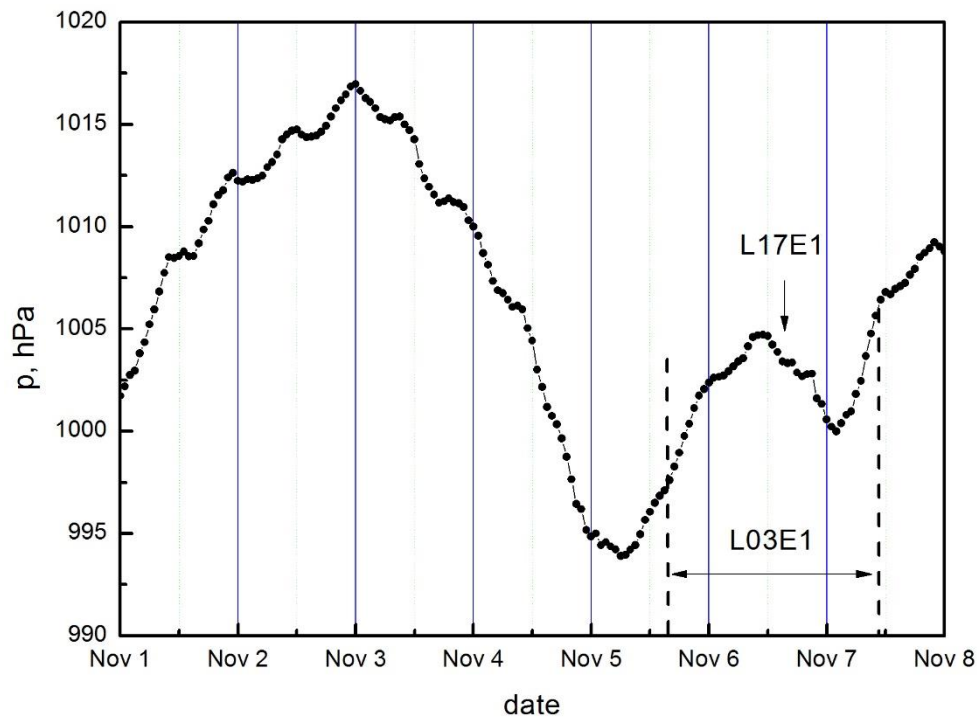
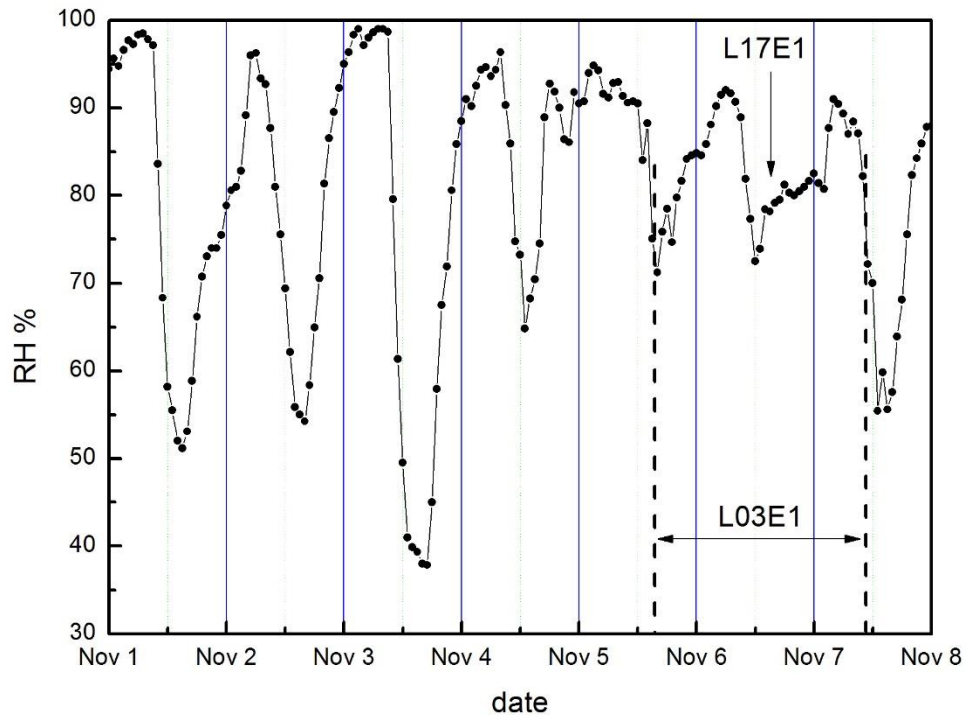
On the other hand, the two cumulative measurements L03E1 and L20E1 could not be compared directly, as they were carried out at different times. The weather was dry and sunny in the week of the L20E1 measurement, in contrast to the weather during the L03E1 measurement. Although no definitive conclusions can be drawn, these results indicate the significant effect of the weather on the radon exhalation rate, which deserves more thorough study. More details about the intercomparison performed at Saelices el Chico are given in [4].

Table 10. Radon exhalation J and its uncertainty $u(J)$ performed in the indicated date with the methodology used by each participant to conduct the test in the “Green Ballesteros”.

Code	J (Bq m ⁻² h ⁻¹)	$u(J)$ (Bq m ⁻² h ⁻¹)	Date (2018)	Methodology
L03E1	361	33	5 Nov 15:30 to 7 Nov 13:30	Gradient method with polycarbonate foils
L17E1	14719	1939	6 Nov 15:30 (approx. 1 hour.)	Accumulation method
L20E1	35100	8200	15 Nov 10:00 to 16 Nov 10:00	Absorption in activated charcoal collector

Fig. 7. Table 10 from the UC report.





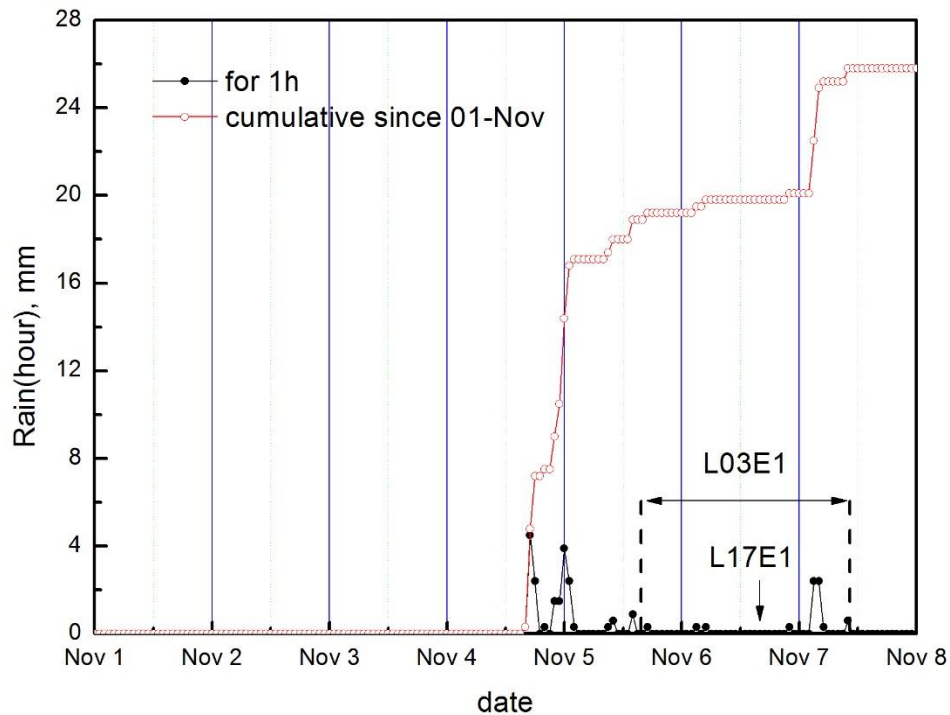


Fig.8. Meteorological data for the period of the intercomparison. The times (or intervals) of the radon exhalation measurements are marked. The temperature increase and the RH decrease indicate sunshine before and during measurement L17E1, which could facilitate radon exhalation. The transient drop in the atmospheric pressure could also facilitate the exhalation.

References:

- [1]. K. Mitev, Ch. Dutsov, S. Georgiev, T. Boshkova, D. Pressyanov, “Unperturbed, high spatial resolution measurement of Radon-222 in soil-gas depth profile”, *Journal of environmental radioactivity*, 196 (2019) 253-258.
- [2]. K. Mitev, P. Cassette, S. Georgiev, I. Dimitrova, B. Sabot, T. Boshkova, I. Tartès, D. Pressyanov, “Determination of ^{222}Rn absorption properties of polycarbonate foils by liquid scintillation counting. Application to ^{222}Rn measurements”, *Applied Radiation and Isotopes*, 109 (2016) 270-275.
- [3]. D. Pressyanov, K. Mitev, S. Georgiev, I. Dimitrova. “Sorption and desorption of radioactive noble gases in polycarbonates”, *Nuclear Instruments and Methods in Physics Research Section A*, 598/2 (2009) 620-627.
- [4]. D. Rabago, L.S. Quindos, J. Quindos, E. Fernandez, A. Fernandez, L. Quindos, S. Celaya, I. Fuente, C. Sainz, Intercomparison of indoor radon and geogenic radon measurements under field conditions, *Metroradon report*, WP.3.3.3., February, 2019.

Table A: Geological characteristics associated with cells identified as hot spots. AM: arithmetic mean; SD: standard deviation.

N° cell	Location		Total data number	Indoor radon concentrations				Geological features					
	County	N° "département"		AM	SD	% of data >300 Bq/m ³	% of data >1000 Bq/m ³	Main lithology (1 000 000 geological map)	Geochemistry	U content	U mines location	Thermal spring location	Major faults
258	Bretagne	29	77	581	1006	40.3	16.9	Monzogranites, granodiorites	Subalkaline	< 8 ppm	No	No	Yes
321	Bretagne	29	124	478	700	43.5	8.1	Leucogranites peraluminous	Peraluminous	< 10 ppm	No	No	Yes
381	Bretagne	29	450	632	1475	39.1	14.9	Leucogranites peraluminous	Peraluminous	< 10 ppm	No	No	Yes
441	Bretagne	29	978	510	916	43.8	10.7	Leucogranites peraluminous	Peraluminous	> 10 ppm	No	No	Yes
502	Bretagne	29	141	387	678	35.5	5.7	Monzogranites, granodiorites	Peraluminous	< 10 ppm	No	No	Yes
562	Bretagne	56	112	340	395	35.7	5.4	Monzogranites, granodiorites	Peraluminous	< 8 ppm	Yes (1)	No	Yes
621	Bretagne	56	44	499	717	38.6	15.9	Leucogranites peraluminous	Peraluminous	> 10 ppm	Yes (9)	No	No
865	Pays de la Loire	44	24	510	893	29.2	16.7	Leucogranites peraluminous	Peraluminous	< 10 ppm	No	No	Yes
1229	Poitou-Charentes	79	127	550	871	51.2	13.4	Leucogranites peraluminous	Peraluminous	< 8 ppm	No	No	Yes
1431	Midi-Pyrénées	31	29	548	624	51.7	13.8	Paleozoic schists	Siliceous, aluminous	< 10 ppm	No	Yes (1)	Yes
1473	Poitou-Charentes	16	94	581	718	50.0	16.0	Monzogranites, granodiorites	Peraluminous, calc-alkaline	< 10 ppm	No	No	No
1533	Limousin	87	87	892	1012	77.0	26.4	Monzogranites, granodiorites	Peraluminous	<10 ppm	No	No	Yes
1593	Limousin	87	329	1082	1650	75.7	28.9	Leucogranites peraluminous	Peraluminous	> 10 ppm	Yes (19)	No	Yes

1594	Limousin	87	558	535	1081	43.9	11.1	Leucogranites peraluminous	Peraluminous	> 10 ppm	No	No	No
1653	Limousin	87/23	103	834	1095	65.0	25.2	Leucogranites peraluminous	Peraluminous	> 10 ppm	Yes (2)	No	Yes
1654	Limousin	87/23	35	489	469	62.9	5.7	Monzogranites, granodiorites	Peraluminous	< 8ppm	No	No	No
1714	Limousin	87/23	125	657	1066	52.8	16.8	Monzogranites, granodiorites	Peraluminous	< 8 ppm	Yes (1)	No	Yes
1715	Limousin	87/19	48	795	1553	52.1	14.6	Leucogranites peraluminous	Peraluminous	> 10 ppm	No	No	Yes
1716	Limousin	19	22	661	758	54.5	18.2	Leucogranites peraluminous	Peraluminous	> 10 ppm	Yes (1)	No	Yes
1775	Limousin	19	80	1222	2267	52.5	26.3	Leucogranites peraluminous	Peraluminous	> 10 ppm	Yes (1)	No	No
1776	Limousin	19	81	835	742	75.3	32.1	Monzogranites, granodiorites	Peraluminous	> 10 ppm	Yes (3)	No	No
1836	Limousin	19	201	453	695	48.8	8.5	Monzogranites, granodiorites	Peraluminous	> 10 ppm	Yes (1)	No	Yes
1960	Auvergne	15/12	11	1747	3527	45.5	36.4	Monzogranites, granodiorites	Peraluminous	< 10 ppm	Yes (1)	Yes (2)	Yes
2019	Auvergne	15	62	606	926	40.3	22.6	Monzogranites, granodiorites	Peraluminous	< 10 ppm	No	Yes (1)	Yes
2020	Auvergne	15/48	78	1279	3640	55.1	17.9	Monzogranites, granodiorites	Peraluminous	< 10 ppm	No	Yes (3)	Yes
2021	Languedoc- Roussillon	48	17	380	543	17.6	17.6	Monzogranites, granodiorites	Peraluminous	< 10 ppm	No	No	Yes
2080	Languedoc- Roussillon	48	137	470	475	57.7	10.2	Monzogranites, granodiorites	Peraluminous	< 10 ppm	Yes (1)	No	Yes
2133	Auvergne	3	31	804	1286	71.0	16.1	Monzogranites, granodiorites	Subalkaline	> 10 ppm	Yes (8)	Yes (3)	Yes
2136	Auvergne	63	12	477	640	33.3	16.7	Leucogranites peraluminous	Peraluminous	< 10 ppm	No	No	Yes

2187	Bourgogne	58	38	598	635	57.9	13.2	Monzogranites, granodiorites	Subalkaline	< 10 ppm	Yes (7)	No	No
2202	Languedoc- Roussillon	48	36	586	531	61.1	19.4	Monzogranites, granodiorites	Subalkaline	< 10 ppm	No	No	Yes
2308	Bourgogne	71	48	530	515	58.3	18.8	Leucogranites peraluminous	Peraluminous	< 10 ppm	No	No	Yes
2309	Bourgogne	71	114	291	390	33.3	5.3	Leucogranites peraluminous	Peraluminous	> 10 ppm	Yes (2)	No	Yes
2311	Bourgogne	71	20	521	497	55.0	10.0	Monzogranites, granodiorites	Subalkaline	< 10 ppm	Yes (1)	No	No
2312	Bourgogne	71/69	43	1150	4296	62.8	11.6	Microgranites, apligranites	Calc-alkaline, subalkaline	< 8 ppm	No	No	No
2726	Franche- Comté	25	149	333	467	37.6	4.0	Limestones, marls	Carbonate	< 2 ppm	No	Yes (1)	Yes
2786	Franche- Comté	25	42	519	446	59.5	11.9	Limestones, marls	Carbonate	< 2 ppm	No	No	No
2844	Franche- Comté	25/70/90	256	408	436	45.7	9.4	Limestones, marls	Carbonate	< 2 ppm	No	No	Yes
2845	Franche- Comté	25	165	357	330	47.3	4.8	Limestones, marls	Carbonate	< 2 ppm	No	No	No
2846	Franche- Comté	25	55	752	1024	65.5	14.5	Limestones, marls	Carbonate	< 2 ppm	No	No	No
3413	Corsica	2A	13	1263	2493	38.5	30.8	Alkaline affinity granites	Peralkaline	> 10 ppm	No	No	No
3532	Corsica	2B	14	560	792	35.7	21.4	Alkaline affinity granites	Peralkaline	< 10 ppm	No	No	Yes

Table B: Geological characteristics associated with identified outliers (grouping by French “départements”)

Location		Data number	Municipalities number	Indoor radon concentration (Bq/m ³)		Geological features						
County	French “département”			Min	Max	Main lithology (1,000,000 geological map)	Geochemistry	System	U content	U mines location	Thermal spring location	Major faults
Auvergne	03 – Allier	4	1	1772	2034	Sand clay	Siliceous aluminous	Neogene	< 4-5 ppm	No	No	No
	43 – Haute-Loire	2	2	1056	1289	Alkaline basalts	Tholeiitic peralkaline	Neogene Quaternary	< 2 ppm	No	No	No
	63 – Puy de Dome	4	3	1090	1579	Basalts basanites	Tholeiitic peralkaline	Paleogene Neogene	< 2 ppm	No	Yes (1)	No
Basse-Normandie	14 – Calvados	2	2	1499	1083	Sandstone claystone	Siliceous aluminous	Neoproterozoic	< 4-5 ppm	No	No	No
	61 – Orne	1	1	1422	1422	Flysch (sandstone claystone)	Siliceous aluminous	Neoproterozoic	< 4-5 ppm	No	No	No
Bourgogne	58 - Nièvre	4	3	1079	2497	Alluvium limestones	Siliceous carbonate	Quaternary Jurassic	<4-5 ppm	No	Yes (1)	No
	71 – Saône et Loire	5	3	1110	4806	Limestones	Carbonate	Jurassic	< 2 ppm	No	No	No
Bretagne	22 – Cotes d’Armor	3	3	1003	1424	Schists	Siliceous aluminous	Cambrian et Carboniferous	<4-5 ppm	No	No	No
	29 – Finistère	4	4	1039	3206	Schists	Siliceous aluminous	Carboniferous	<4-5 ppm	No	No	No
	56 – Morbihan	5	2	1164	7625	Schists	Siliceous aluminous	Cambrian	<4-5 ppm	No	No	No
Centre	37 – Indre-et-Loire	1	1	1036	1036	Limestones clay marls	Siliceous carbonate	Cretaceous	< 2 ppm	No	No	No
Corse	2B – Haute-Corse	1	1	2000	2000	Superficial formations	Siliceous aluminous	Quaternary	<4-5 ppm	No	No	Yes
Franche-Comté	25 – Doubs	56	35	1026	5631	Limestones marls	Carbonate	Jurassic	< 2 ppm	No	No	Yes
	70 – Haute-	21	12	1024	5666	Limestones marls	Siliceous	Triassic	<4-5 ppm	No	Yes (1)	No

	Saône					sandstones clay	carbonate	Jurassic				
	90 – T. de Belfort	3	2	1444	1630	Limestones marls sandstones	Siliceous carbonate	Jurassic Paleogene	<4-5 ppm	No	No	No
Languedoc-Roussillon	48 – Lozère	4	3	1244	2199	Limestones schists	Siliceous carbonate	Jurassic Ordovician	<4-5 ppm	No	No	No
Lorraine	88 – Vosges	2	2	1129	2152	Marls sandstones dolomites	Siliceous carbonate	Triassic	<4-5 ppm	No	Yes (1)	No
Midi-Pyrénées	09 - Ariège	1	1	1922	1922	Limestones marls et sandstones	Siliceous carbonate	Cretaceous	<4-5 ppm	No	No	Yes
Pays-de-la-Loire	44 – Loire-Atlantique	2	1	1508	2061	Schists alluvium	Siliceous	Ordovician	<4-5 ppm	No	No	Yes
	53 - Mayenne	1	1	1170	1170	Schists sandstones arkoses	Siliceous aluminous	Cambrian	<4-5 ppm	No	No	No
	72 - Sarthe	2	2	1001	1535	Limestones marls	carbonate	Jurassic	< 2 ppm	No	No	No
Poitou-Charentes	79 – Deux-Sèvres	2	2	1062	1440	Limestones marls	carbonate	Jurassic	< 2 ppm	No	No	No
Rhône-Alpes	26 – Drôme	1	1	1478	1478	Marls sandstones schists limestones	Siliceous aluminous carbonate	Cretaceous	< 2 ppm	No	No	No
	73 - Savoie	1	1	1229	1229	Limestones and clay	Siliceous carbonate	Cretaceous et Quaternary	<4-5 ppm	No	Yes (1)	No

Table C: Lithostratigraphy associated with categories 1 and 2 digitized. AM: arithmetic mean; SD: standard deviation.

TIPOLOGY	ORIGIN PERMEABILITY	Number of points	Data of >300 Bq/m ³		Data of >1000 Bq/m ³		Bq/m ³						
			N	%	N ^o	%	MIN	MAX	AM	SD	1 st quartile	MEDIAN	3 rd quartile
AREA MORE THAN 400 Bq/m ³ -PENÍNSULA													
Sandstones silts and other conglomerates (Toro-Corrales's arkoses Geroma's silts Villallazán's conglomerates)	DETRITIC-MEDIUM	43	11	26	5	12	15	7400	541	1383.7	41	77	343
Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	META-DETRITIC-LOW	119	15	13	6	5	10	5676	260	726.2	29	64	175
Schist schists-graphites phyllites quartzites ampelites and lydites. Nogueira's Group Paraño and Rábano's Formation	META-DETRITIC-LOW	49	9	18	1	2	52	1184	227	208.6	99	173	269
Schists paragneiss quartzites metav. acid mica-schists. (Ordenes's Schists Malpica-Tuy-Lalín-Forcarey and Ortegá's Quartzites)	META-DETRITIC-LOW	79	18	23	0	0	31	802	219	179.9	88	165	290
Slates shales meta-sandstones and amphibole gneisses. Villalba series	META-DETRITIC-LOW	42	6	14	1	2	26	1181	182	215.6	78	106	168
Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	IGNEOUS-LOW	1341	138	10	10	1	10	11100	163	380.6	56	102	183
Slates grawacks and sporadic carbonated levels. Schist-grawack Complex	META-DETRITIC-LOW	298	22	7	2	1	10	15403	162	896.8	38	70	122
Metamorphized acid rocks	META-DETRITIC-	193	20	10	0	0	10	920	133	140.1	45	91	162

(orthogneisses migmatites) metarriolites (Ollo Sapo) Gn.Pera Gn.Gland.	LOW												
Micaschists quartzites and gneisses	META-DETRITIC- VERY LOW	30	4	13	0	0	14	848	119	169.3	27	60	122
AREA MORE THAN 400 Bq/m ³ -CANARY ISLANDS													
Flows and con. of basaltic tephra crossed by dykes alternating with breccias or with basic intrusive rocks.	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	1	0	0	0	0			199				
Wind sands subordinate	VOLCANICS (PIROCLASTIC AND LAVICS)-HIGHT	76	11	14	3	4	10	1876	181	328	34	70	168
Basaltic flows (and pyroclastics) basaitic and tephritic	VOLCANICS (PIROCLASTIC AND LAVICS)-MEDIUM	3	0	0	0	0	59	240	133	95	69	99	205
Flows and cones of tephra basan nephritic and tephritic. Sub basalts	VOLCANICS (PIROCLASTIC AND LAVICS)-HIGHT	14	0	0	0	0	14	239	64	64.6	17	34	96
Flows and cones of tephra basan nephritic and tephritic. Sub basalts	VOLCANICS (PIROCLASTIC AND LAVICS)-HIGHT	5	0	0	0	0	21	53	38	14.4	24	40	51
Chaotic sliding gaps with clayey- sandy matrix and subordinate epiclastic materials	DETRITIC-MEDIUM	2	0	0	0	0	29	36	33	4.9	29	33	36
Flows and cones of tephra basan pyroclasts subordinates of dispersion	VOLCANICS (PIROCLASTIC AND LAVICS)-VERY HIGHT	1	0	0	0	0			30				
Flows and ignimbr. phonolitic Phonolitic intrusives and pyroclastic flows. Ash & pumice	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	4	0	0	0	0	10	82	28	36	10	10	46

subord														
Tephra-basaltic cones	VOLCANICS (PIROCLASTIC AND LAVICS)-MEDIUM	1	0	0	0	0			18					
Welded polymictic and heterometric gap	DETRITIC-MEDIUM	0	0	0	0	0	0							
Rhyolitic and trachytic flows with subordinate phonolites. Tuffs and ignimbritic gaps peralkaline	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	0	0	0	0	0	0							
Ignimbritic tuffs and rhyolitic-trachytic flows. Salic dykes in more than 65%	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	0	0	0	0	0	0							
Ignimbritic tuffs and rhyolitic-trachytic flows. Salic dykes between 10- 65%	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	0	0	0	0	0	0							
Ignimbritic tuffs and rhyolitic-trachytic flows. Salic dykes subordinates	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	0	0	0	0	0	0							
AREA MORE THAN 301-400 Bq/m ³ -PENÍNSULA (IN CANARY ISLANDS DON T EXIST)														
Schists paragneiss carbonate levels / calcosilicates (Villalcampo's Schists Duero Series Sistema Central's metasediments)	META-DETRITIC-LOW	84	8	10	0	0	13	939	137	155.5	49	81	155	
Limestones sandstones and shales	CARBONATED-LOW	5	1	20	0	0	38	374	118	143.9	40	64	148	
Arkoses with boulders conglomerates and clays	DETRITIC-MEDIUM	81	6	7	1	1	10	1226	114	191.1	28	63	112	
Grawacks and slates. Facies Culm.	META-DETRITIC-LOW	8	1	13	0	0	10	388	104	122.4	30	67	121	
Quartzites conglomerates sandstones and lutites. Purple Series and Constante or Bornova	META-DETRITIC-LOW	18	1	6	0	0	10	461	90	110.9	30	49	122	

Formations														
Phyllites schists quartzites limestones shales and hornfels (metamorphic rocks)	META-DETRITIC- VERY LOW	17	1	6	0	0	10	420	88	101	40	63	82	
Sands and conglomerates	DETRITIC-HIGHT	4	0	0	0	0	80	89	84	3.8	81	83	86	
Quartzites and slates	META-DETRITIC- LOW	25	1	4	0	0	10	323	83	72.7	32	62	109	
Slates	META-DETRITIC- VERY LOW	3	0	0	0	0	13	92	50	39.6	21	46	81	
Quartzites slates and limestones	META-DETRITIC- LOW	4	0	0	0	0	10	47	28	17.1	14	28	43	
Clays marls and limestones	CARBONATED- MEDIUM	0	0	0	0	0	0							
Sands and clays	DETRITIC-LOW	0	0	0	0	0	0							

Table D: Lithologies associated with categories 1 and 2 digitized. AM: arithmetic mean; SD: standard deviation.

TIPOLOGY	Number of points	Data of > 300 Bq/m ³		Data of > 1000 Bq/m ³		Bq/m ³						
		N°	%	N°	%	MIN	MAX	AM	SD	1st quartile	MEDIAN	3rd quartile
AREA MORE THAN 400 Bq/m ³ -PENÍNSULA												
Biotite granitoids	1376	159	12	10	1	10	11100	165	374.2	56	102	192
Slate sandstone quartzite and limestone or volcanoclastic rock	823	58	7	8	1	10	15403	146	613.5	36	68	123
Peraluminous granitoids (collisional type)	124	12	10	0	0	10	939	143	144.6	47	108	165
Felsic gneiss and metabasite	277	27	10	0	0	10	802	131	138.7	41	77	165
Garnet-bearing graphitic mica schist (Veleta)	18	2	11	0	0	14	848	119	197.7	22	59	87
Intermediate and basic igneous rocks	25	2	8	0	0	16	420	103	108.7	28	52	146
Slate schist sandstone limestone bituminous shale and lidite	293	13	4	0	0	10	717	96	104	33	62	115
AREA MORE THAN 400 Bq/m ³ -CANARY ISLANDS												
Calc-alkaline volcanic rocks (andesite dacite riolite). Shoshonite lamproite	148	9	6	3	2	10	1876	114	243.7	28	49	100
Basaltic lavas and pyroclasts	58	2	3	0	0	10	421	73	87.4	20	36	82

Table E: Lithostratigraphy associated with categories 1 and 2 in 10 km x10 km cells system. AM: arithmetic mean; SD: standard deviation.

TIPOLOGY	ORIGIN PERMEABILITY	CELLS		TOTAL DATA NUMBER ON THOSE CELLS	Data of >300 q/m ³		Data of >1000 Bq/m ³		Bq/m ³						
		N°	%		N°	%	N°	%	MIN	MAX	AM	SD	1 st quartile	MEDIAN	3 rd quartile
CELLS MORE THAN 400 Bq/m³-PENÍNSULA (858 CELLS)															
Sandstones silts and ocher conglomerates (Toro-Corrales's arkoses Geroma's silts Villallazán's conglomerates)	DETRITIC-MEDIUM	8	1	23	12	52	5	21.7	29	7400	985	1795.3	69	381	623
Schists paragneiss quartzites metav. acid mica-schists. (Ordenes's Schists Malpica-Tuy-Lalín-Forcarey and Ortegá's Quartzites)	META-DETRITIC-LOW	25	3	64	15	23	1	1.6	27	1083	231	216.5	74	158.5	284
Slates grawacks and sporadic carbonated levels. Schist-grawack Complex	META-DETRITIC-LOW	171	20	300	32	11	8	2.7	10	15403	223	999.9	37	63	126
Schist schists-graphites phyllites quartzites ampelites and lydites. Nogueira's Group Paraño and Rábano's Formation	META-DETRITIC-LOW	29	3	35	4	11	0	0	38	518	175	125.6	80	132	257
Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	IGNEOUS-LOW	400	47	1404	167	12	11	0.8	10	11100	172	373.5	60	110	198
Slates grawacks and sandstones sometimes with limestones and volcanic rocks. Culm facies	META-DETRITIC-LOW	4	0.5	3	1	33	0	0	50	325	168	141.5	70	130	276
Amphibolites metagabbros eclogites mafic granulites metaperidotites serpentinites and green schists.	META-DETRITIC-LOW	2	0.2	5	0	0	0	0	89	201	141	55.5	97	115	201
Sands and gravels (coastal cords and beaches)	DETRITIC (QUATERNARY)-HIGH	1	0.1	4	0	0	0	0	82	214	139	62.3	88	129	190
Metamorphized acid rocks (orthogneisses)	META-DETRITIC-	63	7	234	20	9	1	0.4	10	2068	132	182.9	49	83	149

migmatites) metarriolites (Olo Sapo) Gn.Pera Gn.Gland.	LOW															
Schists paragneiss carbonate levels / calcosilicates (Villalcampo's Schists Duero Series Sistema Central's metasediments)	META-DETRITIC-LOW	3	0.3	4	0	0	0	0	53	238	131	81.1	70	116	192	
Arkoses sometimes with boulders with shales marls limestones and locally flint and gypsum nodules	DETRITIC-MEDIUM	2	0.2	1	0	0	0	0			126					
Slates shales meta-sandstones and amphibole gneisses. Villalba series	META-DETRITIC-LOW	8	1	25	1	4	0	0	26	474	125	95.5	75	99	130	
Gravels sands silts (alluvial deposits valley bottoms and low terraces in rivers mainly)	DETRITIC (QUATERNARY)-VERY HIGHT	3	0.3	18	1	6	0	0	13	333	123	91.2	61	96	165	
Arkoses and silty arkoses white gray-green or ocher with crusts (Pedraja Vill.Adaja Pte.Runel ...)	DETRITIC-MEDIUM	3	0.3	14	1	7	0	0	29	344	113	83	53	97	121	
Phyllites schists quartzites limestones shales and hornfels (metamorphic rocks)	META-DETRITIC-VERY LOW	1	0.1	4	0	0	0	0	39	186	112	64.9	60	111	164	
Marble limestones. Carbonated series of Z. Ossa-Morena and Urda's marbles	CARBONATED-HIGHT	2	0.2	4	1	25	0	0	21	315	106	140	32	43	179	
Conglomerates sandstones and clays red (Roja de Toro and Aspariegos Facies Belver Comglomerates)	DETRITIC-LOW	3	0.3	3	0	0	0	0	29	226	99	110.4	32	41	180	
Quartzite sandstone slate limestone (Calymene and Cantera Ar.Fm Guindo Urbana and Chavera Qz. Mixtos B.)	META-DETRITIC-LOW	1	0.1	2	0	0	0	0	78	111	95	23.3	78	95	111	
Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	META-DETRITIC-LOW	46	5	87	3	3	0	0	10	813	93	112.8	21	58	123	
Micaschists quartzites and gneisses	META-DETRITIC-VERY LOW	29	3	25	2	8	0	0	10	416	83	93.7	36	55	80	
Black schists slates black quartzites migmatites	META-DETRITIC-	3	0.3	4	0	0	0	0	17	150	62	59.5	29	41	96	

amphibolites paragneisses. Black Series	VERY LOW															
Conglomerates gravels sands and red shales	DETRITIC-MEDIUM	10	1	31	0	0	0	0	23	211	75	53.3	32	55	106	
Gravels sands silts clays siltstones limestones (undifferentiated Quaternary)	DETRITIC (QUATERNARY)-HIGH	2	0.2	1	0	0	0	0			16					
Quartzites conglomerates sandstones and lutites. Purple Series and Constante or Bornova Formations	META-DETRITIC-LOW	1	0.1	0	0	0	0	0	0							
Quartzites slates and limestones	META-DETRITIC-LOW	1	0.1	0	0	0	0	0	0							
Sands clays and conglomerates	DETRITIC-MEDIUM	1	0.1	0	0	0	0	0	0							
Gravels sands clays and silts (glacis deposits foothill and surfaces)	DETRITIC (QUATERNARY)-MEDIUM	2	0.2	0	0	0	0	0	0							
Gravels sands silts and clays (medium and high terraces deposits)	DETRITIC (QUATERNARY)-HIGH	1	0.1	0	0	0	0	0	0							
Blocks boulders silts and clays (hillside deposits colluviums moraines)	DETRITIC (QUATERNARY)-HIGH	1	0.1	0	0	0	0	0	0							

CELLS MORE THAN 400 Bq/m³-CANARY ISLANDS (858 CELLS)

Wind sands subordinate	VOLCANICS (PIROCLASTIC AND LAVICS)-HIGH	14	1.6	65	10	15	3	4.6	10	1876	186	352.5	33	66	141
Welded polymictic and heterometric gap	DETRITIC-MEDIUM	1	0.1	3	1	33	0	0	29	421	161	225.4	30	32	324
Flows and cones of tephra basan nephritic and tephritic. Sub basalts	VOLCANICS (PIROCLASTIC AND LAVICS)-HIGH	5	0.6	13	0	0	0	0	14	240	72	77.3	28	40	69
Basaltic flows (and pyroclastics) basaltic and tephritic	VOLCANICS (PIROCLASTIC AND LAVICS)-MEDIUM	1	0.1	6	0	0	0	0	17	136	70	49.8	24	60	123
Flows and ignimbr. phonolitic Phonolitic	VOLCANICS	2	0.2	6	0	0	0	0	10	287	67	108	16	29	32

intrusives and pyroclastic flows. Ash & pumice subord	(PIROCLASTIC AND LAVICS)-LOW															
Flows and cones of tephra basan piroclasts subordinates of dispersion	VOLCANICS (PIROCLASTIC AND LAVICS)-VERY HIGHT	1	0.1	1		0	0	0	0			30				
Detrital alluvial deposits slope (including epiclastic and gravitational) and beaches	DETRITIC (QUATERNARY)-VERY HIGHT	1	0.1	2		0	0	0	0	14	16	15	1.41	14	15	16
Rhyolitic and trachytic flows with subordinate phonolites. Tuffs and ignimbritic gaps peralkaline	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	1	0.1	2		0	0	0	0	10	10	10	10	10	10	10
Ignimbritic tuffs and rhyolitic-trachytic flows. Salic dykes between 10- 65%	VOLCANICS (PIROCLASTIC AND LAVICS)-LOW	1	0.1	0		0	0	0	0							
Wehrlites pyroxenites and gabbros	IGNEOUS-VERY LOW	1	0.1	0		0	0	0	0	0						

CELLS MORE THAN 301-400 Bq/m³-PENÍNSULA (115 CELLS)(IN CANARY ISLANDS DO NOT EXIST)

Quartzites and slates	META-DETRITIC-LOW	37	32.2	32		4	13	0	0	21	717	130	144.4	47	69	173
Gravels sands silts and clays (medium and high terraces deposits)	DETRITIC (QUATERNARY)-HIGHT	1	0.9	2		0	0	0	0	97	106	102	6.3	97	102	106
Grauwakes and slates. Culm Facies	META-DETRITIC-LOW	7	6.1	5		0	0	0	0	20	157	95	55	59	84	144
Metamorphized acid rocks (orthogneisses migmatites) metarriolites (Olo Sapo) Gn.Pera Gn.Gland.	META-DETRITIC-LOW	2	1.7	15		0	0	0	0	10	283	94	69.4	47	82	139
Phyllites schists quartzites limestones shales and hornfels (metamorphic rocks)	META-DETRITIC-VERY LOW	7	6.1	3		0	0	0	0	33	117	89	48.2	54	116	117
Schists paragneiss carbonate levels /	META-DETRITIC-	8	7	11		0	0	0	0	18	201	85	53.6	48	72	114

calcosilicates (Villalcampo's Schists Duero Series Sistema Central's metasediments)	LOW															
Arkoses with boulders conglomerates and clays	DETRITIC-MEDIUM	7	6.1	44	0	0	0	0	11	298	78	59.4	30	67	103	
Limestones sandstones and shales	CARBONATED-LOW	27	23.5	15	1	7	0	0	18	374	77	93.2	26	41	85	
Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	IGNEOUS-LOW	6	5.2	7	0	0	0	0	24	143	74	41	49	60	104	
Sandstones and shales. Gongolaz Sandstones	DETRITIC-LOW	1	0.9	2	0	0	0	0	50	76	63	18.3	50	63	76	
Quartzites conglomerates sandstones and lutites. Purple Series and Constante or Bornova Formations	META-DETRITIC-LOW	11	9.6	4	0	0	0	0	38	71	56	14	45	57	66	
Slates	META-DETRITIC-VERY LOW	2	1.7	3		0	0	0	13	46	27	17	15	22	40	
Quartzites slates and limestones	META-DETRITIC-LOW	6	5.2	1	0	0	0	0			22					
Sandstones conglomerates and shales red (Buntsandstein Facies)	DETRITIC-LOW	1	0.9	1	0	0	0	0			11					
Gravels sands clays and silts (glacis deposits foothill and surfaces)	DETRITIC (QUATERNARY)-MEDIUM	1	0.9	0	0	0	0	0	0							

Table F: Lithologies associated with categories 1 and 2 in 10 kmx10 km cells system. AM: arithmetic mean; SD: standard deviation.

TYPOLOGY	CELLS		TOTAL DATA NUMBER ON THOSE CELLS	Data of > 300 Bq/m ³		Data of > 1000 Bq/m ³		Bq/m ³							
	N°	%		N°	%	N°	%	MIN	MAX	AM	SD	1 st quartile	MEDIAN	3 rd quartile	
CELLS MORE THAN 400 Bq/m ³ -PENÍNSULA AND CANARY ISLANDS (858 CELLS)															
Sandstone slate quartzite limestone and conglomerate	17	2	20	7	35	2	10	20	4969	485	1103.5	57	104	413	
Biotite granitoids	385	45	1404	175	12	17	1.2	10	15403	192	593.4	60	110	202	
Felsic gneiss and metabasite	75	9	151	23	15	2	1.3	10	2068	184	233.4	63	117	208	
Calc-alkaline volcanic rocks (andesite dacite riolite). Shoshonite lamproite	21	2	88	10	11	3	3.4	10	1876	154	309.8	29	56	121	
Peraluminous granitoids (collisional type)	27	3	115	14	12	0	0	10	939	151	164.3	51	95	176	
Intermediate and basic igneous rocks	7	1	5	1	20	0	0	42	390	149	140.2	59	106	206	
Slate sandstone quartzite and limestone or volcaniclastic rock	224	26	406	32	8	5	1.2	10	5676	146	393.8	39	71	125	
Slate schist sandstone limestone bituminous shale and lidite	38	4	53	4	8	0	0	10	457	136	103	61	110	184	
Conglomerate gravel sand sandstone siltstone and mudstone. Fluvial and marine terraces	19	2	70	2	3	0	0	6	344	89	75	35	64	121	
Basaltic lavas and pyroclasts	6	1	11	1	9	0	0	10	421	83	120	19	32	109	
Limestone and dolostone	28	3	61	1	2	0	0	10	315	68	61.9	29	43	93	

Garnet-bearing graphitic mica schist (Veleta)	10	1	9	0	0	0	0	16	87	59	26.9	46	61	80
CELLS MORE THAN 301-400 Bq/m ³ -PENÍNSULA AND CANARY ISLANDS (115 CELLS)														
Intermediate and basic igneous rocks	5	1	4	1	25	0	0	102	402	209	147.2	108	157	310
Slate schist sandstone limestone bituminous shale and lidite	38	4	26	3	12	0	0	33	717	125	145.8	47	69	162
Conglomerate gravel sand sandstone siltstone and mudstone. Fluvial and marine terraces	3	0.3	9	0	0	0	0	19	298	117	78	85	97	135
Peraluminous granitoids (collisional type)	1	0.1	14	0	0	0	0	10	283	100	68.7	54	84	143
Limestone and dolostone	6	1	13	0	0	0	0	28	202	90	58.3	50	72	125
Biotite granitoids	5	1	30	0	0	0	0	24	201	84	45	49	71	117
Sandstone slate quartzite limestone and conglomerate	39	34	40	1	3	0	0	11	374	58	64.7	20	38	72
Slate sandstone quartzite and limestone or volcaniclastic rock	16	2	7	0	0	0	0	18	140	54	43.4	19	52	66
Marl and marly limestone. Turbidic clay marl. Sandy limestone sandstone sand and marl	2	0.2	2	0	0	0	0	24	47	34	18.3	21	34	47

Table G: Radon concentration data > 1000 Bq/m³

CELL	Point number	City	Radon Concentration(Bq/m ³)	GEOLOGY				Gamma Radiation Exposure	Karst 1:1 M		Faults 1:1 M	
				Lithostratigraphies 1:200.000			OneGeology 1:1M		Rate (µR/h)	Karst Area YES/NO	Karst Typology	OneGeology (proximity 2km)
				TYPE	PERM.	ORIGIN	TYPE	Faults YES/NO				Type
SA73	1	Villar de Argañán	15403	Slates grawacks and sporadic carbonated levels. Schist-grawack Complex	LOW	META-DETRITIC	Slate sandstone quartzite and limestone or volcaniclastic rock	11	NO		YES	Strike slip fault
SA116	2	Casillas de Flores	11100	Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	LOW	IGNEOUS	Biotite granitoids	14	NO		NO	
SA102	3	La Alamedilla	7400	Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	MEDIUM	DETRITIC	Sandstone slate quartzite limestone and conglomerate	11	NO		NO	
SA116	4	Casillas de Flores	6374	Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	LOW	IGNEOUS	Biotite granitoids	14	NO		NO	
SA74	5	Saelices el Chico	5676	Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	LOW	META-DETRITIC	Slate sandstone quartzite and limestone or volcaniclastic rock	11	NO		NO	
SA88	6	Espeja	4969	Sandstones silts and ocher conglomerates (Toro-	MEDIUM	DETRITIC	Sandstone slate quartzite limestone and	13	NO		NO	

				Corrales's arkoses Geroma's silts Villallazán's conglomerates)			conglomerate					
SA74	7	Castillejo de Martin Viejo	4070	Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	LOW	META- DETRITIC	Slate sandstone quartzite and limestone or volcanoclastic rock	11	NO		YES	Strike slip fault
SA74	8	Castillejo de Martin Viejo	2375	Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	LOW	META- DETRITIC	Slate sandstone quartzite and limestone or volcanoclastic rock	11	NO		YES	Strike slip fault
SA102	9	La Alamedilla	2294	Sandstones silts and ocher conglomerates (Toro- Corrales's arkoses Geroma's silts Villallazán's conglomerates)	MEDIUM	DETRITIC	Sandstone slate quartzite limestone and conglomerate	11	NO		NO	
SA74	10	Saelices el Chico	2194	Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	LOW	META- DETRITIC	Slate sandstone quartzite and limestone or volcanoclastic rock	11	NO		NO	
SA102	11	La Alamedilla	2150	Sandstones silts and ocher conglomerates (Toro- Corrales's arkoses Geroma's silts Villallazán's conglomerates)	MEDIUM	DETRITIC	Sandstone slate quartzite limestone and conglomerate	11	NO		NO	
PO44	12	O Porriño	2068	Gravels sands silts (alluvial deposits valley bottoms and low terraces in rivers mainly)	VERY HIGH	DETRITIC (QUATER NARY)	Conglomerate gravel sand sandstone siltstone and mudstone. Fluvial and marine terraces	23	NO		YES	Strike slip fault
S42	13	Ramales de la Victoria	1972	Gravels sands silts and clays (medium and high terraces deposits)	HIGH	DETRITIC (QUATER NARY)	Sandstone slate quartzite limestone and conglomerate	6	YES	Mesozoic and Paleogene Tectonize	YES	Strike slip fault

											d Limestones and Dolomites		
SA73	14	Gallegos de Argañan	1937	Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	LOW	META-DETRITIC	Slate sandstone quartzite and limestone or volcanoclastic rock	15	NO			NO	
TF51	15	La Orotava	1876	Wind sands subordinate	HIGH	VOLCANICS (PIROCLASTIC AND LAVICS)-HIGH	Calc-alkaline volcanic rocks (andesite dacite riolite). Shoshonite lamproite	NO DATA	YES	Recent Volcanic Formations (Canary Islands)		NO	
SA73	16	Gallegos de Argañan	1868	Arkosic sandstones schists shales conglomerates and volcanics rocks. Inf. and Sup. detrital series.	LOW	META-DETRITIC	Slate sandstone quartzite and limestone or volcanoclastic rock	15	NO			NO	
T36	17	Gandesa	1776		LOW	DETRITIC	Limestone and dolostone	6	NO			NO	
TF51	18	La Orotava	1628	Wind sands subordinate	HIGH	VOLCANICS (PIROCLASTIC AND LAVICS)-HIGH	Calc-alkaline volcanic rocks (andesite dacite riolite). Shoshonite lamproite	NO DATA	YES	Recent Volcanic Formations (Canary Islands)		NO	
M26	19	Torrelozanes	1545	Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	LOW	IGNEOUS	Biotite granitoids	14	NO			NO	
OR17	20	Ourense	1438	Hercynian acid plutonic	LOW	IGNEOUS	Biotite granitoids	18	NO			YES	Strike slip

							conglomerate					
C88	27	Teo	1184	Schist schists-graphites phyllites quartzites ampelites and lydites. Nogueira's Group Paraño and Rábano's Formation	LOW	META-DETRITIC	Slate schist sandstone limestone bituminous shale and lidite	15	NO		YES	Strike slip fault
LU42	28	Outeiro d Rei	1181	Slates shales meta-sandstones and amphibole gneisses. Villalba series	LOW	META-DETRITIC	Slate sandstone quartzite and limestone or volcaniclastic rock	9	NO		NO	
AV62	29	Navarredonda de Gredos	1153	Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	LOW	IGNEOUS	Biotite granitoids	16	NO		YES	Strike slip fault
AV72	30	Cuevas del Valle	1119	Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	LOW	IGNEOUS	Biotite granitoids	16	NO		NO	
LU52	31	Castroverde	1085	Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	LOW	IGNEOUS	Biotite granitoids	11	NO		YES	Strike slip fault
C34	32	Coruña	1083	Hercynian acid plutonic rocks (granites granodiorites quartzodiorites)	LOW	IGNEOUS	Biotite granitoids	12	NO		NO	
LO10	33	Navarrete	1073	Gravels sands silts and clays (medium and high terraces deposits)	HIGH	DETRITIC (QUATERNARY)	Limestone and dolostone	7	NO		YES	Strike slip fault
PO16	34	Meis	1025	Gravels sands silts (alluvial deposits valley bottoms and low terraces in rivers mainly)	VERY HIGH	DETRITIC (QUATERNARY)	Biotite granitoids	17	NO		YES	Strike slip fault

