

RADON PRIORITY AREAS

Definition, Estimation and Uncertainty

by

Peter BOSSEW

German Federal Office for Radiation Protection, Berlin, Germany

Scientific paper
http://doi.org/10.2298/NTRP180515011B

Posing a substantial risk to human health, indoor radon has increasingly been subject to regulation. One key concept is the one of radon prone or priority areas, which are understood as regions where prevention, mitigation or remediation action should be taken with priority. Radon priority areas must be defined, and once defined, estimated from data. Radon priority area estimation or delineation amounts to a classification problem, as a domain (a country) has to be divided into mutually exclusive zones of different “priorityness”. Classifying areas into priority zones entails decisions about action to be implemented. Touching stakeholder interests, may prove economically and politically costly. Therefore, decisions should be justifiable, which implies, among other, reliability of the radon priority area estimate, meaning statistically, controlling for estimation uncertainty. Some aspects of radon priority area definition, estimation, and uncertainty will be discussed in this paper.

Key words: radon priority area, classification, EURATOM, basic safety standard

INTRODUCTION

Indoor radon (Rn) is acknowledged an important health hazard. According to epidemiological studies, indoor Rn is believed to be the second cause of lung cancer after smoking (WHO 2009 [1]). Therefore, indoor Rn has increasingly been subject to regulation for the last years; In Europe, most importantly the EU directive on basic safety standards for protection against ionizing radiation (BSS; EU 2013 [2]) which must be transposed into national law by the EU Member States. Among many other, it states that national reference levels (RL) for indoor Rn concentration in dwellings and workplaces alike must be set 300 Bqm⁻³ and it requires delineation of Rn priority areas (RPA), i.e., areas in which activities related to prevention and remediation of high indoor Rn concentrations should be taken with priority. The BSS definition is vague, stating (Art.103.3) that an RPA is an area where it is expected that in a significant number of houses the annual mean Rn concentration exceeds the national RL. Therefore, an operable definition of RPA has to be developed in each EU country. It is based on a Rn measure, for example, the mean over a geographical unit (grid cell, municipality, etc.) or the probability that within the unit indoor Rn exceeds a reference level.

Existing solutions are pragmatic in the sense that they have to rely on available data and on external “political” parameters such as reference levels, spatial units to which the term “area” refers and tolerable uncertainty. The task of statistically sound RPA estimation is still under development and will probably remain on the international research agenda for some time.

Once delineated, certain tasks have to be performed in RPA according to the National Rn Action Plan, which has to be developed according to the BSS to “address long-term risks from Rn exposures in dwellings, buildings with public access and workplaces” (Art. 103.1). Items are listed in Annex XVIII. First, Rn concentration must be measured in workplaces and public buildings. This may be followed by mitigation and remediation, which can be expensive. Clearly, stakeholder interests are touched by RPA delineation. Issues discussed here, in particular, QA aspects, are subject of the ongoing EU H2020 Metro RADON project (http://metroradon.eu/).

Concepts, definitions, and estimation methods are shown in this paper. The question of classification errors is discussed in the example of the German binomial approach to estimate RPA. Finally, open problems and research needs are addressed.

CONCEPTS AND DEFINITIONS OF RADON PRIORITY AREAS

RPA definitions differ by the concept and by aspects of the practical implementability. While the target quantity is always – per BSS definition – the annual mean indoor Rn concentration and its limiting value the RL, the meanings of “significant number”,

* Author's e-mail: pbossew@bfs.de
“area” and “exceed” are open to interpretation. Also “annual” is problematic because it is known that annual Rn concentrations in one room or building vary between years. The more appropriate term is, therefore, “estimated long-term”.

For example, one may argue that Rn concentrations in actually existing buildings do not define RPAs, but concentrations that are expected to occur for geogenic reasons in a location, if a building of certain type was there, regardless of whether there is one (a concept analogous to the one of seismic vulnerability of a location). Another rationale is based on a deeper understanding of radiation protection: the RPA concept can relate to cutting extremes, thus individual protection, or reducing the bulk exposure, thus reducing the collective dose. This has been pointed out in [3]. These different concepts would lead to different RPA maps.

The overall goal of the BSS concerning Rn is a sustainable reduction of the risk posed by indoor Rn. This leads to the priority concept, as implicit in the notions of RL and RPA. It implies that Rn exposure should be reduced everywhere, if possible with lower priority (given usually limited resources); after all, Ann. XVIII (13) states as part of the Rn action plan: [Establish] “long-term goals in terms of reducing lung cancer risk attributable to radon exposure”. Some have concluded that graded approaches match this philosophy best; either by defining different RPA level classes requiring the action of different priorities or by subsequent enlargement of RPA according to completed tasks. The idea of optimization as motivation underlying the BSS has been discussed more extensively in [4].

While the EU Member States had to transpose the general concept of the European BSS into national radioprotection law by February 2018, technical issues are left for sub-legislation or ordinances, currently (June 2018) still under elaboration in many countries. Therefore, no authoritative overview on RPA definitions is available so far. It seems, however, that most will choose a definition of the type: An area is labelled RPA if in that area, prob( C > RL ) > probability threshold p0 ( C – estimated long-term mean indoor Rn concentration). Most chose RL = 300 Bqm⁻³ (the highest permitted by BSS), some 200. A common probability threshold is 10 %.

Examples are Finland, Germany, Greece, Montenegro [5] and Spain which chose RL = 300 Bqm⁻³ and p₀ = 10 % (for Germany, derived from ground-floor rooms in buildings with basement only; for Spain, from the ground or first floor rooms only). Ireland has chosen RL = 200 Bqm⁻³, p₀ = 10 %, Belgium and Luxembourg chose RL = 300, but three priority levels, p₁; prob <1 %; p₀ II: prob between 1 and 5 %; p₀ III: prob > 5 %. Note that this information reflects discussions from end 2017 – early 2018 and final legal decisions may turn out different.

Alternatively, some chose definitions of the type, an area is labelled RPA, if the mean indoor concentration in it exceeds the RL. Examples are Austria (proposed two priority levels of RPA, “medium risk” if AM(C) > 150 and “high risk”, if above 300 Bqm⁻³ – also this information is preliminary) and Switzerland, which also opted for two priority levels with thresholds 100 and 200 Bqm⁻³. For comparison, assuming log-normal distribution with GSD = 2 within a 10 km × 10 km cell (about realistic by experience), AM(C) = 300 corresponds to prob( C > 300) = 36 %.

Yet another definition, quite different from the above, and not applied by any European country so far, to my knowledge, could consist in labelling RPA those geographical units (grid cells, municipalities) which represent an upper percentile of Rn measures; for example, say, the 5 % municipalities with the highest mean Rn concentrations. Such strategy would also reflect the prioritization idea implicit in the BSS and limitation of resources: once Rn problems have been largely tackled in the upper 5 % of municipalities, one may start working the second highest 5 %, and so on, as long as found reasonable and feasible.

A final overview of RPA definitions including reference to respective laws or decrees is planned for the Metro Rn project, mentioned above.

ESTIMATION OF RADON PRIORITY AREAS

Once defined, the RPA 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. The decision about whether a geographical unit shall be assigned RPA or not – or in the case of multinomial definition, i.e., grades of “Rn priorityness”, which grade or class should be assigned to that unit, 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 to “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 one of the 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. 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, [6]), 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 [7, 8], quantile regression [9] and local regression [10]. A hierarchical regression model where spatial dependence enters via lithology as a predictor has been proposed in [11] and further developed as a generalized additive mixed model in Austria ([12], not yet fully published). Methods based on the indicator (co-)kriging approach have not been explored yet for RPA estimation, to my knowledge. Literature is growing fast in this field and new techniques pop up in rather a fast pace.

Common estimation supports, i.e. the areas over which statistics are computed, are grid cells (e.g. 10 km × 10 km grids have been chosen for the European indoor Rn map, part of the European Atlas of Natural Radiation and for the current version of the German maps of RPA and of the geogenic Rn potential (GRP, see below)) or municipalities. The latter 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 [13].

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 cokriging family. Block estimates overestimation supports are classified according to RPA definition. (1b) As a simpler variety, parametric estimation of the primary variable from the secondary by regression. For example, estimation of exceedance probability of indoor Rn concentration, prob(C > RL), has been demonstrated in [14] (predictor: GRP) and in [15] (several geogenic predictors) using logistic-type regression. (1c) Another option is modeling the bivariate distribution by a copula, from which desired statistics can be derived [16, 17]. Opposed to 1(a), methods 1(b), and 1(c) 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 fig. 1 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 [18].

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 cokriging type, being parametrical, but not carrying the entire information of the variables.

Figure 1. Top: truth table; “effect”: Area is 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 threshold of 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 below.
DATA

Whatever its definition, estimation of RPA 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 the 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 [6] is based on more than 1.1 million individual measurements in its current state (mid-2018).

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). Precision (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 [19]. 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 predictors 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 the 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 the regression-type analysis) can be a statistical challenge.

The 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.

UNCERTAINTY AND VALIDATION

As said, estimation of RPA 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 the true variability of the mapped quantities on the 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 multinominal 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 the uncertainty of correct assignment of an area to an 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 the occurrence of extremes is measured by other statistics such as high quantiles or dispersion measures. One may, therefore, think of 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.

A task, largely untackled so far, to our knowledge, is a 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.

EXAMPLE

The German approach (at the time of writing, May 2018, still under discussion; the final definition to be laid down in an ordinance still under development) shall be shortly cited as an example. RPA definition is: an area is labelled RPA if prob(C > RL) > γ p₁ otherwise, non-RPA, i.e. a binomial scheme was chosen. C – long-term mean indoor concentration in ground floor dwellings of houses with a basement, RL – reference level; p₁ – the same probability estimated for the entire territory of Germany. RL has been set 300 Bqm⁻³, the multiplier γ = 3, and p₁ ~ 3 %. The definition is in approximate accordance with the one applied by other countries, prob(C > 300) ~ 10 %. Since the German indoor Rn database is fragmentary and insufficient for direct RPA estimation from indoor Rn data, the geogenic Rn potential (GRP) is used as a secondary variable, because a dataset (about 4,500 locations) covering the territory about representativity is available. (This approach has also a conceptual advantage, see above). The task consists in finding a derived or secondary threshold for the GRP, so that classification according to this threshold conforms with the (hypothetical) one in compliance with the primary RPA definition. This has been achieved by cross-tabulation, based on 10 km × 10 km cells for which indoor Rn data are available. The GRP has been estimated by geostatistical means including geology as categorical deterministic trend predictor [20].

Additionally, a constraint on estimation confidence has been imposed: first and second-kind classification error rates shall be below 10 %. Practical implementation was via a ROC-type procedure on the 2 × 2 truth table (fig. 1). The result is factually a trinomial classification, shown in fig. 2, as apart from cells assigned RPA (red) and non-RPA (green) with 90 % confidence, some cells remain un-classified, shown in yellow because confidence is not sufficient.

Being estimates (and hence the RPA being “random objects”), the class limits have some degree of uncertainty. By bootstrapping one finds that the 90 % confidence limits for the upper limit are (41.2, 48.0) and for the lower limit, (14.3, 23.5).

In current reasoning, the final legally binding RPA shall be defined on district or municipality level or even below, taking advantage of locally available knowledge about geology and settlement patterns, which central planning on the federal level cannot deliver. In the non-assigned areas (yellow in fig. 2), further measurements shall clarify the situation. An open question consists in the fact that also in non-RPA areas...
tain risk of indoor concentration above RL is present that would go undetected since in these areas no action is envisaged.

**CHALLENGES**

**Workplaces and dwellings**

So far in all cases (to our knowledge), RPA estimations have been based on indoor Rn concentrations in dwellings, although their primary legal impact is on workplaces and public buildings. However, because of their different physical characteristics, in general, it cannot be assumed, that regionalized estimates of Rn in dwellings and workplaces are equal, or equal between different types of workplaces. In fact, different types of workplaces would afford different RPA maps. Evidently, there are practical limits to this. Currently, databases of Rn in workplaces are still poor in most countries, and discussions about how to manage the problem are ongoing. Logistically and administratively, Rn surveys on workplaces seem to be not less difficult than the ones of dwellings. Given their psychological importance for society, most extensive databases seem to exist for schools and kindergartens, but comparison with dwellings is difficult [21].

**Multivariate estimation**

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 use 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 a 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. Research on the GRHI is ongoing.

**CONCLUSIONS**

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

However, an important message is that there is no "natural" definition of RPA and consequently, no such thing as a "true RPA". Delineated RPAs always depend on their definition – resulting from political decisions, stakeholder interests, availability of resources and of databases – and to some degree also on estimation method. As results of statistical estimation, RPAs are uncertain objects. Communication of RPA uncertainty to the public and to decision makers is another challenge, not to be discussed here, but from experience known not to be easy.

The methodology of statistically sound RPA estimation is still evolving. Particular challenges consist of multivariate estimation and in establishing uncertainty budgets. On the empirical side, databases are often and still not sufficient to fulfill QA requirements satisfyingly.

**ACKNOWLEDGEMENT**

This work is supported by the European Metrology Programme for Innovation and Research (EMPIR), JRP-Contract 16ENV10 MetroRADON (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.

Parts of this text were taken over or adapted from other documents by the author, notably a discussion paper submitted to the EURATOM Art. 31 committee (together with K. Gehrcke, BfS), an abstract submitted to ICHLERA 2018 (9th Intl. Conf. on High Level Environmental Radiation Areas, Hirosaki, Japan, Sept. 2018) and documents produced in course of the Metro RADON project.

**REFERENCES**


Авертабне Лунг Канерс, Radiation Protection Dosiometry, 177 (2017), 1-2, pp. 87-90


Received on May 15, 2018
Accepted on July 12, 2018

Петер БОССЕВ
ПРИОРИТЕТНА ПОДРУЧЈА РАДОНА
Дефиниција, процена и несигурност

Представљајући значајан ризик по људско здравље, радон у затвореном простору је један од најопаснијих радонцама. Насупрот исти на ризику, одређени радонцама се не може превеновати. Математички модел радонцама је једна од најважнијих области истраживања. Математички модели радонцама се могу користити за опредељивање места где је ризик од радонца већи. Математички модели радонцама се могу користити за опредељивање места где је ризик од радонца већи. Математички модели радонцама се могу користити за опредељивање места где је ризик од радонца већи.