LOCAL PROBABILITY OF INDOOR RADON CONCENTRATION TO EXCEED THE THRESHOLD ESTIMATED FROM GEOGENIC RADON POTENTIAL

by

Peter BOSSEW

German Federal Office for Radiation Protection, Berlin, Germany

Scientific paper
http://doi.org/10.2298/NTRP1701070B

Indoor radon has been recognized as an important air pollutant. Based on epidemiological evidence, it is estimated that indoor radon is the second cause of lung cancer after smoking. As a consequence, one tries to limit exposure through regulations concerning the remediation of the existing and prevention of future exposure. In this context, an essential task is the delineation of areas in which it can be expected with certain confidence that time-averaged indoor radon concentrations in dwellings and workplaces exceed the reference level. These are called radon priority areas to denote that these are areas in which remedial and preventive action has to be implemented with priority. There are different definitions of radon priority areas and different methods to estimate them from data. In Germany, the current approach uses the geogenic radon potential as the predictor. However, legal reference levels pertain to indoor radon concentration, not to the geogenic radon potential. One therefore has to identify derived reference levels for geogenic radon potential through statistical association of both quantities.
This paper presents a method to derive the local probability that indoor radon concentration exceeds a threshold, given the local geogenic radon potential. The relationship can be used to derive geogenic radon potential reference levels which in turn serve to define radon priority areas.

Key words: indoor radon, geogenic radon, radon priority areas

INTRODUCTION

By most experts, indoor radon (Rn) is considered an important pollutant which impairs human health. Indeed, epidemiological studies have shown that indoor Rn is the second cause of lung cancer after smoking (for details see WHO [1] and references therein). As a consequence, Rn exposure is increasingly subject to regulation. In Europe, the latest initiative is the European Council Directive "laying down basic safety standards (BSS) for protection against the dangers arising from exposure to ionising radiation" [2] which EU member states are required to implement as national law. Among many other radioprotection issues, BSS address Rn exposure. In this respect, most important are a reference value not exceeding 300 Bqm^{-3} for dwellings and workplaces alike and the requirement to establish a national Rn action plan. This includes identification of regions with a significant number of buildings with elevated Rn concentration exceeding reference levels. Such areas have sometimes been called Rn-prone areas. In view of practical consequences, the term radon priority area has been recently proposed. The BSS (article 103, par. 3) do not provide a specific term*.

The procedure of how to actually define Rn priority areas (RPA) and how to estimate them has to be elaborated by EU Member States (plus others that chose to adopt the BSS). Evidently, one will choose a method that can be implemented using nationally available input data. (There are attempts to perform similar tasks on the European level, *e. g.* [4] and [5].) In this article, I focus on the German approach, but the presented method can be transferred to structurally similar problems.

At the current (late 2016) state of discussions, the approach adopted in Germany is as follows (further discussions may lead to modifications, though). Radon priority areas will be defined through the geogenic Rn potential (GRP, see below). The decision whether an area is RPA shall depend on whether a GRP threshold is exceeded. In other words, not actually observed exceedance of indoor Rn concentration – an approach chosen by other countries – is the criterion, but the one of a geogenic control quantity. The rationale of this approach is twofold:

– *Practical aspect.* The indoor Rn database of Germany is precarious; see details in section „Indoor Rn data" below. Available data are by far insuffi-

* Author's e-mail: pbossew@bfs.de

* This introductory paragraph has partly been take over from [3]

cient to create a RPA map covering the country by simply applying a threshold to – for example – mean indoor concentration or probability to exceed a threshold per municipality. Approaches of this type have been chosen in some other countries.

– *Physical aspect*. Mean indoor Rn concentrations are subject to long-term changes caused by changes in house construction techniques, changing regulations and changing habits of people concerning air exchange. Therefore, a RPA defined from indoor Rn concentration is not temporally constant. On the other hand, geogenic controls do not vary except over the geological time scale.

The critical point of this latter approach is that reference values of the GRP which are linked to ones given in the BSS that apply for indoor Rn have to be found. This is achieved by some kind of a „transfer model" between GRP and indoor Rn. One will notice that by deriving a GRP threshold from that of indoor Rn using observed indoor Rn data, the problem of secular variability of indoor Rn enters through the backdoor, so to say. We are aware of this fact, but it has been decided that the current indoor database shall be used as a baseline which is *frozen* and serves for establishing the said link to the GRP. However, this decision may be reviewed in the future, but it cannot be expected that a better indoor Rn database will be available for years to come.

This paper will present one method for linking the GRP with indoor Rn concentration and for deriving GRP thresholds that define RPA. One possible RPA definition is: a grid cell of a map is assigned RPA if its GRP makes it likely – within a certain tolerated second kind error – that the probability that in this cell (U) indoor Rn concentration (C) exceeds the reference value c (λ) times of the overall probability in Germany.

Formally:

$$\text{Cell U is RPA}: \quad \text{prob}(C > c \mid \text{U}) \\ \text{prob}\,(C > c \mid \text{entire Germany})$$

with c = 100 Bqm^{-3} and λ =3. The overall probability equals about 0.1, so that an RPA is defined over that probability equalling 0.3. The result is almost the same if c = 300 Bqm^{-3} is chosen and again λ = 3. The overall probability is about 0.03 in this case, the one in RPA > > about 0.1. Grid cells are defined 10 km 10 km in the following.

But as the parameters (c, λ) depend on political decisions, they are so to say political parameters, *i. e.* deliberate inputs from the analytical point of view presented here. The role of the tolerated second kind error (another political parameter) will be discussed in more detail below.

The technical difficulty is owed to the relatively poor relation between GRP and indoor Rn. This not only renders estimation of a transfer model difficult, but also induces uncertainty to estimated quantities such as regression parameters or derived GRP thresholds and, hence, to delineation of the resulting RPA.

Some results have been shown at the GeoENV 2016 conference, [6].

DATA AND METHODS

Definitions

Geogenic radon potential (GRP) measures, *what earth delivers* in terms of Rn or, more specifically, the availability of geogenic Rn to exhalation from the ground or for infiltration into buildings. Importantly, GRP is just one of the factors that controls indoor Rn. The other two are construction type (which determines the ease of Rn to infiltrate) and the habits of inhabitants or users of the building regarding air exchange.

GRP depends on the strength of the Rn source and its mobility in the ground. We use a definition by [7] (slightly modified)

$$\text{GRP} = C(\text{ground})/(-_{10}\log(k) - 10)$$

where C(ground) is the Rn concentration in soil gas [kBqm^{-3}] and k – soil permeability in [m^2], both measured through a defined protocol (for Germany: "Kemski protocol", [8]). It can be shown that for mean permeability this is about equal the advective flux, normalized to pressure difference.

GRP data

The German database comprises about 4000 data. From these and a 1:1 M geological map as categorical covariate, a *predictor map* was created by interpolation, (details in [9]). The map consists of 10 km 10 km pixels to whose centres estimated values of the RPA were assigned. (A spatially finer resolved version is currently under work.)

Indoor Rn data

The German database of indoor Rn concentrations consists of 60,700 measurements in dwellings (status early 2016). Their quality in terms of measurement QA, georeferencing and house characterization is very different. More reliable datasets can be generated by data filtering and modelling. Filtering means including only data which conform to given quality criteria, modelling to normalizing suboptimal data to a given standard, *e. g.* by applying seasonal corrections. Certainly, modelling induces additional uncertainty. On the other hand, the number of truly qualified data is too low for further evaluation of the relationship indoor Rn ~ GRP. As a standard situation, we defined the long-term mean of ground-floor living or sleeping rooms in buildings with a basement. Workplaces are excluded because their database is still far too small. An additional problem is strong spatial data clustering, as they originate from different regional surveys with different purposes. Here a set of 39,809 data is used, partly original-filtered, partly modelled. The filtered or modelled data are aggregated into the same 10 km × 10 km cells as for the

GRP. Assuming local log-normality, in each cell the probability prob($C > c$) is estimated, where C is the indoor Rn concentration and c – a reference level. For how this is actually done, see [4, p. 30].

The association model

There are different ways to quantify the association between spatial variables. The model discussed here in detail is essentially a logistic-type regression of collocated data. Two previously explored models will shortly be addressed at the end of the section.

The collocated data which underlie the present model are the cell statistics explained in the previous section, the independent variable or predictor is the GRP, the dependent or response variable, the empirical p: = prob($C > c$).

This simple approach is modified due to the following motivation: there is less interest in an optimal fit – *e. g.* by minimizing least squares – than in an estimate which ensures a given degree of conservativeness, measured by the rate of second kind error, *i. e.* erroneous underestimation of that probability, with a given target value β. Practically, this means that the number of cells in which the probability is underestimated by the model shall not be larger than $\beta \times$ (number of cells).

The actual regression models are:

„logi": $y = [1 + \exp(-a - b\ \psi_q(x))]^{-1}$

„probi": $y = [a + b\ \psi_q(x)]$.

where ψ_q denotes the Box-Cox transform, $\psi_q(x) = (x^q - 1)/q$ if $q \neq 0$, $\psi\ x = \ln(x)$ otherwise; Φ – cumulative normal function; x GRP, y p. The purpose of introducing the Box-Cox transform is to render the model more flexible. After all, the *true* relationship is unknown. Different versions of the Box-Cox transform could be used.

The target function is the number of data points *above the curve* which shall be $\beta \times$ (number of points). Obviously, this has no unique solution. We therefore introduce the constraint that the rate of underestimated y shall be distributed about uniformly over the range of the predictor x. This is achieved as follows: preprocess the data by applying a running window on x (x^* the median of the window), in which y^* is set the $(1 - \beta)$ quantile of the y within each window. Then an least square (LSQ) fit is performed on (x^*, y^*).

Technically, the fit is performed by *brute force*: sample (a, b, q) from a cube in the parameter space by raster search and find the minimal loss function which is the squared deviation from the target function. A new, smaller search cube is centred around the previously found optimal (a_0, b_0, q_0), and so on, until a stopping condition is met. This method is not elegant, but quite fast with modern computers and seems to be robust.

RPA are estimated according to the following procedure: given are the threshold (c) of C exceedance probability, the probability p_0 and the tolerable second kind error rate β. The task is to derive GRP_0, with a GRP threshold corresponding to said requirements. This is achieved by inverting the regression model in the graph (GRP, p_0). Find p_0 on the y-axis and identify GRP_0 so that (GRP_0, p_0) lie on the regression curve. GRP_0 defines RPAs as GRP > GRP_0. (The latter procedure is not entirely correct from the point of view of regression theory; instead, one had to treat p as independent and GRP as a dependent variable for this purpose and invert the entire analysis accordingly. However, here the main target was to find a dependency of p on GRP.)

Computations were performed with homemade software.

Uncertainties

Apart from the uncertainties of the individual input data of GRP and C, generation of cell-assigned GRP and p involves estimation uncertainty. The choice of the regression model involves structural uncertainty. The fitting procedure described above, again, involves estimation uncertainty. It could possibly be assessed by bootstrapping, but this was beyond the scope of this exercise.

Previous attempts

I want to shortly mention previous attempts to quantify the association GRP – C. One consists in constructing a copula of the joint distribution of GRP and C. A Gumbel copula model was used for the purpose, among other things because its parameter is relatively easy to estimate since it is related to the Kendall correlation between the two. This was estimated by extrapolating the lagged Kendall correlation τ[GRP(x), $C(x + h)$] towards $h = 0$. (An alternative is estimating the two on a common grid, or on the locations of the other. The methods agree reasonably well.) Technical problems involve the estimation of true univariate distributions (done by declustering), the estimation of $\tau(0)$ or data collocation, estimating the shapes of the distribution tails, choice of the copula model (implies structural uncertainty) and numerical integration of the copula to retrieve the conditional distribution of one of the other. The practical experience is that the procedure is too complicated to be useful for routine application. The method seems sensitive against misestimating the true distribution. For details see [10, 11].

A second, much simpler approach is based on cross classification using the ROC technique (receiver operating characteristic). The method yields reference levels of GRP derived from a given reference level of indoor Rn (C). For details, see [12]. As a cross-classification method, it conveys less information than a full regression- type model GRP C. (Optimizing cross-

-classification can be understood as a kind of regression between classed variables.)

RESULTS

Model of indoor Rn exceedance probability

As an example, fig. 1 shows the result for threshold $c = 100$ Bqm^{-3} and $\beta = 0.15$ (the degree of conservativeness, *i. e.* second kind error rate of 15 % at most). The noisy grey curve represents the binned data (x^*, y^*), the blue and red ones the „logi" and „probi" fits, respectively. Following the concept given in the section *association model*, the curves are fitted so that 15 % of data points lie above the curves.

The parameters of the fits are given in tab. 1. We recognize that the „probi" fit performs slightly better (lower mean residual sum of squares MRSS, higher explained variance). In the practically relevant range of the GRP, say 10 to 80, the models coincide quite well.

Figure 2 show maps of the empirical (left) and modelled (right) probability that indoor Rn concentration C exceeds threshold $c = 100$ Bqm^{-3}. For the right map, „conservativeness" has been set 15 %.

Radon priority areas

If we set a probability threshold $p_0 = 0.3$, the derived GRP threshold, GRP0, in dependence of conservativeness β is shown in fig. 3, assuming the „probi" model. As menthioned, indoor Rn threshold has been set $c = 100$ Bqm^{-3}. For example, given the allowed second kind error rate $\beta = 0.15$, the threshold prob ($C > 100$) = 0.3 corresponds to GRP threshold $GRP_0 = 25$.

Using GRP_0 as the index for classifying GRP, RPAs depend on three input parameters: indoor Rn threshold c, exceedance probability p_0 and conservativeness β. Evidently, these inputs are not entirely independent, *i. e.* different combinations of (c, p_0, β) can give the same RPA pattern.

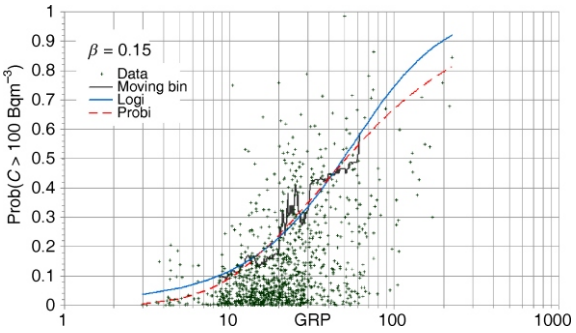

Figure 1. Relation between GRP and the probability that the indoor Rn concentration C exceeds the threshold $c = 100$ Bqm^{-3}. Crosses: data; full and dashed curves: fitted „logi" and „probi" models, respectively, for „degree of conservativeness" $\beta = 0.15$

Table 1. Parameters of the „logi" and „probi" fits corresponding the full and the dashed graphs in fig. 1. MRSS – mean residual sum of squares

	Logi model	Probi model
Box cox q	0.151	–0.196
a	–4.21	–4.10
b	0.796	1.49
MRSS	1.58 E-3	1.30 E-3
Explained variance	92 %	93 %

For $c = 100$ Bqm^{-3}, $p_0 = 0.3$ and different values of β, one finds RPA shown in fig. 4. For example, the map in the upper row, second position, for $\beta = 0.1$ with resulting GRP threshold $GRP_0 = 20$, is to be understood as follows: the orange cells are the ones in which with $1 - \beta = 90$ % confidence one can expect that prob($C >$ > 100 Bqm^{-3}) remains below 30 %. β can be understood as the risk of erroneously classifying a cell as non-RPA, while in reality it is one. Varying threshold c or p_0, instead of β, leads to a similar variety of maps.

It appears that β is a very critical parameter. Clearly, the „more conservative" (lower β) the estimate, the larger the fraction of orange cells, *i. e.*, assigned RPA.

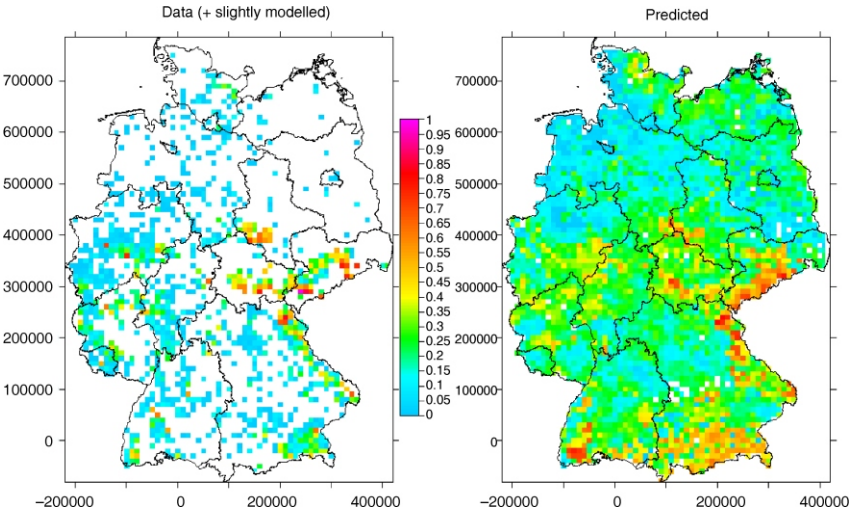

Figure 2. Maps of German territory showing state boundaries. Left map: empirical probability (aggregated in cells through lognormal modelling) that indoor Rn concentration C exceeds threshold $c = 100$ Bqm^{-3}, right map: modelled probability with $\beta = 0.15$. „probi" model assumed. Cells: 10 km × 10 km. Axis units: m, Lambert azimuthal equal area projection

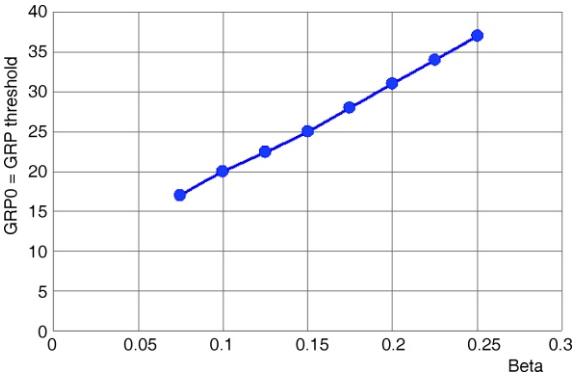

Figure 3. GRP threshold GRP_0 in dependence of β, indoor Rn threshold = 100 Bqm^{-3}. (see text for further explanation)

Comparison with previous results

The results may be compared with ones from the alternative approaches mentioned in section *previous attempts*.

The copula method appears to underestimate exceedance probabilities. A central estimate by the method presented in this paper, setting $\beta = 0.5$, *i. e.* estimating the median, and c = 100 Bqm^{-3}, leads to prob(C > 100 Bqm^{-3} | GRP = 50) = 0.19, for prob(C > > 100 Bqm^{-3} | GRP = 100) = 0.32; the copula model yields 0.10 and 0.25 for the same quantities, respectively. The reason for this discrepancy is unclear, perhaps traceable to the complicated procedure of the copula method.

The cross-classification approach leads to GRP_0 = 20.5, with a (5 %, 95 %) confidence interval, based on 10,000 bootstraps, (17.4 … 21.6), *i. e.* somewhat lower than the value found via the probability model (25). If one is only interested in the GRP threshold, but not in the full regression model p as the function of GRP, in my opinion, cross classification is preferable as it is much simpler and requires fewer technical steps. If interested in the latter, the logi/probi approach is preferable to the copula one because of its comparative simplicity.

CONCLUSION

A relatively simple method has been described to derive thresholds of the geogenic Rn potential GRP which define whether a cell with given GRP has to be considered a Rn priority area. The method is computationally relatively easy. It is flexible in that it allows independently setting the threshold c of indoor Rn concentration, a tolerated exceedance probability p_0, and the *degree of conservativeness*, β. It turns out, however, that different choices of (c, p_0, β) can lead to similar resulting RPA patterns. In particular, this means that in defining RPA one has to define how conservative the estimated RPA shall be (or equivalently, how large the risk of misclassification). It can also be understood as a caveat not to concentrate on one parameter only in the debates about how to define RPA – such as the threshold c, commonly in the focus of discussions.

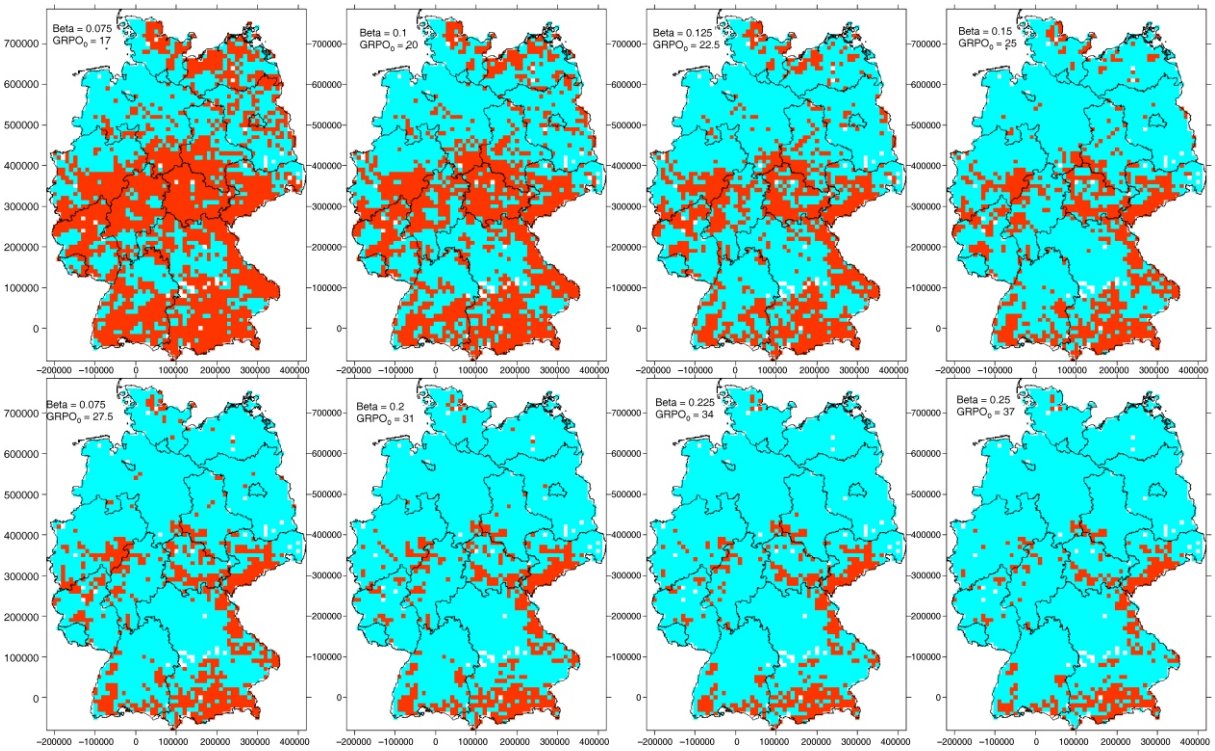

Figure 4. Radon priority areas RPA (dark grey) defined by the probability (C > 100 Bqm^{-3}) = 0.3, for different „degrees of conservativeness" β

European BSS require setting this value to 300 Bqm^{-3}, at most. From current discussions, it appears that many EU countries will indeed adopt this highest possible value. Among other reasons, this seems to be motivated by the wish to reduce costs which result from the Rn action plan, implied by the necessities of Rn remediation and prevention. (Whether this value is reasonable from a radiation protection point of view, instead of opting, *e. g.*, for 100 Bqm^{-3}, as proposed by the WHO [1], is subject to vivid debates, in particular as new Rn dosimetry issued by the ICRP dramatically increases dose conversion factors.) However, as shown, defining RPA may require (depending on the RPA concept) setting parameters apart from threshold *c*. In particular, my plea is that the degree of conservativeness be given more attention than is ordinarily done.

Among questions to be investigated further are the assessment of uncertainties and possible improvement of technical details, such as the inversion necessary to retrieve a GRP threshold. Also, validation by data partition is problematic because this would reduce the already quite low number of data (represented by the green dots in fig. 1) in the calibration subset to a degree that inflates estimation uncertainty.

REFERENCES

[1] ***, WHO Handbook on Indoor Radon – a Public Health Perspective, 2009, World Health Organization, http://www.who.int/ionizing_radiation/env/radon/en/index1.html (accessed 1 Dec 2016)
[2] ***, 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 1 Dec 2016)
[3] Žunić, Z. S., *et al*., The Relation Between Radon in Schools and in Dwellings: a Case Study in a Rural Region of Southern Serbia, *J. Environmental Radioactivity 167* (2017), Feb., pp. 188-200, http://dx.doi.org/10.1016/j.jenvrad.2016.11.024
[4] Bossew, P.*, et al*., Status of the European Atlas of Natural Radiation, R*adiation Protection Dosimetry, 167* (2015), 1-3, pp. 29-36
[5] Cinelli, G., *et al*., Status of The European Atlas of Natural Radiation; https://my.cloudme.com/#radoneurope/IWEANR Verbania 2015 (part 1) (accessed 1 Dec 2015)
[6] Bossew, P., Local Probability of Indoor Radon Concentration to Exceed a Threshold, Estimated from the Geogenic Radon Potential, GeoENV 2016, 11th International Conference on Geostatistics for Environmental Application Conference, Lisbon, July 6-8, 2016
[7] Neznal, M., *et al*., The New Method for Assessing the Radon Risk of Building Sites, *Czech Geol., Survey Special Papers, 16* (2004); Czech Geol., Survey, Prague, http://www.radon-vos.cz/pdf/metodika.pdf (accessed 1 Dec 2016)
[8] Kemski, J., *et al*., Transfer Function between Radon Concentration in Soil air and Indoor air in Dwellings, Final Report of the Research Projects St. Sch. 4186 and St. Sch. 4187: Determination of a Transfer Function between Radon Concentration in Soil air an Indoor air in Dwellings Including Radon Measurements in Buildings and Validation of the Geogenically Induced Radon Potential. Part A: Soil Investigations Related to the Geogenic Potential; Part B: Validation of the Geological Prediction through Measurements of Radon Concentration in Buildings. Publication Series Reactor Safety and Radiation Protection, BMU-2002-598
[9] Bossew, P., Mapping the Geogenic Radon Potential and Estimation of Radon Prone Areas in Germany, *Radiation Emergency Medicine, 4* (2015) 2, pp. 13-20 (accessed 1 Dec 2016)
[10] Bossew, P., A Radon Risk Map of Germany Based on the Geogenic Radon Potential, in: E. Pardo-Iguzquiza *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.
[11] Bossew, P., Stochastic Dependence of Rn-Related Quantities, *Romanian J. Physics 68* (2013), Suppl., S44, www.nipne.ro/rjp/2013_58_Suppl.html (accessed 1 Dec 2016)
[12] Bossew, P., Determination of Radon Prone Areas by Optimized Binary Classification, *J. Environmental Radioactivity, 129* (2014), March, pp. 121-132, http://dx.doi.org/10.1016/j.jenvrad.2013.12.015

Received on December 5, 2016
Accepted on January 30, 2017

Петер БОСЕВ

ПРОЦЕНА ЛОКАЛНЕ ВЕРОВАТНОЋЕ ЗА ПРЕКОРАЧЕЊЕ ГРАНИЧНЕ ВРЕДНОСТИ КОНЦЕНТРАЦИЈЕ РАДОНА У ЗАТВОРЕНОЈ СРЕДИНИ НА ОСНОВУ ГЕОГЕНОГ ПОТЕНЦИЈАЛА РАДОНА

Радон у затвореној средини сврстан је у значајне загађиваче ваздуха. На основу епидемиолошких доказа, процењено је да је радон у затвореној средини, после пушења, други узрок за настанак карцинома плућа. Као последица тога, води се брига о ограничавању излагања радону регулативном за смањењ постојећег и спречавање будућег излгања. У ту сврху, од важности је одредити области у којима се са одређеном сигурношћу може очекивати да концентрација радона у стамбеним и пословним објектима пређе референтне вредности. Ове области се називају приоритетним областима за радон како би се означило да је у њима потребно спровести акције за ублажавање постојећег стања и превенцију. Постоје различите дефиниције приоритетних области за радон и различите методе за њихову процену на основу података. У Немачкој, тренутно се за предвиђање користи геогени потенцијал радона. Међутим, званични референтни нивои се односе на концентрацију радона у затвореној средини, а не на геогени потенцијал радона. Стога је потребно извести референтне нивое за геогени потенцијал радона статистичким повезивањем обе величине.

Овај рад представља метод за добијање локалне вероватноће за прекорачење граничне вредности концентрације радона у затвореној средини на основу геогеног потенцијала радона. Добијена веза може се употребити за формирање референтних нивоа геогеног потенцијала радона који се потом могу користити за одређивање приоритетних области за радон.

Кључне речи: радон у затвореној средини, геогени радон, приоритетна област за радон