

# PREDICTION OF INDOOR RADON RISK FROM RADIUM CONCENTRATION IN SOIL: REPUBLIC OF MACEDONIA CASE STUDY\*

PETER BOSSEW<sup>1</sup>, ZDENKA STOJANOVSKA<sup>2</sup>, ZORA S. ZUNIC<sup>3</sup>, MIMOZA RISTOVA<sup>4</sup>

<sup>1</sup> German Federal Office for Radiation Protection, Köpenicker Allee 120-130, 10318 Berlin, Germany  
e: pbossew@bfs.de, peter.bossew@reflex.at

<sup>2</sup> Faculty of Medical Sciences, Goce Delcev University, Stip, Republic of Macedonia

<sup>3</sup> Institute of Nuclear Sciences “Vinca”, Electro Chemical Etching Laboratory – ECE LAB,  
Laboratory for Radiobiology and Molecular Genetics, P.O Box 522, 11000, University of Belgrade,  
Serbia

<sup>4</sup> Institute of Physics, Faculty of Natural Sciences and Mathematics, University in Skopje,  
Republic of Macedonia

*Received November 15, 2012*

Geo-referenced datasets of indoor radon concentrations and radium concentrations in soil are available for the Republic of Macedonia. However, the indoor  $^{222}\text{Rn}$  data are spatially strongly clustered as the measurements were essentially confined to major towns and cities. Hence, the estimation of the geographical distribution of  $^{222}\text{Rn}$  concentration based only on the  $^{222}\text{Rn}$  data is difficult to be made. On the other hand, geochemical measurements  $^{226}\text{Ra}$  are quite well distributed over the country. Since  $^{226}\text{Ra}$  is the source of  $^{222}\text{Rn}$ , one may think of using  $^{226}\text{Ra}$  as a predictor for  $^{222}\text{Rn}$ .

In this paper we present a method for modelling the stochastic dependency of indoor  $^{222}\text{Rn}$  of soil  $^{226}\text{Ra}$ . The method is new in the area on  $^{222}\text{Rn}$  assessment and still needs to be validated by more case studies.

It must be born in mind that the indoor  $^{222}\text{Rn}$  depends, in some cases more strongly, on controlling factors other than the  $^{226}\text{Ra}$  in soil, so that its estimation from  $^{226}\text{Ra}$  alone is inevitably imperfect. The results must therefore be understood as estimates in absence of other information, and as a motivation to carry out measurements in regions where the model predicts higher  $^{222}\text{Rn}$  levels, but for which no measurements are available so far.

*Key words:* Republic of Macedonia, indoor radon, radium in soil, probabilistic prediction.

## 1. INTRODUCTION

It is not a rare situation when a geographical distribution of a quantity Y shall be estimated, that the number of observations is not sufficient to cover the entire

\* Paper presented at the First East European Radon Symposium – FERAS 2012, September 2–5, 2012, Cluj-Napoca, Romania.

territory. On the other hand, sufficient data of a quantity  $Z$  are available which is to some extent related to the first one, for physical reasons. In such case, one can use the relationship between the two, and estimate the geographical distribution of the first quantity  $Y$ , predicted by the one of the “proxy” quantity  $Z$ . One should be aware, however, that the resulting estimate can be correct only insofar as it reflects the properties of the predictor  $Z$  and the relationship between the two, but not those properties of the response variable  $Y$  which are not accounted for by the predictor  $Z$ . (Otherwise a perfect correlation would be required.)

For this kind of problems, in the field of environmental assessment and in  $^{222}\text{Rn}$  research in particular, transfer models have been proposed by a number of authors. Normally these models are based on conventional regression  $Y$  by  $Z$ . The “calibration” of the models (*i.e.* estimation of its parameters, for example, the intercept and the slope in a linear model) requires a dataset consisting of pairs  $\{(z^*_i, y^*_i)\}$ , where each pair is understood to be fixed at one point, or at one geographical unit. In many cases, however, the original data (measurements or observations) are not located at the same spot, and first need to be “collocated” by estimating one at the location of the other one, or both on a common grid, or by aggregating them into spatial units (windows, grid cells, geological or administrative units). The viability of such a procedure has to be checked in each individual case; in any case it involves uncertainty stemming from the collocation or aggregation procedure.

A disadvantage of a conventional (Gaussian) regression is that normally it does not allow probabilistic estimates because, in real situations, the distribution of residuals cannot be assumed normal. Hereafter we apply a different method, which can be called generalized regression: it operates on the level of the joint probability distribution of the investigated quantities, from which desired statistics can be derived as statistics of the conditional distributions, for example, an exceedance probability of variable  $Y$ , given  $Z$ , or expected  $Y$ , given  $Z$ .

The “methods” section of this paper will not explain the method for the sake of saving space, with reference to this paper and bibliographic details. However, additional technical issues are discussed and the results given for the sake of estimation of the variable  $Y = \text{indoor } ^{222}\text{Rn concentration}$  from  $Z = ^{226}\text{Ra concentration in soil}$ , for the available dataset from the Republic of Macedonia.

## 2. METHODS

### 2.1. DATA

A total number of 213 undisturbed surface soils on 20 cm depth were collected during the period of 2008-2010. The sampling locations shown in Figure 1,

are concentrated in the populated areas within the country. The  $^{226}\text{Ra}$  activity concentration in soil samples were determined by gamma spectrometry. Soil sampling, sample preparation, measuring and analysis procedure were done according to ISO 18589 methods [1-2].

The nationwide survey of indoor radon concentration in the country has been performed in the period December 2008–November 2009. The methodology of the survey, seasonal and regional variability as well the influence of building characteristics of the indoor radon  $^{222}\text{Rn}$  concentration has already been described in more detail by Stojanovska *et al.* [3-4]. To allow data for this study for indoor radon concentration we extracted 125 measurements made in the ground floor of the dwellings with basement (Figure 1).

Table 1 summarizes the results for  $^{226}\text{Ra}$  concentration in soil samples and indoor radon  $^{222}\text{Rn}$  concentration, measured in the dwellings on the ground floor, in houses with basement.

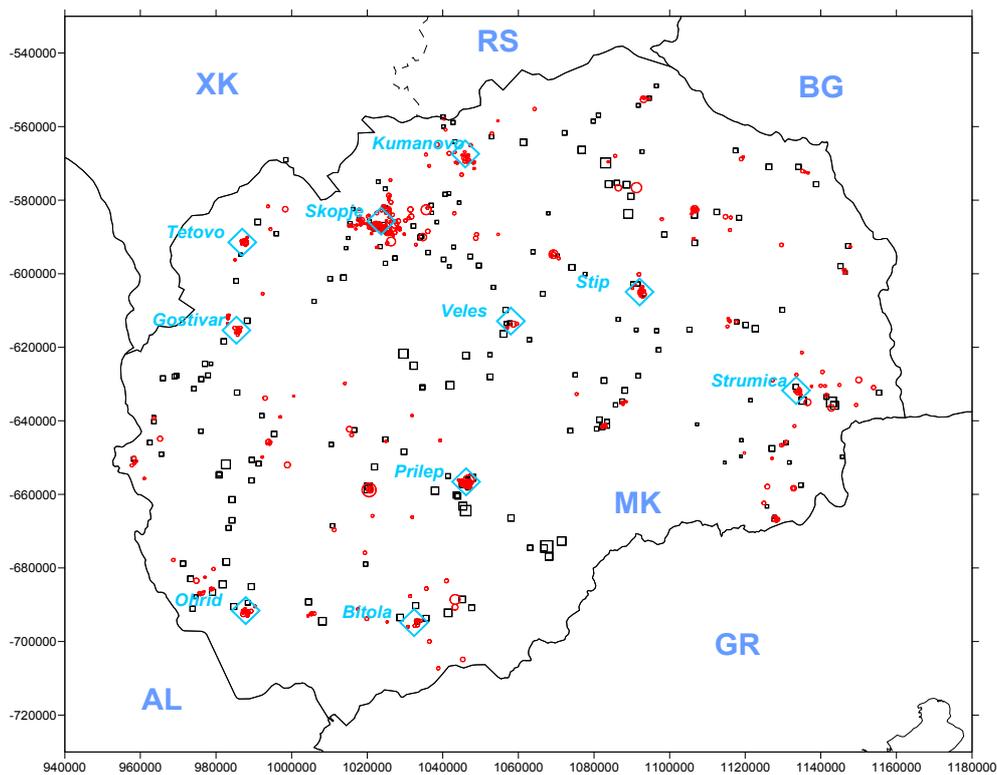


Fig. 1 – Locations of the measurement point. Black squares:  $^{226}\text{Ra}$  concentration in soil; red circles: indoor  $^{222}\text{Rn}$ . The linear size of the symbols is proportional to the values. - This and the following figures can be found in colour in the electronic version.

Table 1

<sup>226</sup>Ra concentration in soil and indoor radon concentration

	<sup>226</sup> Ra (Bq/kg)	<sup>222</sup> Rn (Bq/m <sup>3</sup> )
No. of measurements	213	125
Minimum	9	24
Median	37	79
Maximum	123	502
Arithmetic mean	41	102
Standard deviation	18	76
Geometric mean	37	84
Geometric standard deviation	1.53	1.84

## 2.2. STOCHASTIC DEPENDENCE

The theory behind the approach applied hereby has been presented in [5-6]. Essentially one states a model of the joint probability distribution of the investigated variables. In this paper the variable is the quantity  $C$  = indoor <sup>222</sup>Rn concentration in Bq/m<sup>3</sup>, (estimated long-term concentration in ground floor rooms of dwellings in houses with basement), and  $Z$  = <sup>226</sup>Ra concentration in soil in Bq/kg, considered as spatial random variables or random fields.

We wanted to estimate, at location  $x$ :

$$p(z; c_0) := \text{prob}[C(x) > c_0 \mid Z(x) = z],$$

the probability that the indoor <sup>222</sup>Rn concentration exceeds threshold  $c_0$  (for example 100 Bq/m<sup>3</sup>). From the conditional distribution of  $C$  (given the  $Z=z$ ) at location  $x$ , we also estimate the local conditional expectation of  $C$ ,

$$E(C(x) \mid z) = \int_{(c=0 \dots \infty)} c \, dF(c \mid z(x)), \quad \text{with } F(c \mid z(x)) := 1 - p(z(x), c) \quad (1)$$

The variance is  $\text{Var}(C \mid z) = E(C^2 \mid z) - E^2(C \mid z)$ ; the GM and GSD are calculated by analogy. In practice, the integral was evaluated numerically (homemade software was used for all calculations).

What remains to be done is modelling the joint distribution of  $C$  and  $Z$ , for which a Gumbel copula has been chosen. It is parameterized through the lagged Kendall  $\tau$  correlation, estimated from the non-located data (see the theory section of Bossew [5-6]).

### 2.3. SPATIAL ESTIMATION OF THE PREDICTOR

The scattered observations of  $Z$  ( $^{226}\text{Ra}$  concentrations) have to be estimated on a grid by geostatistical methods. This could be done by ordinary kriging (OK; not further discussed here as this is routine) or by sequential simulation (SGS or DSSIM etc.). The first method has the advantage of being a straightforward technique, but results in a local expectation  $E_{\text{spat}}(Z) = Z^*(x) = E[Z](x^*)$  (where  $x^*$  is the location where it was estimated); its transform into the mean,  $t_m: z \rightarrow E(C(x)|z)$  suffers from the fact that  $t_m$  may not be linear, hence  $E(C(x)|E_{\text{spat}}(z)) \neq E_{\text{spat}}(E(C(x)|z))$  in general. Estimating  $E(C)$  from  $Z^*_{\text{OK}}$  therefore involves a bias. This can be fixed by applying the second method whose output is a set  $\{z^*(x, \omega)\}$ , where  $\omega$  indexes realizations ( $\omega \in \Omega$ , “ergodic” sample space). The spatial expectation at  $x$  is the “ergodic” mean over  $\omega \in \Omega$ . Hence,

$$E(C(x)) = \int_{\Omega} dG_Z \int_{z=0, \dots, \infty} z dF(c | Z(x) = z(x, \omega)), \quad (2)$$

with  $G_Z$  the estimated local conditional distribution of  $Z(x)$  (*i.e.* conditional to the data  $\{z_i\}$ ). Again, the previous requires numerical integration. The first integral is in practice simply the mean over  $\omega$  of replications  $z(x, \omega)$ . Simulation was performed with SGeMS [7]. The method was DSSIM with Soares correction (*ibid.*, p.144). 100 realizations were generated. For the variogram, estimation and modelling, Surfer 8 was used.

### 2.4. DE-CLUSTERING

The method which uses the joint distribution is sensitive against correct estimation of the true distributions of  $^{222}\text{Rn}$  and  $^{226}\text{Ra}$  concentrations, which enter as  $F_Z \equiv F_{\text{Ra}}$  and  $F_C \equiv F_{\text{Rn}}$ . Since the observations are spatially clustered, the raw data cannot be used for estimating the distribution, because this would cause biases for areas which have been sampled more densely.

De-clustering is done here as follows. Overlay a grid of chosen grid constant (in this work between 5 and 20 km) with random offset over the domain. In each cell choose one observation randomly. Compute the distribution of these random points. Repeat the procedure many times (in this work 100 was chosen) with varying grid offset and observations within cells. Compute the statistics over realizations. The result is an estimate of the true distribution of the data. It depends on the grid size to some extent; the best choice of grid size can only be guessed. We have chosen 20 km for the  $^{226}\text{Ra}$  data and 10 km for the  $^{222}\text{Rn}$  data. Also for de-clustering home-made software was used.

### 3. RESULTS

#### 3.1. DE-CLUSTERING AND ESTIMATED DISTRIBUTIONS

Fig. 2 shows the locations of the indoor  $^{222}\text{Rn}$  data (only ground floor, houses with basement;  $n=125$ ) in the upper left plot, and three examples of realizations of de-clustering for three different grid geometries. Note that each of these graphs is only one realization of many (100 chosen), of which the mean distribution is estimated.

The cumulative distributions of the original indoor  $^{222}\text{Rn}$  data and the mean distributions of the de-clustered data are shown in Fig. 3. The differences do not appear large at first sight, but in some sections of the curve they are substantial. The deviations between original and de-clustered distribution reflect the removal of the effect of preferential sampling, which is implied by clustered observations. In this case, the main bias was caused by high number of samples in the capital Skopje (the cluster in the NW), where the mean  $^{222}\text{Rn}$  concentration is relatively low.

#### 3.2. ESTIMATED DEPENDENCE MODEL

In Fig. 4 the Pearson correlation of the ln-transformed data, the Spearman  $\rho$  and the Kendall  $\tau$  coefficients were plotted against lag  $h$  (see Bossew [5-6] for the theoretical background). For  $\tau$  uncertainties ( $1\sigma$ ) are given, resulting from the simulation (they were omitted for the other coefficients for legibility of the graph). Spearman  $\rho$  and Pearson correlation are not further used in this study. In any case, one can see that essentially the different correlation measures follow a similar dependence pattern of lag  $h$ .

As a technical detail, lag tolerance bands were set equal to lag distances (1000 m), as is also a common practice in variogram estimation.

We can observe that for small lags, correlations between 0.2 and 0.35 can be found, but within about  $h = 5$  km the correlations disappear. One can thus say that indoor  $^{222}\text{Rn}$  and soil  $^{226}\text{Ra}$  are correlated within about 4 km. From the graph we can estimate  $\tau(h=0) \approx 0.24$  and the corresponding  $\mathcal{G}(\text{Gumbel})=1.32$ , which enters the dependency model.

The resulting estimates for the exceedance probability are shown in Fig. 5, and the expectations (AM and GM as well as SD and GSD bands) in Fig. 6, according to eq. 1.

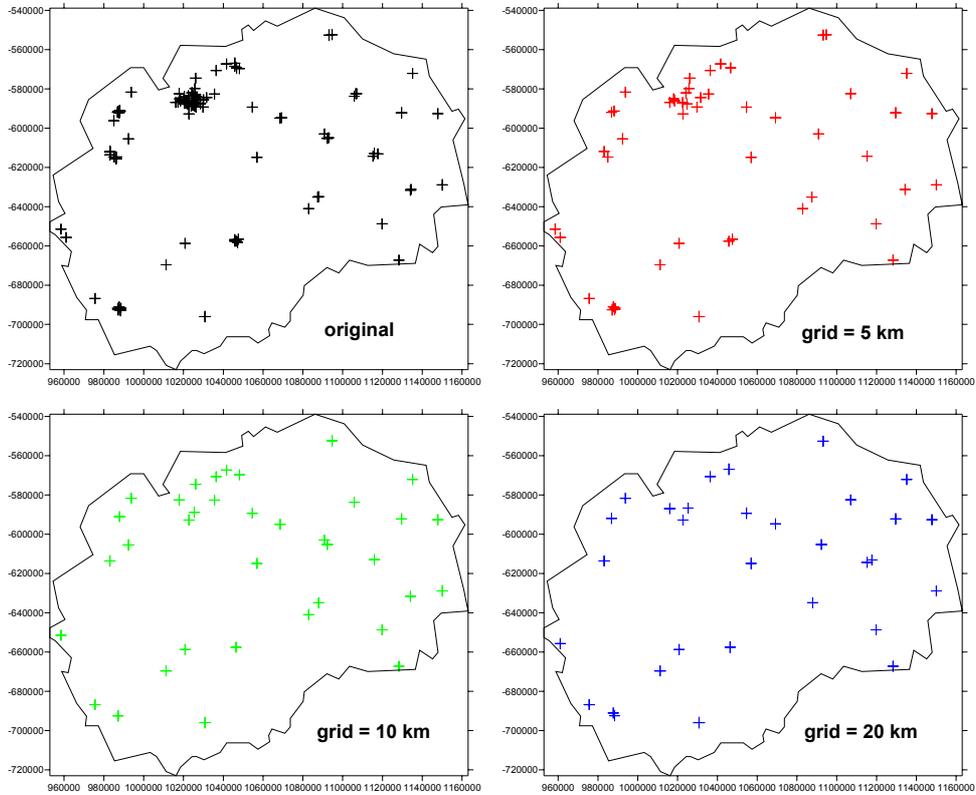


Fig. 2 – Data de-clustering: locations of original  $^{222}\text{Rn}$  data and three examples of realizations of random de-clustering for three grid sites.

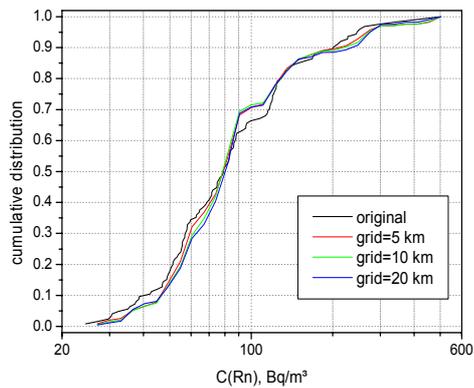


Fig. 3 – Cumulative distribution of original data,  $^{222}\text{Rn}$  concentration in ground floor rooms of houses with a basement and of means over 100 realizations of de-clustering, for three grid sizes.

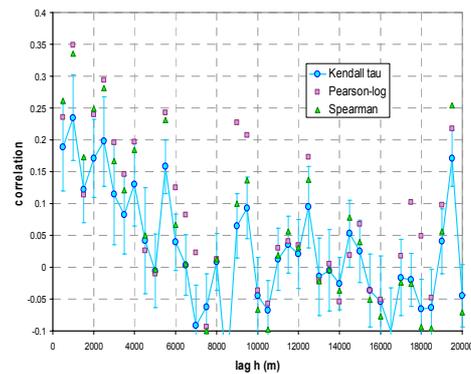


Fig. 4 – Cross-correlation of indoor radon (estimated long-term mean, ground floor rooms in houses with a basement) and  $^{226}\text{Ra}$  concentration in soil.

The “unsmooth” shape of the “transfer functions” between soil  $^{226}\text{Ra}$  and the statistics of  $^{222}\text{Rn}$  is caused by the discontinuities in the estimated marginal distribution of soil  $^{226}\text{Ra}$  (not shown, but with similar “edges” as the one of indoor  $^{222}\text{Rn}$ , Fig. 3) which is in turn a consequence of the relatively low number of observations.

Physical reasoning suggests linear true dependence of the indoor  $^{222}\text{Rn}$  on the soil  $^{226}\text{Ra}$  (if other controlling factors are ignored). Fig. 6 shows deviation from this behaviour (remember that they have been estimated from the actual observation without anticipating an analytical transfer model, apart from the stochastic dependency model controlled by  $\vartheta$ , again estimated from the data).

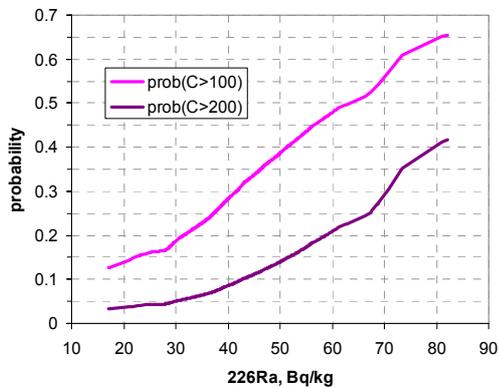


Fig. 5 – Estimated probabilities that indoor  $^{222}\text{Rn}$  exceeds a threshold, as function of  $^{226}\text{Ra}$  concentration in soil.

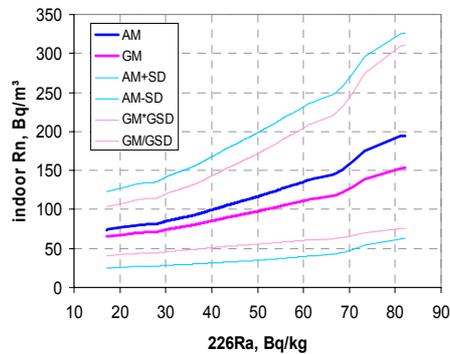


Fig. 6 – “Transfer function”  $t_m$  between  $^{226}\text{Ra}$  in soil and indoor  $^{222}\text{Rn}$ , estimated from the stochastic model.

### 3.3. ESTIMATED GEOGRAPHICAL DISTRIBUTION OF RADIUM AND DERIVED EXPECTED RADON RISK

The distribution pattern of  $^{226}\text{Ra}$  concentration in soil is shown in Fig. 7. It has been created using the two methods, described in section 2.3.

The OK method uses a variogram, modelled as short-range spherical (as a continuous nugget model, in order to avoid too strong smoothing), one exponential and one hole effect component. For simulation, the variogram was a short-range exponential and a long-range Gaussian. The upper tail was modelled with hyperbolic exponent 4.5, derived from the empirical distribution (not to be further explained here; see *e.g.* Remy *et al.* (2009), p. 106, [7]).

A relation to geology can be recognized: The low-Ra zone running NW-SE from Skopje is the alluvial valley of river Vardar; the high-Ra zone N Stip lies in a region of volcanic rocks; near Strumica and S Prilep there are granitic rocks.

The structure is reproduced in the two radon maps (Fig. 8 and Fig. 9) of the probabilities to exceed 100 and 200 Bq/m<sup>3</sup>, and the expectation, respectively. Since soil <sup>226</sup>Ra has been used as the only predictor, the pattern of indoor <sup>222</sup>Rn is necessarily the same as the one for <sup>226</sup>Ra; this is of course not the *true* pattern, but the one estimated from <sup>226</sup>Ra.

The estimation uncertainty is high: for the OK predictor we find coefficients of variation CV between 66 and 70%, or GSD = 1.6 to 2, derived from the local G(c|z), see section 2.1, after eq. 1. This reflects the relatively weak dependence between <sup>226</sup>Ra in soil and indoor <sup>222</sup>Rn, since the latter is controlled by other factors as well, probably even more importantly than by soil-<sup>226</sup>Ra.

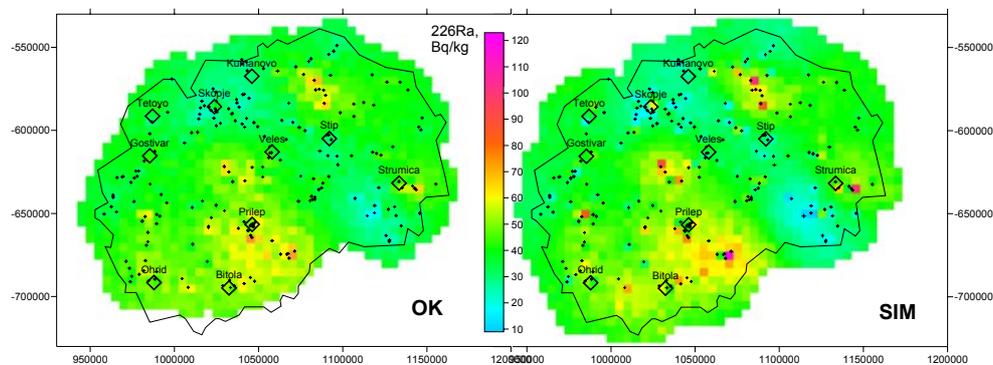


Fig. 7 – Geographical distribution of estimated <sup>226</sup>Ra concentrations in soil. Grid size 5 km. Crosses: Measurement points of <sup>226</sup>Ra. Left: by OK, levels: kriging estimate; right: by simulation, levels: AMs over realizations.

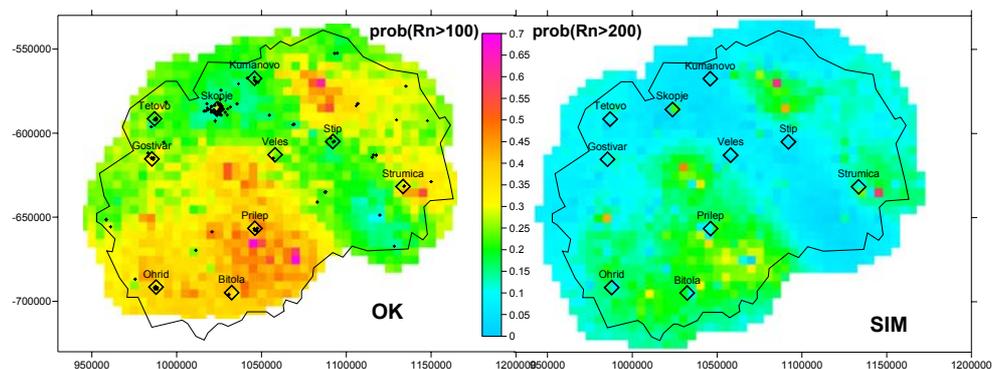


Fig. 8 – Probability that the <sup>222</sup>Rn concentration is ground floor rooms of houses with basement exceeds 100 Bq/m<sup>3</sup> (left) or 200 Bq/m<sup>3</sup> (right), estimated from <sup>226</sup>Ra concentration in soil. Crosses: <sup>222</sup>Rn measurement points. Left: by OK, levels: transformed OK-means; right: by simulation, levels: quantities of transformed realizations.

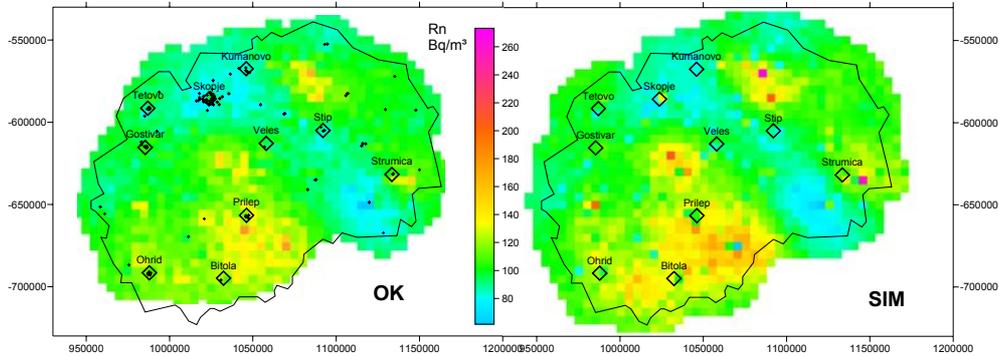


Fig. 9 – Expectation of  $^{222}\text{Rn}$  concentration in ground floor rooms of houses with basement, estimated from  $^{226}\text{Ra}$  concentration in soil. Crosses: Rn measurement points. Left: by OK, levels: transformed OK-means; right: by simulation, levels: AMs over transformed realizations.

The OK-versions of the maps, Fig. 8 and Fig. 9, rely in eq. 1, substituting the OK-mean  $z^*$  for  $z$  and therefore include the bias mentioned on section 2.3. The simulation-based versions use eq. 2. Possible biases may stem from the choice of the method (DSSIM, Soares correction, target variogram = the empirical one); this has not been investigated further.

One can recognize slight differences in the results between the two methods. The simulation tends to reproduce extremes better than kriging and the pictures look therefore a bit “rougher”. The effect of the non-linearity of transform  $t_m$  (section 2.3) can best be recognized in Fig. 9: it causes the simulation method to yield higher estimates, which seem more realistic.

### 3.4. RADON PRONE AREAS

There is no authoritative definition of a radon prone area (RPA). Qualitatively, the concept denotes areas where observed or expected values of a  $^{222}\text{Rn}$ -related variable are high with respect to reference values or with respect to the mean over the domain (the country in this case).

The radon prone areas may be defined as grid cells where a certain statistical criterion is fulfilled. For example one can define a RPA as a cell (or union of cells) in which the expected indoor  $^{222}\text{Rn}$  concentration exceeds a threshold, such as 100 or 200  $\text{Bq/m}^3$ . Another possibility is to declare a cell RPA if the probability that a threshold is exceeded, is greater than a probability threshold. For the sake of demonstration only, we shall use the second (more flexible) definition, and set  $c_0=200 \text{ Bq/m}^3$  and  $p_0=10\%$ :

Definition of RPA: “cell U is RPA if  $\text{prob}(C > 200) > 10\%$  in cell U”.

The cells are coded easily as RPA or non-RPA by simple indicator transform of  $p(z;c_0)$ , such as in Fig. 10. Again, as an effect of the mathematical pitfalls discussed above, the pictures look slightly different depending on the method.

Finally we investigate the percentage of total area of the country designed RPA by this criterion. The percentage is simply the fraction of the cells labelled RPA, of the total number of cells. As an example, again for  $c_0=200$  Bq/m<sup>3</sup>, this percentage is shown as function of the probability threshold  $p_0$  in Fig. 11, again for the two methods.

For example, according to the above definition ( $c_0 = 200$ ,  $p_0 = 0.1$ ), about 37% and 64%, respectively, of the country would be assigned RPA, depending on the method. The rather big difference shows that methodical differences should not be taken lightly. It is caused by the fact that with certain probability, local realizations lead to high  $z(\omega)$ .

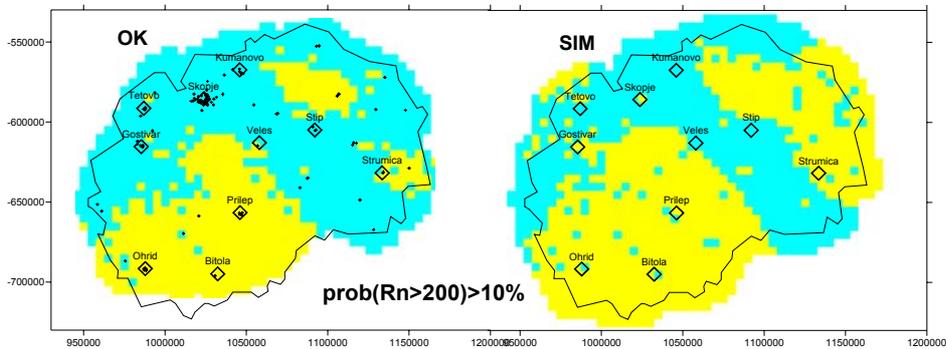


Fig. 10 – Parts of the country in which the estimated probability exceeds 10% that indoor <sup>222</sup>Rn exceeds 200 Bq/m<sup>3</sup> (yellow); derived from <sup>226</sup>Ra concentration in soil. Left: by OK, levels: transformed OK-means; right: by simulation, levels: quantiles of transformed realizations, in both cases indicator (0.1) classified.

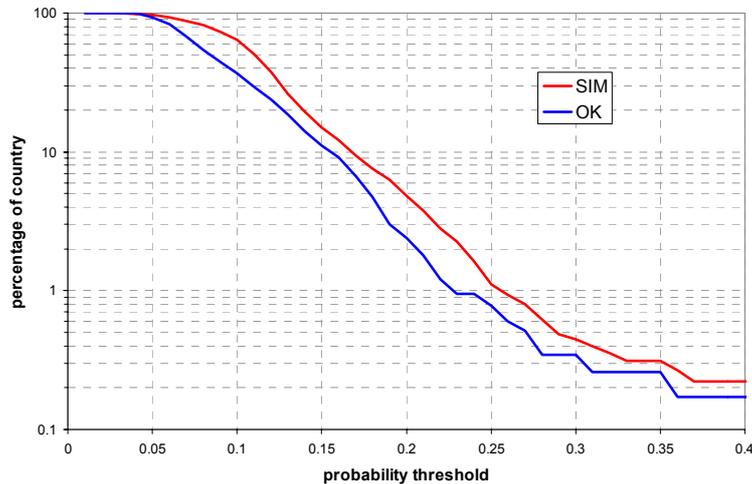


Fig. 11 – Percentage of the country in which the indoor  $Rn>200$  Bq/m<sup>3</sup>, in dependence of probability threshold.

(Note that for the simulation method, this probability is not defined as the fraction of realizations which have  $C(\omega) = t_m(z(\omega)) > 200$ , but as means over the probability-transformed realizations  $z(\omega)$ )

### 3.5. COMPARISON OF MODELLED AND EMPIRICAL INDOOR $^{222}\text{Rn}$ VALUES

Finally one would like to validate the model through comparison with empirical values. Table 2 compares empirical and estimated exceedance probabilities and mean values of indoor  $^{222}\text{Rn}$  (ground floor, houses with basement) for different cities, for which at least 3 indoor  $^{222}\text{Rn}$  data are available. The data are shown as scatter plots in Fig. 12. All considerations of this section are limited to the OK estimates.

As one can notice, there is no really visible relation between empirical and modelled statistics. Reasons may be:

- The model is insufficient. It is certainly true that  $^{226}\text{Ra}$  in soil is a very imperfect predictor of indoor  $^{222}\text{Rn}$  which is controlled by other factors too, which may be even more important than  $^{226}\text{Ra}$  concentration in soil. For example, different soil permeability leads to very different indoor  $^{222}\text{Rn}$  concentrations for the same source term ( $^{226}\text{Ra}$ ).
- It is unlikely that the mathematical base of the model is completely flawed, as it is merely a generalization of traditional regression, just more flexible through accounting (though limited by the chosen structure of the copula model) for the joint distribution.
- Apart from the city of Skopje, there are too few data per city to calculate robust empirical statistics.
- Most  $^{222}\text{Rn}$  concentrations are in the range from 60 to 160 Bq/m<sup>3</sup>, while low and high values are missing. Given the notoriously high local variability of  $^{222}\text{Rn}$  the factor “city” does not represent a meaningful grouping factor for  $^{222}\text{Rn}$ .

Table 2

Empirical and estimated probabilities that the indoor  $^{222}\text{Rn}$  concentrations in ground floor rooms of houses with basement exceed 100 Bq/m<sup>3</sup>, and empirical arithmetic mean ( $\pm 1$  SD) and modelled expectation. n = number of observations; Ra =  $^{226}\text{Ra}$  concentration in soil.

Red figures: strong under-estimation

City	n	Empirical Probability	Probability Estimated from Ra	AM(Rn), Empirical	E(Rn), Estimated from Ra
Gostivar	12	0.33	0.31	91 ± 55	103 ± 71
Kocani	3	0.33	0.36	160 ± 169	112 ± 78
Kumanovo	7	0.57	0.19	137 ± 77	85 ± 57

Table 2 (continued)

Negotino	3	0	0.25	69 ± 19	94 ± 64
Ohrid	7	0.29	0.35	87 ± 63	110 ± 77
Prilep	5	0.6	0.41	141 ± 77	120 ± 84
Radovis	3	0	0.23	88 ± 7	91 ± 61
Skopje	54	0.35	0.25	99 ± 75	94 ± 64
Stip	3	0.33	0.19	84 ± 30	86 ± 57
Strumica	3	0.33	0.34	127 ± 118	109 ± 76
Tetovo	7	0.29	0.22	70 ± 42	89 ± 60

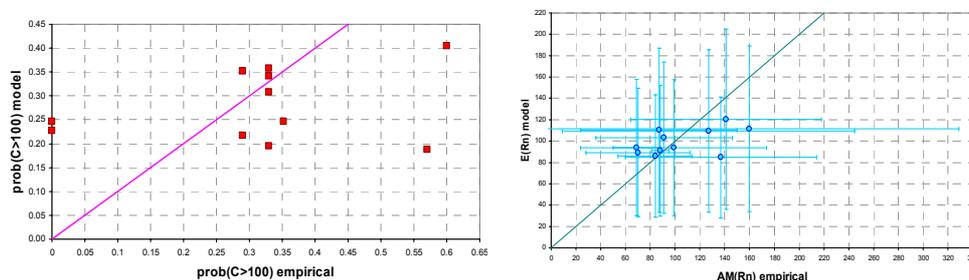


Fig. 12 – Scatter plots of estimated vs. empirical values of exceedance probability (prob (Rn>100)) and local means (uncertainty bars: 1 SD).

High values tend to be underestimated and low values to be overestimated. This results probably from the smoothing property of kriging, with which the  $^{226}\text{Ra}$  map was created (Fig. 7), on which this estimation was based.

The estimation is particularly poor for the city of Kumanovo, for which the empirical probability of high indoor  $^{222}\text{Rn}$  is much higher than the estimated one. The reasons are not clear; as a hypothesis, other more important factors which control the indoor  $^{222}\text{Rn}$  obscure its relation to  $^{226}\text{Ra}$  in soil.

#### 4. FURTHER DISCUSSION AND CONCLUSIONS

A method for estimating indoor  $^{222}\text{Rn}$  from  $^{226}\text{Ra}$  concentration in soil is presented. The method may be particularly useful if the radon data cover the domain less completely than the observations of the predictor (here  $^{226}\text{Ra}$ ), as is the case for Macedonia.

As a consequence of the method, the geographical distribution of estimated indoor  $^{222}\text{Rn}$  is the same as the one of the predictor, since other factors are not accounted for, which also influence indoor  $^{222}\text{Rn}$ . The different soil permeability of each region was not considered, nor was the possibly regionally different building styles. Both factors are known to have strong influence on indoor  $^{222}\text{Rn}$  levels.

Since the model was derived from all data, these (like the other) factors enter implicitly as means over the domain; but locally they can be very different.

Comparison of modelled with empirical means and exceedance probabilities is not very informative in this case, because (1) local variability is high (the high standard deviations in Table 2), (2) the range of values is relatively narrow, so that identification of tendencies is difficult, and (3) the number of empirical data is too low for most cities.

This means that proper validation is currently not possible with the given data.

Critical issues in applying the method are:

- Estimation of the predictor: It must be spatially modelled carefully. In this case the result depended sensitively on the choice of variogram model, in particular on the nugget effect. It was forcefully set to zero and small range dependence modelled as spherical or exponential with a small range. This allows more realistic reproduction of maxima and minima which are smoothed away when setting nugget  $> 0$ . (Not further discussed here.) A more reliable estimation could be achieved by using geology as categorical deterministic predictor and stochastic simulation as an estimation method. For simulation, certain parameters may be sensitive, such as choice of simulation algorithm, histogram and upper tail modelling.
- The model of stochastic dependency has to be estimated. This includes the selection of the model (here the Gumbel copula, not further discussed in this article) and the dependency parameter, here the Kendall  $\tau$  correlation. Both the choices of dependency model and estimation of the controlling parameter are sensitive to the result. Some more experience with real data will be required here.
- Estimation of the true distributions of the predictor ( $^{226}\text{Ra}$  concentration) and the response variable (indoor  $^{222}\text{Rn}$ ): Some experience in assessing de-clustering is needed; alternative de-clustering methods whose results might be easier to assess objectively should be investigated for this purpose.
- The method implicitly assumes the same spatial autocorrelation structure for the indoor  $^{222}\text{Rn}$  and for  $^{226}\text{Ra}$  in soil. Since they are physically related, the assumption is probably true to some extent; but not entirely, because other spatial trends of  $^{222}\text{Rn}$  can be overlaid. These may result from regionally different geogenic factors which are independent of  $^{226}\text{Ra}$  (such as permeability), or anthropogenic factors, as discussed above.

## REFERENCES

1. ISO 18589-2, 2007, Measurement of radioactivity in the environment-Soil-Guidance for the selection of the sampling strategy, sampling and pre-treatment of samples (2007).
2. ISO 18589-3.2007, Measurement of radioactivity in the environment-Soil-Measurement of gamma-emitting radionuclides (2007).

3. Z. Stojanovska, J. Januseski, P. Bossew, Z.S. Zunic, T. Tollefsen, M. Ristova. Seasonal indoor radon concentration in FYR of Macedonia. *Radiat. Meas.* 46:5-6, 602-10 (2011);
4. Z. Stojanovska, J. Januseski, B. Boev, M. Ristova. Indoor exposure of population to radon in The FYR of Macedonia. *Radiat. Prot. Dosim.* 148:2, 162-7 (2012).
5. Bossew P. (2012a): A probabilistic indoor radon map of Germany. In preparation.
6. Bossew P. (2012b): Stochastic dependence of Rn-related quantities. Presentation, FERAS 2012, 2–5 September 2012, Cluj-Napoca, Romania, *Rom. J. Phys.* **58**, S44–S55, 2013.
7. N. Rémy, A. Boucher, Wu Jianbing, *Applied Geostatistics with SGeMS*. Cambridge University Press, ISBN 978-0-521-51414-9 (2009).