NUKLEONIKA 2010;55(4):429–432

ORIGINAL PAPER

Selected statistical problems in spatial evaluation of Rn related variables

Harry Friedmann,
Peter Bossew

Abstract. Since indoor radon is considered a potential hazard to health, Rn prevention and mitigation are necessary in certain areas. In this article we address the issues of mapping support and resolution, and conceptually discuss two common ways of generating maps from given information. Further, a short overview is given on the sources of uncertainties which are inevitably associated to every estimate and how to treat them. Finally, some possibilities of generating classified risk indices are outlined, since it is most often necessary to classify regions by estimated hazard.

Key words: radon • mapping • geostatistics • uncertainties • classification

H. Friedmann✉
Faculty of Nuclear Physics,
University of Vienna,
17 Währingerstr., A 1090 Vienna, Austria,
Tel.: +43 1 4277 51760, Fax: +43 1 4277 51752,
E-mail: harry.friedmann@univie.ac.at

P. Bossew
German Federal Office for Radiation Protection,
120/130 Köpenicker Allee, D-10318 Berlin, Germany

Received: 7 July 2009
Accepted: 30 December 2009

Introduction

Because radon (at least in high concentrations) is known as a health risk, it is necessary to reduce high indoor radon concentrations as well as to take precautions to avoid high indoor concentrations in new buildings. For mitigating existing high radon concentrations, it is necessary to know where the most affected buildings are located. Therefore, the administrations of many countries funded or performed themselves radon surveys to determine the radon exposure of the public and to reveal radon prone areas. But high indoor radon concentrations are not only found in areas where high soil gas radon concentrations exist and *vice versa* even in areas with a high geogenic radon hazard the indoor radon concentrations need not to be very high. The reason is the different construction-types of houses and partially different living styles of the inhabitants. So, there are two questions and consequently working programmes:

– Where to mitigate high indoor radon concentrations?
– Where to take special precautions for new buildings?

For both questions a spatial visualization by a map with predictive power is necessary. A map is the display of local estimates of a target variable. Estimating the geographical distribution of the Rn hazard and visualizing it with maps involves understanding of the statistical properties of Rn, interpreted as spatial variable. Statistical treatment of Rn is not trivial mainly due to the high spatial fluctuation of Rn related variables and the difficulty to accurately define "target variables", which are subject to measurement, analysis and subsequent mapping. The target variable can be a radon concentra-

tion (in a certain location in a certain season, a mean over a building, …), a standardized indoor concentration, a soil gas concentration, an excess probability, a class etc. Because usually the available information (e.g. measurements, geological data, …) is not sufficient for a prediction, a model is needed too.

During the creation of a map one has to consider the following issues:

– What means spatial? What is the smallest reasonable unit on a map to which an estimate can be given? This is the question about support and resolution of a map.
– How to produce estimates from existing information and models?
– How reliable is an estimate? This is the question of the uncertainty of the estimates.

Support, resolution

The mapping support of a target variable can either be irregularly bordered areas (e.g. municipalities, geological units) or regular geometrical areas (grid). The coverage of an area depends on the available information as well as on the map support. Unpopulated areas will be given the same value as the populated part of a municipality in the case of a map based on municipalities, however in a map based on a grid some grid-cells may stay empty. As an example, the results of the Austrian radon survey can be seen in two different types of maps. In the upper panel of Fig. 1, measurements are spatially assigned to municipalities, while in the lower part, the same measurements are assigned to grid-cells. A number of grid-cells representing sparsely populated areas remain empty because the strategy of the survey was to choose the number of measurements proportional to the population density.

The resolution of a map is of major importance. A high resolution reduces the variability within the map support; however, the uncertainty can be substantially increased because of a reduced number of information (measurements). However, if the resolution is reduced, the estimates (e.g. mean) will become more precise, but the variability within the map support will become much larger which greatly reduces the worth of information from the map.

Producing a map from information and models

At certain locations xi exist measurements $m(x_i)$ and some knowledge (e.g. geology, house-type, floor-level) $k(x_i)$ is available. From this information an estimate of a target variable $Z(A)$ at the map support (map unit) A should be derived. The usual way to determine $Z(A)$ is:

1. z_i: $= z(x_i) = f(m(x_i), k(x_i))$ – physical model,
2. $\{z_i\} \rightarrow Z(A)$ – mathematical model.

The first step depends on the available data and shall not be discussed here. However, the mathematical model shall be treated in a more general way. Generally, the z_i can be

a) considered as mutually independent: random samples of a probability distribution within the map support unit (cell, municipality,…), or

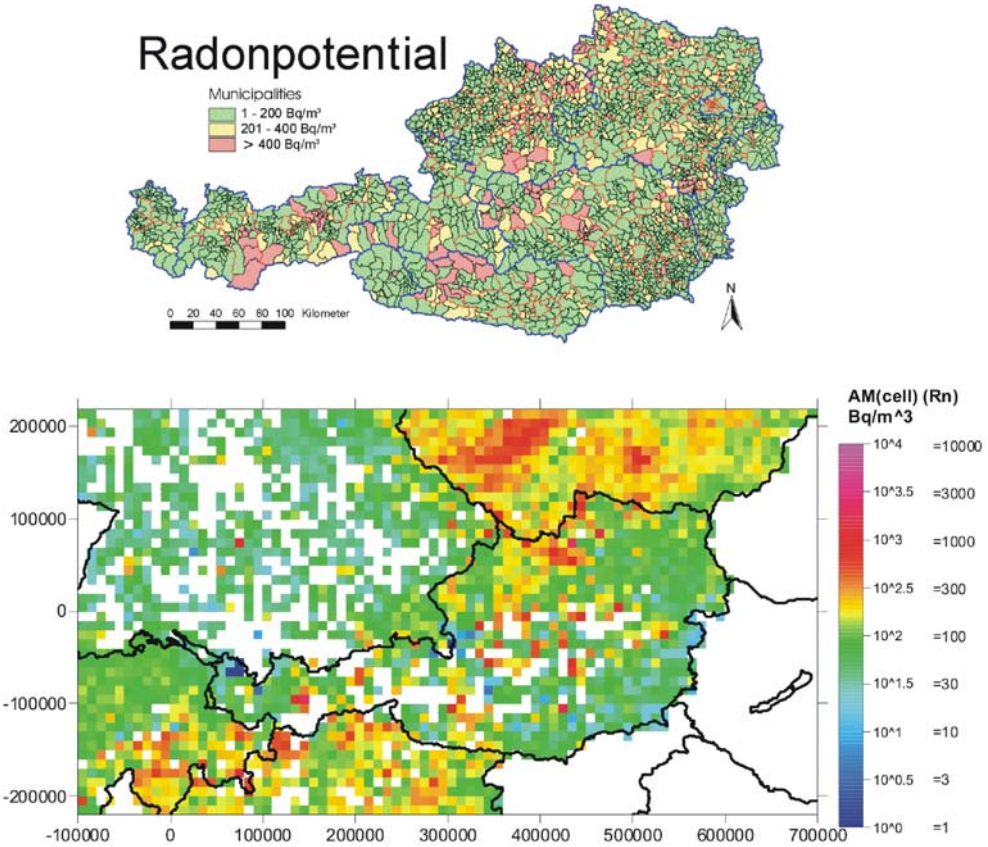

Fig. 1. Top: the radon potential map of Austria based on municipalities [2, 3]. Bottom: small section (covering Austria) of the European radon map: AM of Rn-concentration at ground floor (10 × 10 km^2 grid cells) [5].

b) considered as mutually dependent (spatially autocorrelated).

In the first approach the $Z(A)$ can be the arithmetic mean (AM), the geometric mean (GM), the median (MED). Modifications are possible, e.g. the introduction of weights (typically: weighting according to the population density), but also other types of quantities derived from the $\{z_i\}$ are common, e.g. excess probabilities $\text{prob}(Z(x) > Z_o : x \in A)$ assuming a log-normal distribution or a classification scheme with fixed class limits.

In the case of a mutually dependency, nearby locations have more similar values for the target variable than more distant locations. This property can be measured by a variogram (semivariogram: $\gamma(h) = < (Z(x) - Z(x + h))2 >/2)$ or by the covariance. The estimation of Z at an unsampled location x^* (interpolation) under the condition of minimisation of the spatial variance (requires $\gamma(h)$) is called kriging (named after the South-African mining engineer Daniel Krige). In addition to such point estimates $Z(x)$, also block estimates $Z(A)$ can be computed with all the possible modifications and extension as in the case of mutually independent values of the target variable.

Uncertainties

The estimation of uncertainties is of crucial importance for the use of radon maps. Uncertainties can be separated in random uncertainties and uncertainties which do not contribute on a random basis to the uncertainty budget (systematic uncertainties). Here only the sources for the random type of uncertainties shall be discussed. These uncertainties are caused by:

- the variability of true Z:
  - "lateral": because $Z(x)$ varies with x within A (variability caused by underlying soil/rock material);
  - "vertical" or "longitudinal": different realizations of Z at x (e.g. different values in different seasons, in different floors, during different usage of a room, variability within a room);
- the variability of $m(x_i)$:
  - measurement uncertainty;
  - "dispersion variance": m is used as an estimate to a location or an area even though the measurement was only performed in certain parts (e.g. certain rooms in a house) of a location/area;
- the variability due to the sampling design: different spatial distributions of $\{x_i\}$ in A lead to different estimates for $Z(A)$;
- the variable definition: certain $Z(x_i)$ may not be adequate representations of $Z(A)$.

Figure 2 visualizes the above-mentioned sources of uncertainties (for details see, e.g. [1]). To keep the uncertainty of $Z(A)$ small, the variability of the above-mentioned components must be kept small. This means:

1. The map support A should be chosen in a way that the variability of Z in A is small. This means either small areas or areas with very similar "radon related" characteristics.
2. The "vertical" variability can be reduced by a normalization to a standard situation (which of course introduces an additional model uncertainty) or by restricting to a very stringent rule where and how to measure (e.g. measurements only at ground floor in winter).

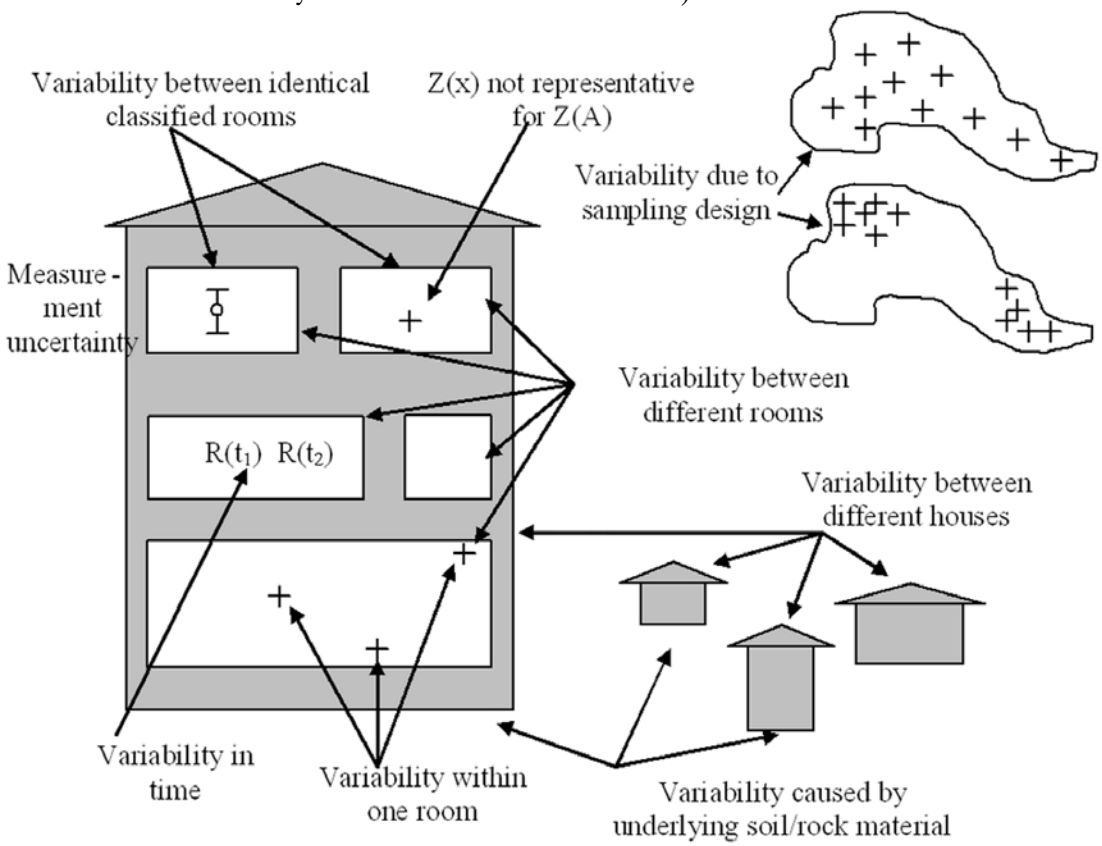

Fig. 2. Contribution to the random part of the uncertainties.

3. The variance in the measurement results can be reduced by increasing the number of measurements. This is partly in contradiction to the first point, because of economical reasons the number of measurements can often only be increased by enlarging the map support A.
4. The sampling design must be well chosen: a bias can be easily introduced by an incorrect sampling design which may misleadingly appear to cause a small uncertainty. Therefore, it is recommended to test the sampling design for randomness by an independent method.
5. The reliability of the data as representatives of the target variable must be kept as high as possible. This point has also some connection with the "vertical" uncertainty in Z, but it must be seen in a much broader context.
6. Applying adequate measurement methods.

Finally, it is necessary to keep the systematic uncertainties small, which predominantly means to check very critically the used models.

Classification of regions

It is necessary to classify areas for a decision policy concerning prevention and mitigation measures. The classification scheme can be based on a single target variable with fixed class limits (class(j): = $\{A: Z_j \le Z(A) < Z_j + 1\}$) or fixed probability limits, (class(j) = $\{x: P_j \le \text{prob}(Z(x) > Z_o) < P_j + 1\}$) above a certain level Z_o of the target variable.

But also classification schemes based on multiple target variables are possible, e.g. by cross tabulation and rules how to assign class-values to the table cells. So, in the Czech Republic a combination of soil-gas Rn concentration and permeability is used to classify localities and areas [4].

It is not yet clear which scheme will become mostly accepted because the different countries already have introduced classification schemes based on the special type of information gathered for radon mapping, i.e., the classification schemes are different and are based on different target variables. The best method of locating individual houses where radon mitigation is necessary is to determine the indoor radon concentration in all houses. However, houses with high indoor radon concentration will not necessarily be situated in areas with high geogenic radon hazard and *vice versa*, low indoor radon concentration will not ensure a low geogenic risk because of house construction type and living style. To determine the geogenic radon hazard a closer look to the soil gas radon concentration and the geological situation seems more beneficial. Therefore, harmonizing different approaches is a challenge for the future. The introduction of a Geogenic Radon Index is in discussion which should be deducible from very different data sets like indoor Rn concentration, soil gas Rn concentration, U or Ra concentration in rock or soil, permeability, external dose, etc. It is clear, that such different input data will have different reliability for the final target variable (e.g. geogenic Rn hazard) and, therefore, will cause different sizes for the uncertainties in the classifications.

Conclusion

The creation of a map which visualizes the "radon hazard" is crucial in view of the responsibility for the health of the public. Many different approaches are possible and implemented in different countries. There is no simple way to create such a map, which should be deducible from very different sources of information. Different methods for the estimation of radon related target variables in mapping units exist, assuming spatially uncorrelated or spatially correlated input data. Very different sources of uncertainties exist which are partly difficult to identify and to quantify, but a reliable estimate of the uncertainties is a key point in the usefulness of a radon hazard map. Finally, a classification scheme for the radon hazard is necessary which should be acceptable for most countries. Based on such a harmonized classification scheme decisions for prevention and mitigation measures should be made.

References

1. Bossew P, Dubois G (2008) Geostatistical challenges encountered in mapping indoor radon concentrations. In: Ortiz J, Emery X (eds) Proc of the 8th Int Geostatistics Congress, Geostats 2008, 1–5 December 2008, Santiago, Chile. Gecamin Ltd, Santiago, pp 851–860, http://downloads.gecamin.cl/cierre_eventos/geostats2008/prsntcns/00162_00731_pr.pdf
2. Friedmann H (2005) Final results of the Austrian radon project. Health Phys 89:339–348
3. Friedmann H, Gröller J (2010) An approach to improve the Austrian radon potential map by Bayesian statistics. J Environ Radioact 101:804–808
4. Neznal M, Neznal M (2006) Reliability of the new method for assessing the radon risk – gas permeability classification. In: Barnet I, Neznal M, Pacherová P (eds) Radon investigations in the Czech Republic XI and the 8th Int Workshop on the Geological Aspects of Radon Risk Mapping. Czech Geological Survey, Radon v.o.s, Joint Research Centre, Prague, pp 166–172
5. Tollefsen T, De Cort M, Bossew P (2009) Status of the European indoor radon map. Int Conf Radon in Environment, 10–14 May 2009, Zakopane, Poland. Presentation A18, http://radon.ifj.edu.pl/zakopane2009/