Towards a multivariate geogenic radon hazard index

P. Bossew¹, G. Cinelli², T. Tollefsen², M. DeCort²

¹ German Federal Office for Radiation Protection, Berlin
² European Commission, Joint Research Centre (JRC), Directorate for Nuclear Safety and Security, Ispra, Italy

8th Conference on Protection against Radon at Home and at Work (8th Radon conference) &
13th International Workshop on the Geological Aspects of Radon Risk Mapping (GARRM 13th)

Prague, Czech Republic, 12-16th of September 2016
Rationale & objective

• Identify areas where for geogenic reasons the probability is elevated to encounter elevated indoor Rn concentrations.

• These are geogenic Rn prone areas / Rn priority areas / Rn hazard areas... or however called – see our presentation about Rn prone (or whatever) areas.

• BSS! ... Article 103 § 3: “Member States shall identify areas where the radon concentration (as annual average) in a significant number of buildings is expected to exceed the relevant national reference level”.

• The “Rn proneness” shall be measured by a quantity which can be calculated from available geogenic data.

• Across European countries different sets of geogenic quantities are available. It shall therefore be possible to calculate the “Rn proneness quantity” from different sets of input quantities.

• We call this quantity the Rn hazard index, RHI term originally proposed by H. Friedmann, in a document for the report to the European Geogenic Radon Map, 2008.

• Target in the future: Create a RHI map of Europe based on data which are available country wise. Therefore, the method shall be as simple and flexible as possible, to allow application to different data situations.
Content

This presentation:

• Input quantities
• Conditions to a RHI
• Ideas how to construct a RHI, incl. a bit of math. (sorry)
• First examples; using DE data
• So far only ideas and trials, no final conclusion about methodology!

Note:
• work on generating a RHI is ongoing. Here we show just another step in this endeavour which reflects our current wisdom (if any).
• A first version of this presentation has been shown at the “V. Terrestrial Radionuclides in Environment International Conference on Environmental Protection / VIII. Hungarian Radon Forum and Radon In Environment”, Veszprém, Hungary, 17 – 20 May 2016.
• This modified version includes some new ideas and methodical developments. It is certainly not the last word … more can be expected to follow in the future!
Reminder: Rn - From rock to risk

Radon – a complex system

Often factors are
- by themselves heterogeneous
- interact in complicated way, sometimes not well known

Result: complicated dependence of Rn quantities.

Further, often factors are
- fuzzy or ill defined;
- not well known;

Result: difficult to understand the source of variability
Reminder 2: The geogenic radon potential

• The geogenic radon potential (GRP):
  measures “what earth delivers” in terms of Rn; a measure of the availability to exhalation from the ground, or for ingestion into buildings.
  Several possible definitions; a common one (used also here): “Neznal-RP” (slightly modified),
  \[ \text{GRP} := \frac{C(\text{soil})}{(-10 \log(k) - 10)} \]
  \( C(\text{soil}) \) Rn concentration in soil or rock (kBq/m³), \( k \): gas permeability (m²),
  both measured acc. a defined protocol.
  \textit{Many other proposals! – See our presentation “Definition and estimation of radon prone areas – a review”}

• Geogenic Rn prone area:
  An area in which the probability of elevated indoor Rn concentrations is increased for geogenic reasons. In most cases most important control: the geogenic RP. Rn concentration in an actual building depends on building characteristics (presence of basement, isolation against ground) and of ventilation habits of users or inhabitants.
The geogenic radon potential, 2

Wanted:
Multivariate definition of Geogenic Radon Risk Index

Geogenic Rn map = independent of anthropogenic factors

- defined only in a house
- indoor Rn
- subject to human activity, temporally variable
- living habits
- variability at different time scales
- meteor
- house construction
- defined everywhere on solid earth

"what earth delivers", without influence of human interference, temporally constant over geologic timescale

Geogenic Radon Risk Index

- Outdoor Rn
- Indoor Rn
- Terrestrial gamma dose
- U in soil
- Soil Rn
- Soil Permeability
- U in bedrock
- Faults and Fractures
- Aquifers
geogenic and anthropogenic compartments

- The GRP quantifies availability of Rn for infiltration
- Anthropogenic factors determine, to which extent available geogenic Rn leads to indoor Rn concentration... “infiltration and accumulation potential”

observable quantities that can be used for constructing the RHI

e.g. EURDEP database
Geochemical data, e.g. GEAMAS, FOREGS, in situ-gamma, aero-gamma

exhalation: outdoor Rn, Rn prog.

exhalation rate

permeability

Rn-prog

U

Ra

GRP

indoor Rn

anthropogenic compartment

usage patterns
building characteristics

transport: permeability
source: U

geogenic quantities

GRP

geogenic compartment

terrestrial gamma DR

soil Rn surveys

exhalation:
outdoor Rn, Rn prog.
Input quantities

- **Geogenic Rn potential GRP**: measures the availability of Rn for exhalation into the atmosphere and infiltration into a building
- **Rn concentration in soil air**
- **Rn exhalation rate from the surface**
- **U, Ra concentrations in the ground (soil, rock, creek sediments)**: from samples, in situ-gamma, aero gamma
- **Terrestrial component of ambient dose rate**
- **Standardized indoor Rn concentration** (e.g. Friedmann Rn potential)
- **Geological units, soil types**
- **Hydrological quantities**: Groundwater recharge coefficient GWRC
- **“Special features”**: indicate the presence of objects which may influence the GRP: tectonic lines, caves, mines, anthropogenic sources,...

**Lack of harmonization:**

- Countries (sometimes even regions of one country) have different datasets;
- Same nominal quantity defined or measured by different method

**Therefore: RHI !, measure of “geogenic radon proneness”**
Initial idea (Cinelli et al. 2015)

European Geogenic Radon Map: Multivariate classification approach

Grid 10 km x 10 km

Input Variable

SOIL GAS Rn (1)
INDOOR Rn (2)
GEOLOGY (3)
U_{soil} (4)
U_{rock} (5)
SOTI Properties (5)
TGDR (7)

Classification

s_1
s_2
s_3
s_4, s_5
s_6
s_7

Weight

\omega_1(n)
\omega_2(n)
\omega_3
\omega_4(n), \omega_5(n)
\omega_6(n)
\omega_7(n)

Geogenic Radon Risk Index

n – number of samples per grid cell

Low
High

of 19
desired properties of RHI

• Its value at a location must be independent on which quantities it has been estimated from.
  I.e., RHI calculated from U concentration in soil should have approximately the same value as if calculated from dose rate or GRP, etc.

  This follows from the requirement to be consistent across borders, or regions in which different input quantities are available.

  consistency requirement

• desirable: a RHI estimator which includes as much information as possible (≈”sufficient estimator”)

• calculation shall be as simple as possible.
**Different concepts**

Rn hazard index RHI can be:

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

---

**Input quantities**

- continuous
- categorical / discrete, nominal (unordered)

**RHI**

- continuous
- categorical ordinal

---

this pres.: continuous RHI proposed
Consistency condition

Given input quantities (U, DR, geol. class). Then should be:
\[ RHI(U,.,.) \cong RHI(.,DR,.) \cong RHI(U,.,Geo) \cong RHI(U,DR,Geo) \]
etc.
\[ \cong \] means “up to deviations which are due to the imperfect correlation between geogenic quantities & statistical uncertainty”
Why?
Because it shall be applicable independent of the input quantities in a region.
This is the most difficult condition!
some options

original quantities

Z₁

Z₂

Z₃

classify

Z₁', Z₂', Z₃'

combine

e.g. weighted mean

RHI

rescale

Z₁', Z₂', Z₃'

combine

RHI

dimensional reduction

construct new variable which contains most of joint variability

some varieties shown here

proposal Cinelli et al. 2015

some options

Z₁

Z₂

Z₃

classify

Z₁', Z₂', Z₃'

combine

e.g. weighted mean

RHI

rescale

Z₁', Z₂', Z₃'

combine

RHI

extract

e.g. first principal component

RHI

original quantities

Z₁

Z₂

Z₃

classify

Z₁', Z₂', Z₃'

combine

e.g. weighted mean

RHI

rescale

Z₁', Z₂', Z₃'

combine

RHI

extract

e.g. first principal component

RHI
Proposal: 3 families of methods

• **Methods “F”**:  
  Transform $\text{GRP}$, $U$, $\text{DR} \rightarrow F_{\text{GRP}}(\text{grp})$, $F_{U}(u)$, $F_{\text{DR}}(dr)$; inclusion of geology (or other categorical quantities): $F(\text{geo}) := F_{\text{GRP}}(\text{GM}[\text{GRP}|\text{geo}])$  
  $\text{RHI}:=\text{AM}[F]$ (equal weights for now)  
  **rationale**: for $Y=g(X)$ strictly monotonous, $F_X(x)=F_Y(y)$  
  **problem**: estimation of distributions $F$; data must be representative for the population; here by estimation on a common grid (10 km $\times$ 10 km). Difficult for region in which no representative data of a quantity are available.

• **Methods “R”**:  
  Regression model $\text{GRP}^i=g^i(Z_i)$, $Z^i=U$, $\text{DR}$, ...; inclusion of geology: $Z^{\text{geo}}:=\text{GM}[\text{GRP}|\text{geo}]$  
  $\text{RHI}:=\text{AM}[\ln \text{GRP}, \ln \text{GRP}^i, \ln Z^{\text{geo}}]$ (equal weights for now)  
  **rationale**: treat $Z^i$ as GRP proxies  
  **problem**: bad relations $Z^i \sim \text{GRP}$ (see from rock to risk!); here by quantile regression

• **Methods “P”**:  
  Perform $\text{PCA}($ln GRP, ln U,..., ln$Z^{\text{geo}})$; $\text{RHI}:=\text{first component}$ (i.e. weights= 1. eigenvector)  
  option: alternative inclusion of geology: generate indicators $Z^{\text{geo}}(\text{geo}'):=1(\text{geo}=\text{geo}')$, treat as #(geo)-1 independent variables  
  **rationale**: reflects approx. correlation  
  **problem**: physical justification?  

• for “R” and “P” rescale to (0,1). Here done by tgh transform parameterized such that $\text{AM} \rightarrow 0.5$, $\text{Min} \rightarrow 0.01.$
Proposal, 2:

RHI models chosen here:

<table>
<thead>
<tr>
<th></th>
<th>GRP</th>
<th>U-conc</th>
<th>terr. DR</th>
<th>indoor-C</th>
<th>geology</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>F2</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>R1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>R5</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>P3</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>P4</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>P5</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>P6</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

idea: see effect of including or excluding GRP (available in few countries only) or indoor Rn or geology explicitly
RHI & RPA maps

(only some shown as examples)

Rn prone areas (RPA) by cross-classification against prob($C_{\text{indoor}}$>$100$)$>0.1$ by maximizing Y-stat on ROC graph
Performance of RHI

2 measures chosen: RHI as predictor of $p = \text{prob}(C_{\text{indoor}} > 100 \text{ Bq/m}^3)$

1) AUC (area under ROC curve = measure of deviation from random, AUC=0.5) in cross-classification; threshold for $p$: 0.1 chosen.

2) $r^2$ (coeff. of determination) of logistical regression

both: the higher the better!

RHIs perform a bit better than individual quantities, but not really a dramatic difference!

(including indoor-C naturally improves prediction capacity)
Conclusions

• Possible approaches shown for defining RHI;
• technically relatively simple;
• results not yet entirely convincing: predictive capacity not really increased as expected

To do:
• Further develop method; also the score-type approach
• Uncertainties?
• Test on different datasets
• Include additional quantities: tectonic, hydrologic, soil type and texture, surface geology, “special features”
• how to reasonably include small features with very high RP?
• Recent new idea: link RHI or existing RPA definitions to an “action catalogue”
  → next Rn conference!
Thank you for your attention!