Towards a multivariate geogenic radon hazard index

P. Bossew[1], G. Cinelli[2], T. Tollefsen[2], M. DeCort[2]

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

v.8.9.16

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

<u>This presentation:</u>

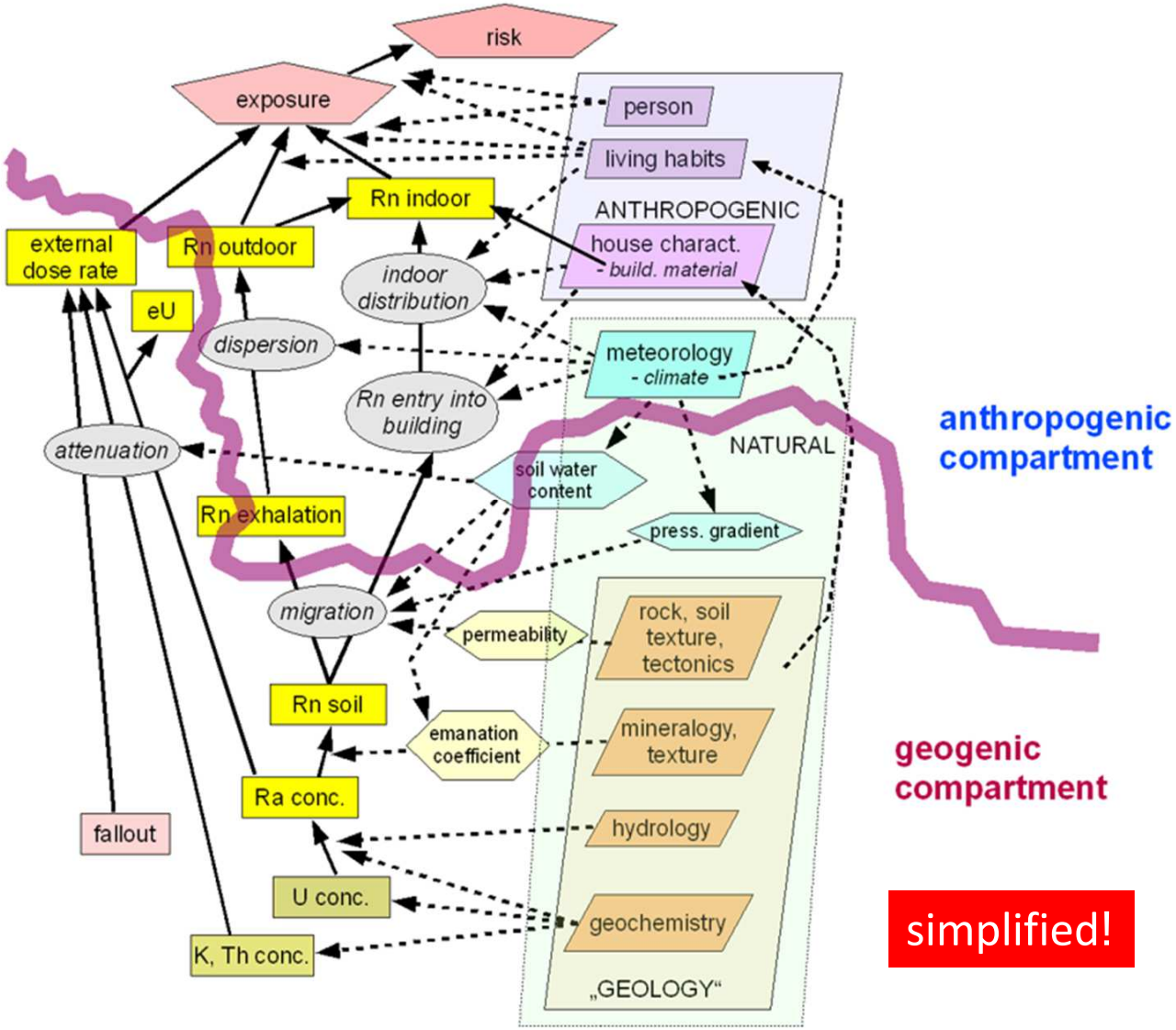
- Input quantities
- Conditions to a RHI
- Ideas how to construct a RHI, incl. a bit of math. (sorry)
- First examples; using DE data
- So far only ideas and trials,
  no final conclusion about methodology!

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

<u>Result:</u> complicated dependence of Rn quantities.

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

<u>Result:</u> difficult to understand the source of variability

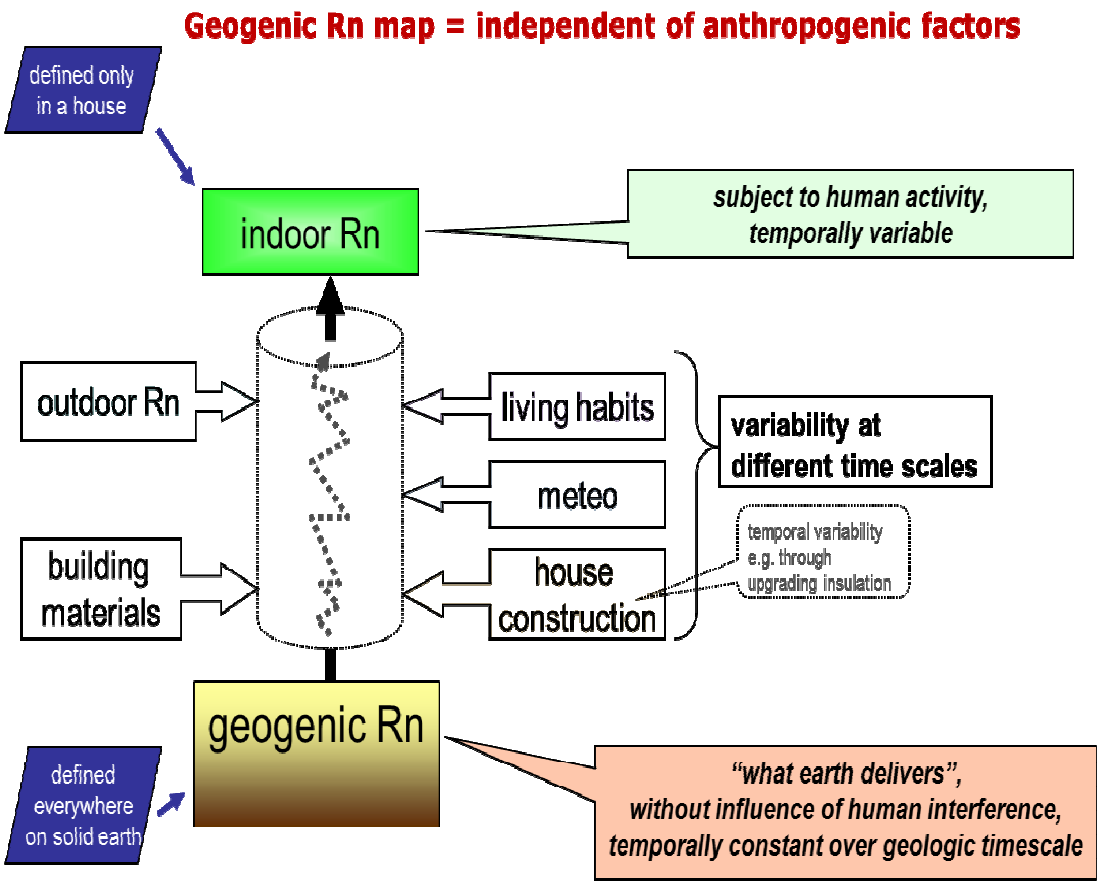
simplified!

Reminder 2: The geogenic radon potential

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

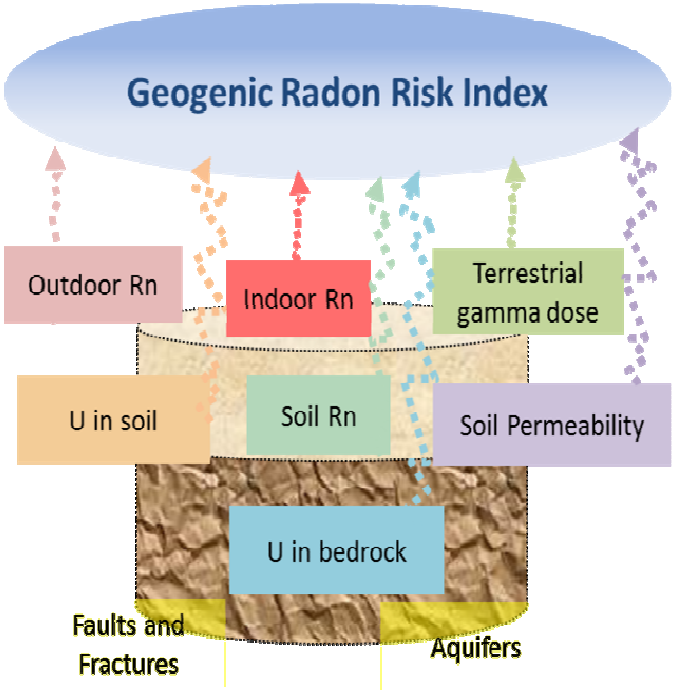
- <u>Geogenic Rn prone area:</u>
  An area in which the probability of elevated indoor Rn concentrations is increased for geogenic reasons. In most cases most important control: the geogenic RP. Rn concentration in an actual building depends on building characteristics (presence of basement, isolation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

The geogenic radon potential, 2

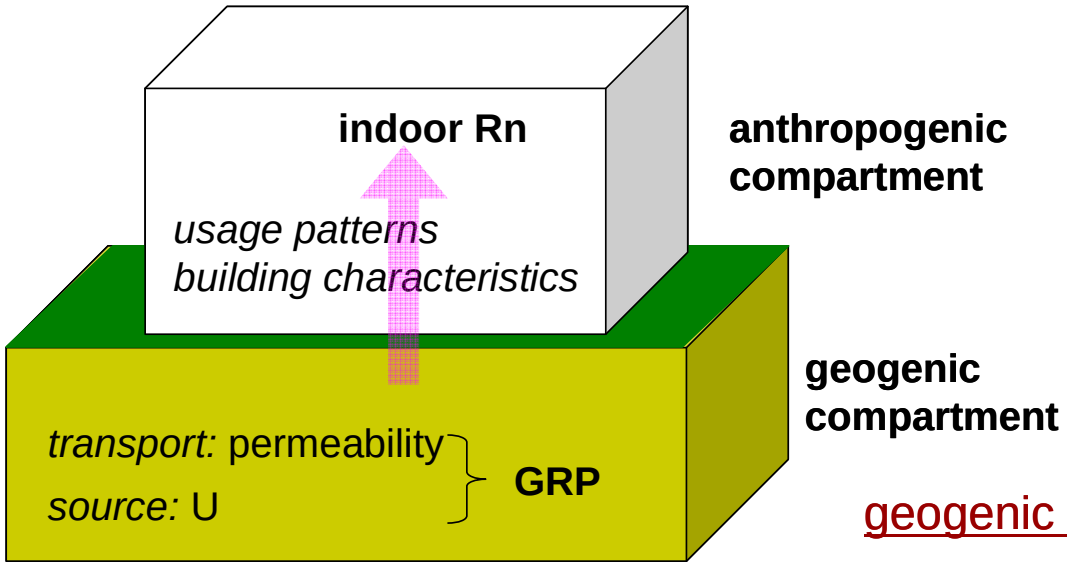


Wanted:

Multivariate definition of Geogenic Radon Risk Index

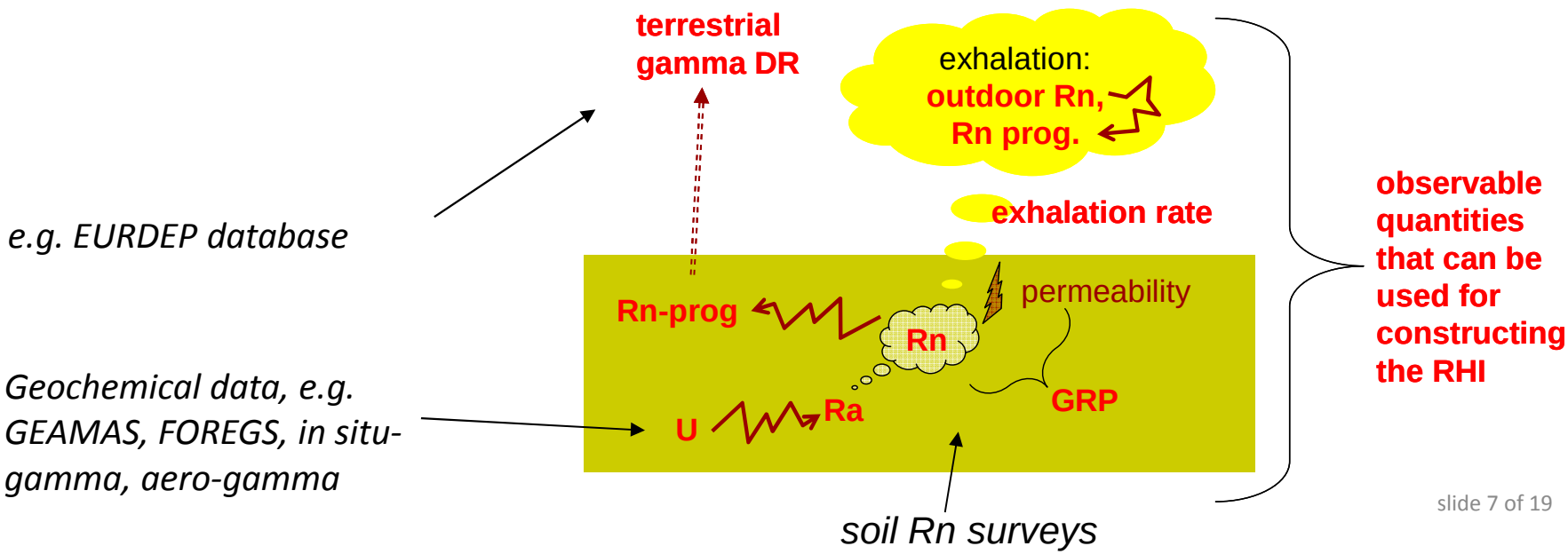


geogenic and anthropogenic compartments



- 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"

geogenic quantities



Input quantities

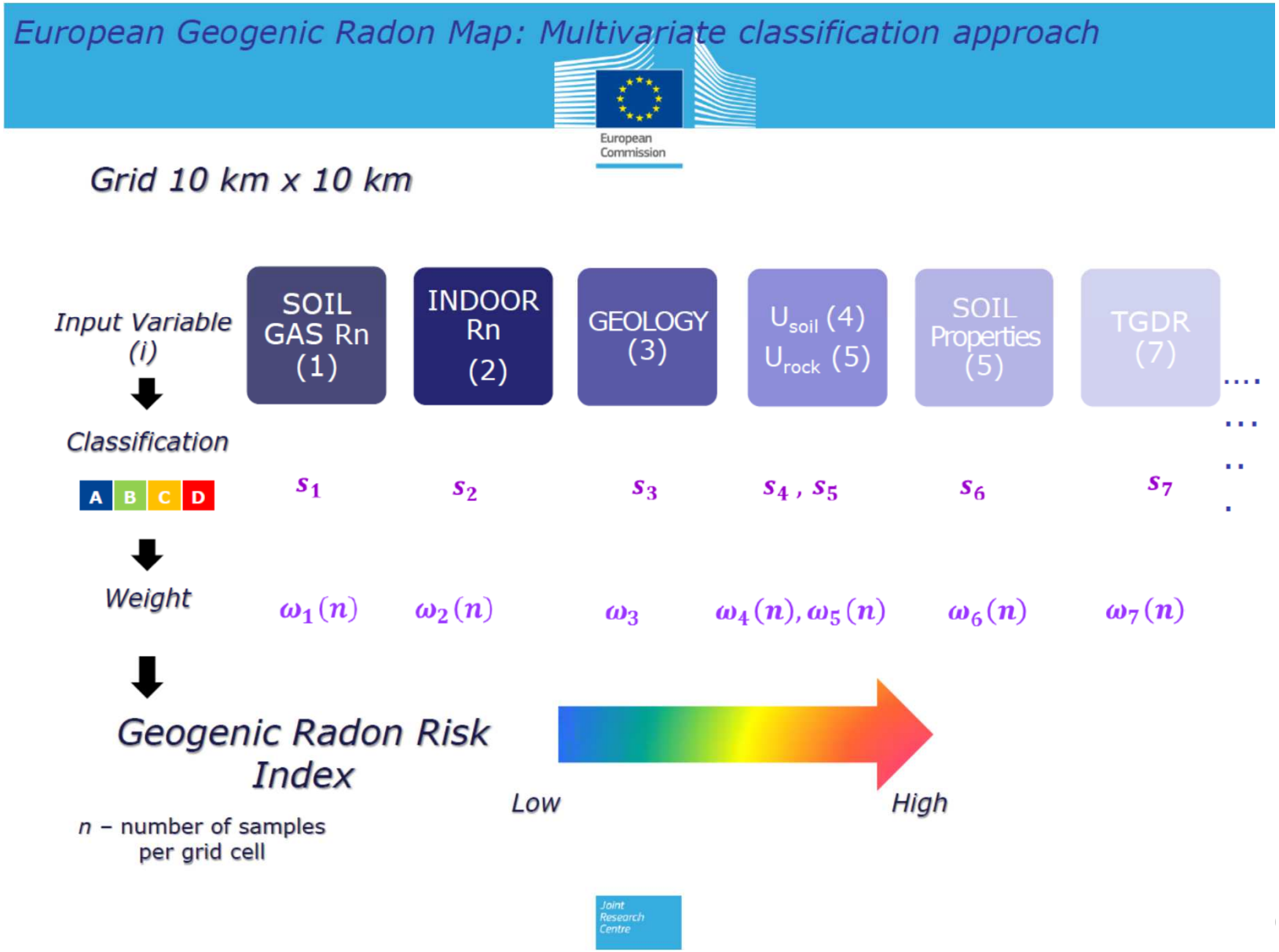
- Geogenic Rn potential GRP: measures the availability of Rn for exhalation into the atmosphere and infiltration into a building
- Rn concentration in soil air
- Rn exhalation rate from the surface
- U, Ra concentrations in the ground (soil, rock, creek sediments): from samples, in situ-gamma, aero gamma
- 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

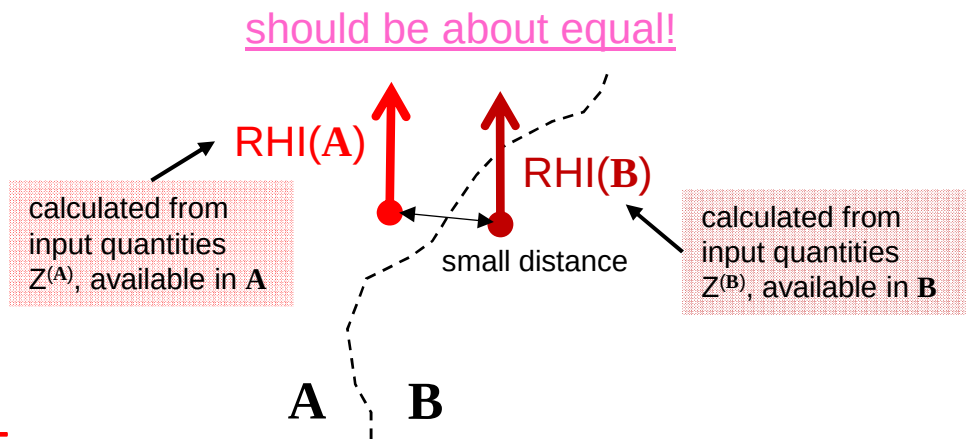
Therefore: **RHI !,** measure of "geogenic radon proneness"

Initial idea (Cinelli et al. 2015)



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

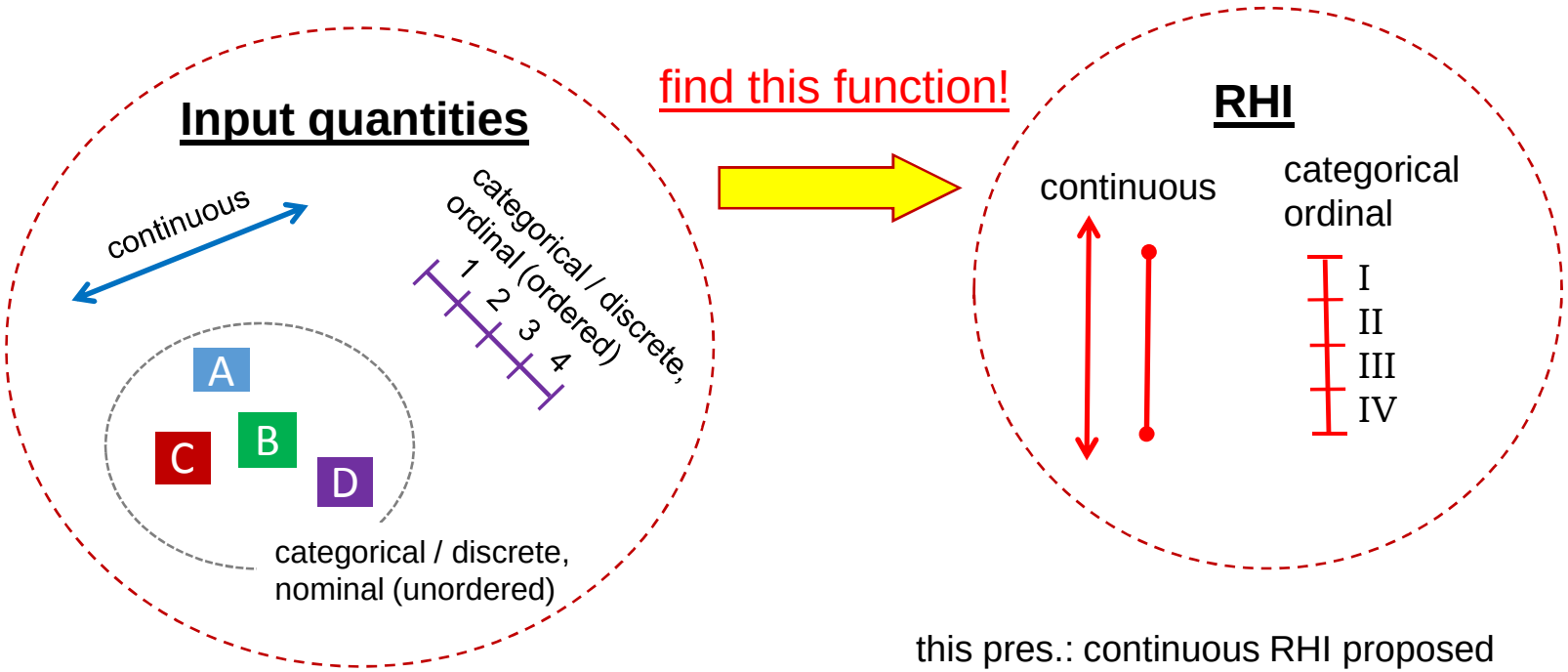


- desirable: a RHI estimator which includes as much information as possible (~"sufficient estimator")
- calculation shall be as simple as possible.

Different concepts

Rn hazard index RHI can be:

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



this pres.: continuous RHI proposed

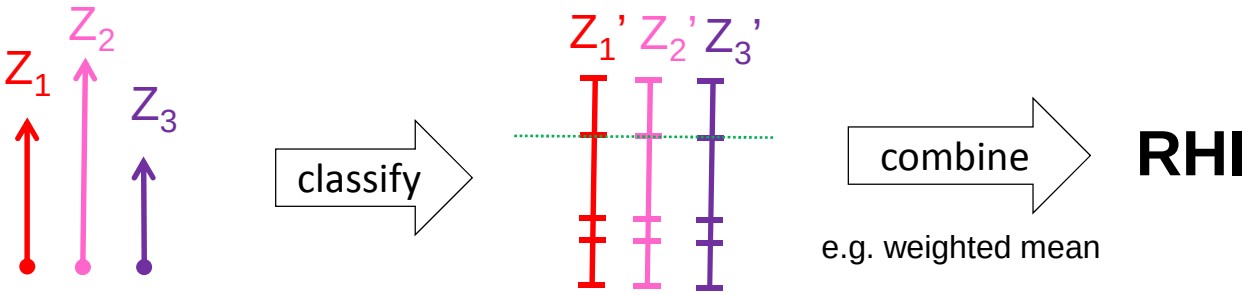
Consistency condition

Given input quantities (U, DR, geol. class).
Then should be:
RHI(U,.,.) ≅ RHI(.,DR,.) ≅ RHI(U,.,Geo) ≅ RHI(U,DR,Geo)
etc.
≅ 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!

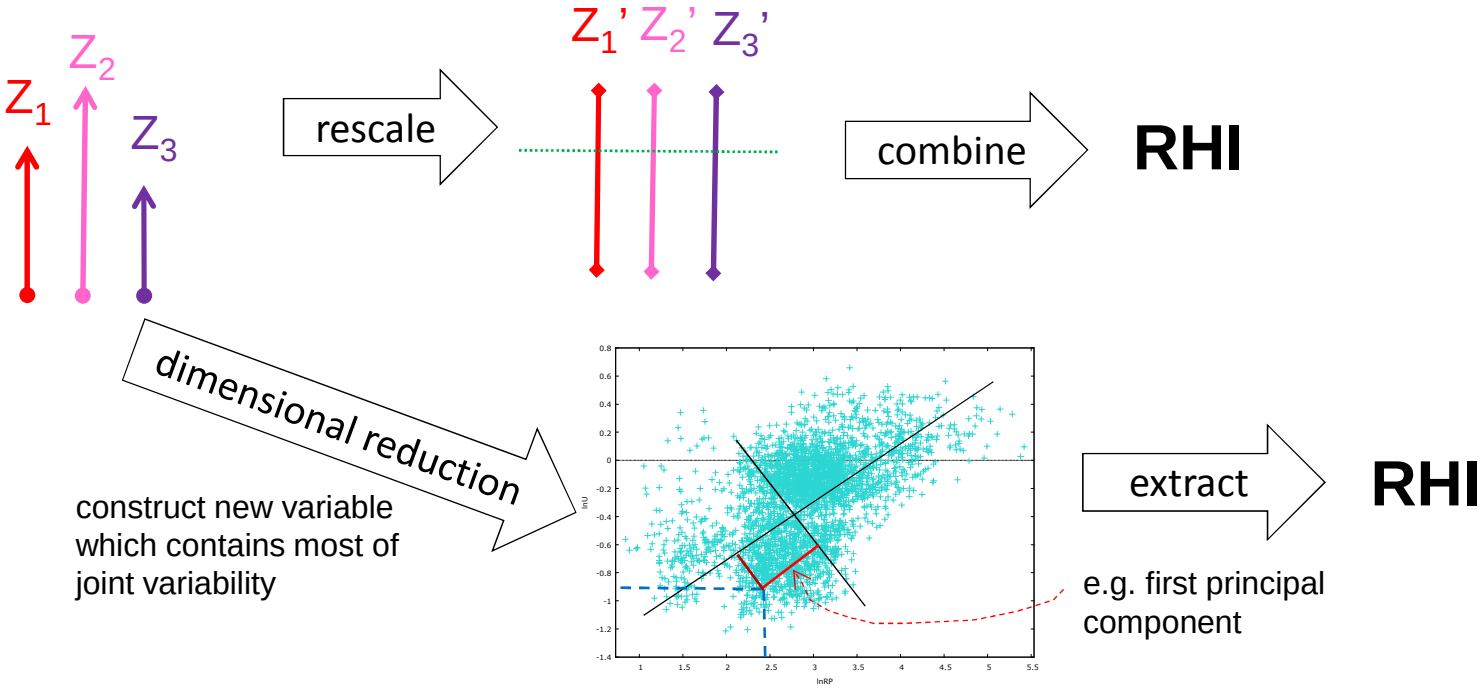
some options

original quantities

proposal Cinelli et al. 2015



some varieties shown here



Proposal: 3 families of methods

- <u>Methods "F":</u>
  Transform GRP, U, DR $\rightarrow$ F_{GRP}(grp), F_U(u), F_{DR}(dr); inclusion of geology (or other categorical quantities): F(geo) := F_{GRP}(GM[GRP|geo])
  RHI:=AM[F^i] (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.

- <u>Methods "R":</u>
  Regression model $GRP^i=g^i(Zi)$, Z^i=U, DR,…;
  inclusion of geology: Z^{geo}:=GM[GRP|geo]
  RHI:=AM[ln GRP, ln GRP^i, ln Z^{geo}] (equal weights for now)
  rationale: treat Z^i as GRP proxies
  problem: bad relations Z^i ~ GRP (see from rock to risk!); here by quantile regression

- <u>Methods "P":</u>
  Perform PCA(ln GRP, ln U,…, lnZ^{geo}); RHI:=first component (i.e. weights= 1. eigenvector)
  option: alternative inclusion of geology: generate indicators Z^{geo}(geo'):=**1**(geo=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 AM$\rightarrow$0.5, Min$\rightarrow$0.01.

Proposal, 2:

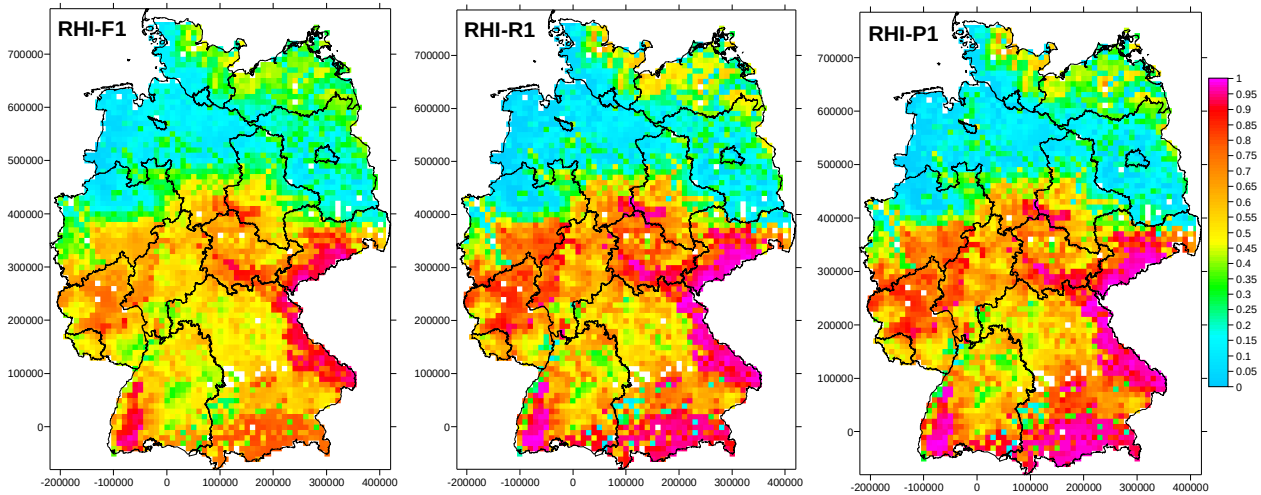
RHI models chosen here:

	GRP	U-conc	terr. DR	indoor-C	geology
F1	x	x	x		x
F2		x	x		x
F3	x	x	x		
R1	x	x	x		x
R2		x	x		x
R3	x	x	x		
R4	x	x	x	x	x
R5		x	x	x	x
P1	x	x	x		x
P2	x	x	x	x	x
P3	x	x	x		
P4	x	x	x	x	
P5		x	x		x
P6		x	x	x	x

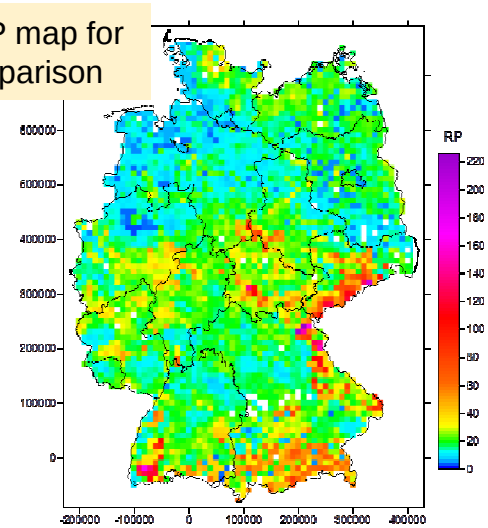
idea: see effect of including or excluding GRP (available in few countries only) or indoor Rn or geology explicitly

RHI & RPA maps

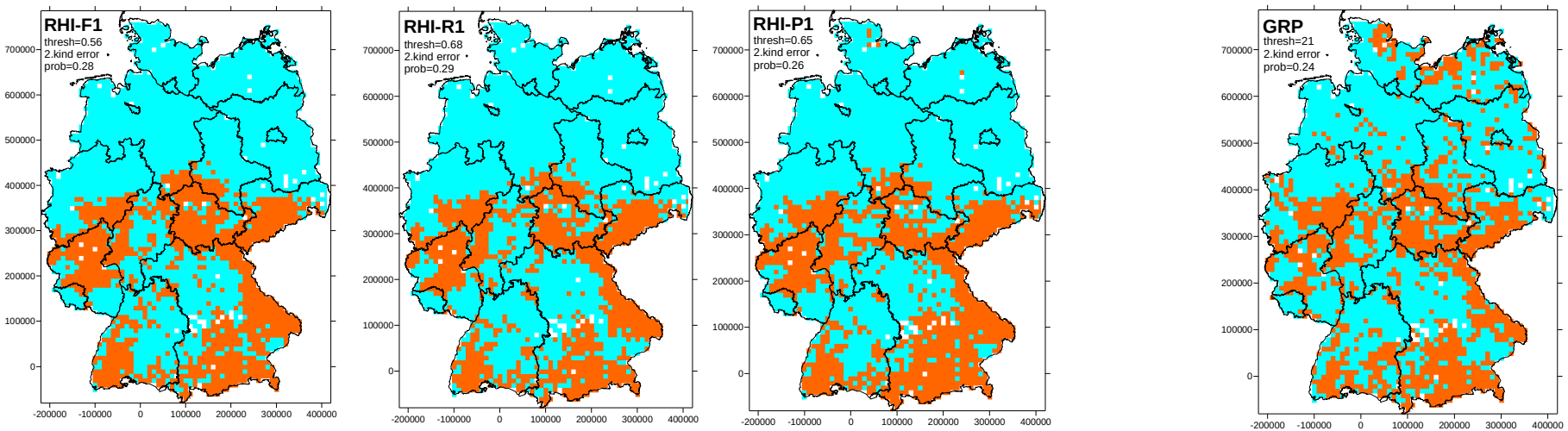
(only some shown as examples)



GRP map for comparison



<u>Rn prone areas</u> (RPA) by cross-classification against prob(C_{indoor}>100)>0.1 by maximizing Y-stat on ROC graph

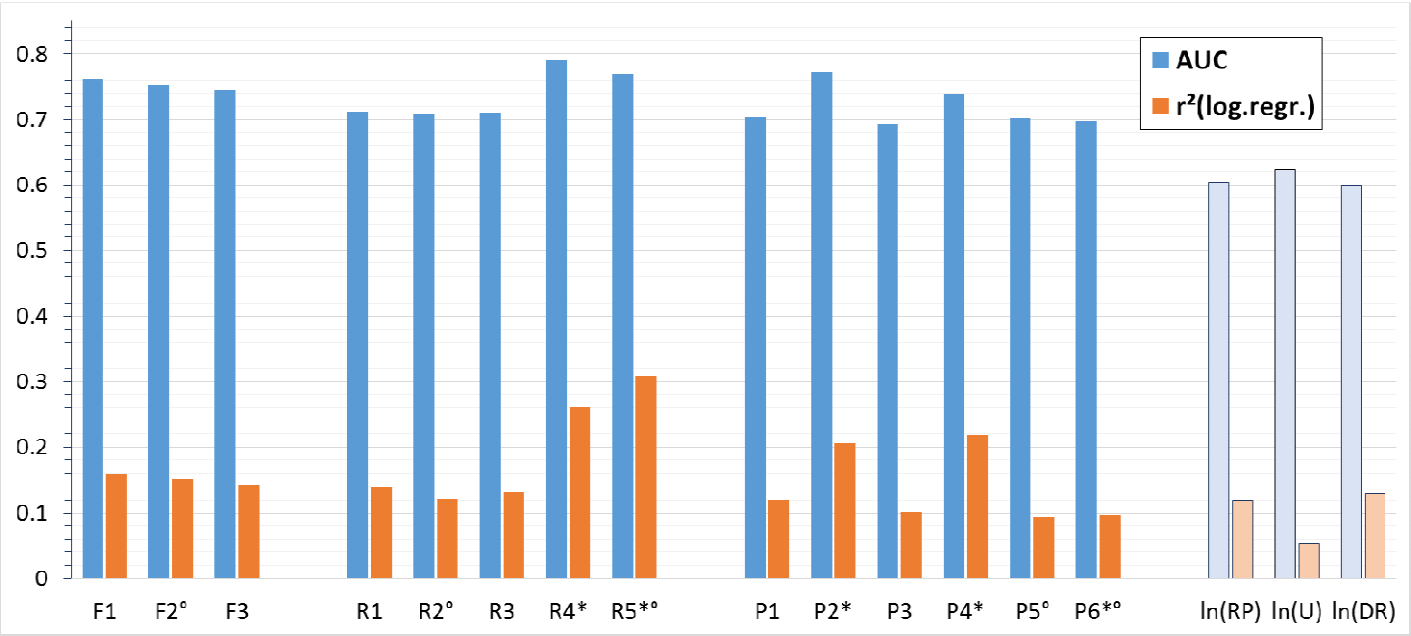


Performance of RHI

2 measures chosen: RHI as predictor of p=prob(C_{indoor}>100 Bq/m³)

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!



°…. GRP <u>not</u> included; *… GM(C_{indoor}) included

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”

  $\rightarrow$ next Rn conference!

Thank you for your attention!

Bundesamt für Strahlenschutz

European Commission