Radon priority areas –
definition, estimation and uncertainty

Peter Bossew

German Federal Office for Radiation Protection (BfS), Berlin

12th
CONFERENCE ON
GEOSTATISTICS FOR
ENVIRONMENTAL
APPLICATIONS

geoENV 2018

BELFAST

QUEEN'S UNIVERSITY BELFAST

v.2.7.18

Content

- Rationale:
  - Importance of indoor radon
  - Legal background
- Definitions of Radon priority areas
- Data sources
- Estimation
- Uncertainties

Indoor radon - essentials

Indoor radon – most important contribution to dose!
Second most important cause of lung cancer after smoking!
In Europe estimated about 62,000 lung cancer fatalities per year attributed to Rn.
(Gaskin et al., Envir. Health Perspectives 125, 5 (2018); incl. RU, TR; missing: BiH, LV, MD, MK, MT, RS, UA)

(figure appears a bit overestimated to me)

<u>Sources of indoor Rn:</u>

1. Geogenic Rn (most important in most cases)
2. Building materials
3. Tap water, natural gas

<u>Concentrations of indoor Rn controlled by</u>

Geogenic factors:
Geology, soil type, U concentration in topsoil, permeability, granulometry,…

Anthropogenic factors:
Construction type (tightness of structures in contact with the ground),
life or usage patterns (ventilation)

Very high local and temporal variability → makes prediction very difficult.

Legal background

<u>Basic Safety Standards (BSS)</u>

Council Directive 2013/59/Euratom laying down basic safety standards for protection against the dangers arising from exposure to ionizing radiation

http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2014:013:TOC (OJ L, 17.01.2014)

Official Journal
of the European Union
L 13
Legislation
EN

<u>Art. 103,3; RPA:</u>
"Member States shall identify areas where the radon concentration (as an annual average) in a significant number of buildings is expected to exceed the relevant national reference level."

Conceptual definition, which has to be translated into an *operable* definition.

<u>Art. 54, 74, annex XVIII; Radon Action Plan:</u>
In areas according Art.103,3: Buildings with public access and workplaces must be measured and if above RL, remediated. New buildings: particular Rn prevention. Strategy to reduce Rn in dwellings.

Reference level (RL): must be ≤ 300 Bq/m³ (BSS Art 54,1 & 74,1). Most countries chose 300, Ireland and others: 200

These areas are called <u>Radon Priority Areas (RPA)</u>, to indicate priority in taking action.
Formerly also "Radon Prone Areas"

RPA definitions, 1

Some examples of operable RPA definitions, based on different Rn measures:

- An area B (grid cell, municipality…), in which the mean population-weighted indoor concentration C exceeds the reference level (RL); $AM_B(C)>RL$; measure = AM_B
- same, but indoor concentration in *dwellings on ground floor*
- An area B, in which the probability that C exceeds the RL, is greater than p (typically 10%); $prob_B(C>RL)>p$; measure = $prob_B$
- The areas B which represent the upper 10% of $AM_B(C)$; measure = percentile
- An area, in which the collective exposure (e.g., $AM_B(C)\times$population) is among the upper 10%

Important:
There is no "natural" definition of RPA! Therefore, also no "true" RPA!
RPAs always depend on definition and to some extent, on estimation method.
This is partly a political decision, partly a pragmatic one (i.e., availability of data).

Consequence:
RPAs may, in general, not be comparable across borders. This may create communication and credibility problems. Discussing this and proposing solutions is another subject of the Metro Radon project. One way may be a map of the Rn hazard index (RHI – currently under development) as "universal" (but still to an extent deliberate) measure of Rn "priorityness".

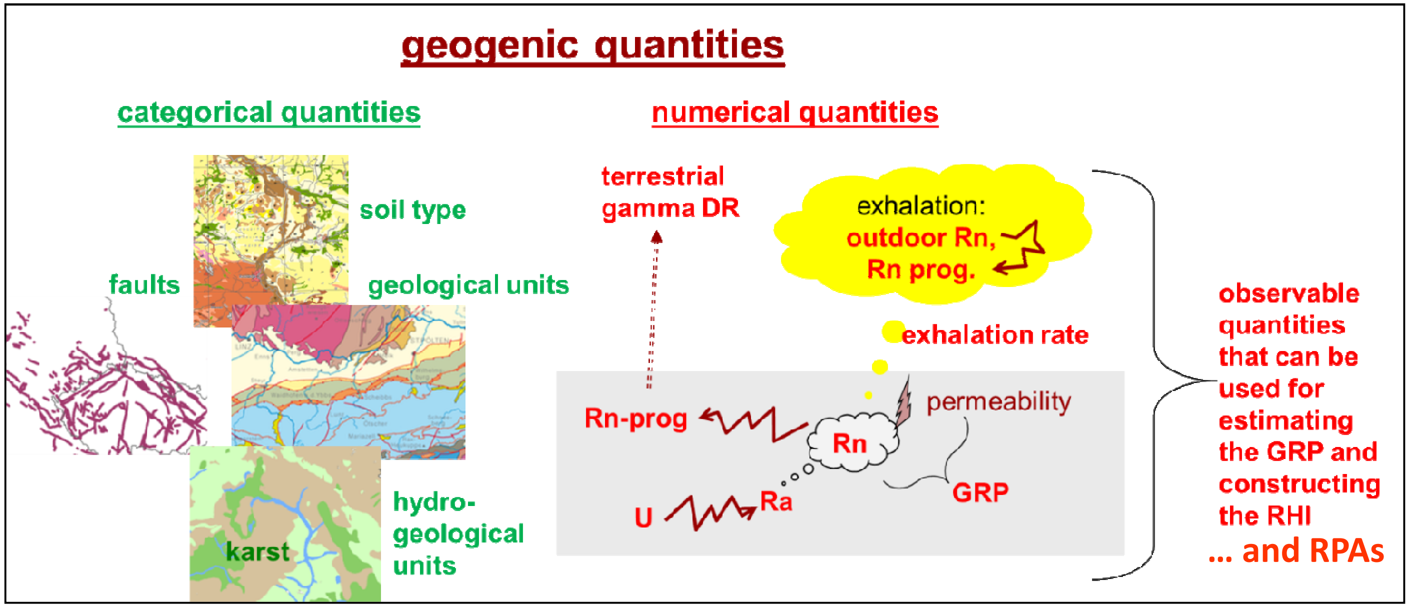
RPA definitions, 2

Multinomial:

Instead of 2 classes (RPA / non-RPA), several classes of “Rn-priorityness”; approach chosen by some countries.

Multivariate:

Although the BSS definition relies on indoor Rn concentration, one may chose to base estimation on other Rn-related variables instead or additionally. Examples: geogenic Rn potential, U concentration in the ground, terrestrial gamma dose rate, geological unit, tectonic features etc.



GRP = geogenic Rn potential
RHI = geogenic Rn hazard index

RPA estimation – a classification problem

Decision about whether a geographical unit shall be labelled RPA or not (in the case of multinomial definition, which grade of “Rn priorityness” it should be assigned): a **classification problem**.

If estimated from secondary quantities: conditional and cross classification.

Existing solutions are pragmatic in the sense that they have to rely on available data and on external “political” parameters such as reference levels and tolerable uncertainty.

Classification uncertainty

Whereas the uncertainties of the estimated actual levels of the Rn measure are commonly quantified by confidence intervals, the ones of classes are given by **first and second kind classification error** probabilities.

The complication consists in the large spatial variability of indoor Rn, also in small scale (~high nugget). Whether estimated from indoor Rn directly or from secondary quantities, this may lead to large classification uncertainty.

In particular:
for geographical units whose Rn measure is close to the class limits.

Data 1

passive
Rn monitors

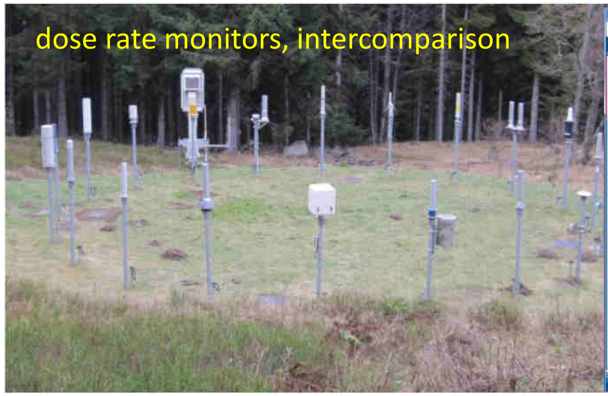
passive Rn progeny monitor

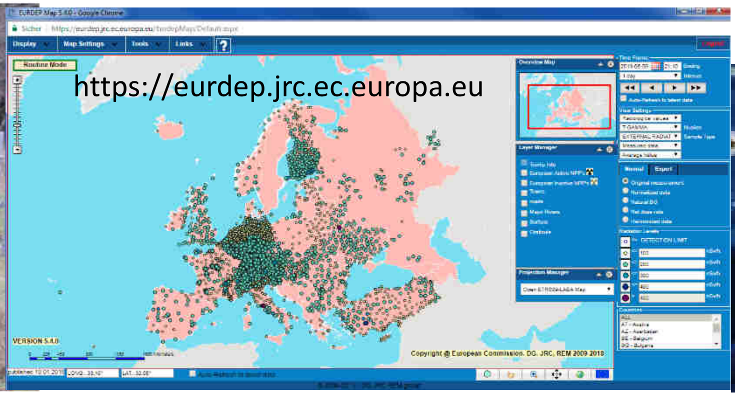
<u>Primary quantity:</u>

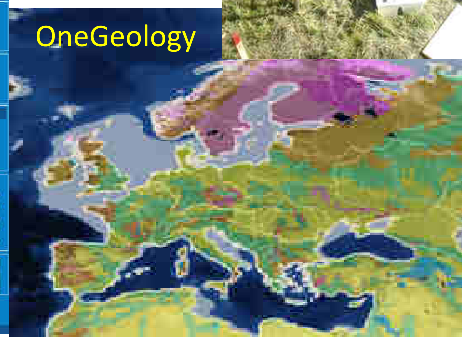
- Indoor radon: mostly measured with passive solid-state ("track-etch") detectors, accumulation time 3 months – 1 year.

<u>Secondary quantities:</u>

- Soil radon, geogenic Rn potential: surveys in a few countries. Mostly grab sampling with probes, 70 cm – 1.2 m deep.
- Geochemistry (U concentration in the ground): European databases, e.g. GEMAS & FOREGS, national and regional surveys.
- Terrestrial gamma dose: EURDEP system (European emergency monitoring system, ca. 5500 stations in continuous operation); national and regional surveys.
- Geology: e.g. OneGeology map; national geological maps.







Data, 2 – Example: European indoor Rn map

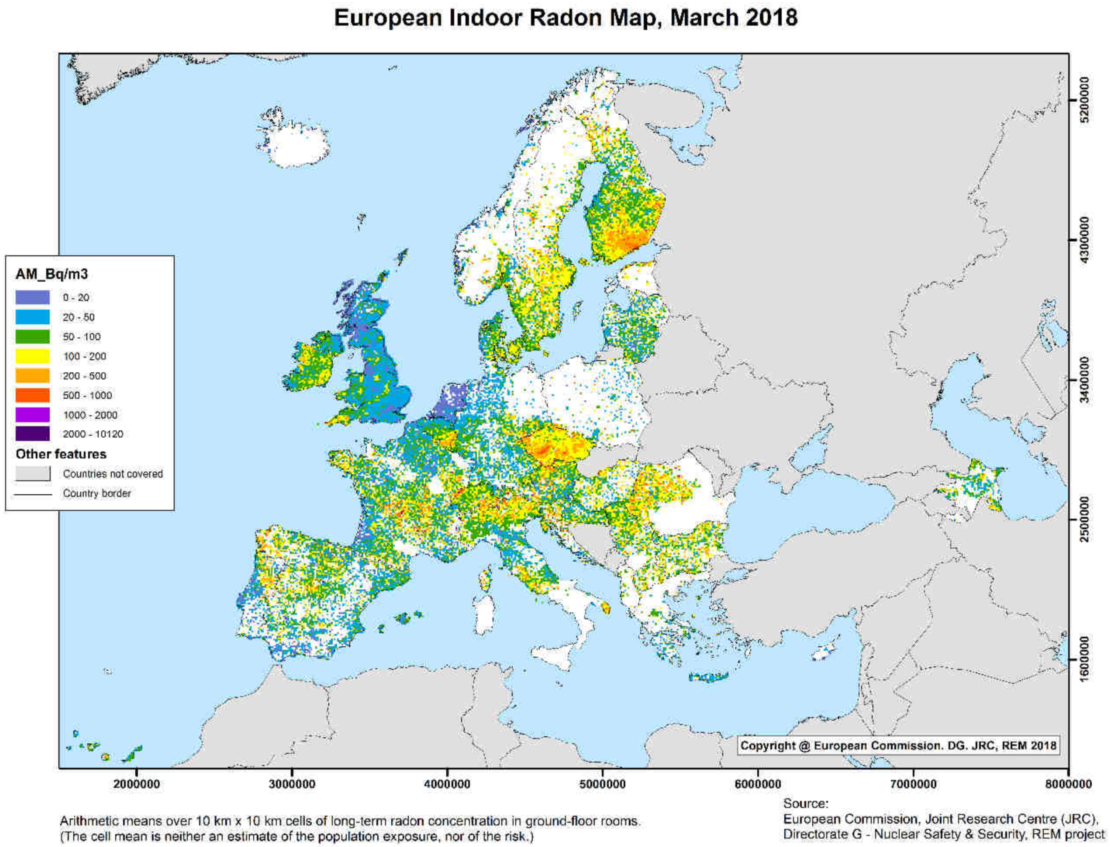
<u>Participating countries</u> aggregate indoor Rn data (residential, ground floor, long term measurements) into 10 km x 10 km cells, aligned to a given grid.

<u>In each cell</u>, statistics are calculated:

- number of measurements
- arithmetic mean (AM)
- standard deviation (SD)
- AM(ln data)
  - → GM = exp(AM ln)
- SD(ln data)
  - → GSD = exp(SD ln)
- median
- minimum
- maximum

Cell statistics are sent to the JRC. Original data remain with suppliers – data protection!

<u>JRC:</u> plausibility checks, statistics over cells, map



Status (March 2018):
34 countries participate
~28,000 non-empty cells
~1,150,000 original measurements

All EU Member States except one have provided data.

Tentative Taxonomy of estimation approaches

certainly not complete!

	spatial correlation not considered	spatial correlation considered
univariate	• sample stat, stat(z) • enhanced by assuming univar. distribution, e.g. LN, stat'(z) $\rightarrow$ cut-off, I(stat(z))	• geostat. model & cut-off, I(Z*) • indicator kriging (hard/soft), I*(z)
multivariate	• ANOVA type • logistic-type regression, logi(z)=g(y) • geographically weighted, local regression etc. $\rightarrow$ cut-off, I(f(y)) • full bivariate through copula • cross-classification	• co-kriging et al. & cut-off, I(Z*;y) • regression kriging & cut-off, I(f*(y)) • indicator regression kriging, I*(f(y)) • indicator co-kriging, I*(z;y) • machine learning

I – indicator according RPA class definition; more complicated for multinomial class $\rightarrow I_1, I_2$,..
z, y – primary and secondary variables

Example 1: Cross-classification

Binomial classification, one secondary variable. Idea:

- RPA definition -> threshold for primary variable (Z);
- Find threshold of secondary variable (Y) on which the RPA map is then based
- Construct truth table & ROC graph
- Perform statistic in ROC space, according constraints, e.g. tolerated 1. and 2. kind errors or optimized classification strength.

Procedure is easy and robust; <u>drawbacks:</u> ignores actual levels of the variables (similar to indicator kriging); ignores spatial correlation. <u>Advantage:</u> easy control over classification error probabilities.

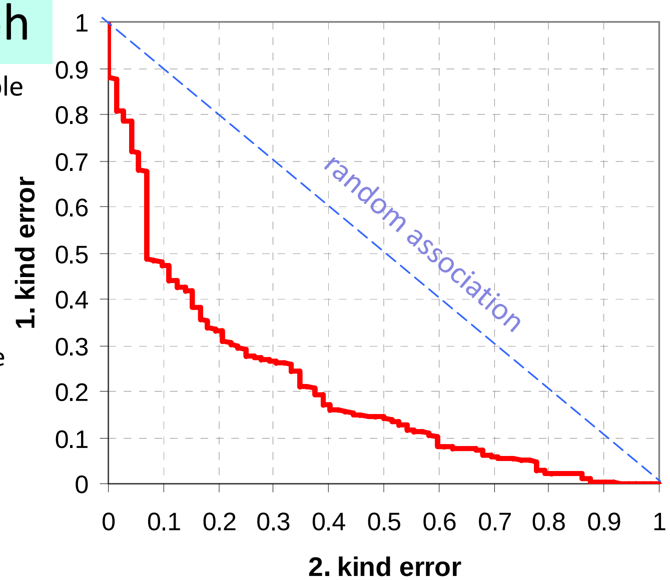
truth table

secondary variable	**primary variable**: no effect	**primary variable**: effect
estimated no effect	correct estimate true negative (TN)	wrong estimate false negative (FN) **second kind error**
estimated effect	wrong estimate false positive (FP) **first kind error**	correct estimate true positive (TP)

ROC graph

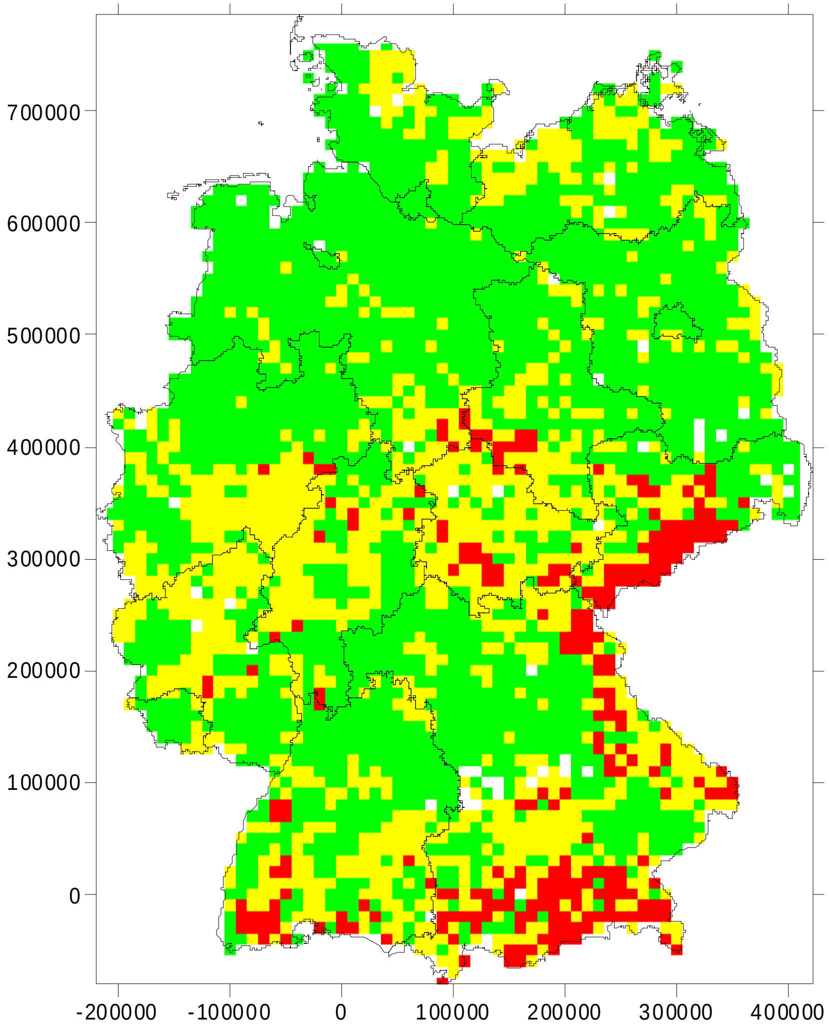
(for example next slide)

classification strength measured by statistics such as AUC (area between curve and diagonal), max. distance from diagonal or min. distance from (0,0)



Example 1, cont.

Suggestion for RPAs, Germany, based on cross-classification method.



Primary variable: Z= indoor Rn concentration in ground floor dwellings, houses with basement;

Secondary variable: Y= Geogenic Rn potential (GRP). Modelled by SGS on 10 km × 10 km grid, geology as deterministic predictor.

RPA definition: grid cell B = RPA, if $p:=prob_B(Z>300$ Bq/m³$) > 3 \times$ German average ≈ 10%.

p estimated by enhanced empirical exceedance prob., assuming LN within cells, GSD=2.

Cell labelled RPA or non-RPA with confidence 90%, i.e. 1. and 2. kind error probability <0.1.

RPA: Y>44.5 CI_{90} by bootstrap (41.2, 48.0)
Non-RPA: Y<20.2 CI_{90}: (14.3, 23.5)
Yellow: undecided

Example, 2a: European indoor Rn data

Enhanced empirical exceedance probability in cell B by LN modelling, given n data Z in U:

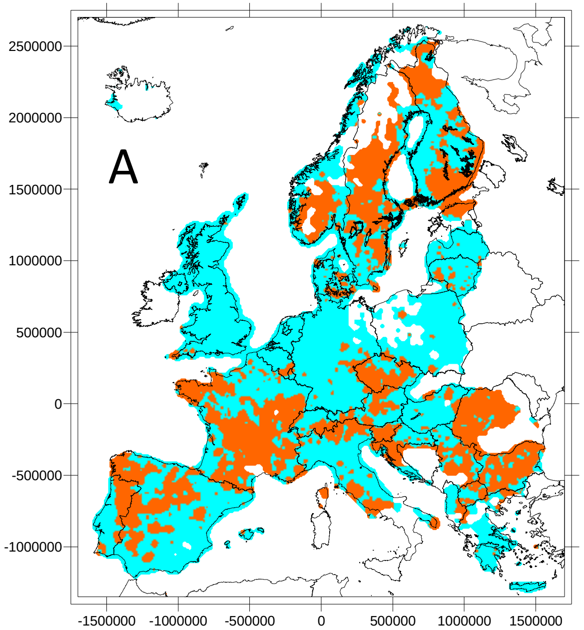
p:=prob(Z>300 Bq/m³); RPA criterion: p>0.1 estimated from cell statistics:

$$p' := \text{prob}_B(Z > z) = t_{n-1}\left(\zeta\sqrt{\tfrac{n}{n+1}}\right);\ \zeta := \frac{\ln z - AM_B(\ln Z)}{SD_B(\ln Z)}$$

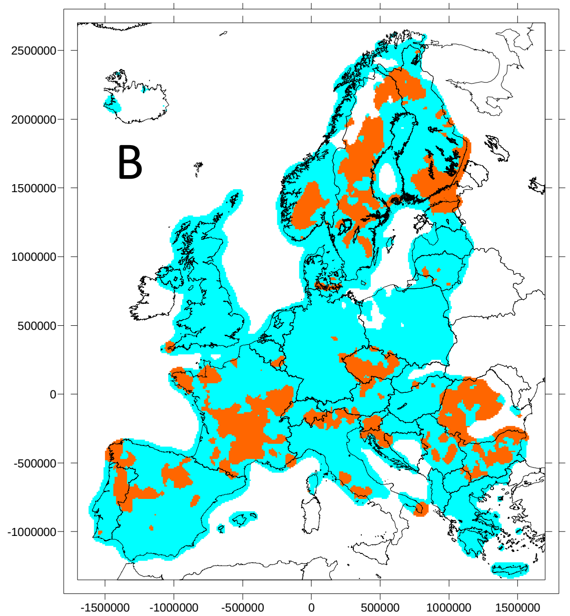
Z = long-term Rn concentration in ground floor dwellings;
z = 300 Bq/m³,
B = 10 km × 10 km cells
* ... OK modelling
# ... classified p< / >0.1

univariate

$p^{*\#}=I_{0.1}(p'^{*}(Z))$



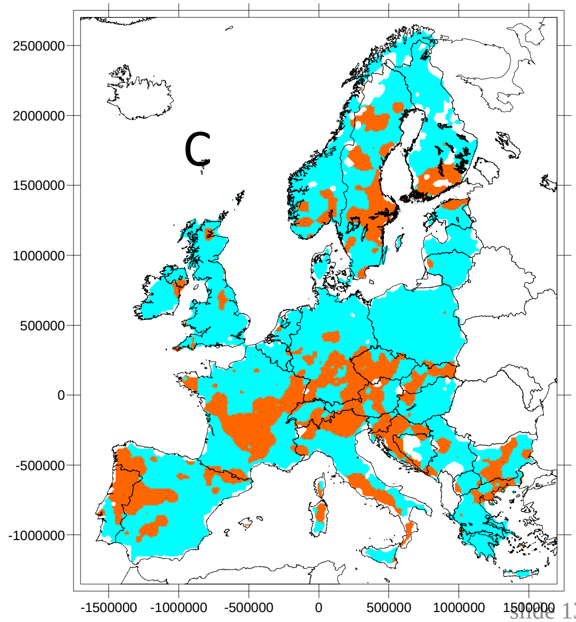
$p^{*\#}=I_{0.5}(I^{*}_{0.1}(p'(Z)))$



bivariate

$p^{*\#}=I_{0.1}(p^{*}(f_{logi}(Y)))$

f_{logi} - logistic regression

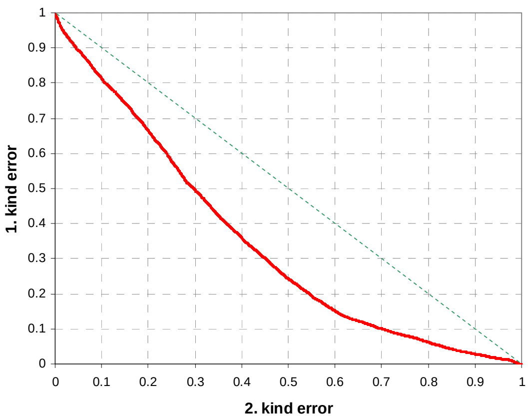


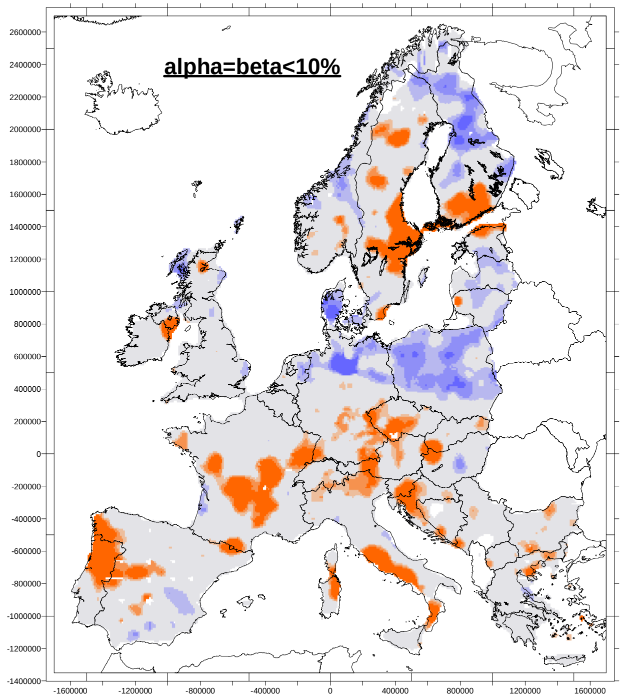
Example, 2b: European indoor Rn data

bivariate cross-classification p' against U concentration in cells B

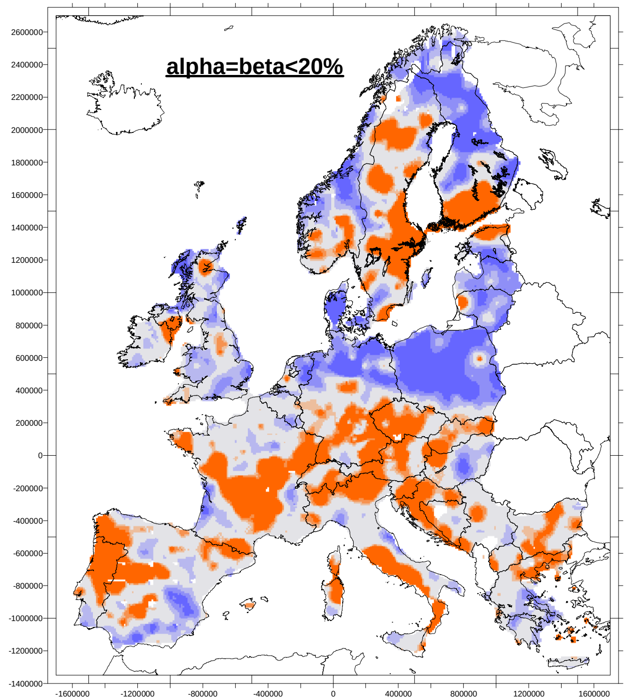
blue: prob(Z>300)<10% with 1-α confidence
orange: >10% with 1-β confidence
grey: undecided
… large areas because of weak association
class limits with 90%C.I. (by bootstrap)

ROC curve close to diagonal: association between p' and U not very strong!





lower: U=1.43 ppm (1.28…1.57)
upper: U=3.27 ppm (3.00…3.73)



lower: U=1.66 ppm (1.50…1.88)
upper: U=2.71 ppm (2.55…2.89)

Uncertainties

- Data uncertainty
  - Radiometric data: counting uncertainty;
  - "Semantic" data uncertainty: possibly erroneous attribute of measurement, e.g. room recorded to be in ground but in reality first floor;
  - Uncertainty about representativeness of data used for inference.

- Model uncertainty
  - "minimal model": sample statistics, e.g. mean or exceedance probability from raw data → sampling statistic → uncertainty (SD of the mean, bias of SD)
  - Data uncertainty inflates dispersion → bias e.g. of exceedance probability
  - Structural uncertainty: choice of model
  - Estimation uncertainty, e.g. of regression parameters
  - $\Rightarrow$ Prediction uncertainty

- Problem, specific to RPA mapping: dwellings vs. workplaces
  - RPA estimation mostly based on dwellings, because large datasets available; but legally relevant for public buildings and workplaces (BSS Art.54,2);
  - Latter and former types of buildings have different Rn characteristic, in general $\Rightarrow$ what has been estimated based on dwellings, may not be true for workplaces.

- RPAs are "random objects". Understood as realizations of a stochastic process, all realizations of RPA maps look differently.

To-do

- Estimation methodology: improvements still possible!
- Uncertainty budgets: largely missing!
- Practically:
  - how to communicate the fact that RPAs are "random objects"?
  - how to deal with RPA uncertainty in administrative decision-making?

The European H2020 Metro Radon project deals with some of these questions.

Conclusions

- RPA definition and estimation: not only academic exercise, but practically important. May have severe economic & political impact. Heavy stakeholder interest!
  Therefore: QA very important!
- Classification theoretically and practically less trivial than it may seem at first glance; in particular because of high small-scale variability of Rn, "irreducible" by modelling or increased sampling rate.
- Methods available, but development still ongoing.

RPA – a sensitive subject!
Action required in RPA can be costly → political disputes

Thank you!

Bundesamt für Strahlenschutz

This work is supported by the European Metrology Programme for Innovation and Research (EMPIR), JRP-Contract 16ENV10 MetroRADON (www.euramet.com). The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States.

MetroRADON