

Designing an indoor radon survey - results of a recent IAEA workshop on survey planning in Brazil

Bossew P.1, Da Silva N.2, Fleury A.P.3, Froidinho 4

1 retired, Vienna, Austria

2 CNEN / LAPOC, Poços de Caldas, MG, Brazil

3 consultant, São Paulo, Brazil

4 bagunceiro, São Paulo, Brazil

v.25.3.24



Vienna, Austria, 14 – 19.4.2024

background and motivation

Brazil is envisaging a large scale plan for indoor radon assessment. Radon levels shall be mapped and priority areas identified. Given the size of the country and its diversity in natural and socio-economical respects, this is a challenging project. Pilot studies and local surveys have been performed in the past but no country-wide assessment exists.

In November 2023, the IAEA organized a workshop on radon survey planning in Poços de Caldas, Minas Gerais, Brazil, to support the project. The objective was to identify items which have to be resolved before starting the actual experimental, i.e., field and laboratory work.

Experience with past surveys:

Certain statistical requirements – as part of QA – often poorly considered.

⇒ Results may be difficult to interpret and their relevance may be questionable.

topics of survey design \Rightarrow QA/QC

- Which is the **objective**?
- **Target quantity**?
- **Mapping support**?
- **Spatial estimation strategy: design based or model based?**
- **How to generate a representative sample, and how to verify it?**
- **If chosen a design based strategy: which sample size is required to achieve a given precision of the result?**
- How should an operational database be structured, which **metadata** should be included?
- **Harmonization:**
 - How should a "cooking recipe" look like, to which generation of new data should follow? ("Bottom-up harmonisation")
 - How can existing data be integrated into the database ("Top-down harmonisation")?

red: \rightarrow following slides, details in supplementary material;

green: \rightarrow supplementary material

1) target quantity

Examples:

- Mean within an area (municipality, region, grid cell, geological unit,...) of:
 - Rn concentration to which inhabitants are exposed
 - Rn concentration in apartments
 - Rn concentration in ground floor living rooms of residential houses
 - Rn concentration in ground floor living rooms of residential houses which have basement
- Probability that in an area a reference value is exceeded
- Status of an area as radon priority area (RPA)

Survey methodology and results depend on the choice of the target variable!

2) design vs. model based approach

Design

- Create a sample such that it guarantees a correct result (i.e. accurate: bias tolerable; precise: random uncertainty tolerable) calculated from the values of the sample.
- Bias depends on the choice of physical instances of the sample, e.g. their location
- Precision depends on the sample size

Model

- Value of the variable Z at location x :
 $Z(x) = f(\text{predictors})$
- predictors:
 - (a) independent (geology,...) \rightarrow regression;
 - (b) the same process $Z \rightarrow$ geostatistics.
- Calibration of the model:
 - in the investigated region (geographical unit), possibly very small and not representative sample
 - in different regions \rightarrow upscaling
- Uncertainty budget: difficult!

Advantages / Disadvantages

	design-based	model-based
Advantages	Conceptually simpler	<ul style="list-style-type: none">• Fewer measurements• Representativeness relaxed
Dis-advantages	<ul style="list-style-type: none">• Many measurements!• Representativity very strict: the most difficult issue!	<ul style="list-style-type: none">• Complex procedure• Predictors in high resolution necessary• Unc. budget complicated; additional unc. of the model itself

3) representative

- Means that the distribution of the sample which has been drawn from the true quantity is the same as the (unknown) one of the latter (up to tolerance).
- In practice, the first two moments should coincide (mean, standard deviation)
- **Probably, taking a representative sample is the most difficult challenge in a survey!**
- Verification through metadata.

Verification of representativeness

Essential step of QC!

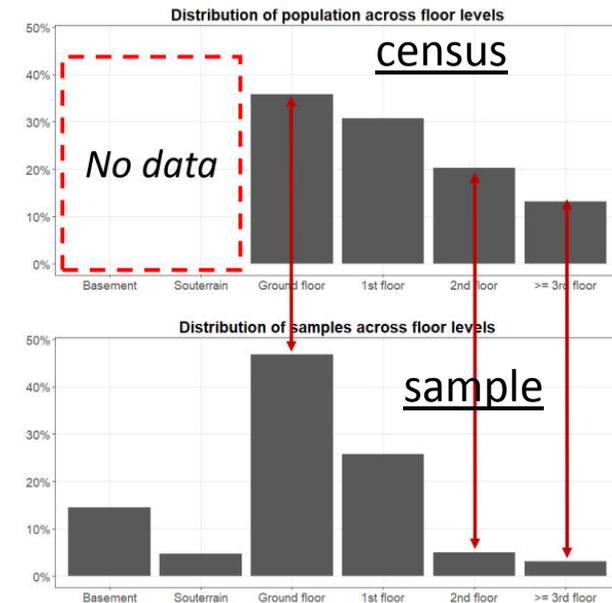
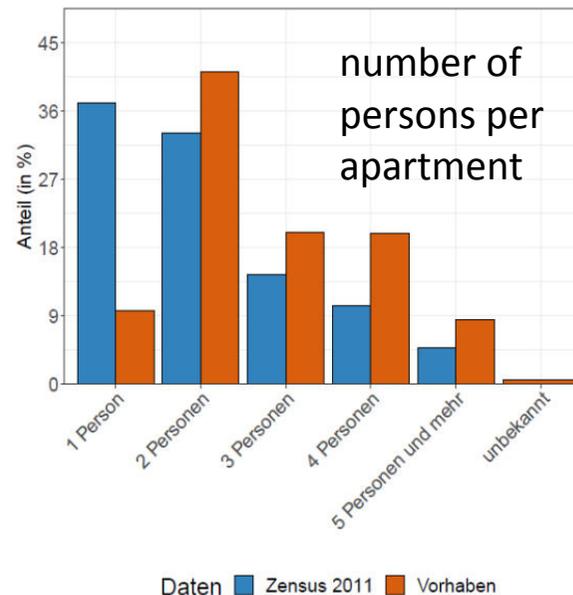
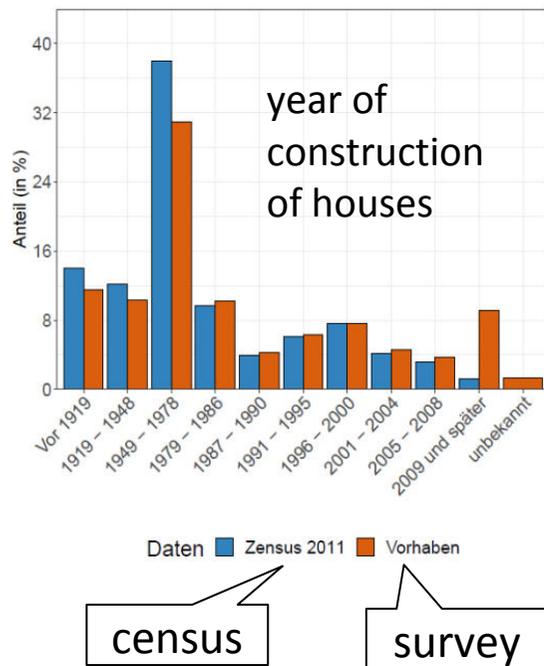
- Done by evaluation of metadata. By comparison of statistics (mean, median, distribution,...) of metadata acquired during the survey with the ones taken from independent reference sources, such as census data.
- Ex. 1: If the sample shall be demographically representative → the geographical distribution of the sample (its regional density) must coincide with the population density.
- Ex. 2: The distribution of the age of buildings measured in the survey must coincide with the one in external databases, best: per region.

Example

- Petermann et al., indoor Rn map of Germany, based on an *almost* representative sample (n = ca. 6500);

<https://arxiv.org/abs/2310.11143>

- Comparison of metadata:



⇒ not optimal in spite of big effort

sources: Kemski & Gruber: Strahlenschutzgespräch Radon, Berlin, 2.11.21; Hoffmann et al., Pres. AIR, Berlin, 15.11.23

4) sample size and precision

- Considerations strictly applicable only for design-based approach
- Sample size such that the target quantity can be estimated with requested precision
- Precision typically expressed as (half confidence interval at significance α) / (mean of the quantity)
- For classification: e.g. by probability of misclassification.
- Sample size the bigger, the larger the statistical dispersion of the data.
- **Some formulas in the supplemental material!**
- Model based: more complicated! Depends on the uncertainty budget of the model output. QA of this family of approaches is still subject of research!

Conclusions

- Many professionals who perform radon surveys are experimental physicists who often do not like statistics.
- But some statistical nut cracking is necessary for a carefully QAed survey!
- Properly designed surveying project means: planning with appropriate QA, which includes that afterwards verification (as part of QC) is possible.
- An experiment is never 100% perfect, but one should know where possible weaknesses are.

Thank you!

For more detailed information including formulas please see the supplementary material which can be downloaded from the EGU platform:

- This presentation incl. supplement
- Material prepared for the IAEA workshop in Brazil (partly identical with the one shown here)

A more detailed paper is planned.



Supplemental slides

Objective of an indoor radon survey

- Mapping the geographical distribution of a target quantity (mean Rn concentration, probability of exceeding a reference level RL)
- Assessment of local or regional radon hazard
- Assessment of the collective risk per area
 - measured as exposure or dose
 - measured as detriment attributable to radon
- establishing a decision base for prevention or mitigation action, e.g. through delineation of radon priority areas

Mapping support

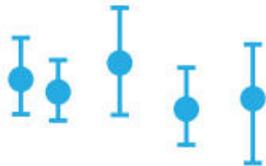
- Support: the geographical area to which a value of the target quantity shall be assigned
- Examples:
 - Mean of the target quantity or its exceedance probability per municipality, administrative region, grid cell, geological unit,...
 - Compound of adjacent areas classified as radon priority areas (RPAs)
- Often difficult: Re-mapping between different supports

Accuracy // Precision

Data are:

1. Numbers or items or instances which should be accurate and precise individually. This concerns metrological quality assurance.
2. A sample from a population which should represent the population accurately and precisely, i.e. be representative for the population. This concerns quality assurance of the survey design and its implementation.

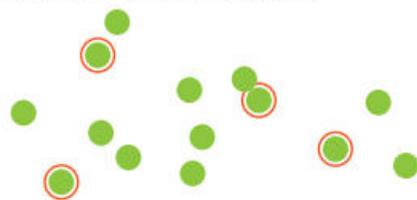
Individual data:



Issue: uncertainty

**high uncertainty
⇒ low precision**

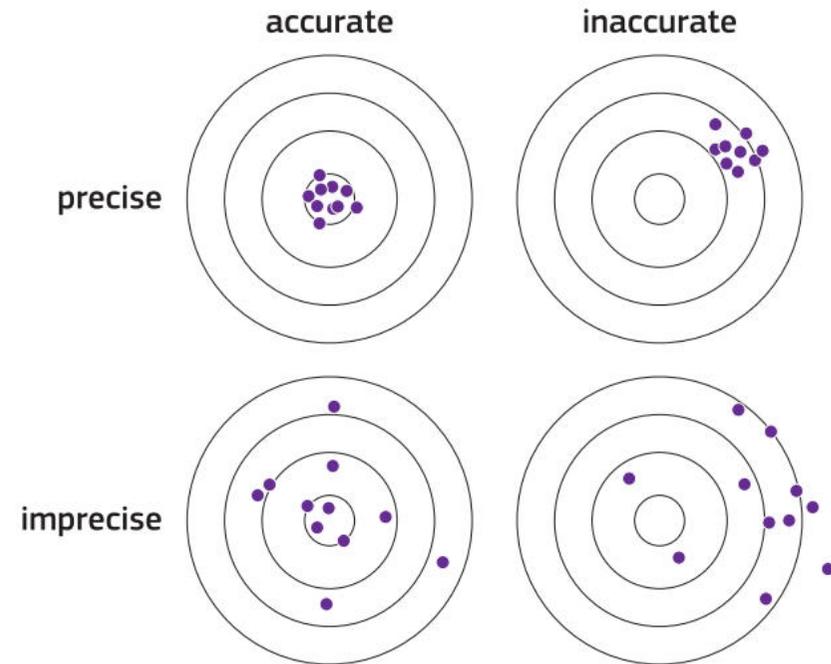
Sample from a population:



Issue: representativeness

● population
○ samples

missing representativeness ⇒ bias (inaccuracy)



Sample size:

mainly a matter of costs

Sample representativeness:

this is the most complicated part!!

LN distribution

- $X \sim \text{LN}(\mu, \sigma^2) \Leftrightarrow \ln X \sim \text{N}(\mu, \sigma^2)$

- $\exp(\mu)$ = geometrical mean GM, $\exp(\sigma)$ =geometrical standard deviation GSD

- $$pdf(x) = \frac{1}{\sqrt{2\pi} \sigma x} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right)$$

- Mean=AM= $\exp(\mu + \sigma^2/2)$ =GM $\exp(\sigma^2/2)$ =GM $\sqrt{1+CV^2}$,
CV= $\sqrt{\exp(\sigma^2)-1}$; SD=AM.CV

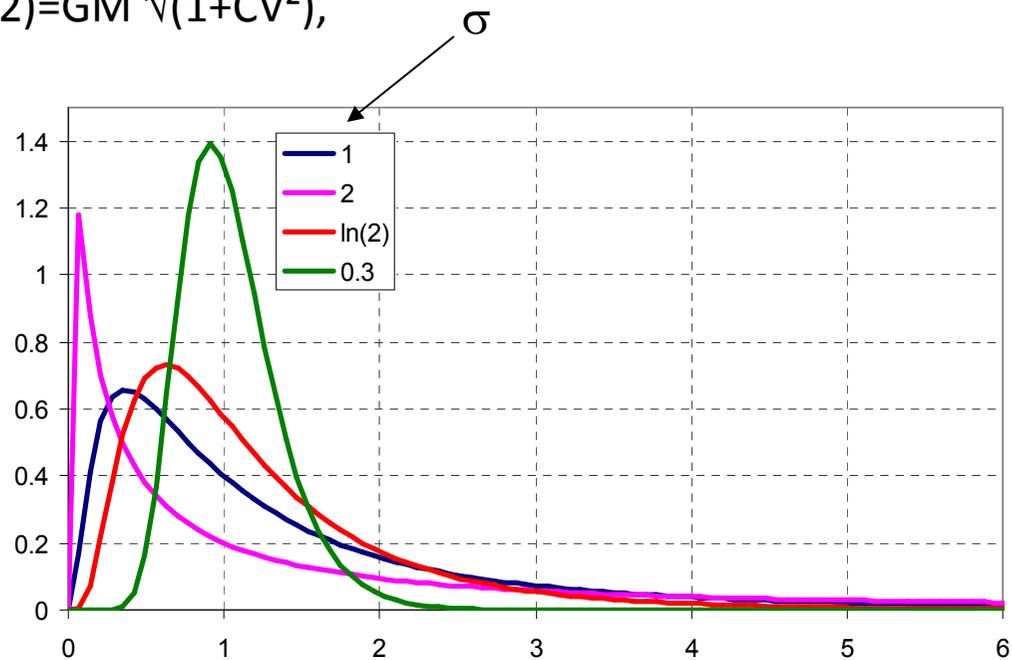
Mode= $\exp(\mu - \sigma^2)$

skew= $CV^3 + 3 CV$

kurt= $CV^8 + 6 CV^6 + 25 CV^4 + 16 CV^2$

- For $X \sim \text{LN}(\mu, \sigma^2)$:
 $aX^b \sim \text{LN}(\ln(a) + b\mu, b^2\sigma^2)$

- For $X_i \sim \text{LN}(\mu_i, \sigma_i^2)$:
 $a \prod X_i^{b_i} \sim \text{LN}(\ln(a) + \sum b_i \mu_i, \sum b_i^2 \sigma_i^2)$



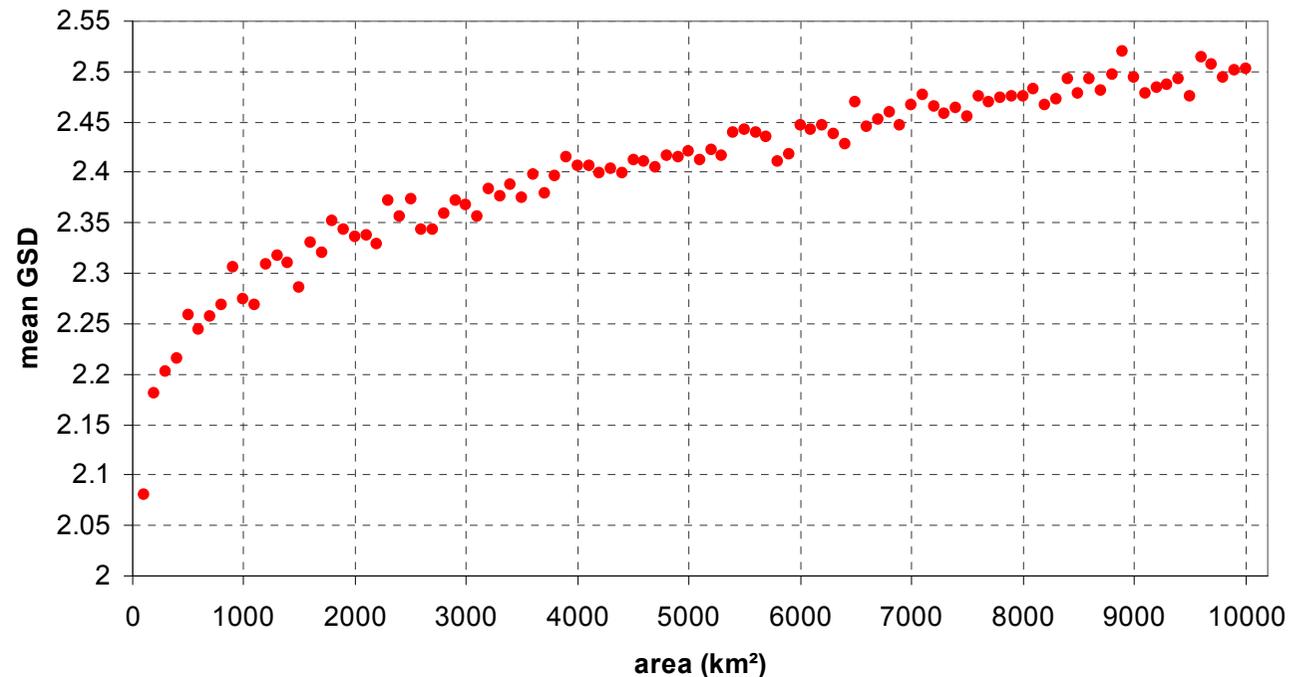
GSD

Teóricamente: dentro da unidade U, $GSD \approx \exp\left(\sqrt{\gamma(\langle s \rangle)}\right)$

γ = variogram of $\ln(\text{data})$, $\langle s \rangle$ = mean distance between two random points within U; depends on the area and on the shape of U. U=circle: $\langle s \rangle \approx 0.511 \sqrt{\text{area}}$; U=quadrat: $\langle s \rangle \approx 0.521 \sqrt{\text{area}}$.

Empírica dependence of the GSD of indoor Rn concentration on aggregation area, Europe. (Data from the indoor radon map, European Atlas of Natural Radiation)

(Bossew 2021, GeoENV Parma, in <https://hdl.handle.net/1889/4373>, <https://www.repository.unipr.it/handle/1889/4373>)



Precision \Rightarrow sample size

1. sample size for estimating the GM, AM
2. Sample size for estimating the exceedance probability
 - a) Empirical proportion
 - b) Modelled proportion
3. Sample size for estimating the RPA status

1) sample size for estimating the GM, AM for LN process

Hale formula (Hale 1972):

Condition: |observed μ - true μ | < TOL

prec := (obs GM - true GM)/true GM \rightarrow TOL = $\ln(\text{prec} + 1)$

$$n_{\min} = \left(\frac{z_{1-\alpha/2} \ln(GSD)}{\ln(PREC + 1)} \right)^2$$

Approximative confidence limits (α) of the AM (Olsson 2005):

$$CL_{high,low} = AM \cdot \exp \left[\pm t_{1-\alpha/2, n-1} \sigma \sqrt{\frac{1}{n} + \frac{\sigma^2}{2(n-1)}} \right] \quad \sigma = \ln(GSD)$$

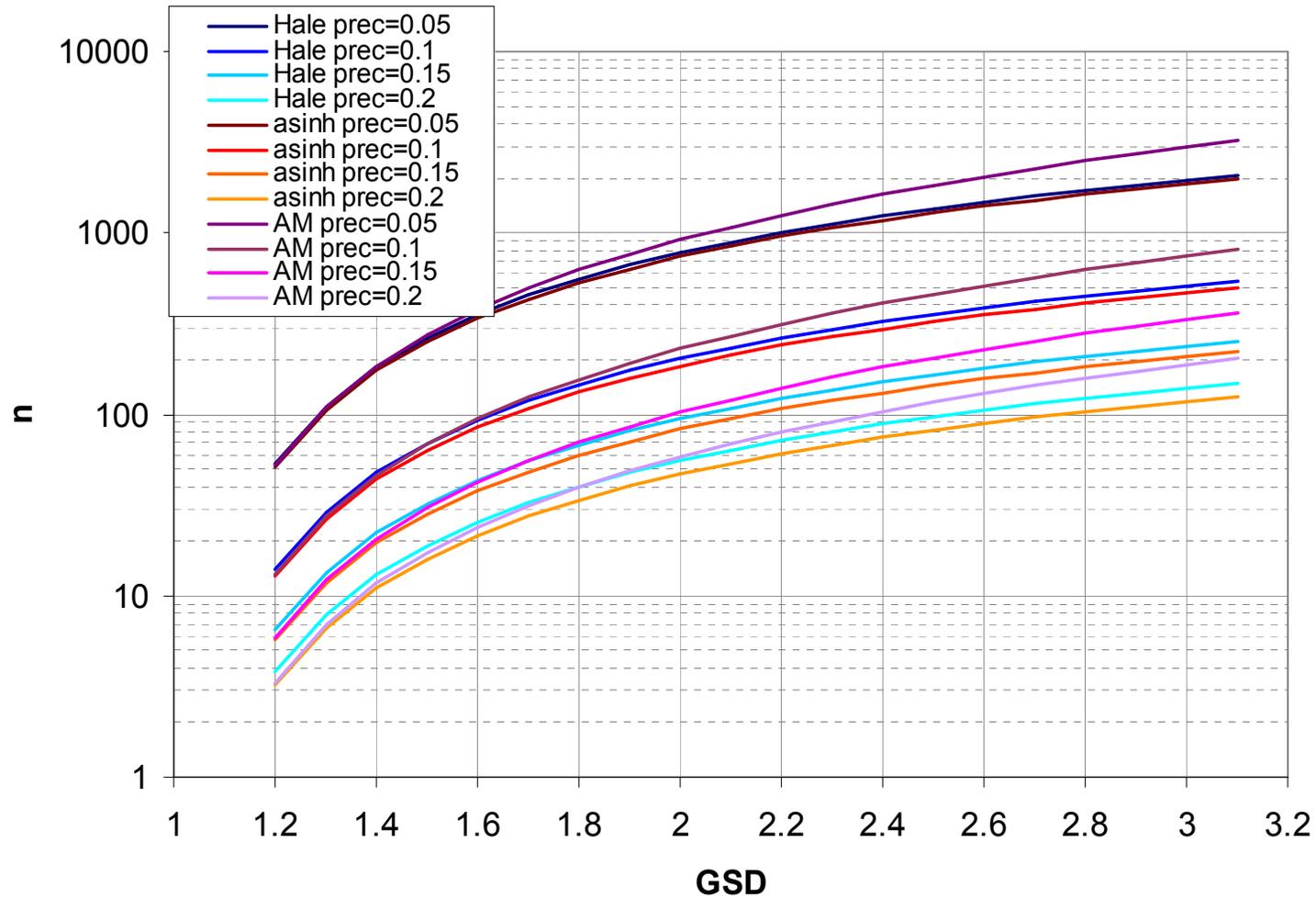
with prec := $(CL_{high} - CL_{low}) / (2 AM) \rightarrow$

inversion, approximating $t_{1-\alpha/2, n}$ by normal $z_{1-\alpha/2}$ for sufficiently large n (>ca.20):

$$n_{\min} = \frac{1}{4A} \left[2A + 2 + \sigma^2 + \sqrt{(2A + 2 + \sigma^2)^2 - 16A} \right], \quad A := \left(\frac{a \sinh(prec)}{z_{1-\alpha/2} \sigma} \right)^2$$

Hale 1972, [https://doi.org/10.1016/0004-6981\(72\)90138-2](https://doi.org/10.1016/0004-6981(72)90138-2)

Olsson 2005: <https://jse.amstat.org/v13n1/olsson.html>



sample size necessary to achieve a given precision (0,05; 0.1; 0.15; 0.2 chosen) at $\alpha=0.05$, for 3 models: Hale formula; confidence interval of the GM (formula not shown); confidence interval of the AM.

2) Sample size for estimating the exceedance probability

Target variable = $\text{prob}(Z > RL)$

Options:

1. Confidence intervals of the empirical proportion: requires a sufficient sample size.
2. Model: assume LN, estimate the distribution tail and its confidence limits. However, inversion to recover the minimal sample size may be feasible only numerically.

a) empirical proportion

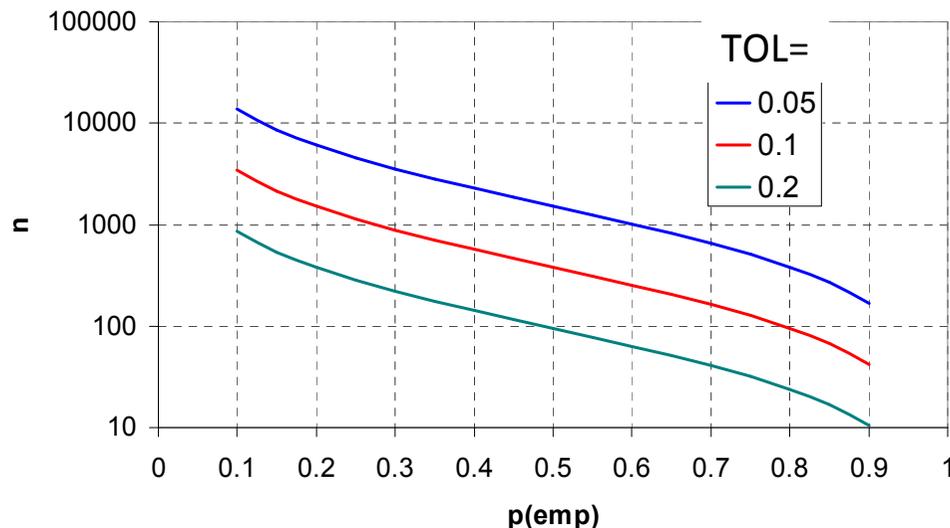
- $p(\text{emp}) := \#(z > \text{RL})/n$
- **simple version:** “Wald interval” (Laplace 1812); asymptotic with $n \rightarrow \infty$!!
- assumption: binomial process
- several problems (overshoot etc.), therefore not recommended for exact purposes, e.g. Brown et al. 2001, [doi:10.1214/ss/1009213286](https://doi.org/10.1214/ss/1009213286)
- **inversion:** set p^\wedge as criterion in place of $p(\text{emp})$, invert formula

$$C.I. \approx \hat{p} \pm x_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$$

$$\text{Def. } \text{PREC} = \left(x_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right) / \hat{p} < TOL$$

$$\Rightarrow n > \approx \frac{\hat{p}(1-\hat{p})}{(TOL \hat{p} / x_{1-\alpha/2})^2}$$

$\alpha=0.05 \rightarrow x_{\alpha/2}=1.96:$



Differently from the case of $n_{\min}(\text{GM}, \text{AM})$, $n_{\min}(p)$ depends on the actual size p . For low $p^\wedge \rightarrow$ high n necessary for achieving tolerable precision!

Ex.: $p^\wedge = 20\%$ of houses $>$ RL, as criterion e.g. for RPA $\rightarrow n=380$ necessary for determining whether $p > p^\wedge$ with precision 20%.

Ex. inverse: sample size $n=20$, 4 values $>$ RL $\rightarrow p(\text{emp})=4/20=0.2$;
C.I. $\approx 0.2 \pm 0.17 \rightarrow \text{PREC}=87\%!!$

- **Wilson score** interval:

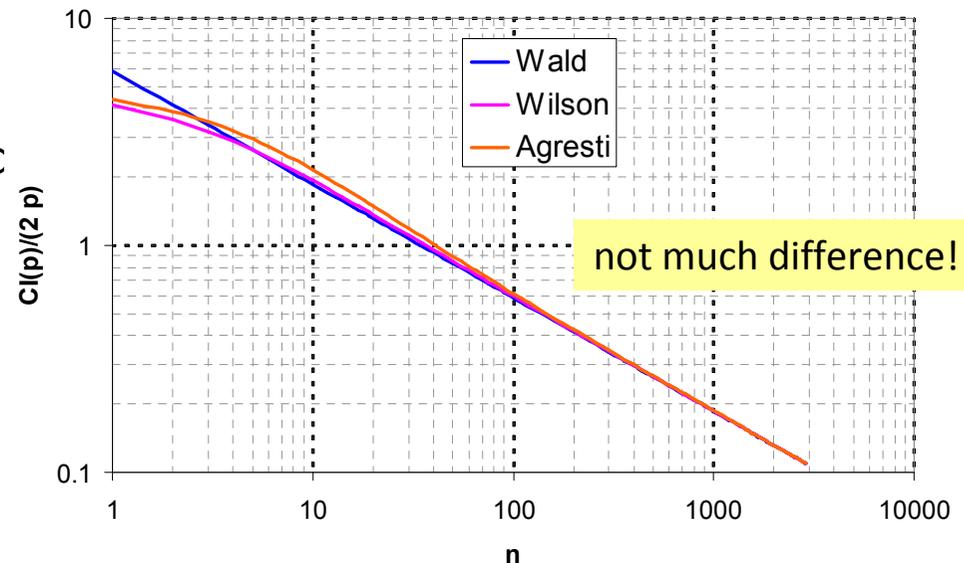
$$CL_{L,U} = \frac{1}{1 + x^2 / n} \left(\hat{p} + \frac{x^2}{2n} \mp \frac{x}{n} \sqrt{n\hat{p}(1 - \hat{p}) + x^2 / 4} \right)$$

($x \equiv x_{1-\alpha/2}$; correction possible to avoid overshoot)

- **Agresti** interval:
like Wald, but substitute:
 $n \rightarrow n' = n + x^2$, $p \rightarrow p' = (np + x^2/2) / n'$

$x=1.96$, $p(\text{emp})=0.1$

- **“Rule of three”**
If no “sucess” has been observec
i.e. $np=0$,
interval ($\alpha=0.05$) $\approx (0 , 3/n)$



b) Modelled proportion

- Assumption: LN process
- Tail area = $p = \text{prob}(Z > RL) = 1 - \Phi(y)$, $y = (\ln(RL) - \mu) / \sigma$
- How to estimate this quantity, if μ and σ are unknown and have to be estimated from data?
- **“naive”**: in Φ , use empirical μ , σ = AML, SDL, but leads to bias because no linear function.
- **Exact**: $p = \text{Beta}(u_b, \beta)$, Beta – the cumulative symmetric beta distribution with $\beta = (n-2)/2$ and $u_b = \max[0, 0.5 \cdot (1 - (\ln RL - \text{AML}) \sqrt{n} / (\text{SDL} (n-1)))]$, AML, SDL = (μ, σ) empirical.

$$\text{Cum}B(x, a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt, \quad B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

symmetric: $a=b=\beta$

Lieberman GJ, Resnikoff GJ (1955): Sampling plans for inspection by variables. J Am Stat Assoc 50(270):457. <https://doi.org/10.2307/2280972>

- Andersen et al. (2001), Murphy & Organo (2008)

DOI: 10.1016/s0048-9697(01)00697-0, DOI: 10.1088/0952-4746/28/3/001

$p=1-\Phi(u)+B$, where:

- $u=(\ln(RL-Rn_{out})-AML)/SDL$,
 Rn_{out} =mean outdoor Rn conc.
- B =bias term = $(-u/(2n\sqrt{2\pi})) \exp(-u^2/2)$

- Beard (1960) suggested simpler

$p = t_{n-1}[u \sqrt{n/(n+1)}]$, where:

$u = (\ln RL-AML)/SDL$

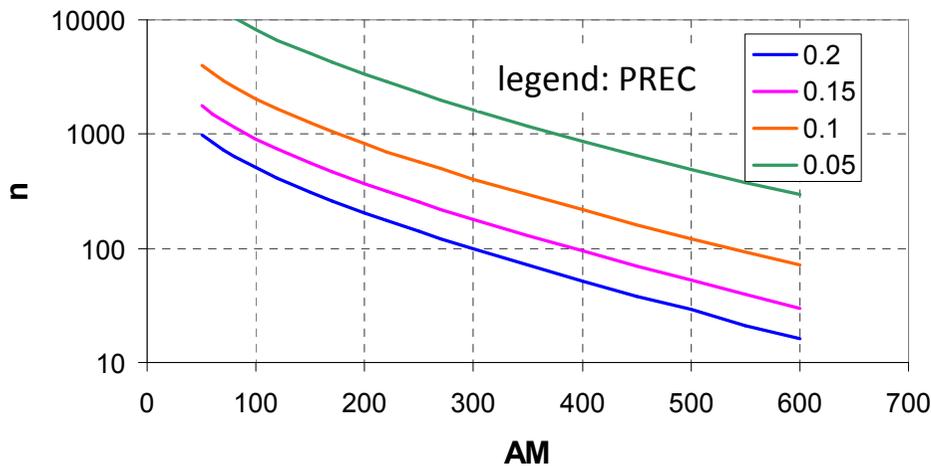
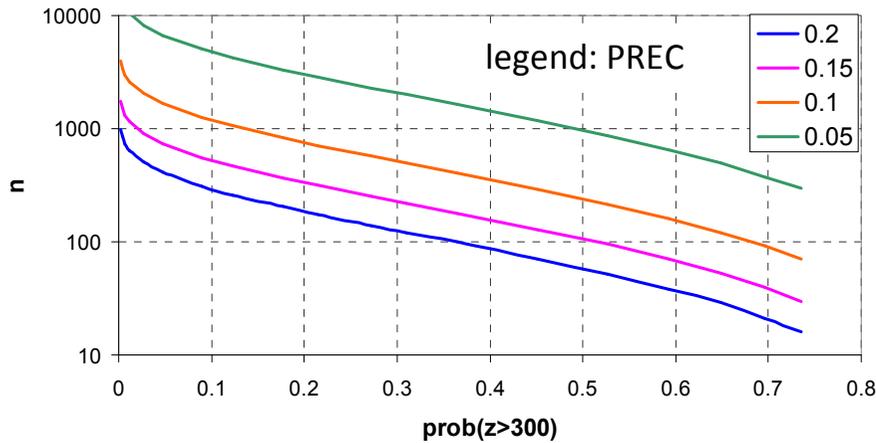
Beard (1960), <https://doi.org/10.1029/JZ065i007p02143> ;

Proschan (1953), <http://www.jstor.org/stable/2281009> ;

Wilks (1941), <http://www.jstor.org/stable/2235627>

Still to do: comparison between the methods! Bias (accuracy), precision, coverage probability

n = sample size such that the relative uncertainty of p = prob(z>300) is below PREC.



n as function of p or AM, for given PREC: calculated numerically (Excel solver), no explicit inversion formula available

GSD=2 fixed (about realistic)

$\alpha=0.05$

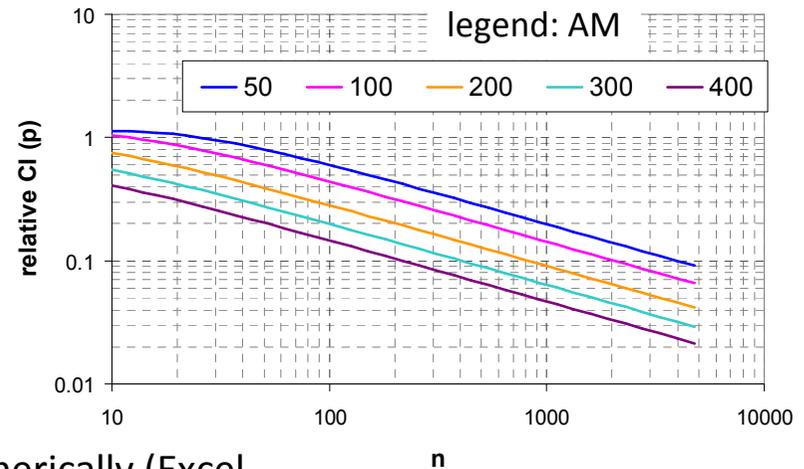
RL=300 Bq/m³

Based on Beard formula

$$\text{Prec} = \delta p / (2p) = (p_+ - p_-) / (2p)$$

$$p_{\pm} = t_{n-1} \left(u_{\pm} \sqrt{\frac{n}{n+1}} \right), u_{\pm} = \frac{\ln(RL) - AML_{\pm}}{SDL}$$

$$AML_{\pm} = AML \pm z_{1-\alpha/2} SDL / \sqrt{n}$$



3) Sample size for estimating the RPA status

- RPA status of an area $\in \{\text{yes, no}\} \rightarrow$ classification task. Fuzzy classification: assign a probability for an area to be RPA.
- Precision in this case: misclassification probability.
- Power of a test = probability that an effect is discovered if in reality there is one. Effect in this case: area = RPA.

Classification precision of RPAs

Example: RPA definition: area in which $\text{prob}(Z > RL) > p_0$.

$RL=300$ (but does not matter in the following), $p_0=0.1$.

Based on Wald confidence interval of proportion.

p^\wedge = empirical or observed probability.

Questions:

1. Given observed p^\wedge , based on sample size n : how big is the chance that in fact the true $p > p_0$?
2. how big must the sample size n be, that the classification error is below a tolerance $\alpha/2$ (TOL), i.e., that $p^\wedge < p_0$ while true $p > p_0$ (second kind error)?
3. given effect $\text{eff} := p^\wedge - p_0$ and an actual sample size n , how big is the second kind error $\alpha/2$, i.e. the power curve or operational characteristic?

All by reordering or inverting the same formula!

Wald CL: $CL_{\pm}(\hat{p}) = \pm z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$

Question 1:

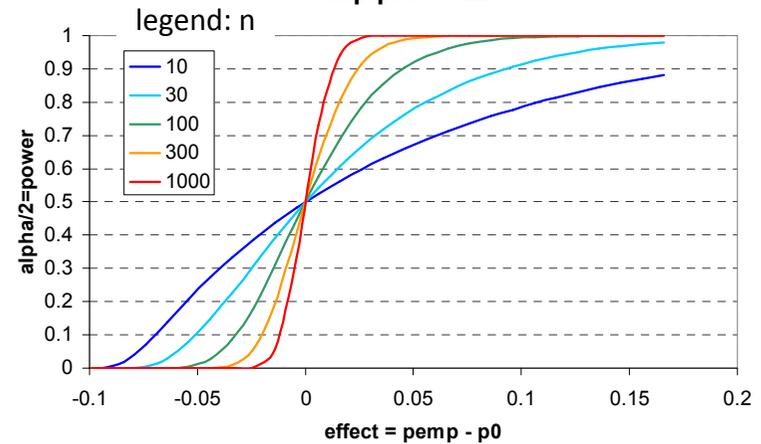
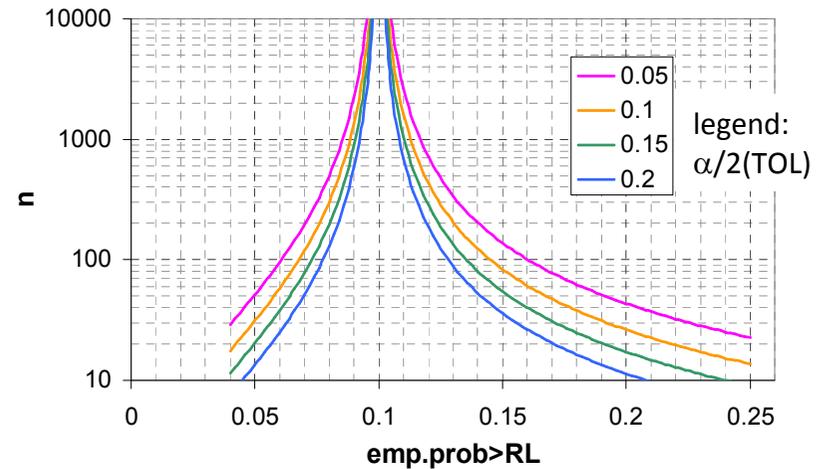
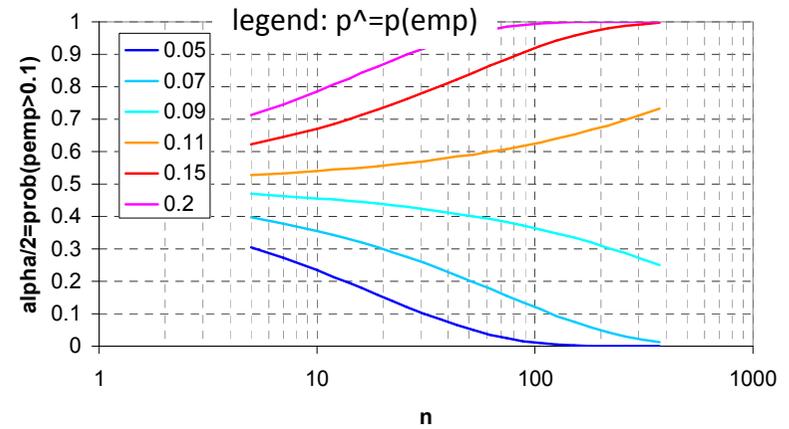
$$\alpha/2 = 1 - \Phi\left(\frac{p - \hat{p}}{\sqrt{\hat{p}(1-\hat{p})/n}}\right)$$

Question 2:

$$n = \frac{\hat{p}(1-\hat{p})}{\left(\frac{p - \hat{p}}{z_{1-\alpha/2(TOL)}}\right)^2}$$

Question 3:

$$\alpha/2 = 1 - \Phi\left(\frac{-\text{eff}}{\sqrt{(p + \text{eff})(1-p-\text{eff})/n}}\right)$$



Harmonization bottom up // top down

If a survey project is performed by different working groups \Rightarrow all should use the same methodology = **bottom-up** harmonization.

Often: data from different data bases (different projects, different methodology) \Rightarrow **top down** harmonization

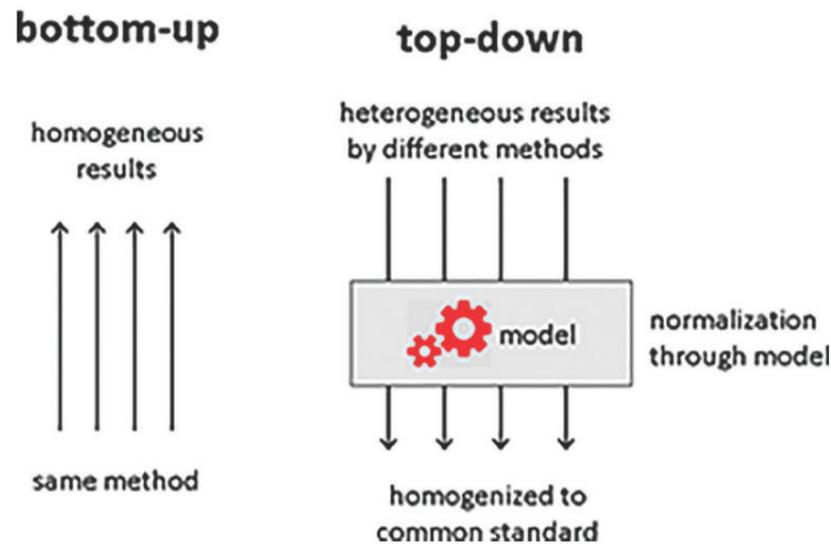


Fig. 1. Concepts of bottom-up and top-down harmonization.

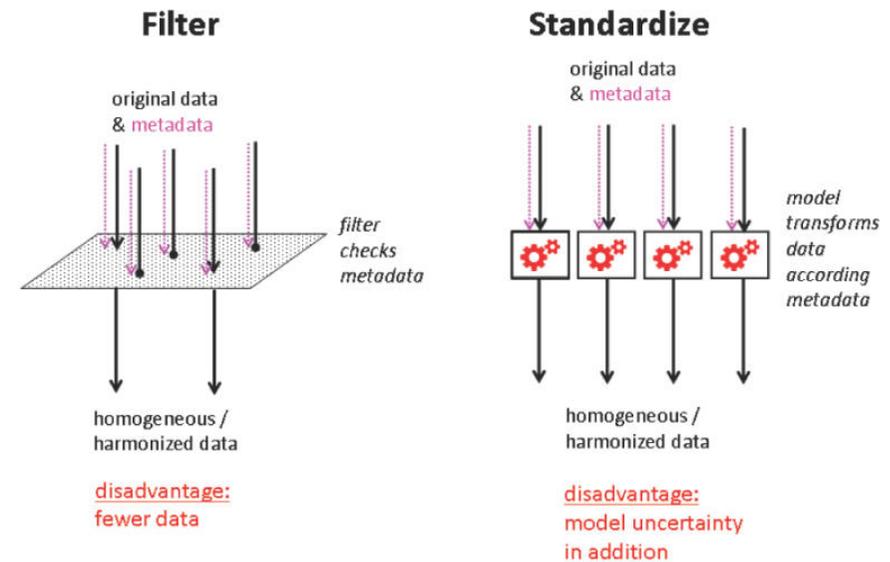


Fig. 2. Top-down harmonization by filtering and modeling.

Metadata

- Factors which influence the measured variable or inform about the conditions and circumstances of the measurement; for indoor Rn:
 - Type and characteristics of a building, floor level, isolation, etc.
 - geology, U content in the ground, soil type, climate, etc.
- **Necessary for:**
 - **Interpretation** of results;
 - Input for **modelling**: metadata are model parameters;
 - **Verification** (comparing the metadata acquired in the survey with analogue data from independent databases)

The quality assurance chain

1. **Design QA:** Capability of an effort (experiment, survey) to enable a decision: sampling design;
2. **Data QA:**
 - a) classical metrology
 - b) protocols & actual performing an experiment;
3. **Evaluation QA:** proper methodology, adequate models, uncertainty budgeting;
4. **Decision QA:** Assessment of the chance of a wrong decision, given data and methods and their uncertainties.

Participants of the workshop

Christie Helouise Engelmann de Oliveira, UNISINOS

Fabiano Cabañas Navarro, UNIFAL-MG

Gabriel Bias Fortes, IBGE

Heber Luiz Caponi Alberti, CNEN

Jacson Jung, UNISINOS

Janine Nicolosi Correa, UTFPR

Laura Cardoso Takahashi, CNEN

Marcela Rabelo de Lima, SGB/CPRM

Marco Antonio da Silva, CNEN

Nivaldo Carlos da Silva, CNEN

Oderson Antonio de Souza Filho, SGB/CPRM

Patricia Pereyra, PUCPERU

Peter Bossew

Peter Marshall Fleming, CNEN

Ricardo Alberto Giannoni, CNEN

Rildo Borges Duarte, IF SUL DE MINAS

Sergei Pascuk, UTFPR

Talita de Oliveira Santos, UFMG

Thomaz A. de Oliveira, IF SUL DE MINAS

Tiago Antonioni Arauji de Olioveira, CNEN

Ubirani Otero, INCA

Vanderlei Vilaca De Moura, CBMM

Vera Paniz, UNISINOS

Wanilson Luiz Silva, UNICAMP

Yasmin Duarte Siqueira, CNEN

Yuniel Tejeda Mazola, GLP

Supplement 2

Background material of the IAEA workshop, Poços de
Caldas, Brazil, November 2023
(partly in Portuguese)

Experimental design of surveys aimed to assess radon exposure of the population at different geographical scales

-- background material --

Bossew P.

retired from BfS (German Federal Office of Radiation Protection),
Vienna, Austria

v. 12.11.23

**EX-BRA9061-2303817 – IAEA Expert Mission,
13 – 17 November 2023, Poços de Caldas, MG, Brasil**

Estrutura *temática* tentativa

A. Parte geral

1. Definição do problema; rationale / justificativa da pesquisa; hazard vs. risk; o que é “decisão”?
2. Documentos existentes IAEA, WHO,... : sumário, discussão; até que ponto são uteis / relevantes para esse projeto
3. Pesquisas análogas em outros países ou regiões (esp. Europa, porque temos muita experiência embora as condições sejam diferentes -> identificação das diferenças!)

B. Assuntos técnicos

1. Metrologia: laboratórios e capacidades deles; questões QA/QC; noção de "observação" => "cadeia QA"
2. "front end": desenho de survey (locação dos detetores, sample size em dep. do objetivo, etc); hazard vs. risk
3. "tail end": avaliação / análise, em particular mapeamento; incerteza e sua consequencia para decisões

C. Assuntos políticos

1. fatores sociais: quais? constraints / limitações económicas, logísticas, sociológicas
2. Radon action plan (RAP) BR; implementação dos resultados do survey; inclusão dos stake holders; possível papel da Citizen Science

D. Parte específica – Poços de Caldas

1. Situação Poços de Caldas: dados, resultados existentes, conclusões, ações?
2. Micro-survey: Investigação de um hot spot, várias quantidades; avliação preliminar
3. Upscaling Poços de Caldas -> Brasil: bases de dados BR; no ambito do RAP

E. Conclusões

1. Sumário do workshop
2. Consequências possíveis

Estrutura *real* do workshop

- A decidir pelos participantes em acordo das necessidades, interesses, ...
- **Proposta: Agenda elaborada por Nivaldo**
- Temos 4.5 – 5 dias: Sequencia segundo importância dos assuntos
- No entanto, na minha opinião tem certos pré-requisitos que devem ser abordadas.

A seguir, preparei algum material... opcional!

A.1.

Básicos e pré-requisitos

Soluções universais / regionais?

- Não (ou raramente) tem metodologia universal válida para todas situações & condições!
- Tem que
 - identificar as condições;
 - desenhar estratégia adaptada / apropriada;
 - pode ser diferente entre regiões!
- Parece trivial, mas na realidade \Rightarrow trabalho inicial!
- Senão: projeto não vai ser aceitado, pode fracassar.
- Estratégia não pode ser estipulada de cima para baixo \Rightarrow inclusão de pessoas & instituições regionalmente reconhecidas.

Definição do problema

- Ideia geral: **Avaliação da situação Rn no Brasil**
 - Em forma de mapa? Tabela?
 - Recomendações? Consequências?
- A definir:
- Qual é a **base geográfica**?
 - Município, census tract,... ?
- Qual é a **“target variable”**?
 - Valor médio por unidade geográfica? $\rightarrow (0, \infty)$
 - Probabilidade de excesso do nível referencial (reference level, RL)? $\rightarrow (0, 1)$
 - Status como Área de Prioridade (Radon Priority Area, RAP)?
 \rightarrow (sim, não) = quantidade binária

Como decidir?

Depende de:

- Legislação, National Radon Action Plan (RAP)
- Possibilidades logísticas, limitações / constraints

Consequências

Cada objetivo tem metodología particular dele!

@ desenho da pesquisa

@ método de medição, incl. logística necessária!

@ disponibilidade / necessidade de bases de dados adicionais / auxiliares

RAP ↔ survey ?

Documento muito importante: A. Canoba, Planning a radon survey; Workshop on Radon Analysis methods and labs for environmental and occupational purposes IAEA – CNEN, Río de Janeiro, Brasil, 2-6.10.2023

Mas acho que a ordem deveria ser parcialmente invertida:

*Una vez que se cuente con la **información confiable** basado en un **relevamiento representativo** sobre el **gas radon** se podrá establecer un Plan de Acción de Radón (documento gubernamental en donde se definen los objetivos, las estrategias, lineamientos, acciones a seguir)*

(Canoba, slide 2)

Porque o desenho do survey depende do RAP (objetivo da estratégia Rn), não vice-versa.

No outro lado é correto que as ações são consequência dos resultados do survey.

RAP tem 2 aspectos:

1. RAP: definição do objetivo → survey, definição de RL (reference level) e da RPA (Rn prone / priority area);
2. resultado do survey → ações a definir no RAP

Terminologia

(não tão trivial que parece)

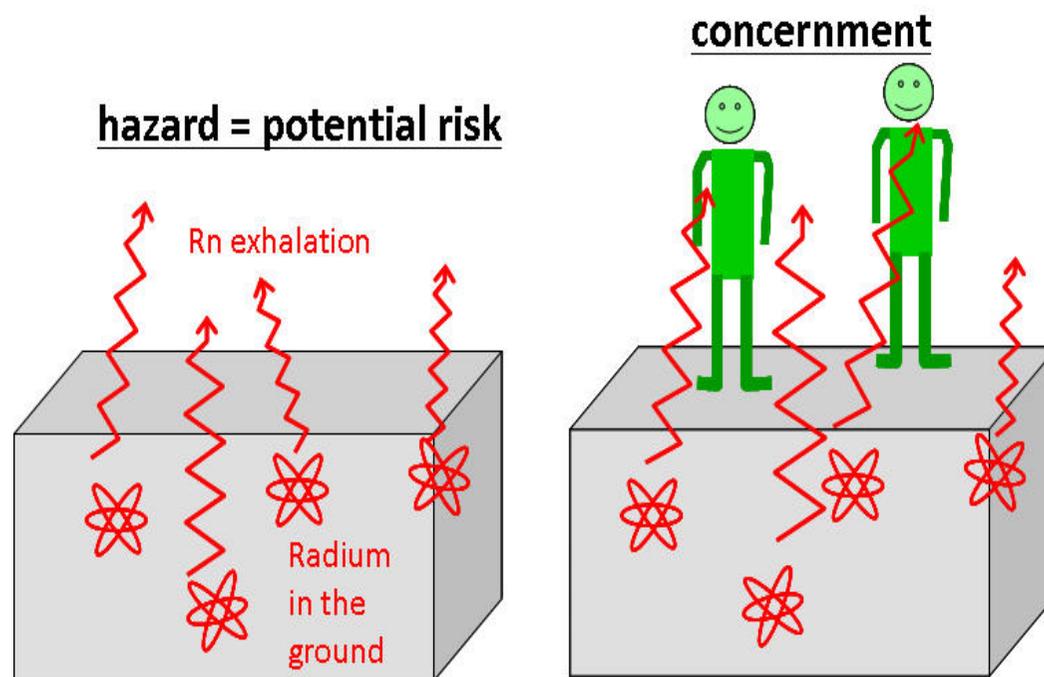
- **Survey:**
Pesquisa com objetivo de avaliação de uma situação.
Pesquisas ambientais, políticas, sociológicas,...
- **Sample / amostra:**
 - Fisicamente: um objeto físico, como amostra do solo, certa medição de R_n indoor, ... → valor z , incerteza $u(z)$, metadata como localização geográfica, tipo e circunstâncias de medição, ...
 - Estatisticamente: um conjunto de valores, cada um representante de objeto físico, p.ex. o conjunto de medições realizado dentro de um município → $\{z_1, z_2, \dots, z_n\}$, n = tamanho da amostra (sample size)
 - Muitas vezes confusão entre essas noções!
- **Nível referencial (RL), Área de prioridade:**
→ BSS (IAEA, U.E.)... a discutir depois!

Hazard vs. risk

- **Hazard** exists also if nobody is affected, exposed or concerned;
- It becomes a **risk**, (= a certain probability or size of damage or detriment), if there is somebody who can be harmed. If there is nobody, evidently there is no risk, even if hazard, i.e. a physical cause exists.

(Or in general, any being or thing whose damage or detriment should be avoided.)

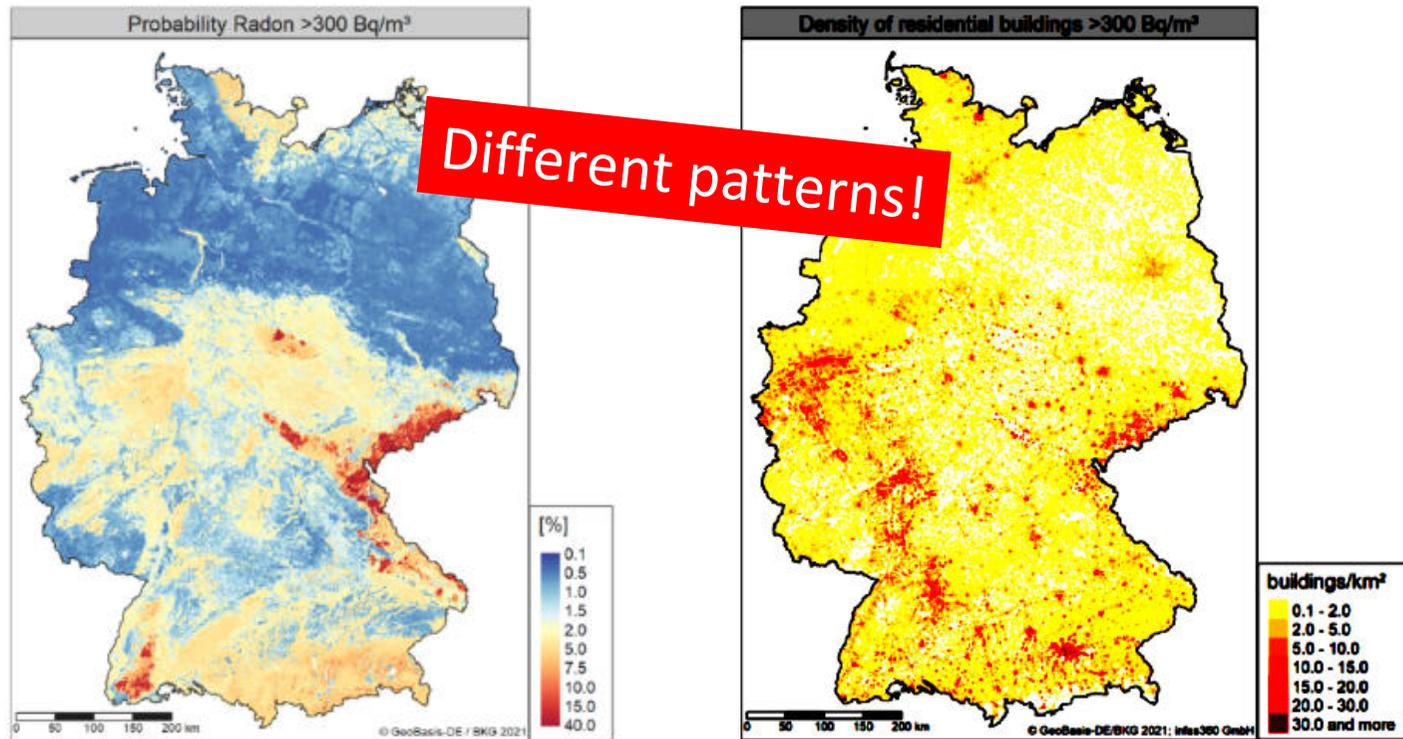
$$\underline{\text{risk}} = \text{hazard} \times \text{concernment}$$



quantity	meaning	example from Rn
risk =	probability or size of detriment	prob. or number of lung cancer fatalities
hazard ×	physical cause	geogenic radon potential GRP, derived quantities
vulnerability ×	exposure circumstances exposure specific factors	e.g. floor level, building properties e.g. equilibrium factor Rn progeny per Rn conc.
exposure exposure factor risk factor	number of persons exposed exposure resulting from hazard under certain vulnerability factors person specific vulnerability detriment per exposure	(Bq*h/m ³ ²²² Rn) per unit GRP e.g. age, breathing rate, smoking habits (prob. lung cancer fatality or ERR) / (Bq*h/m ³ Rn progeny) or ERR=f(dose)

concernment

Ex: Alemanha



prob($IRC > 300 \text{ Bq/m}^3$ in dwellings,
ground floor)
= **hazard map**

Residential buildings with $IRC > RL = 300$
 Bq/m^3 ; [buildings/ km^2]
= **risk map**

It has also been shown (among other, Petermann et al. 2021, 2022) that confining Rn abatement measures to hazard-RPAs has very little effect on target = reducing detriment!

O que é “decisão” ?

- Exemplos na área do Rn:
 - Uma medição está acima ou abaixo do valor referencial?
 - Tem que remediar um prédio?
 - Uma área (município etc.) é Rn Priority Area?
- Quantidade que é base da decisão (medição, outra informação quantitativa ou qualitativa; valor numérico ou categorial) está com incerteza!
⇒ também a decisão!
- Decisão é avaliação binária de uma quantidade, o que representa um efeito.
 - First kind error: Decisão: sim, efeito está presente, enquanto na verdade (que não se conhece) o efeito não está presente
 - Second kind error: Decisão: não, não tem efeito, enquanto na verdade tem.
 - Na maioria dos casos não independentes, por isso não podem ser minimizados independentemente. Para minimizar juntos → otimização do método (desenho diferente, mais observações, instrumento mais sensível,...)
 - No mundo ideal, se deveriam estimar os dois erros; na realidade não é fácil!
 - Specificity = taxa de estimativas corretamente negativas = $1 - \text{taxa de erro tipo 1}$
.... capacidade de evitar decisão positiva (=tem efeito) errada.
 - Sensitivity = taxa de estimativas corretamente positivas = $1 - \text{taxa de erro tipo 2}$
.... capacidade de evitar decisão negativa (= não tem efeito) errada.
 - https://en.wikipedia.org/wiki/Sensitivity_and_specificity
 - Decision theory = área da matemática, tem a ver com fuzzy logic

Decision making

DECIDE model (Guo 2008; cited after Wikipedia*)

- **D**efine the problem
- **E**stablish or **E**numerate all the criteria (constraints)
- **C**onsider or **C**ollect all the alternatives
- **I**dentify the best alternative
- **D**evelop and implement a plan of action
- **E**valuate and monitor the solution and examine feedback when necessary

* Guo, K. L. (2008). DECIDE. *The Health Care Manager*, 27(2), 118–127.
doi:10.1097/01.hcm.0000285046.27290.90 ; <https://en.wikipedia.org/wiki/Decision-making>

A.2.

Documentos relevantes

Documentos relevantes

- Documentos + relatórios internacionais
 - BSS-IAEA
 - BSS-U.E.
 - WHO radon handbook
 - European Atlas of Natural Radiation
 - ...
- Documentos + relatórios nacionais

IAERA Basic Safety Standards, BSS

→ Apresentações da Analia Canoba no workshop em Outubro!

IAEA Safety Standards for protecting people and the environment

Protection of the Public
against Exposure Indoors
due to Radon and Other
Natural Sources of
Radiation

Jointly sponsored by the IAEA, WHO



Specific Safety Guide
No. SSG-32



IAEA Safety Standards for protecting people and the environment

Radiation Protection and
Safety of Radiation Sources:
International Basic
Safety Standards

Jointly sponsored by
EC, FAO, IAEA, ILO, OECD/NEA, PAHO, UNEP, WHO

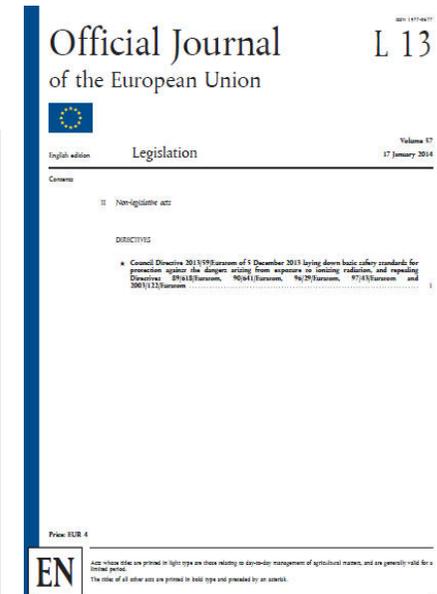


General Safety Requirements Part 3
No. GSR Part 3



O documento chave da legislação Europeia

- EURATOM Basic Safety Standards (normas de segurança de base)
- Diretiva da União Europeia: Todos Estados-Membros devem transferir e implementar a diretiva na legislação nacional.
(Senão: multa bem alta)
- Parecido a IAEA-BSS



European Council Directive 2013/59/Euratom laying down basic safety standards, etc. Official Journal of the European Union 2014; 57(L13), 1 – 73. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2014:013:FULL&from=EN>

Português:

<https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=OJ:L:2014:013:TOC>

Essencias dos EU-BSS

- **Plano de ação nacional de radônio** deve ser estabelecido (Art. 103/1)
- **Níveis de referência (RL) max. 300 Bq/m³** média longo prazo; válido para habitações, locais de trabalho e edifícios abertos ao público (Art. 74/1, 54/1)
- Princípio de prioridade
- Edifícios novos: prevenção (Art. 103/2, An. XVIII (8))
- Identificar **Radon Priority Areas (RPA)** (Art. 103/3): Nos RPAs, medição obrigatória nos locais de trabalho (Art. 54/2a); Habitações : prioridade em estratégia de redução (An. XVIII (6))
- Locais de trabalho >RL: remediação. Caso redução embaixo de RL não possível: calcular dose (Art. 54/3).
- Habitações : “promover” identificação; se >RL: “fomentar” redução (Art.74/2); estratégia para reduzir Rn (An. XVIII (6))
- Informar o público sobre Rn (Art. 74/3)

Reference and Action Level

- It must be emphasized that RL (as defined in the BSS, IAEA and EU alike) are no AL!
- AL: if exceeded, action must be taken, if not exceeded, no action required.
- RL: exceedance is “inappropriate” (BSS Art.4 (84)), but also if $IRC < RL$, minimization should be attempted (BSS Art. 7/1: *“Optimisation of protection shall give priority to exposures above the reference level and shall continue to be implemented below the reference level”*).
- However, it seems that in regulatory practice (as laid down in legislation), RL is often practically treated as AL; RL is a juristically complicated thing.
(Ex.: In German legislation, for workplaces, the RL is understood as AL.)

The famous BSS Article 103/3

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.

Annex XVIII

List of items to be considered in preparing the national action plan to address long-term risks from radon exposures as referred to in Articles 54, 74 and 103:

(2) Approach, data and criteria used for the **delineation of areas** or for the definition of other parameters that can be used as specific indicators of situations with potentially high exposure to radon.

(6) Strategy for reducing radon exposure in dwellings and for giving **priority** to addressing the situations identified under point 2.



“significant number”

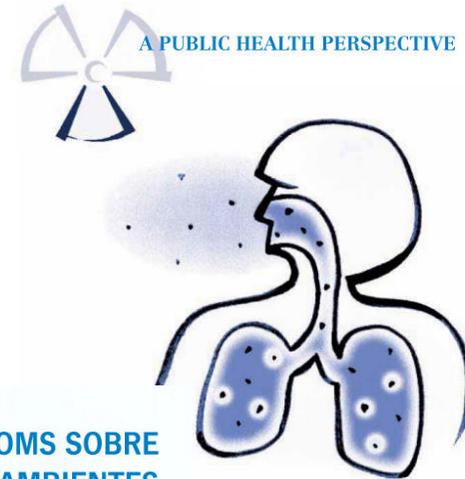
- From the beginning (2013) there were discussions about the meaning of this apparently cryptic formulation.
- Perhaps it was put like this on purpose to allow flexible interpretation?
- Mostly it was interpreted as
 - “significant fraction” of buildings in an area $> RL$;
 - “mean over buildings” in an area $> RL$
- But this ignores the number of persons affected, which is relevant for the detriment to society (number of lung cancer fatalities caused by R_n); after all, the objective is to reduce the detriment.... discussion ongoing.

WHO Radon Handbook

1. EFEITOS DO RADÔNIO SOBRE A SAÚDE.....	27
1.1 O Risco de câncer de pulmão em mineiros expostos ao radônio.....	28
1.2 Risco de câncer de pulmão para a população geral devido ao radônio em ambientes internos.....	31
1.3 O radônio e outras doenças além do câncer de pulmão.....	38
1.4 O peso do câncer de pulmão ocasionado pelo radônio em ambientes internos.....	39
2. MEDIÇÕES DE RADÔNIO.....	45
2.1 Dispositivos de medição.....	47
2.2 Protocolos de medição.....	52
2.3 Garantia da qualidade para medições de radônio.....	55
3. PREVENÇÃO E MITIGAÇÃO DO RADÔNIO.....	65
3.1 Organização das ações de prevenção e mitigação do radônio.....	65
3.2 Estratégias para prevenção de radônio em novas construções.....	69
3.3 Estratégias para mitigação de radônio em imóveis existentes.....	73
4. CUSTO-EFETIVIDADE DO CONTROLE DE RADÔNIO.....	81
4.1 A estrutura da análise de custo-efetividade.....	81
4.2 Avaliações econômicas anteriores à prevenção e mitigação do radônio.....	86
4.3 Exemplo de análise de custo-efetividade.....	87
5. COMUNICAÇÃO DE RISCO DO RADÔNIO.....	97
5.1 Aspectos fundamentais, estratégias e canais.....	98
5.2 Contextualização das questões relacionadas com o risco do radônio na comunicação de risco.....	99
5.3 Mensagens principais na comunicação de risco do radônio.....	102
5.4 Campanhas de Comunicação.....	103
6. PROGRAMAS NACIONAIS DE RADÔNIO.....	107
6.1 Organização de um programa nacional de radônio.....	108
6.2 Estudos nacionais de radônio.....	110
6.3 Níveis nacionais de referência.....	113
6.4 Regulamentos e códigos de construção.....	114
6.5 Identificação e remediação de casas com altas concentrações de radônio.....	115

WHO HANDBOOK ON INDOOR RADON

A PUBLIC HEALTH PERSPECTIVE



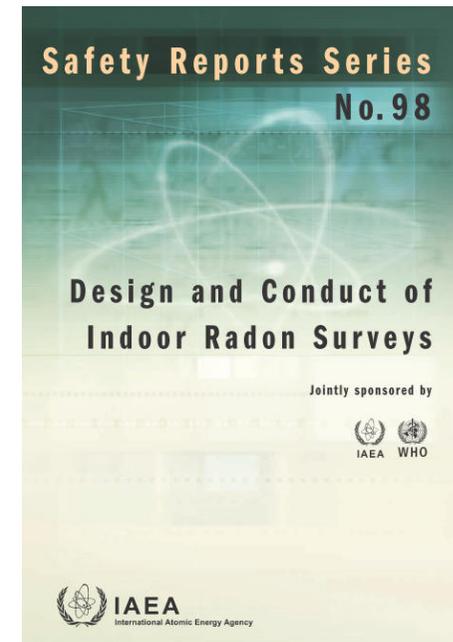
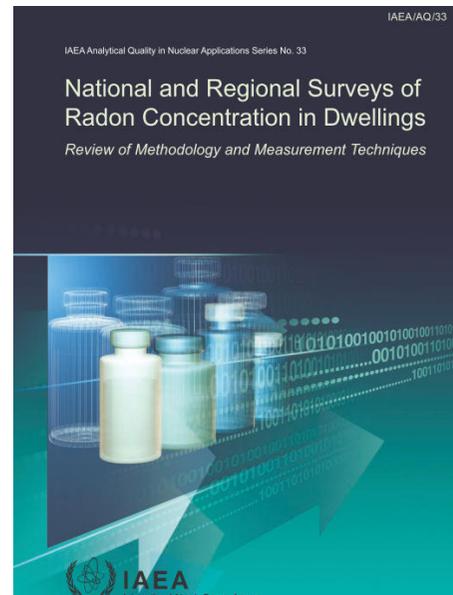
MANUAL DA OMS SOBRE RADÔNIO EM AMBIENTES INTERNOS

UMA PERSPECTIVA DE SAÚDE PÚBLICA

Documentos técnicos da IAEA relacionados a surveys

- importante: cap. 4.1: Sampling strategy
- problemas no cap. 4.1.1, p.16:
 - ref [16] (Chen et al. 2008) e ref [17] (Friedmann et al. 2018) não são geralmente aplicáveis;
 - ref [17] tem erro na eq. (3)
- Annexes I – XII: exemplos internacionais

Mais material sobre técnicas de medição;
não tanto sobre aspectos quantitativos do desenho do survey.



European Atlas of Natural Radiation

Chapters:

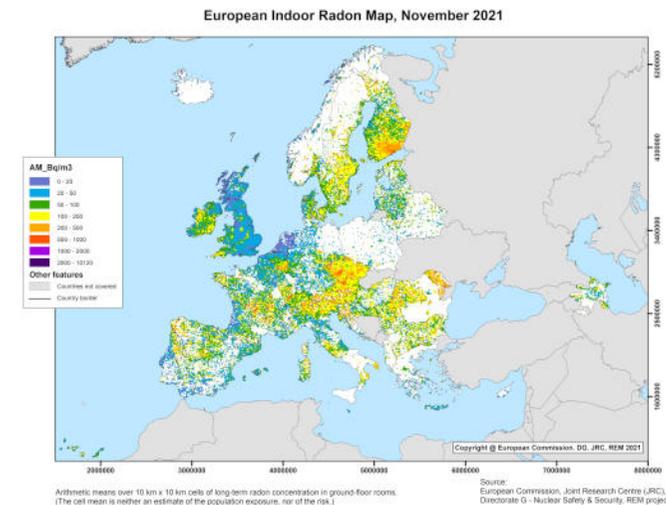
1. Rationale, legal basis etc.
2. Basics: radiation, effects, radon, exposure pathways, geology, statistics, survey design, mapping, measurement techniques, etc.
3. Terrestrial radionuclides
4. Terrestrial radiation: Ambient dose rate
5. Radon: soil, indoor, outdoor, exhalation, dose
6. Radionuclides in water and sediments
7. Radionuclides in food
8. Cosmic radiation, cosmogenic radionuclides
9. Annual effective dose



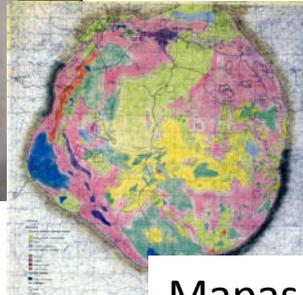
<https://remon.jrc.ec.europa.eu/About/Atlas-of-Natural-Radiation>

Versão online + livro (download gratuito)

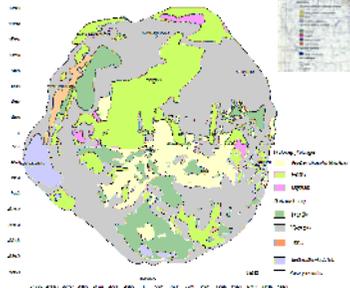
Mapa Rn indoor, AM em células 10×10 km²



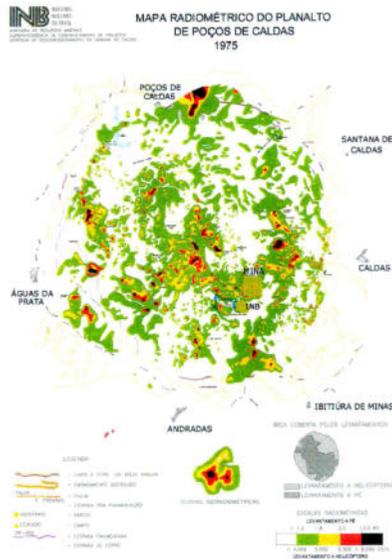
Documentos relativo a Poços de Caldas (pequena seleção)



Mapas geológicos

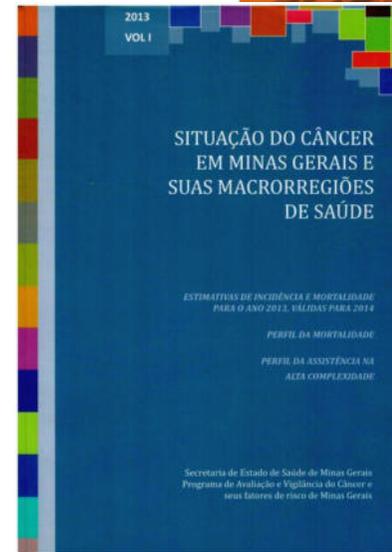
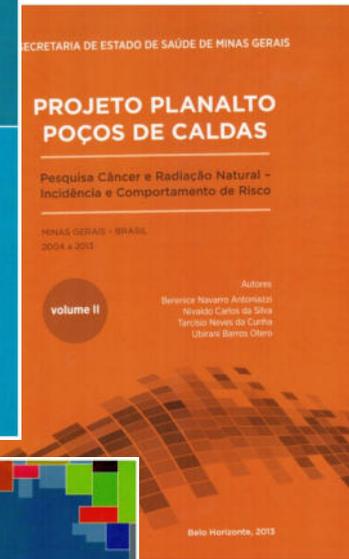
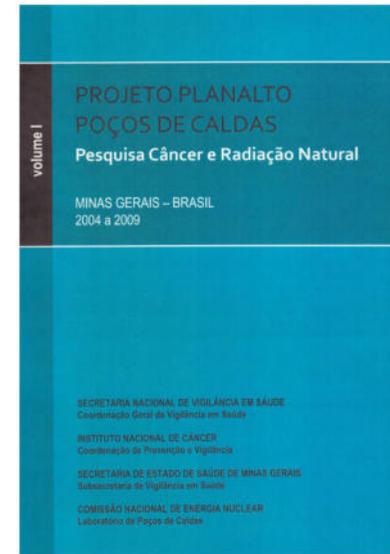


Mapa radiométrico



pesquisas geológicas

1. Classificação (CNPQ-04A, 1997 C.O.D. 014.004.031.000104.10)	2. Perfil	4. Distrito de origem
3. Palavras Chaves (selecionadas pelo autor)	INTERNO <input type="checkbox"/>	EXTERNO <input type="checkbox"/>
5. Número de	5. Data	7. Orientado por
8. Título e Sub-Título	9. Autorizado por	
10. Setor (DDD)	11. Nº de cópias	12. Nº de páginas
13. Assinatura Responsável	14. Proq.	
16. Sumário/Notas		
17. Observação		



A.3.

Situação internacional

Europa: Veja apresentações na Radio-22

Fundamentals of European radon abatement policy; German example

Round table RT 04

Bossew P., Suhr N.

German Federal Office for Radiation Protection (BfS), Berlin; pb: retired



v.15.8.

IX Congresso Internacional de Radioproteção Industrial
VIII Congresso de Proteção Contra as Radiações dos Países de Língua Portuguesa
VII Congresso Brasileiro de Proteção Radiológica
I Encontro Nacional de Mulheres do Setor Nuclear

1

European radon abatement policy; state and ongoing discussions

Bossew P., Suhr N.

German Federal Office for Radiation Protection (BfS), Berlin; pb: retired



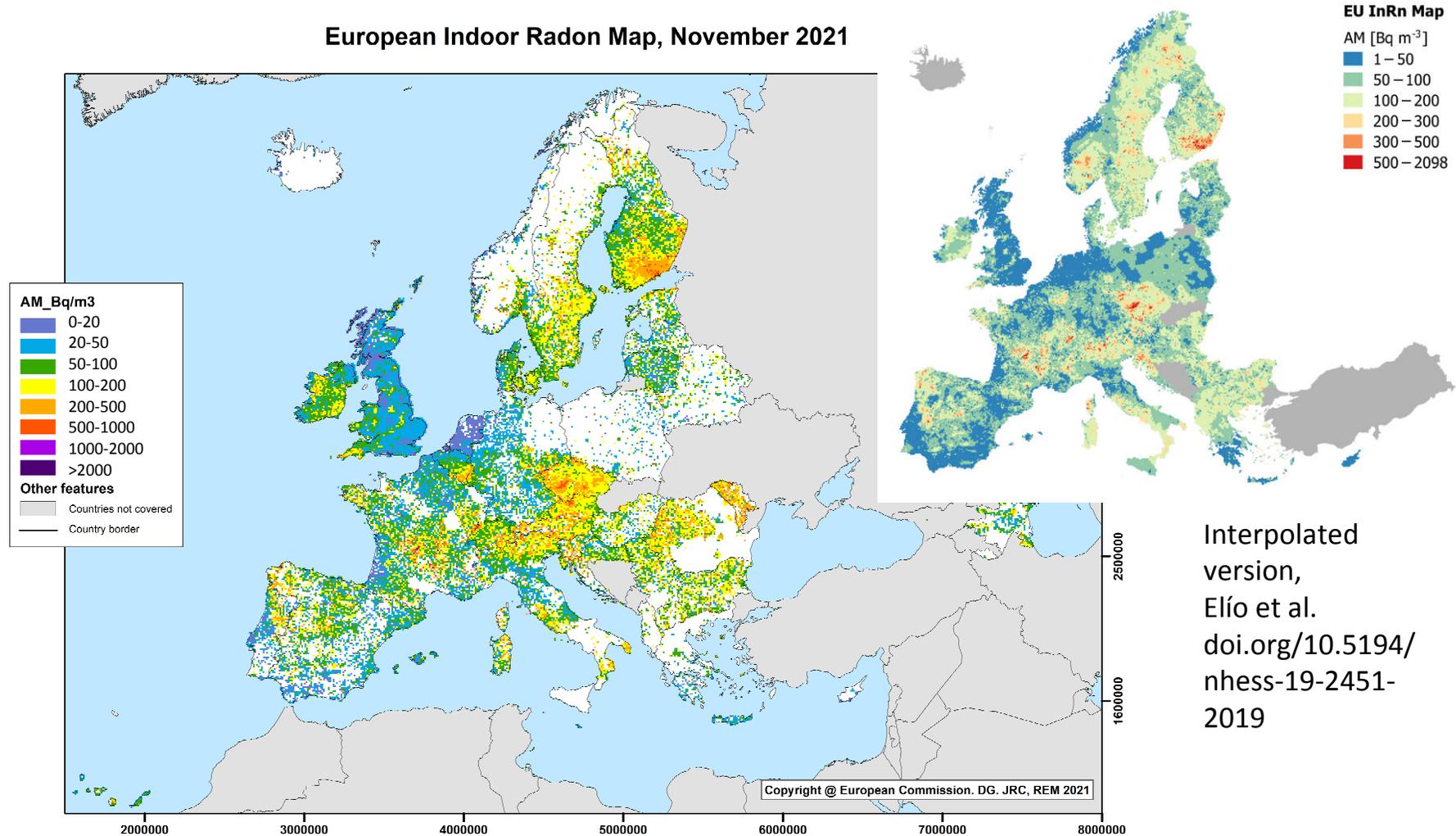
v.16.8.22b

IX Congresso Internacional de Radioproteção Industrial
VIII Congresso de Proteção Contra as Radiações dos Países de Língua Portuguesa
VII Congresso Brasileiro de Proteção Radiológica
I Encontro Nacional de Mulheres do Setor Nuclear

1

Situação: European indoor radon map

European Indoor Radon Map, November 2021



Arithmetic means over 10 km x 10 km cells of long-term radon concentration in ground-floor rooms.
(The cell mean is neither an estimate of the population exposure, nor of the risk.)

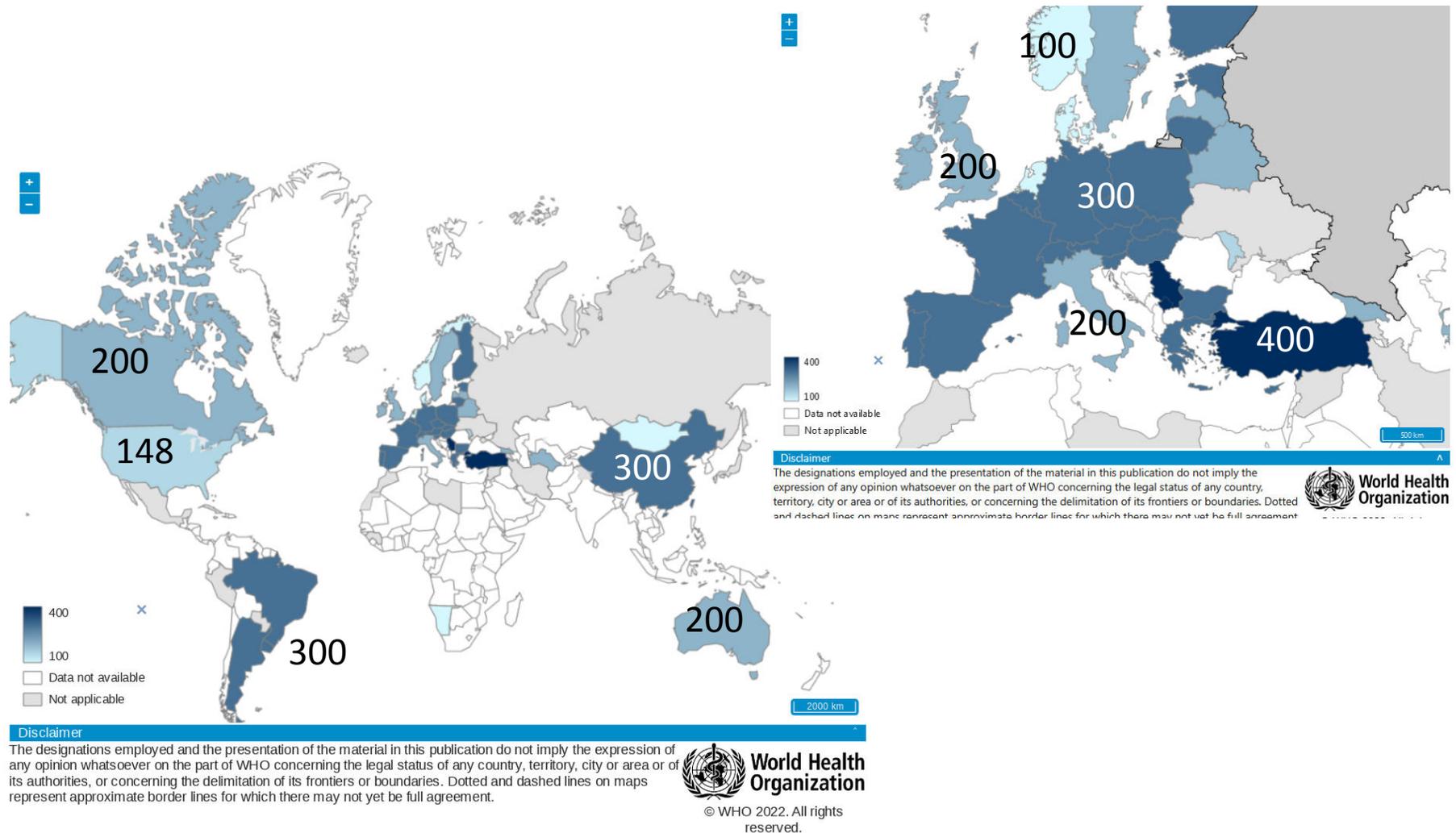
Source:
European Commission, Joint Research Centre (JRC),
Directorate G - Nuclear Safety & Security, REM project

Interpolated
version,
Elío et al.
[doi.org/10.5194/
nhess-19-2451-
2019](https://doi.org/10.5194/nhess-19-2451-2019)

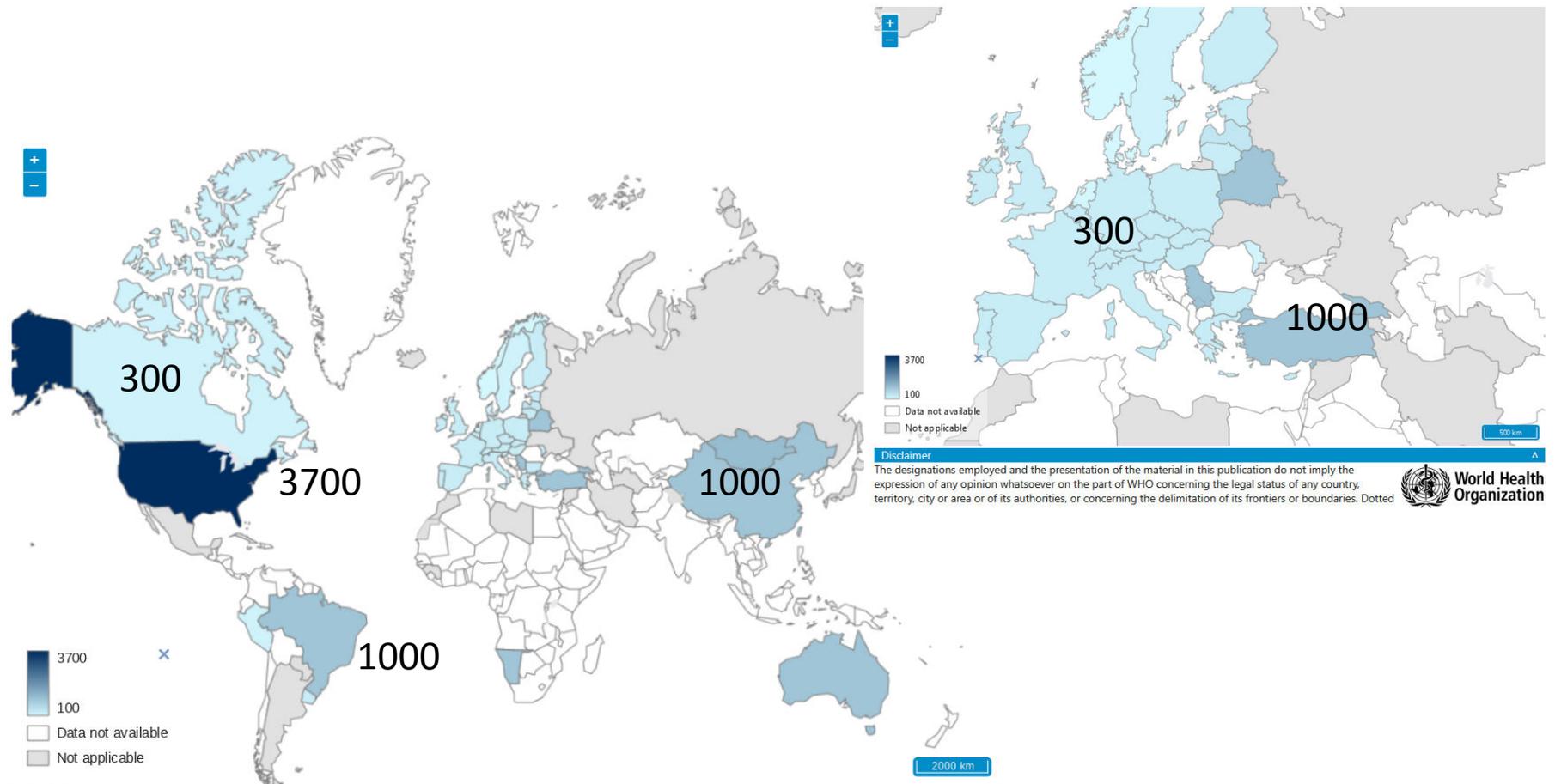
Indoor radon concentration, dwellings, ground floor living rooms, long-term mean.

<https://remon.jrc.ec.europa.eu/About/Atlas-of-Natural-Radiation/Digital-Atlas/Indoor-radon-AM/Indoor-radon-concentration>

Níveis de referência - habitações



Níveis de referência – locais de trabalho



Disclaimer

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.



© WHO 2022. All rights reserved.

Existe plano de ação nacional de radônio ?

By May 2022:

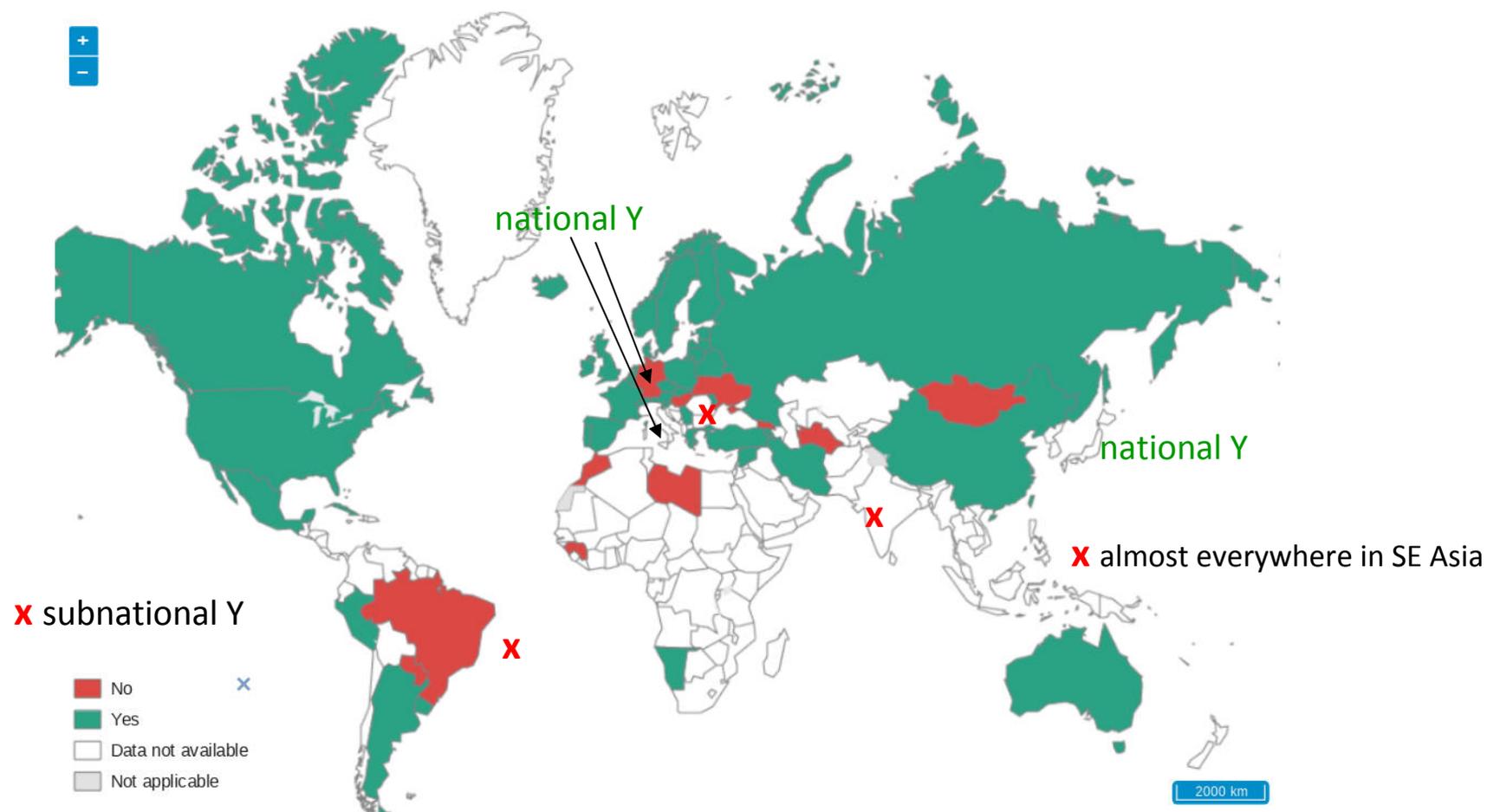
- RAP approved in 23 countries
- 'Draft document' in 3 countries (ES, IT, PT)
- No dedicated RAP in 2 countries (but radon actions are taken) (LV, LT)

Fonte:

Rovenská K. et al., Radon Action Plans in European Member States and the UK. Dealing with Radon Risk on Workplaces; NORM X symposium & RICOMET conference, Utrecht, the Netherlands on May 9 – 13, 2022



Existem estudos (surveys) nacionais?



Disclaimer

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.



© WHO 2022. All rights reserved.

RPA

- IAEA, ICRP: **Radon prone area**
- Europa: **Radon priority area**
(mas o termo não aparece nas EU-BSS)

Radon prone areas: áreas con mayor incidencia

En algunos países, estas áreas se denominan “áreas con alto contenido de radón” o “áreas afectadas por el radón”.

ICRP 65: las áreas propensas al radón podrían definirse como aquellas donde más de un cierto porcentaje (por ejemplo, 1%) de viviendas tienen una concentración de radón que excede diez veces el valor promedio nacional.

ICRP 103: un área en la que es probable que la concentración de radón en los edificios sea mayor que el promedio nacional.

ICRP 115/126 - Un área geográfica o una región administrativa definida sobre la base de relevamientos que indican un nivel significativamente más alto de concentración de radón que en otras partes del país.

(Canoba, slide 5)

Radon **priority** area: O termo indica que a ação nestas áreas deve ser priorizada porque os recursos são limitados.

Definição RPA

- BSS: *“Os Estados-Membros identificam as zonas em que se prevê que a concentração de radão (em média anual) venha a exceder, num número significativo de edifícios, o nível nacional de referência pertinente.”* (Art 103/3)
- Interpretação commum:
probabilidade de exceder RL numa zona.
Problema: não considera número de pessoas afetadas → vj.
discussão mais tarde!
- Vários métodos:
 - baseado em medidas em habitações... Austria
 - em medidas de potencial geogênico Rn (GRP)... Alemanha
 - combinação de fatores geogênicos & outros (geologia, teor U no solo, taxa de radiação externa,...) ... Suécia, França
 - base de estimação: Rn indoor, Rn geogênico, taxa de radiação externa, teor U no solo, geologia, física do solo, hidrologia, tectônica,...

Definição RPA

azul: excesso RL (nas habitações, na maioria dos casos), percentagem; atenção: RL diferentes!

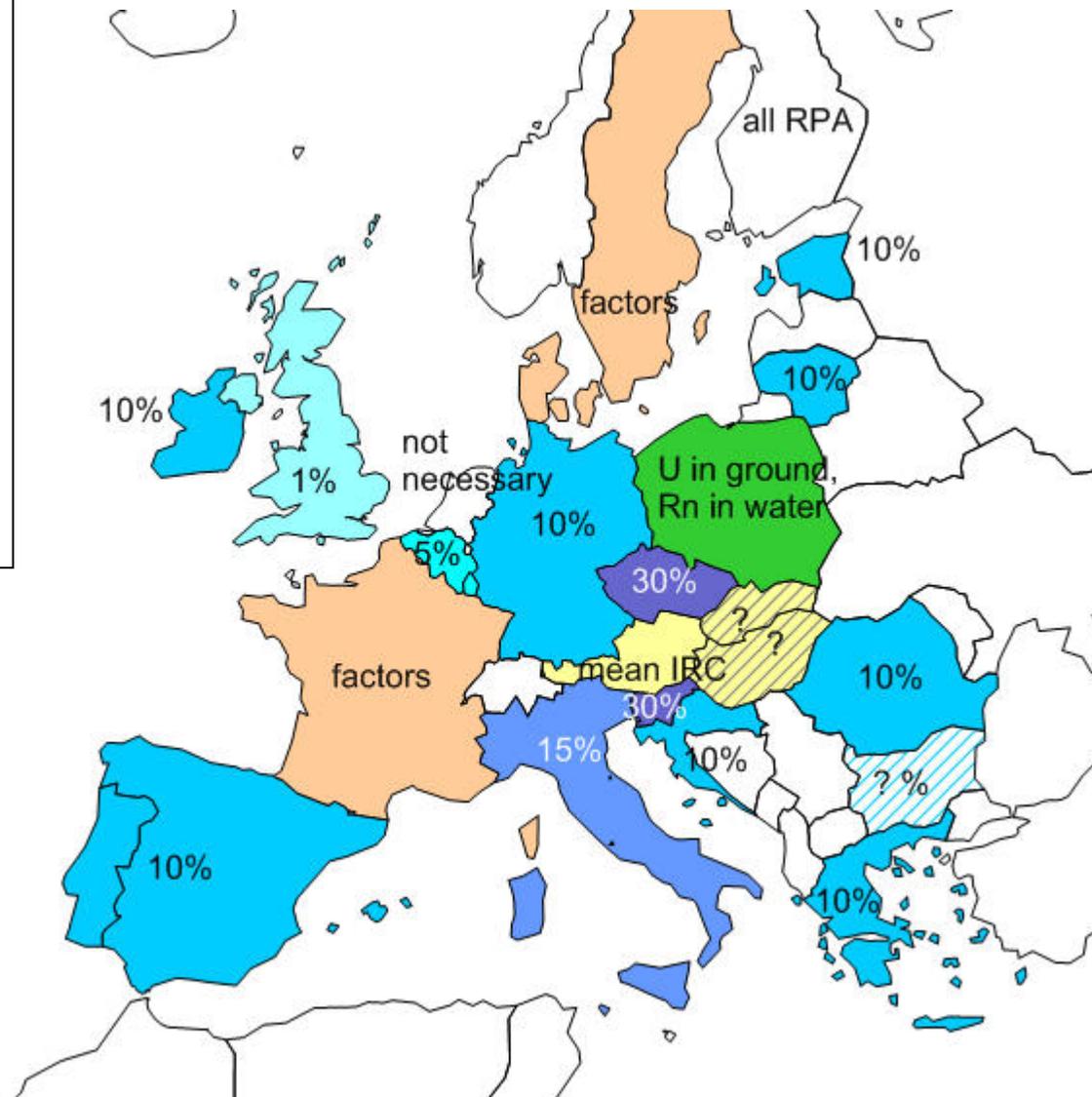
amarelo: baseado em média por município

cor laranja: combinação de fatores geogênicos

verde: diferente

fonte de dados:

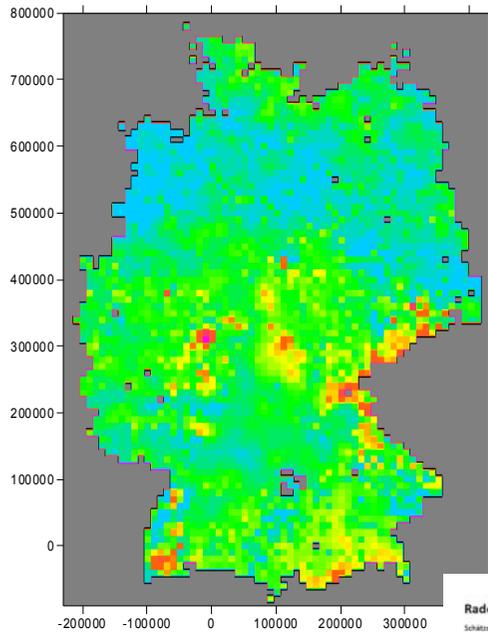
Perko T., Martell M., Rovenska K., Fojtikova I., Paridaens J., Geysmans R. (2022): Report on review and analysis of national radon action plans and their implementation, updated intermediate report, D2 (updated): SCK CEN, MERIENICE, SURO for EC DG Energy; EU-RAP project, Ref. Ares (2020)2496502



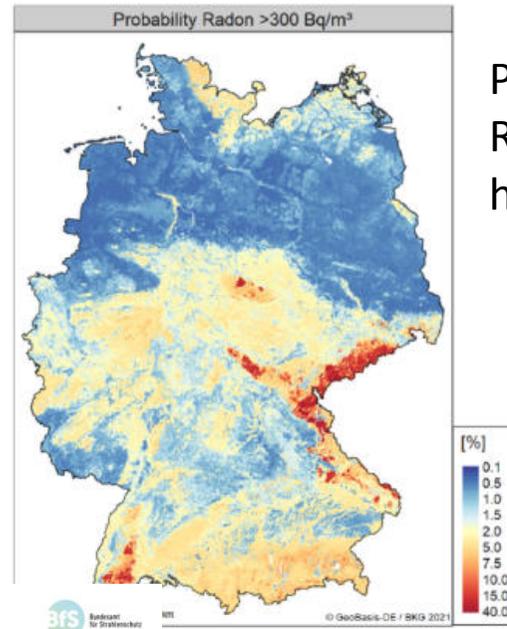
Exemplo: Alemanha

- Estrutura federal: Lei de radioproteção = Lei Federal, mas implementação: em grande parte competência dos Estados (*Bundesländer*).
- Lei (*Strahlenschutzgesetz*): essenciais, básicos
Decreto (ordinance, *Strahlenschutzverordnung*): detalhes técnicos
- RL = 300 Bq/m³ habitações, locais de trabalho + edifícios públicos
- RPA guia de orientação: áreas nos quais $\text{prob}(\text{IRC} > \text{RL}) > 10\%$. Definição exata pelos Estados. Na Lei: excesso em $> 75\%$ do território do município ou distrito.
- Até hoje, 7 de 16 Estados definiram RPAs.
Negociações duras!
Administrações receiam consequências negativas em relações econômicas & políticas: custo de estudos e da atenuação do Rn; pode desanimar investidores; pode ser publicidade negativa para a região.
- No mínimo todos 10 anos deve ter revisão dos RPAs.

Mapeamento radônio, Alemanha

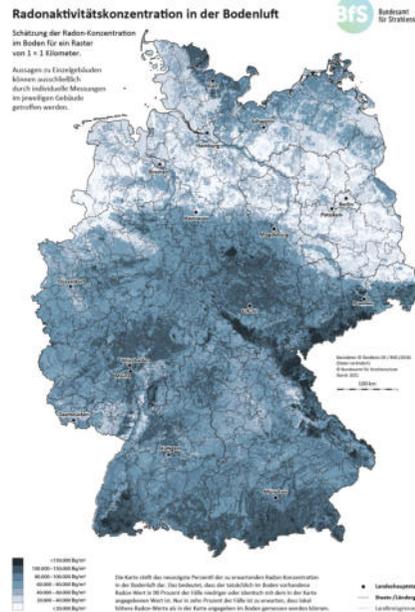


Geogenic radon potential (definição Neznal)

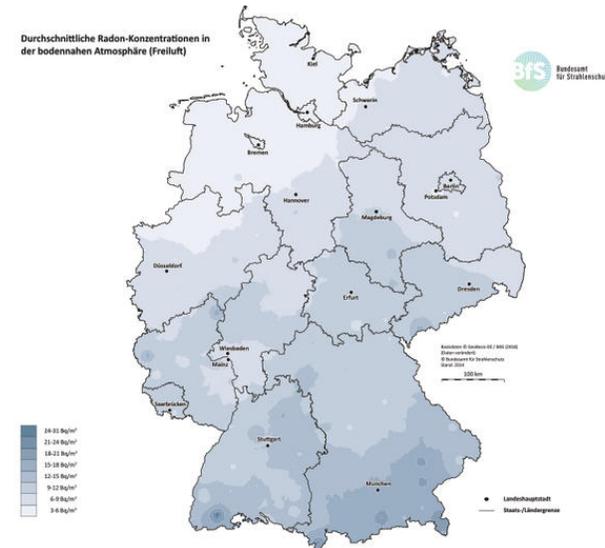


Probabilidade de exceder $RL=300 \text{ Bq/m}^3$ em habitações

concentração Rn no ar livre



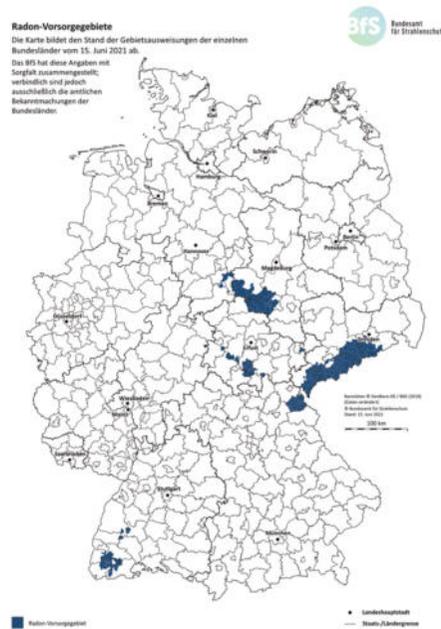
concentração Rn no ar do solo



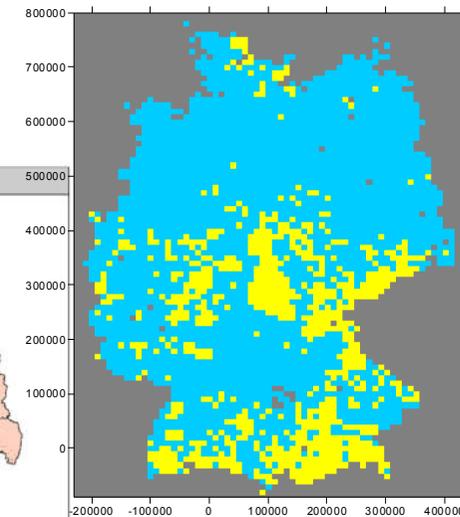
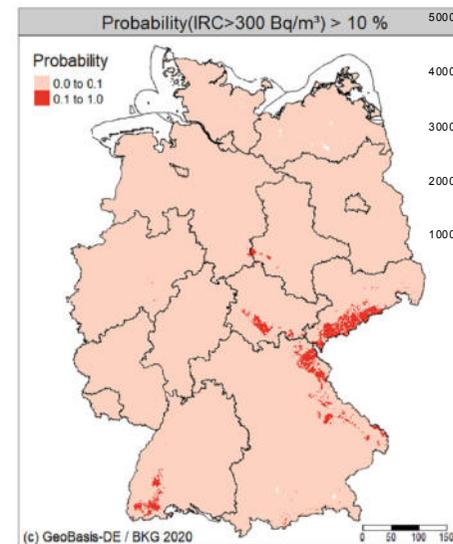
Definição RPA

- Os RPAs devem ser definidos com base em metodologia científica e dados apropriados. (Para evitar „free style“ por políticos regionais.)
- Dados apropriados: medições de Rn no solo & indoor, geologia, permeabilidade do solo, remote sensing.

(Entretanto desenvolvimos métodos de regressão machine learning que incluem dúzias de preditores incl. clima, topografia, etc. etc. – remote sensing ainda não incluído)



RPAs delineadas pelos Estados



prob(Rn>RL)>10%
métodos diferentes

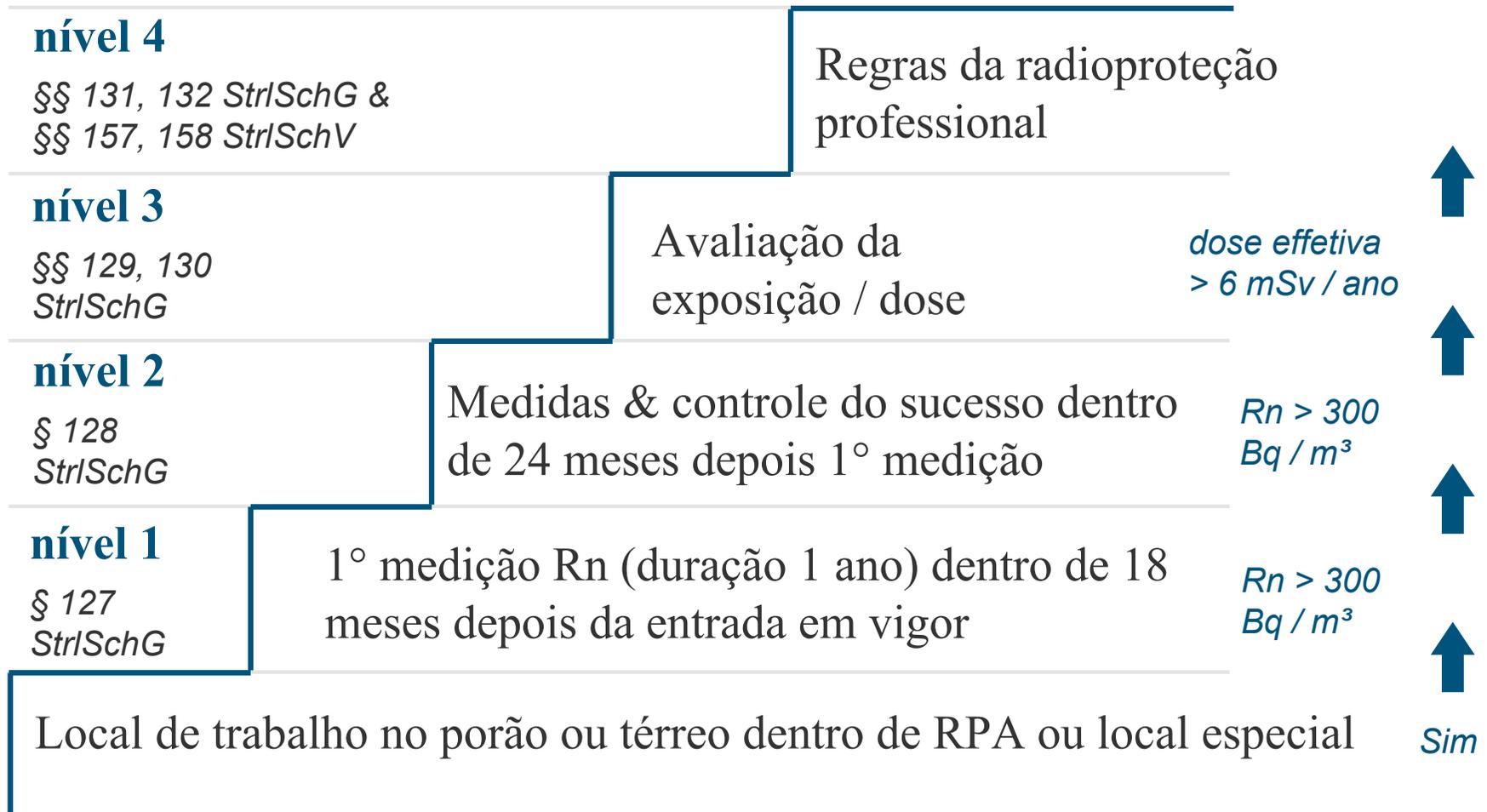
Habitações

- **Habitações existentes:** a lei de radioproteção não exige ação. Apenas voluntariamente. Administrações devem informar sobre opções. (Intervenção do estado na esfera privada é assunto muito sensível. Infelizmente a prontidão do público para reduzir Rn é muito limitada.)
- **Habitações novas:** impedir „em grande parte“ a infiltração de Rn;
 - **Fora de RPA:** Isolação contra intrusão de umidade do solo (presumido suficiente tb. contra Rn);
 - **Dentro de RPA:** medidas adicionais (isolação especial por folhas etc.)

Locais de trabalho

- **Dentro das RPAs:** medições obrigatórias no porão & térreo (ground floor)
- Caso concentração > RL: redução obrigatória.
- Conc. > RL depois da redução: BSS Art 54/3 → Art 35/2:
No caso dos locais de trabalho especificados no artigo 54. o , n. o 3, e sempre que a exposição dos trabalhadores seja suscetível de exceder uma dose efetiva de 6 mSv por ano ou um valor de exposição ao radão correspondente integrado no tempo, determinado pelo Estado-Membro, estes são geridos como situações de exposição planeada e os Estados-Membros determinam quais os requisitos estabelecidos no presente capítulo que são apropriados. Para os locais de trabalho especificados no artigo 54. o , n. o 3, em que a dose efetiva para os trabalhadores seja igual ou inferior a 6 mSv por ano ou a exposição seja inferior ao valor de exposição ao radão correspondente integrado no tempo, a autoridade competente exige que a empresa mantenha as exposições sob observação.
- **Independente de RPA:** Locais especiais com conc. Rn alta, p.ex. minas, cavernas turísticas, instalações de tratamento de água, usinas que tratam com material com teor elevado de U ou Th, etc.

Proteção contra Rn, locais de trabalho: Plano escalonado



B.1.

Front end: estatística,
desenho do survey

Valor médio

- Média aritmética, AM = unbiased estimate of expectation E !
É a quantidade preferencial + relevante!
- $E(Z) = \int_{(z \in \text{range}(Z))} z f(z) dz$, $f(z)$ = distribuição; $\text{range}(Z)$ = p.ex. $(0, \infty)$
estima $AM = (\sum z_i) / n$ n = sample size
com $n \rightarrow \infty$: $AM \rightarrow E$.
(No caso de população limitada, e.g. número de casas ou pessoas, $N < \infty$:
com $n \rightarrow N$)
- No entanto pode ter incerteza grande, em caso de distribuição assimétrica / oblíqua ou presença de valores extremos
- Média geométrica: adequado para *descrever* média de dados, mas não vale como estimação de E !!
- Caso $f(z)$ = lognormal: $AM = GM \exp(\sigma^2/2)$,
 $\sigma(z)$ = SD (desvio padrão) de $\ln(z)$ \rightarrow pode servir para estimar o AM, mais é sensível para σ .

LN distribution

- $X \sim \text{LN}(\mu, \sigma^2) \Leftrightarrow \ln X \sim \text{N}(\mu, \sigma^2)$
- $\exp(\mu)$ = geometrical mean GM, $\exp(\sigma)$ =geometrical standard deviation GSD

$$pdf(x) = \frac{1}{\sqrt{2\pi} \sigma x} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right)$$

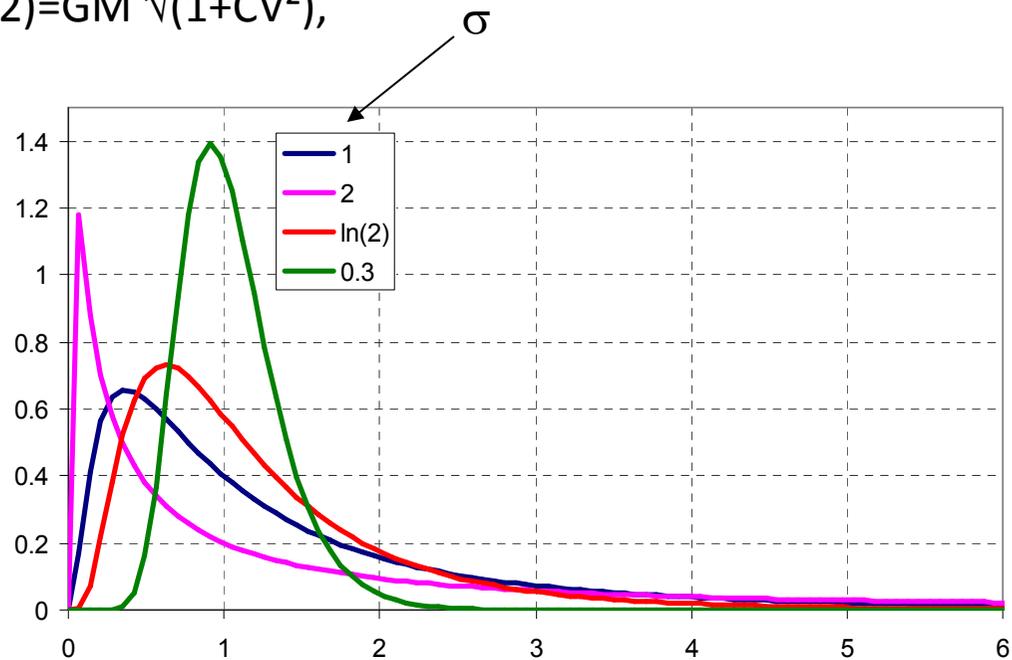
- Mean=AM= $\exp(\mu + \sigma^2/2)$ =GM $\exp(\sigma^2/2)$ =GM $\sqrt{1+CV^2}$,
CV= $\sqrt{\exp(\sigma^2)-1}$; SD=AM.CV

$$\text{Mode} = \exp(\mu - \sigma^2)$$

$$\text{skew} = CV^3 + 3 CV$$

$$\text{kurt} = CV^8 + 6 CV^6 + 25 CV^4 + 16 CV^2$$

- For $X \sim \text{LN}(\mu, \sigma^2)$:
 $aX^b \sim \text{LN}(\ln(a) + b\mu, b^2\sigma^2)$
- For $X_i \sim \text{LN}(\mu_i, \sigma_i^2)$:
 $a \prod X_i^{b_i} \sim \text{LN}(\ln(a) + \sum b_i \mu_i, \sum b_i^2 \sigma_i^2)$



Sample size

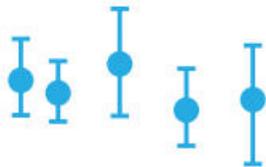
- Várias abordagens! → slides seguintes
- Depende da target variable! (média, probabilidade de excesso, estado RPA)
- Depende do objetivo: média etc. com certa incerteza definida anteriormente; diferencia significativa entre duas médias,...
- Input essencial: dispersão da população!
 - coefficient of variation, $CV=SD/AM$
 - geometrical standard deviation $GSD(Z) = \exp(SD(\ln Z))$
 - quantile deviation, p.ex.
 $Qdev_p = (Q_p - Q_{1-p}) / (Q_p + Q_{1-p})$, p.ex. $p=0.95$, $Q_p=p$ -quantile
- problema não trivial: como estimar GSD ?
 - no meu conhecimento, não se conhece a sampling distribution do GSD → bias / accuracy? random error / precision?
 - por simulação: caso true $GSD \approx 2$, $\sigma \approx \ln 2$ (realista para Rn), GSD empírica aproximadamente unbiased também para tamanho pequena da amostra.

Accuracy // Precision

Data are:

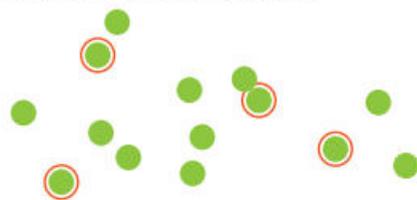
1. Numbers or items or instances which should be accurate and precise individually. This concerns metrological quality assurance.
2. A sample from a population which should represent the population accurately and precisely, i.e. be representative for the population. This concerns quality assurance of the survey design and its implementation.

Individual data:



Issue: uncertainty

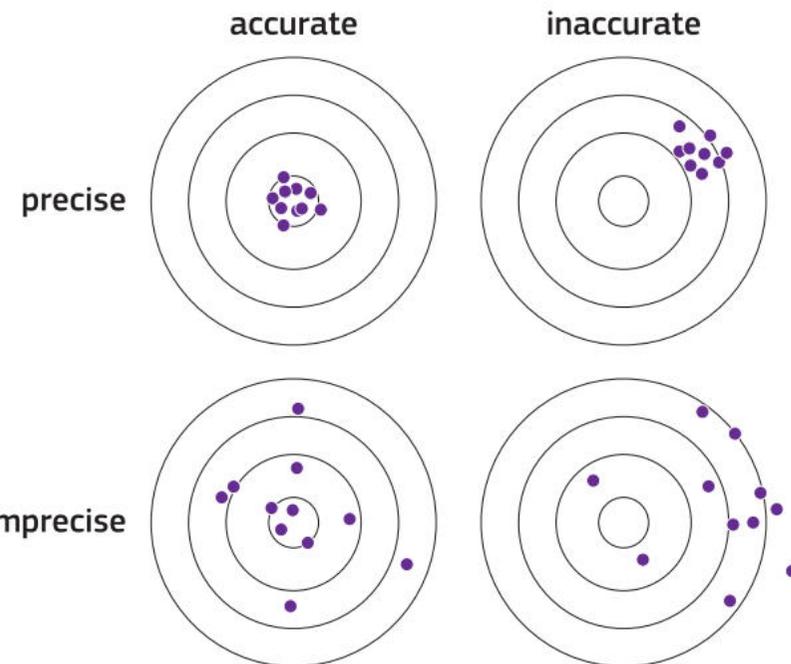
Sample from a population:



Issue: representativeness

● population
○ samples

falta de representatividade \Rightarrow bias (inaccuracy)



Sample size n: abordagens

n depende de:

1. Target variable:

- AM Rn por unidade geográfica U
- exceedance probability $\text{prob}(R_n > R_L)$
- RPA status

2. Design // model based approach

- Design based → representatividade!
- Model based → disponibilidade + qualidade de preditores; disponibilidades de modelos apropriados

Design // model based

Design

- Gerar a amostra assim que garanta resultado correto (acurado: bias tolerável; preciso: incerteza aleatória tolerável) calculado pelos valores da amostra.
- Bias depende da seleção / locação da amostra
- Incerteza depende do tamanho da amostra

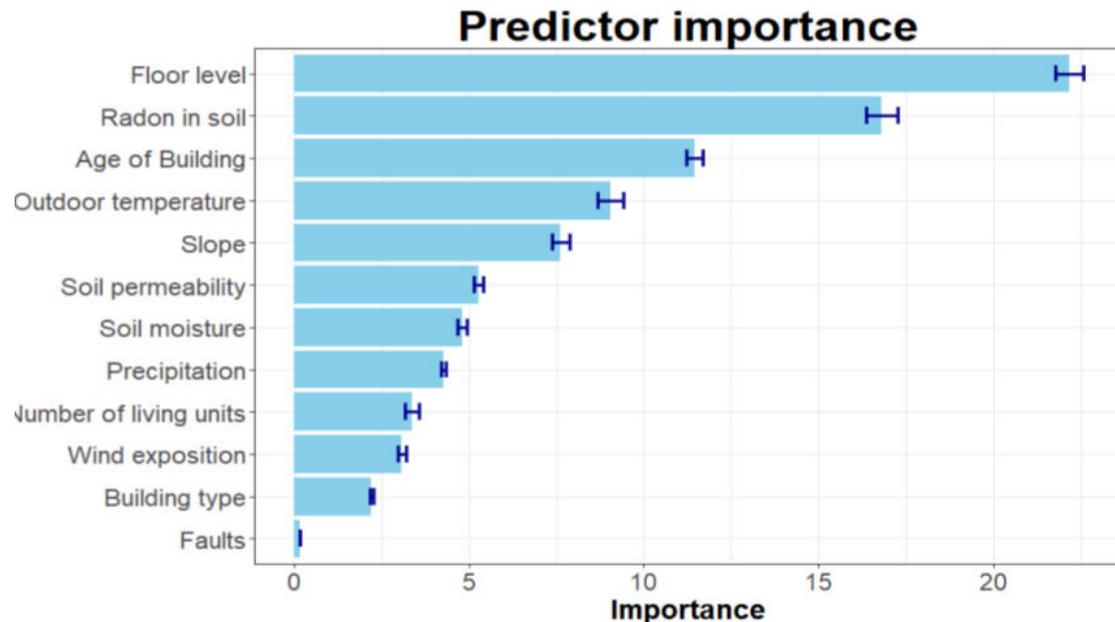
Model

- Valor da variável Z num lugar x:
 $Z(x) = f(\text{preditores})$
- preditores:
 - (a) independentes (geologia,...) → regressão;
 - (b) o mesmo proceso Z → geoestatística.
- Calibração do modelo:
 - na região (unidade geográfica) em questão, possivelmente amostra pequena, não representativo
 - em regiões diferentes → upscaling
- Budget da incerteza: complicado!

Model based, exemplo recente

- Mapeamento fo Rn indoor em Alemanha
 - Resolução: município (tipicamente poucos km)
 - **Apenas 7500 valores** (casas) Rn, do survey *quase* representativo.
 - Preditores inicialmente: 15, relevantes: 12 (→ VIP plot)
-
- Método: ML
 - Procedimento bem complicado!
 - Petermann et al.,
subm. Environ.
Health Persp.

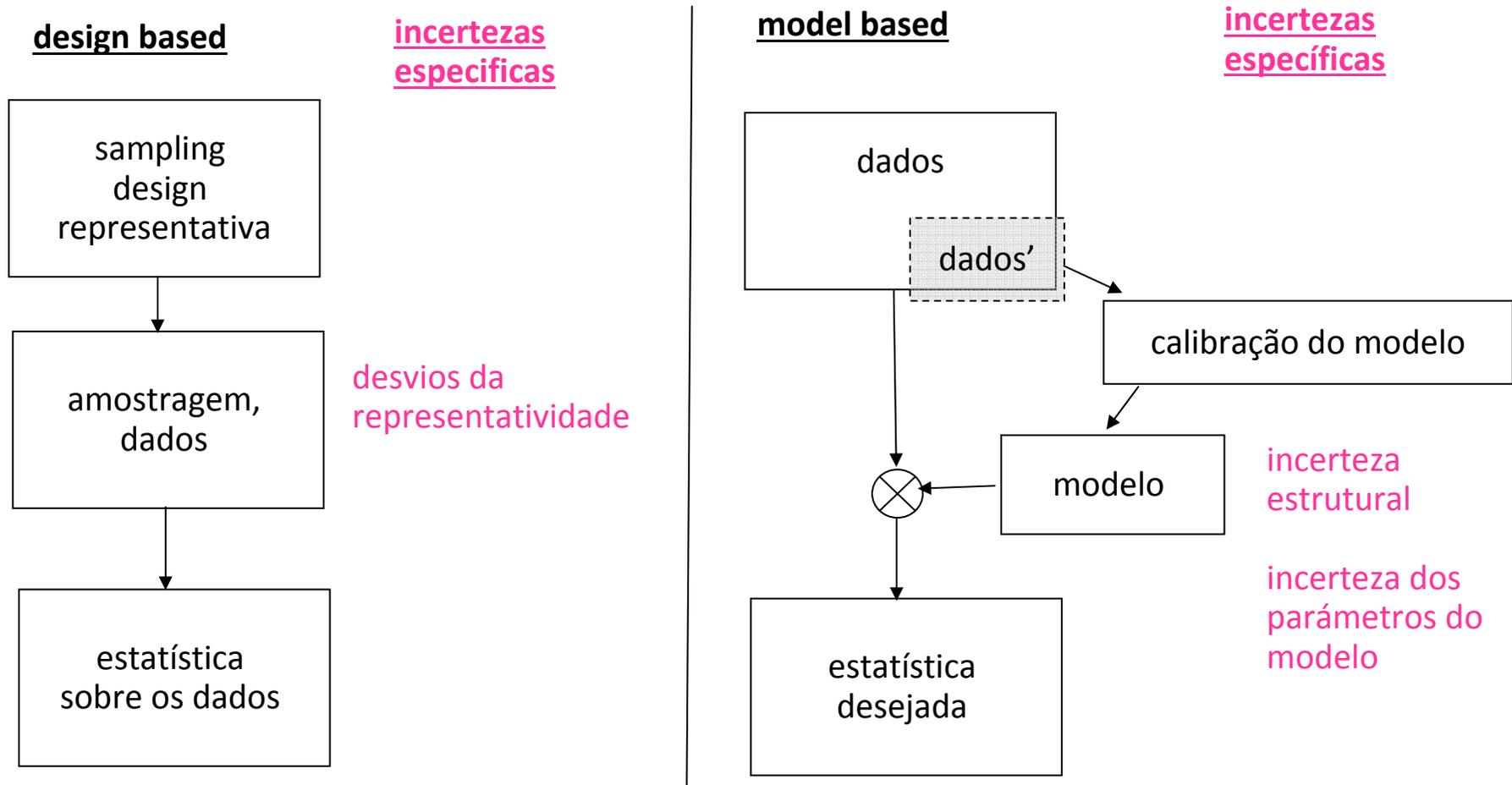
<https://arxiv.org/abs/2310.11143>



Vantagens / Desvantagens

	design-based	model-based
Vantagens	Conceitualmente mais fácil	<ul style="list-style-type: none">• Menos medições• Representatividade relaxada
Desvantagens	<ul style="list-style-type: none">• Muitíssimas medições!• Representatividade estrita: complicado	<ul style="list-style-type: none">• Procedimento complexo• Necessidade de preditores em resolução alta• Unc. budget mais complicado; incerteza adicional relacionada ao modelo

design vs. model based approach - 2



incertezas comuns:

- incert. dos dados mesmos (resultados de medições)
- incert. de estimativa ~ número de dados

representativo

- significa que a distribuição da quantidade da qual foi retirada uma amostra, é igual à distribuição da quantidade verdadeira (que não é conhecida)
- Na prática, os primeiros 2 momentos devem coincidir (Média, desvio padrão)
- Provavelmente o desafio mais difícil do survey!
- Verificação através de metadados, vj. depois.

Tamanho de amostra n

em caso de design based

- Passo inicial: definição da precisão (error aleatório) toleravel TOL e da significância p
- Target variable = **AM ou GM**:
 - Calcula o intervalo da confiança do AM ou GM empírico em dependencia de n, TOL, p e da característica estatística assumida do proceso Z
 - Invertir → $n_{\min} = f(\text{TOL}, p, \text{carateristica})$
 - Como ? → slide seguinte
- Target variable = **prob(Z>RL)**:
 - Calcula o intervalo da confiança da probabilidade empírica = proporção empírica $\#(Z>RL)/n$ em dependencia de N, TOL e p;
 - Invertir, igual antes
 - Como? → 2. slide seguinte.
- Target variable = **RPA status** {sim, não}
 - Mais complicado! Tem que estimar 1. and 2. kind error probabilities e comparar com aqueles toleráveis. Possivelmente por simulação numérica.
 - Alternativa: derivar de AM ou $\text{prob}(Z>RL)$, dependente da definição do RPA.

Target variable = GM

- Assumpção: Processo $Z = \text{lognormal} \Rightarrow \ln(Z) \sim N(\mu, \sigma^2)$,
 $\mu = \ln(\text{GM})$, $\sigma = \ln(\text{GSD})$
- Intervalo de confiança: $\mu \pm x_{\alpha} \sigma / \sqrt{n}$, x_{α} = standard normal quantile,
p.ex. $\alpha = 0.05$: $x_{1-\alpha/2} = 1.96$, $1 - 2 \times 0.05 = 90\%$ = prob que μ = dentro do intervalo.
- Condição: $|\text{observed } \mu - \text{true } \mu| < \text{TOL}$
- Def.: $(\text{obs GM} - \text{true GM}) / \text{true GM} =: \text{PREC} \rightarrow \text{TOL} = \ln(\text{PREC} + 1)$
- \Rightarrow Hale formula:

$$n_{\min} = \left(\frac{x_{1-\alpha/2} \ln(\text{GSD})}{\ln(\text{PREC} + 1)} \right)^2$$

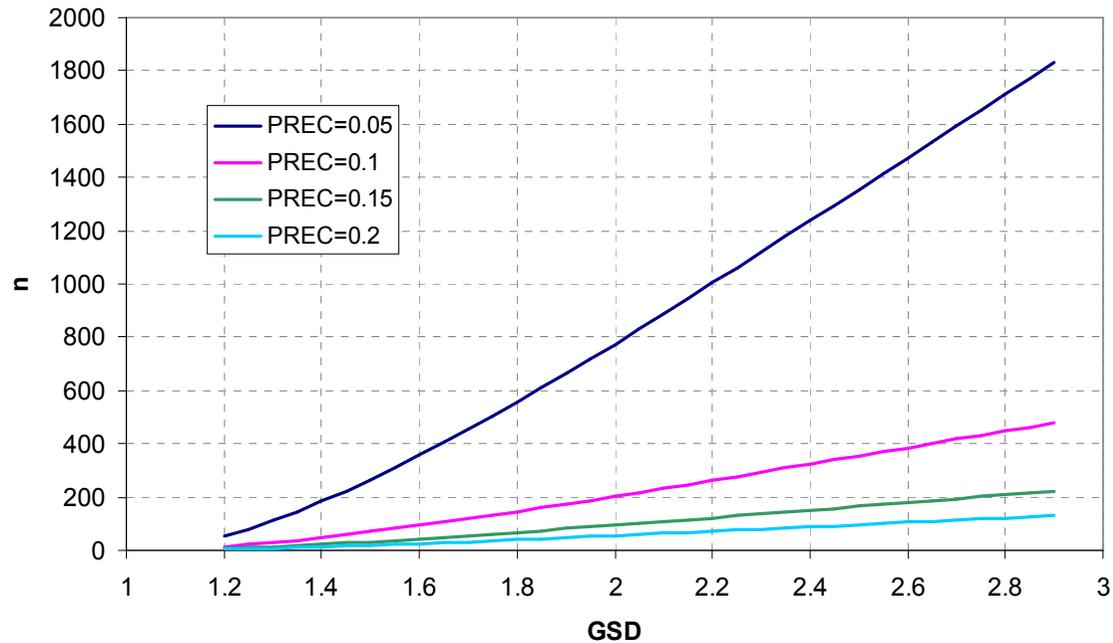
Valido para processo Z contínuo, p.ex. amostra de solo $\rightarrow n_{\min}^{\infty}$.
População limitada N , p.ex. casas: $n_{\min} = n_{\min}^{\infty} / (1 + n_{\min}^{\infty} / N)$

Fonte: Hale 1972, [https://doi.org/10.1016/0004-6981\(72\)90138-2](https://doi.org/10.1016/0004-6981(72)90138-2)

Problemas:

- Processo realmente log-normal? Não se sabe.
- Como saber GSD?
 - GSD empírico, estimado da amostra \rightarrow substituir $x_{1-\alpha/2}$ por $t_{1-\alpha/2, N-1}$ \rightarrow inversão não possível facilmente \rightarrow iteração.
 - Caso Rn: Valor típico GSD ≈ 2 ; mas na realidade depende da área da região e do valor da média AM.
- Na prática: n_{\min} é valor indicativo, por isso acho aproximações admissíveis:
 - Processo = log-normal
 - GSD = 2
 - Calcular n_{\min} para GM em vez do AM
- Há fórmula teoricamente correta para amostra = tiragens independentes da mesma distribuição (i.i.d.); na verdade Z = auto-correlated, por isso não i.i.d.
Por simulação assumindo auto-correlação (variogram) realista e área da região > uns km: Erro não muito alto, acho tolerável.

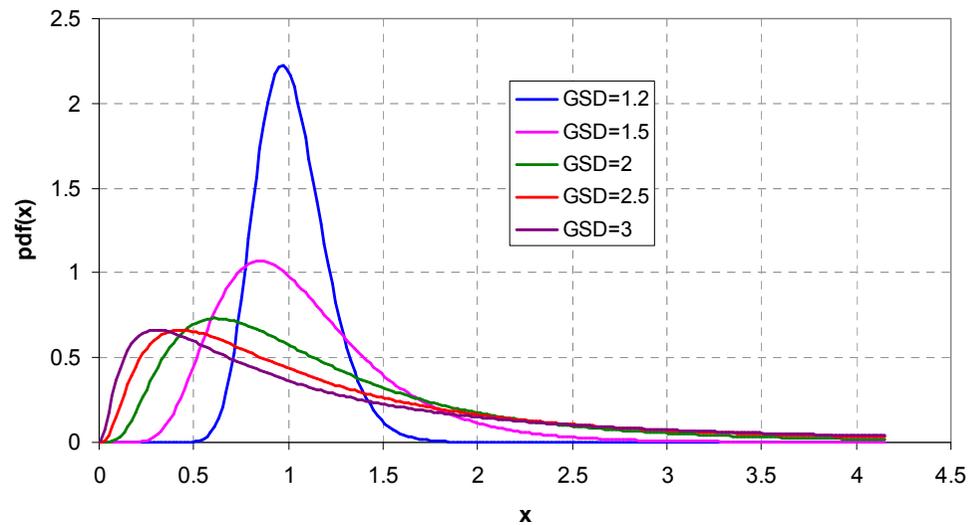
Hale formula



application of Hale formula

for high precision (low PREC value): high sample number!

($\alpha=0.05$)



reason: long right tail of distributions with high GSD

(GM set to = 1)

Consequência para survey: exemplo

São Paulo City
has 1523 km²,
2019: 12.3 Mill.
inhabitants
(2023: 11.5 Mill);
Agglomeration
ca. 21 Mill.



Administrative structure:
Zonas (colours)
Subprefeituras
Distrito

Question:

How many dwellings have to be measured, in order to determine, with precision (PREC)=0.1 at $\alpha=0.05$, the mean indoor Rn concentration

- (1) Of the entire city;
- (2) Of each zona;
- (3) Of each subprefeitura;
- (4) Of each distrito?

- Use of Hale formula;
- GSD acc. power model;
- Shape of units assumed as circles with area of the unit;
- Finite sample correction necessary only for one small unit.

- Result:

administrative level	dwellings sampling rate (%)	total number of samples
(1) São Paulo City	0.016	427
(2) Zonas	0.081	2131
(3) Subprefeituras	0.20	5163
(4) Distritos	0.44	11504

The more detail (higher resolution) wanted, the more measurements are required!

Sampling rates:

- Zielinski (2008) proposed 0.5% of dwellings for cities > 100,000 inhabitants → For São Paulo: 15,400 dwellings (occupation factor OF=4 pers/dwell assumed).
- In the first Austrian survey, an overall rate of 0.5% was targeted → 40,000 measurements. This was almost achieved.

The São Paulo survey has never been performed, unfortunately!

Target variable = AM

Confidence interval of the LN mean

There is no closed formula for the confidence limits of a LN mean.

A commonly used approximation (!) of the confidence limits of a LN mean is given (e.g. Olsson 2005) by:

$$CL_{L,U} = \exp \left\{ y + \frac{s^2}{2} \mp t_{1-\alpha/2; n-1} s \sqrt{\frac{1}{n} + \frac{s^2}{2(n-1)}} \right\}$$

with $y = \text{AM}(\ln Z)$ and $s = \text{SD}(\ln Z)$, empirical logarithmic sample mean and standard deviation, $s = \ln(\text{GSD}) \equiv \text{SDL}$.

Evidently, this cannot be solved easily for n , such that $CL_U - CL_L < \text{tolerance}$; iterative algorithms would be necessary. Therefore practical use is limited.

Olsson (2005): <https://jse.amstat.org/v13n1/olsson.html>

slide from: Design of a radon survey. IAEA Regional Workshop on Good Laboratory Practice in the Measurement of Radon. RER9153-EVT1903873. BfS-Berlin, 18.-20. 11. 2019

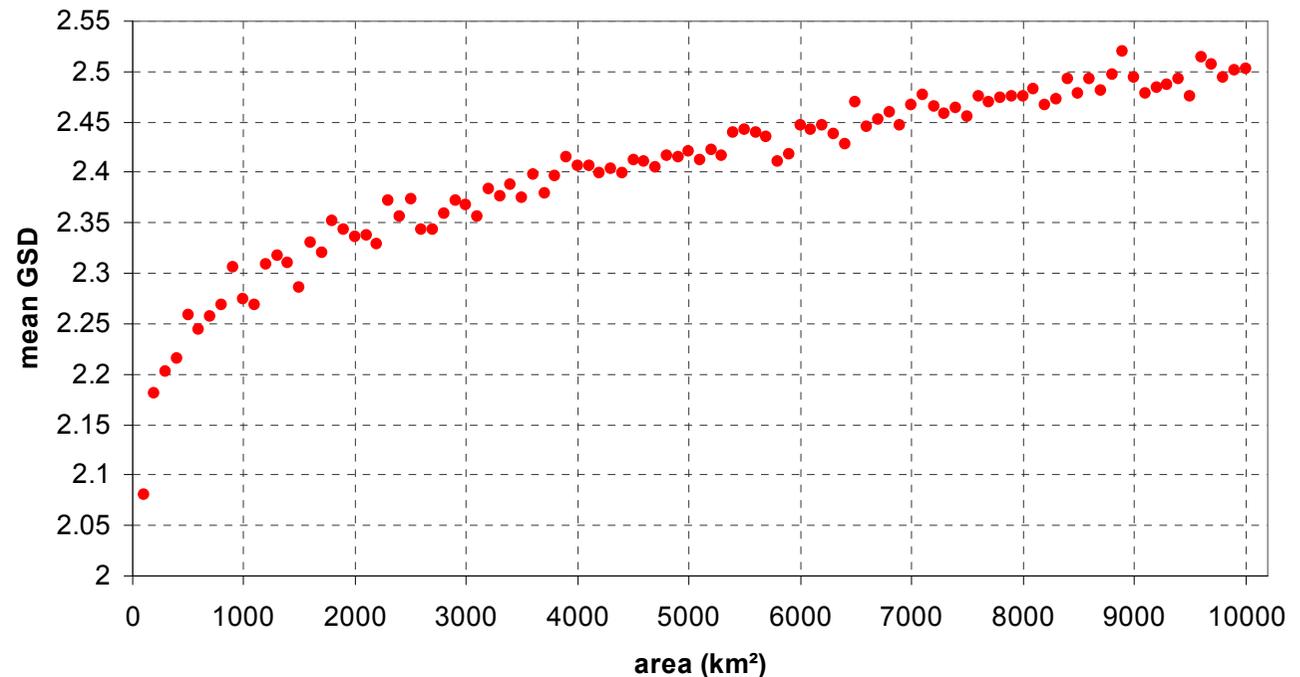
GSD

Teóricamente: dentro da unidade U, $GSD \approx \exp\left(\sqrt{\gamma(\langle s \rangle)}\right)$

γ = variogram, $\langle s \rangle$ = distância média entre dois pontos aleatórios dentro de U;
depende da área e da forma de U. U=círculo: $\langle s \rangle \approx 0.511 \sqrt{\text{área}}$; U=quadrado:
 $\langle s \rangle \approx 0.521 \sqrt{\text{área}}$.

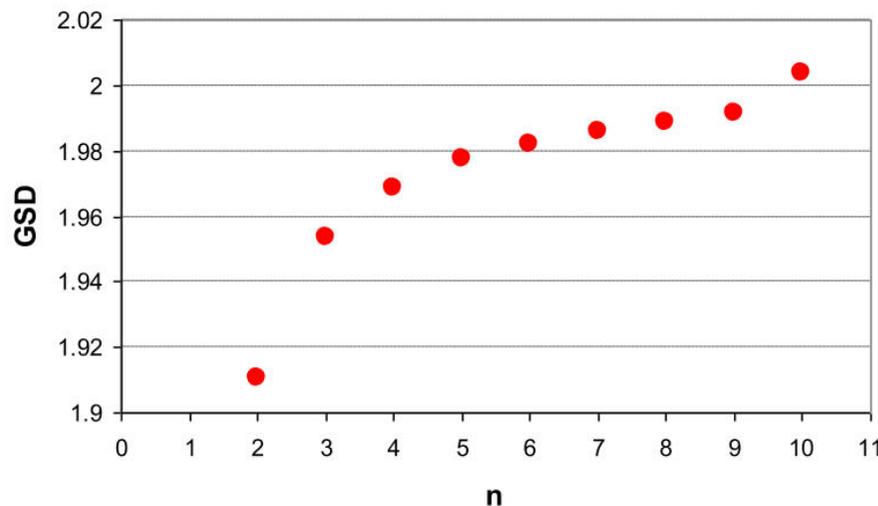
Dependência **empírica**
de GSD de Rn-indoor da
área, Europa. (Dados do
mapa Rn, European
Atlas of Natural
Radiation)

(Bossew 2021, GeoENV Parma, in
<https://hdl.handle.net/1889/4373> ,
<https://www.repository.unipr.it/handle/1889/4373>)



Alternativa prática

- GSD empírico dos dados: bias depende do GSD verdadeiro
- GSD verdadeiro = 2 (realista para Rn), assumption LN: GSD empírico **quase** unbiased também para pequeno tamanho de amostra (n).



baseado em 10,000
simulações

Dependence of the empirical GSD on sample size. True population GSD = 2 assumed

Estimar desvio padrão SD

- Assumpção: processo normal
- exacto:
p.ex., Stackexchange (2012) <https://stats.stackexchange.com/q/27984>

$$SD(corr) = SD(simple) \sqrt{\frac{n-1}{2} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}}$$

- aproximações:
 - substituir fator Bessel (n-1) na formula commum por (n-3/2)
p.ex., Brugger (1969), <https://doi.org/10.1080/00031305.1969.10481865>
 - ou corr. factor $(1+1/(4(n-1)))$
Gurland & Tripathi (1971), <https://doi.org/10.2307/2682923>
- $SD^2(\text{simple}) = \text{Var} = \text{unbiased!}$

Target variable = $\text{prob}(Z > RL)$

Opções:

1. Intervalos de confiança da proporção empírica. Preciso de número suficiente de valores.
2. Modelo: assumir LN, estimar o distribution tail. Difícil calcular o intervalo de confiança desse “tail”, provavelmente apenas por simulação.

Opção:

- (i) Usar o distribution tail como estimação da proporção;
- (ii) Procedimento 1)

1) Proporção empírica

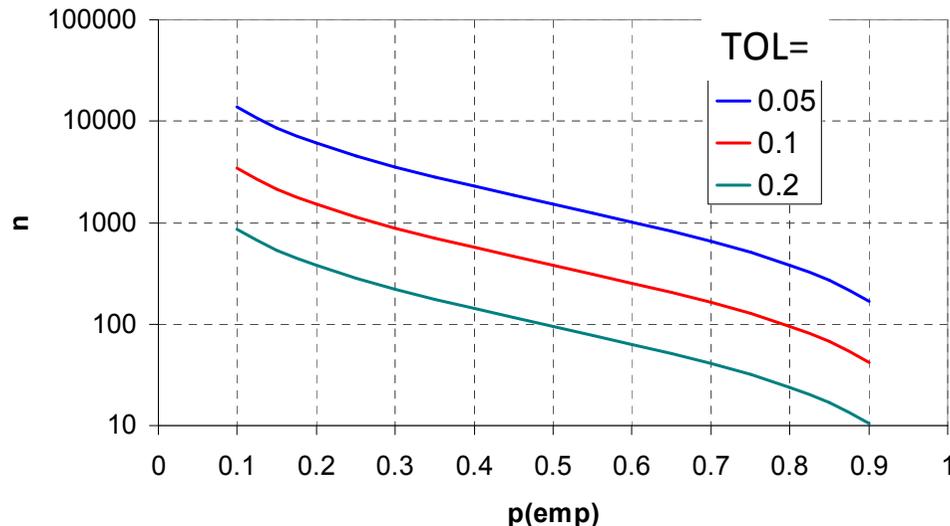
- $p(\text{emp}) := \#(z > \text{RL})/n$
- **versão simples:** “Wald interval” (Laplace 1812): ($p^{\wedge}=p(\text{emp})$)
- assintótico com $n \rightarrow \infty$!!
- **assumpção:** processo binomial
- several problems (overshoot etc.), therefore not recommended for exact purposes, e.g. Brown et al. 2001, [doi:10.1214/ss/1009213286](https://doi.org/10.1214/ss/1009213286)

$$C.I. \approx \hat{p} \pm x_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$$

$$\text{Def. } PREC = \left(x_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} \right) / \hat{p} < TOL$$

$$\Rightarrow n > \approx \frac{\hat{p}(1-\hat{p})}{(TOL \hat{p} / x_{1-\alpha/2})^2}$$

$\alpha=0.05 \rightarrow x_{\alpha/2}=1.96$:



$p(\text{emp})$ baixa \rightarrow n bastante alto para obter precisão tolerável!

p.ex.: $p(\text{emp}) = 20\%$ das casas $>$ RL $\rightarrow n=380$ para estimar p com precisão relativa 20%.

Ex. inverso: $n=20$ amostras, 4 valores $>$ RL $\rightarrow p(\text{emp})=4/20=0.2$; C.I. $\approx 0.2 \pm 0.17 \rightarrow PREC=87\%!!$

- **Wilson score** interval:

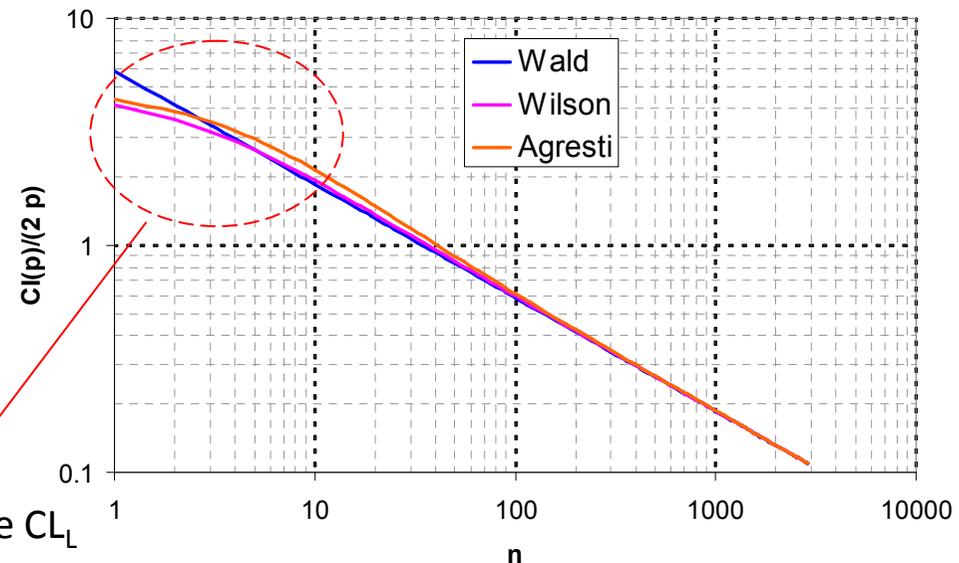
$$CL_{L,U} = \frac{1}{1 + x^2 / n} \left(\hat{p} + \frac{x^2}{2n} \mp \frac{x}{n} \sqrt{n\hat{p}(1 - \hat{p}) + x^2 / 4} \right)$$

($x \equiv x_{1-\alpha/2}$; possível corrigir para evitar overshoot)

- **Agresti** interval:
 igual Wald, mas substitui:
 $n \rightarrow n' = n + x^2$, $p \rightarrow p' = (np + x^2/2) / n'$

$x=1.96$, $p(\text{emp})=0.1$

- **“Rule of three”**
 Se nenhum “sucesso” for observado, i.e. $np=0$, intervalo ($\alpha=0.05$) $\approx (0, 3/n)$



absurdo, porque CL_L
 não pode ser <0

2) Proporção modelada

- Assumpção: processo LN
- Tail area = $p = \text{prob}(Z > RL) = 1 - \Phi(y)$, $y := (\ln(RL) - \mu) / \sigma$
- Como estimar essa quantidade, se μ e σ foram desconhecidos e devem ser estimados pelos dados?
- **“naive”**: em Φ , usar μ , σ empíricos = AML, SDL, mas \rightarrow bias porque não é linear.
- **Exato**: $p = \text{Beta}(u_b, \beta)$, Beta – the cumulative symmetric beta distribution with $\beta = (n-2)/2$ and $u_b = \max[0, 0.5 \cdot (1 - (\ln RL - \text{AML}) \sqrt{n} / (\text{SDL} (n-1)))]$, AML, $\text{SDL} = (\mu, \sigma)$ empíricos.

$$\text{Cum}B(x, a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt, \quad B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

symmetric: $a=b=\beta$

Lieberman GJ, Resnikoff GJ (1955): Sampling plans for inspection by variables. J Am Stat Assoc 50(270):457. <https://doi.org/10.2307/2280972>

- Andersen et al. (2001), Murphy & Organo (2008)

DOI: 10.1016/s0048-9697(01)00697-0, DOI: 10.1088/0952-4746/28/3/001

$p = 1 - \Phi(u) + B$, where:

- $u = (\ln(RL - Rn_{out}) - AML) / SDL$,
 Rn_{out} = mean outdoor Rn conc.
- $B = \text{bias term} = (-u / (2n\sqrt{2\pi})) \exp(-u^2/2)$

- Beard (1960) suggested simpler

$p = t_{n-1}[u \sqrt{n/(n+1)}]$, where:

- $u = (\ln RL - AML) / SDL$
- Beard (1960), <https://doi.org/10.1029/JZ065i007p02143> ;
 Proschan (1953), <http://www.jstor.org/stable/2281009> ;
 Wilks (1941), <http://www.jstor.org/stable/2235627>

Still to do: comparison between the methods! Bias (accuracy), precision, coverage probability

- Slide 10 en Canoba 2023:

← IAEA SRS 98

Tamaño de la muestra

1) Método 1

- Para 100 viviendas elegidas aleatoriamente en un área, el porcentaje de viviendas por encima del NR puede determinarse con una precisión del 25%.
- Para 50 viviendas, este porcentaje es del 30%.

Health Physics 94(4):362-5 Determining the sample size required for a community radon survey-May 2008

2) Método 2

- Si para 5 viviendas la $[Rn] + 1\sigma < RL$, el área no es de prioridad

não acho correto!

Baseado em Chen et al. (2008); Apenas applicavel para conc. Rn considerada neste artigo!

Baseado em Friedmann et al. (2017), contém erro

surveys: exemplos Austria, Alemanha

	carater	base da amostra	n	AM / Med (Bq/m ³) / %>300
AT-1	≈ dem. representativo	lista telefónica, visitas individuais, 1992-2002	≈10 k domicílios, 16,096 quartos	99 / 61 / ?
AT-2	geográfico	bombeiros, ≥12 loc./município, 2014-19	≈25 k dom.	166 / 99 / 12%
DE-1	vários projetos, corr. por modelo	< 2006 ?	32,336	49 / 37 (GM) / ?
DE-2 "Burg"	vários projetos, voluntários	ca. 1980-2022	44,629, corr. por modelos	70 / 45 (GM) / 2.8%; modelo 1.9%
DE-3	≈ dem. representativo	convite por carta, mídia, 2019-21	7500 casa, 6500 utilizavel	dados: 78 / 44 / 3.5% modelo: 63 / 41 (GM) / 2.2%

AT-1: Friedmann (2005), doi:10.1097/01.hp.0000167228.18113.27

AT-2: Gruber et al. (2021), doi:10.1016/j.jenvrad.2021.106618

DE-1: Menzler et al. (2008), doi:10.1097/01.hp.0000309769.55126.03

DE-2: Petermann & Bossew (2021), <https://doi.org/10.1016/j.scitotenv.2021.1>

DE-3: Petermann et al. (2023), <https://arxiv.org/abs/2310.11143>

AT: Austria, 84,000 km²,

9 milhões habitantes, HDI 0.916

DE: Alemanha, 358,000 km², 84 mil.

hab., HDI 0.942

Brasil: 8.5 mil. km², 215 mil. hab.,

HDI 0.754

leção:

- Metodologia = fator muito sensível!
- Representatividade demográfica = muito difícil!
- Modelos = indispensáveis em várias etapas do procedimento.
- Importante: verificação, comparação com resultados de projetos / abordagens diferentes – QA/QC!

B.2

Tail end: Avaliação

Mapeamento

- **Interpolação** = inferência de informação puntual (pequenos pontos ou unidades discretas) para informação areal (pixels que cobrem uma certa área).
- Como?
 - Modelos que assumem continuidade da quantidade mapeada: geo-estatística, vários métodos.
Problema: modelo da continuidade.
 - Modelos de regressão contra preditores (geologia, permeabilidade, teor urânio, fatores climáticas,...). Cada vez mais por **machine learning**, que provou ser mais flexível e poderosa de que a regressão tradicional (incl. general linear model).
Problema: disponibilidade e resolução espacial de dados preditores; cobrança de anomalias locais.
 - Combinação – tipo regression kriging; ainda campo de pesquisa ativa!

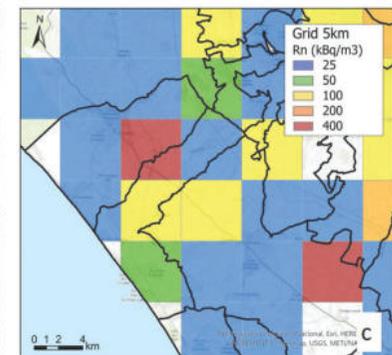
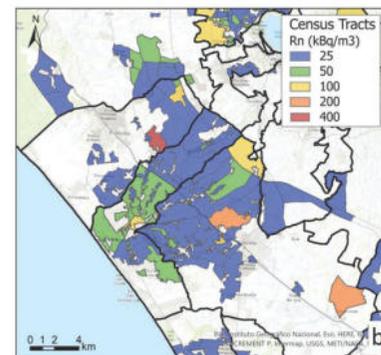
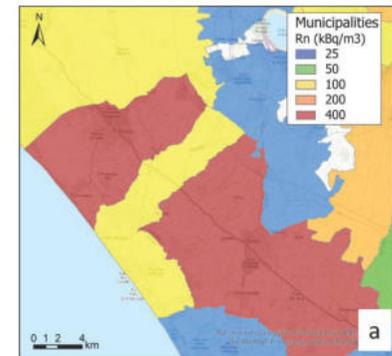
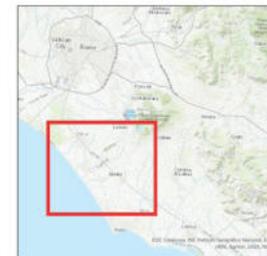
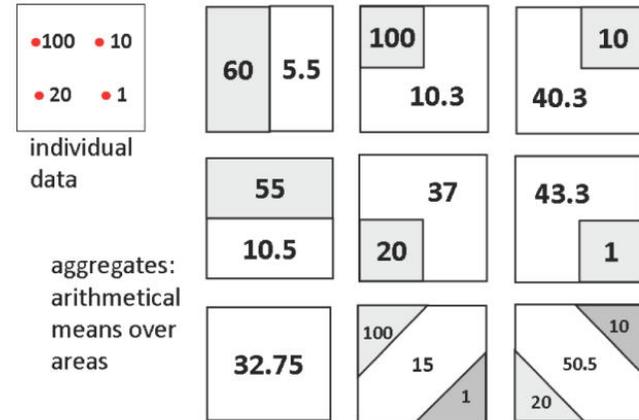
Efeito da agregação

Exemplo artificial:

área dividida em diferente maneira

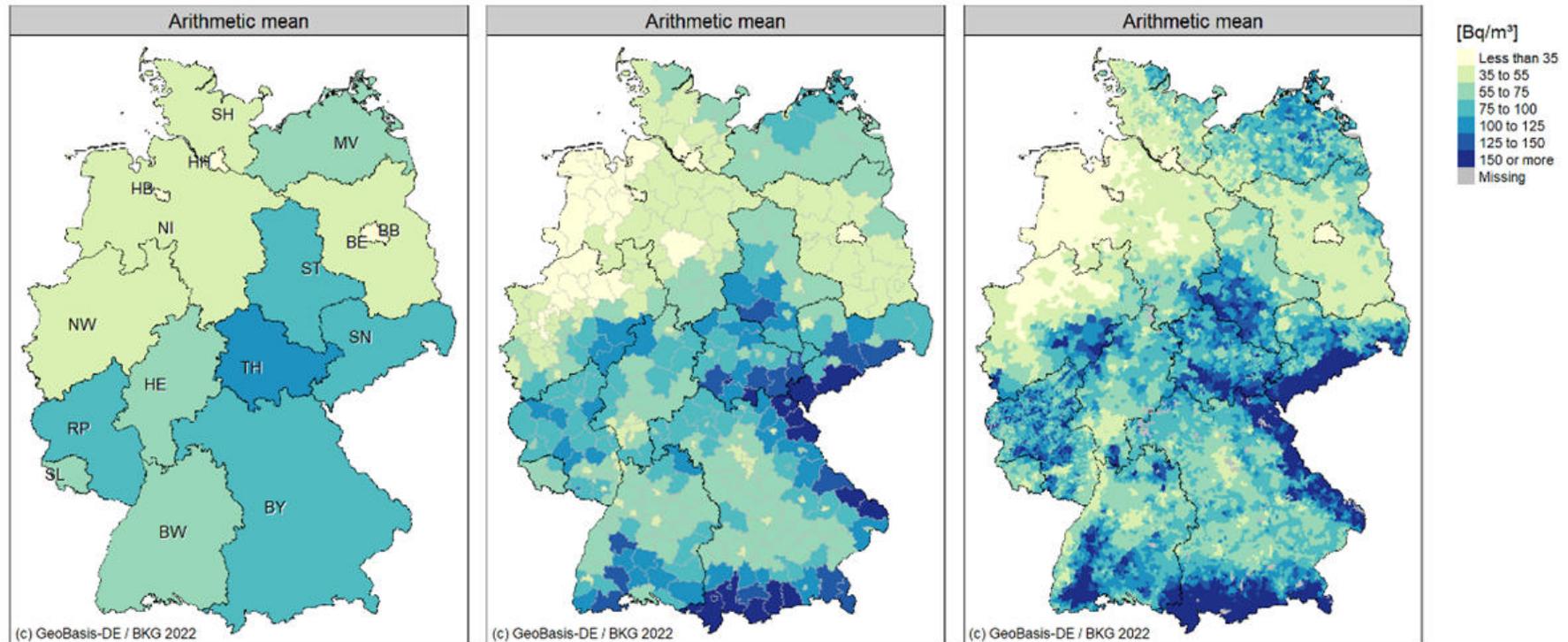
“Modifiable areal unit problem (MAUP)” – pode ser muito complicado!

Exemplo real de Lazio / Itália:



Fonte: Bossew, Čeliković, Cinelli et al.: On harmonization of radon maps. JERA 2022, 3: 7554 <http://dx.doi.org/10.35815/radon.v3.7554>

scale matters!



Rn indoor Alemanha: agregação por estados, distritos, municípios

Petermann et al. (2023): <https://arxiv.org/abs/2310.11143>

Incerteza

- Todos resultados tem incerteza
- Impacto nas decisões! → veja B.3
- Este aspecto deve ser levado em conta porque decisões erradas podem prejudicar a credibilidade; pode ter consequências jurídicas.
- Estratégia como tratar com situações que permitem apenas decisões incertas.

B.3.

QA & cadeia QA

Metrologia

Aspectos importantes

- QA/QC
- medição = parte da observação → cadeia de QA que cobre a sequência a partir do desenho → medição → avaliação
- uncertainty budget!! = muito mais que a incerteza da medição mesma!

QA/QC

- **Medição no sentido estrito:**
 - Valor reportado deve (em princípio) ser rastreável (traceable) a padrões primários (primary standards)
 - ⇒ Laboratório referencial certificado!
câmara de calibração
 - Intercalibration, intercomparison!
- **Obervação no sentido generalizado:**
 - desenho experimental (survey design) cuidadoso, bem documentado
 - avaliação cuidadoso, métodos apropriados para o objetivo

QA: intercalibration // intercomparison

- Intercalibration
 - Participants compare with calibrated standard (“true” values)
 - Statistics derived from participant’s results relative to the “true” values
- Intercomparison:
 - Participants compare between them
 - Statistics derived from participant’s results
 - From point of view of QA: weaker than intercalibration; but sometimes “true” value cannot be determined. (Typically: soil Rn)



Quality assurance

- Radon abatement policy consists of decisions about actions
- Decisions should be **reliable**, i.e. chance of erroneous decision low.
Important because
 - Rn policy can be expensive;
 - Credibility to stakeholders is important!
 - Must be legally proof
- Decisions rest on data, methods, maps,... therefore these must be QAed too. → QA chain
- Important: understanding of variability and uncertainty, because propagate into uncertainty of maps and of decisions!
Sources of variability and uncertainty are multiple and often difficult to identify and quantify... among the most difficult tasks.

The quality assurance chain

1. **Design QA:** Capability of an effort (experiment, survey) to enable a decision: sampling design;
2. **Data QA:**
 - a) classical metrology
 - b) protocols & actual performing an experiment;
3. **Evaluation QA:** proper methodology, adequate models, uncertainty budgeting;
4. **Decision QA:** Assessment of the chance of wrong decision, given data and methods and their uncertainties.

QA chain step 1: design

geral: desenho deve ser capaz de gerar amostra representativa!

- **Planning a survey:**
Experiment must be set up that a geographical or temporal tendency can be identified reliably, if it exists, while avoiding spurious tendencies.
- **Detecting differences and changes:**
Experiment must be set up such, that an effect of given size can be detected with set confidence, if it exists. Spurious effects shall be avoided with set confidence.

Power of a test = probability that an effect is detected, if it exists = $(1 - \text{prob})$ that an effect is missed, although it exists = $1 - 2\text{nd kind error probability}$.

In a rigid test or experiment, the desired power should be set beforehand and the test or experiment designed such, that this is fulfilled!

Design, 2

Implications for Rn surveying:

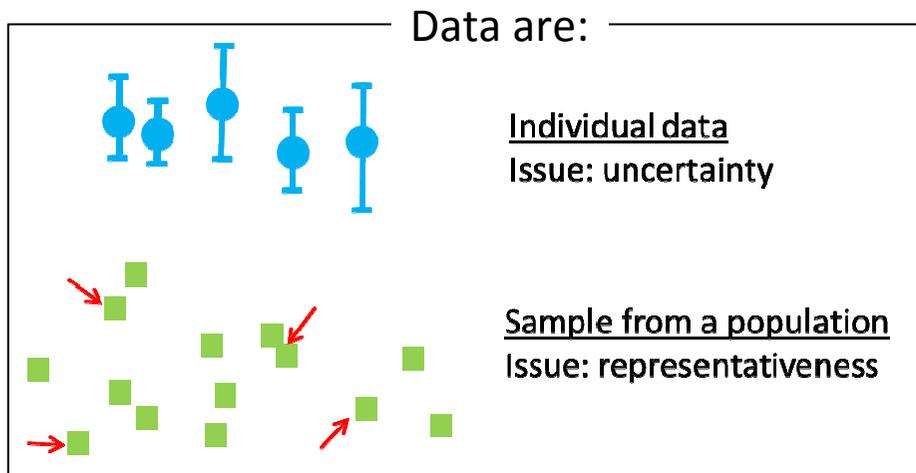
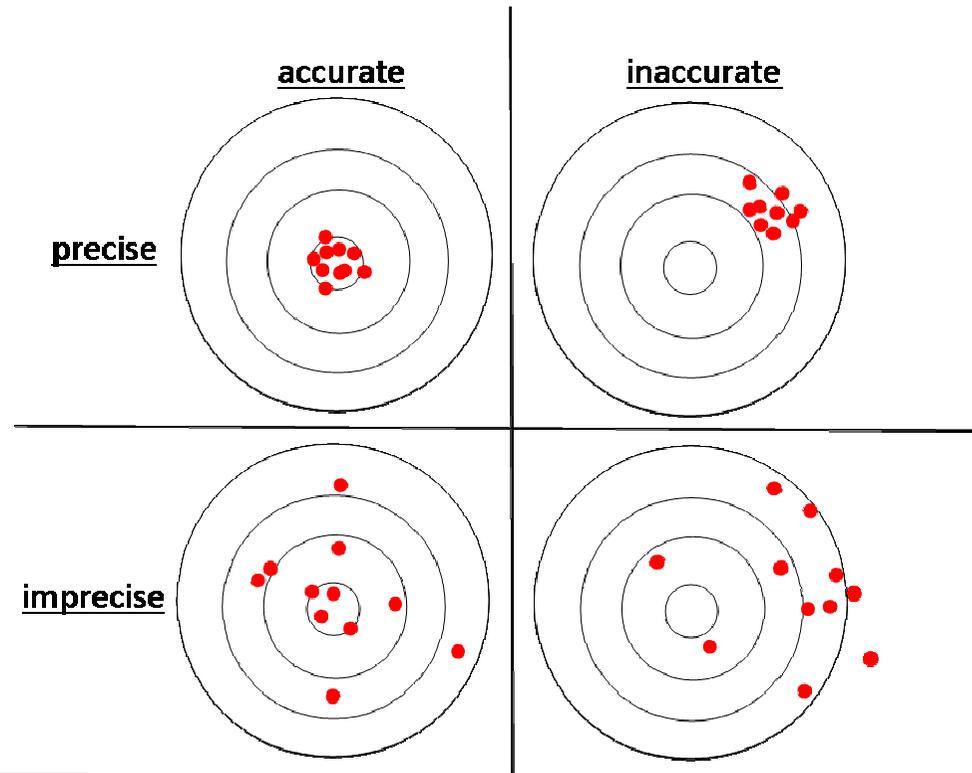
- Sample size: how many locations, how many detectors?
- Criteria for where and when to locate detectors?
- Representativeness of the sample!
(A sample is representative, if its statistical distribution = the true distribution of the sampled quantity.)
- Exposure time
- Metadata: which are needed for an objective?
- How to deal with constraints: limited budget, limited lab. resources, data protection

These are un-trivial questions!

**⇒ Careful discussions and planning to avoid sub-optimal results
(and waste of money and work)**

Design, 3: Precision and accuracy

- Precision: related to stochastic uncertainty; sources:
 - individual values: measurement statistics etc.
 - aggregates (AM,...): sample size
- Accuracy: related to bias or systematic uncertainty; sources:
 - wrong calibration
 - lack of representativeness of the sample



Sample size:
mainly a matter of costs
Sample representativeness:
this is the most complicated part!!

Design, 4: QC

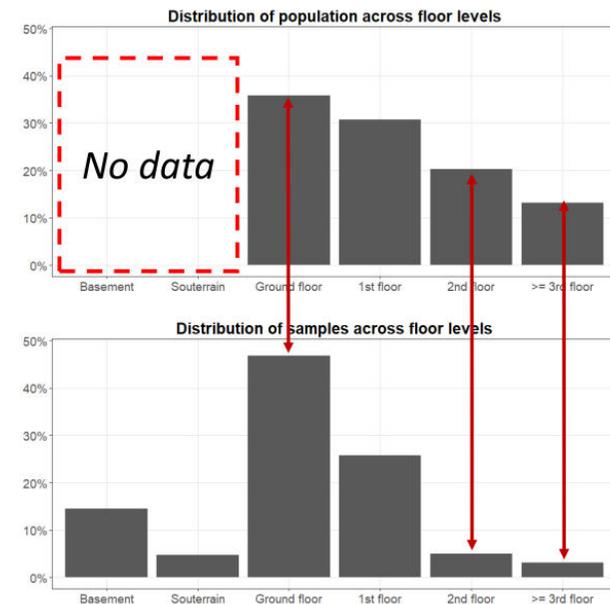
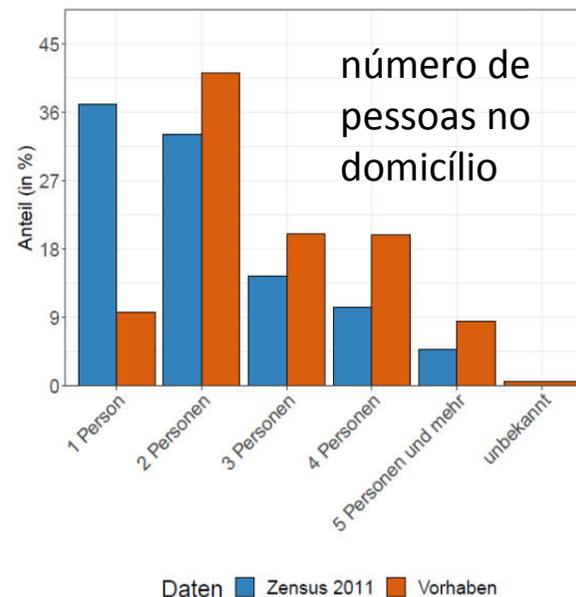
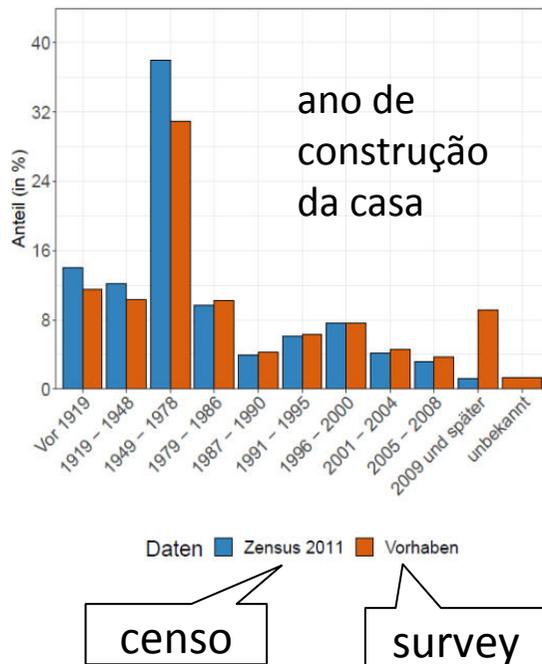
Muito importante:

- Verificação da representatividade do amostra!
- Por avaliação dos **metadados**: comparam-se estatísticas (media, median, distribuição,...) dos metadados da amostra e aqueles em bases de dados independentes.
- Ex. 1: caso amostra demograficamente representativa → distribuição geográfica da amostra deve coincidir com a distribuição demográfica.
- Ex. 2: Distribuição da idade de casas na amostra deve coincidir com aquela em registros.

Exemplos: Antignani et al., <http://dx.doi.org/10.1016/j.radmeas.2012.06.015> ; outro ex: proximo slide

Exemplo

- Estudo Petermann et al., mapa Rn indoor da Alemanha, baseado em amostra quase representativa (n = ca. 6500); <https://arxiv.org/abs/2310.11143>
- Comparação dos metadados:



fontes: Kemski & Gruber: Strahlenschutzgespräch Radon, Berlin, 2.11.21; Hoffmann et al., Apres. AIR, Berlin, 15.11.23

⇒ não é ótimo, apesar de grande esforço

Desenho não representativo

Type of sampling	explanation	resulting bias
Volunteer sampling	People who are interested in the subject or concerned about their health participate preferentially.	Sociological bias: more educated people are more likely to participate; preferential bias: houses in regions with higher Rn levels are probably over-represented.
Quota sampling	Certain categories, or types of people or environments, are sampled independently of their fraction relative to the total population	Inference to total population is difficult
Preferential sampling	Certain categories or regions are sampled with higher density, e.g. regions with known or assumed higher Rn levels, for detecting extremes	Participants from that region contribute more to the total mean, unless corrected by model
Pooled samples	Samples from different studies with different objectives are pooled	All kind of above biases, difficult to correct.

Type of bias	Cause
Non-response bias	Respondents and non-respondents are statistically different, e.g. socio-economically.
Coverage bias	The database from which samples are drawn does not include all possible samples, or individuals. For example an electoral register does not include residents without citizenship; telephone book: today fewer people have fixed tel.
Selection or proxy bias	The sampling criterion does not assign the same probability to all members of the targeted population. For example, if for the Rn survey samples are selected per dwelling or per address (as proxy to individuals), members of larger households (e.g. of families with more children) will be underrepresented. If this effect is in turn correlated to sociological factors and these on the other hand to building style and hence to Rn, the effect can cause a bias.

Biases due to shortcomings of otherwise intended representative sampling schemes

Types of shortcomings, what to do against?

just to repeat, because this is so important:

In environmental research, representative sampling is extremely important!

Often this is taken too lightly.

Shortcomings:	Bias (lack of accuracy)	Uncertainty (lack of precision)
What is it?	systematic deviation of the empirical from true mean	random deviation of the true from the empirical mean
Sources?	(1) samples not representative with respect to the variably distributed quantity	(1) true variability of the sampled quantity
(1) related to the phenomenon which is being sampled;		
(2) related to the sampling technique	(2) systematic measurement error e.g. due to calibration error	(2) Statistical measurement (counting) uncertainty
What to do against?	(1) representative sampling	(1) sufficient number of samples
	(2) proper QA	(2) appropriate measurement device

QA chain step 2: metrology

a) “traditional” QA

- Measured values must be traceable to primary standards
- Detectors must be appropriate to the objective
- Calibration
- Uncertainty budget of the measurement

b) Protocol, actual measurement

- Placing of detectors: choice of room, distance from wall, away from the cat which likes to play with it,...
- Correct assignment of metadata which are necessary for interpretation, usually through questionnaires
- Documentação da amostragem!

Importância do protocolo

- **O protocolo da observação define a quantidade, a qual pretende observar.**
- \Rightarrow Diferentes protocolos produzem, em geral, sistematicamente (i.e. além da incerteza da observação) valores diferentes p.ex. da quantidade nominal “Rn conc. no solo”, mesmo sobre condições objectivas idênticas.
- \Rightarrow É preciso investigar essas diferenças, para poder comparar valores de “Rn no solo”.
 \Rightarrow é preciso entender em detalhe os métodos.

Metadados

- Fatores que afetam a variável de medição e/ou informam sobre as circunstâncias da medição; caso Rn indoor:
 - Tipo & característica da casa, andar,...
 - geologia, teor U, tipo de solo, clima,...
- **Necessários para:**
 - **Interpretação** dos resultados
 - Input para **modelagem**: metadata são parâmetros de modelo
 - **Verificação** (comparar com os mesmos dados em bases de dados independentes)

QA chain step 3: evaluation

- **Correct models**
 - specification of models, e.g. is a functional dependence linear?
 - correct choice of parameters, e.g. variogram model in geostatistics, hyper-parameters in ML
- **Appropriate mapping**
 - reasonable spatial resolution
 - interpolation method
 - classed or continuous level scale
- **Uncertainty budget**
 - often very complicated, but at least sources of uncertainty and error should be identified;
e.g. uncertainty of choice of model; of parameterization of a model; estimation and prediction confidence intervals; ...
 - whenever possible, quantify
- **Documentação transparente do procedimento estatístico!**

QA chain step 4: decision

- From statistical perspective: **decision = classification**
- The **predictor space** is divided into classes according to the decisions taken for a value in that space
- Decision: assigning a “status variable” which denotes the class into which the decision result belongs
 - E.g., Decision about the RPA status of a municipality = yes / no / undecided
- The status variable has uncertainty, because it results from estimation
- How to quantify this uncertainty?
 - Probability of wrong decision, i.e. which is the probability that status has been estimated “yes” while in reality it is “no”?
 - Multinomial decisions: For ordinal classes, which is the probability that correct class has been estimated / error = ± 1 class / error = ± 2 classes / etc.

Decision QA

- Decisões com alvo de cumprir com regulação podem ter consequências económicas & políticas importantes.
- Por isso devem ser juridicamente seguros \Rightarrow todos elementos devem ser sujeito a QA \Rightarrow QA em todos passos cujo resultado é a decisão.
Por essa razão se fala da “**cadeia de QA**”.
- Parte integral de QA é avaliação da incerteza; para decisão (pode ser interpretado como ato de classificação) significa avaliação do risco de decisão errada.

Decision QA, 2

- Estratégia para reduzir Rn é baseada em decisões sobre implementação de ação.
- Exemplos:
 - Certo município deve ser declarado RPA?
 - É preciso fazer survey ou medições adicionais?
 - Um certo prédio deve ser recuperado?
- Decisão depende de
 - Conhecimento: dados, condições físicos, geológicos etc.;
 - Interesses dos Stakeholders
- Também decisões baseados em dados podem ser erradas!
 - Incerteza dos dados
 - Quantidade limitada de dados (pesquisa custa \$\$\$)
- First kind error (erro 1º tipo): Ação desnecessária, p.ex. município classificado RPA, baseado em informação disponível, embora não é.
- Second kind error (erro 2º tipo): Ação omitida erradamente, embora seja necessária. P.ex. município classificado non-RPA, embora é.
- Ambos erros podem sair caro economicamente, politicamente & juricamente.
- Podem prejudicar credibilidade & prontidão dos stakeholders para seguir as regras.
- ⇒ importante avaliar fidedignidade & credibilidade de decisões. Parece que metodologia ainda não existe.

Reliability of decisions

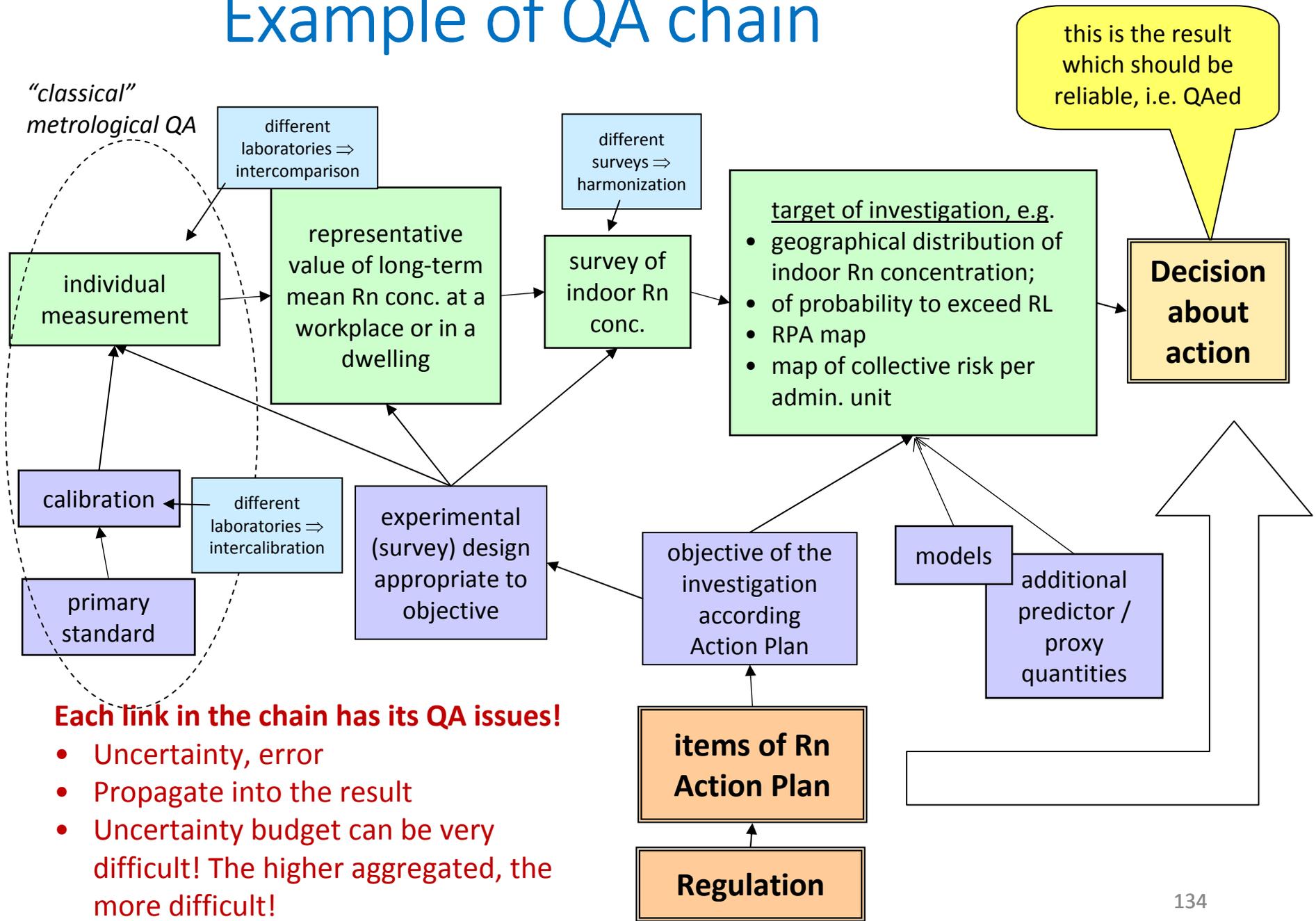
When is a decision reliable?

- A decision is reliable, if the chance of being a wrong decision, is below a preset tolerance.
- This amounts to misclassification:
 - Binary classification:
 - **First kind error** = false alarm: a phenomenon (effect, signal) has been estimated as present, although in reality it is not.
Ex.: a municipality is labelled RPA, although it is not.
 - **Second kind error** = falsely omitted alarm: A phenomenon has been estimated as not present, although in reality it is.
Ex.: A municipality has been labelled non-RPA, although in reality it is.
- **Given data**, upon which a decision (classification) relies: 1st and 2nd kind errors are **not independent!**
- Reversely: If tolerances for 1st and 2nd kind errors are **targeted independently**, the experiment (survey etc.) must be **designed appropriately from the beginning**, so that the target can be achieved! (Sample size, location of measurements)
- In practice, this can be a very difficult problem which can be solved only roughly. In any case, one should try to assess to 1st and 2nd error probabilities afterwards, to see to which degree a target has been achieved.
- The concept can be expanded to multinomial classification, but quite complicated!

QA de decisão

- Decisions to comply with regulation can have far-reaching economical and political consequences.
- They must therefore be legally proof, which means that their elements must be QAed. This implies QA of all steps whose result is a decision; therefore one speaks of the **QA chain**.
- Part of QA is estimation of uncertainty, which for decisions means assessment of the chance of ill-decision.

Example of QA chain



Each link in the chain has its QA issues!

- Uncertainty, error
- Propagate into the result
- Uncertainty budget can be very difficult! The higher aggregated, the more difficult!

Uncertainty budget

- Muitas vezes se considera apenas a incerteza da medição mesma
- Mas normalmente é fonte menos relevante de incerteza!
- Veja cadeia QA!
- Incertezas relacionadas ao desenho do survey (locação dos detectores, representatividade,...) : difícil quantificar!
- Incertezas relacionadas à avaliação
 - mapeamento: \pm difícil
 - decisão sobre ações: difícil!

Reproducibility, repeatability, interpretability

- Reproducibility: A result can be reproduced by other experimental or theoretical means
- Repeatability: Repeating the experiment with same device, same persons, leads to the same result
- Interpretability: Can a pattern or signal that is visible in the result be physically interpreted?
 - Does the observed effect point to a real physical effect or is it a statistical artefact?
 - Is the difference between repeated measurements due to change in true conditions or due to statistical fluctuation?
 - **Interpreting observations under presence of statistical noise can be challenging and requires experience!**

Challenges in implementing QA

- **Uncertainty budget**, especially related to
 - Survey design; impact of lack of representativeness
 - Semantic uncertainty (correct assignment of metadata, uncertainty of geological maps)
 - Model uncertainty
- Presence of **extremes and anomalies**
Generates troubles with geostatistical methods
- Sometimes wild spatial and temporal **variability**
Variability itself not constant over space and time
- Understanding **decisions** as functions of variables of different type that predict hazard or risk
- **Multivariate** estimation can be methodically challenging
- Including **“soft” data** as decision base
- Measuring the **reliability of decisions**, validating the efficiency of action based on decisions

Definitely much space for future research! Work will not run out!

C.1.

Fatores políticos, sociológicos

Stakeholders

- A comunicação com os stakeholders pode ser crucial para o sucesso do projeto!
- Comunicação diferente:
 - Especialistas de radiología
 - Profissionais:
 - health industry
 - building industry
 - Público em geral
 - Pessoas afetadas em casas com Rn alto & nas Rn Priority Areas
 - Midia
 - Administração, políticos
- Acho que faz sentido envolver especialistas de comunicação num programa deste.
- Não se esqueça que nos também somos stakeholders com interesses específicos!

alguns stakeholders:

quem ?	por que ?
público geral	afetado pelo Rn
setor da saúde	esclarecimento sobre Rn, lidar com as consequências
políticos	tem que decidir & pagar
administração	tem que implementar a pesquisa & possíveis medidas em consequência
mídia	tem que comunicar ao público
indústria da construção	implementar & efetuar tecnicamente normas de construção
especialistas Rn	tem bastante trabalho

→ *tem que incluir todos de modo apropriado !*

Da Silva & Bossew (2014): O projeto radônio do Brasil: Desafios particulares; II seminário sobre radônio no Brasil, Poços de Caldas, 19.-23.5.2014 (modificado)

comunicação em todas fases do projeto:

fase	assunto
preparatória	explicar o que é Rn, porque é importante etc.
antes da implementação	fatos básicos da técnica
implementação	porque escolher certas casas, como tratar os detetores etc.; talvez: resultados preliminares
depois da amostragem	resultados: interpretação
depois do próprio projeto	medidas de remediação & prevenção; o que fazer caso for encontrado Rn alto?

→ *são necessários especialistas na área de comunicação !*

Weighing between societal factors

- Criteria and arguments which determine decision about certain action in the framework of Rn Action Plans are not only Rn exposure, but also constraints from the economical or political sphere.
- Question: how to weigh conflicting stakeholder interests?
- Unlike physical quantities (Rn concentration etc.), political interests cannot be easily given numerical values.
- This does not belong to the sphere of physics, but to political science; nevertheless it is crucial and should be given more attention!
- Public discussion about it is nearly inexistent. Stakeholder interests assert themselves through political and economical power --- little satisfying from a democratic perspective!
- It is little probable that those who dominate the procedure would be inclined to change it.
- What is the role of natural scientists, Rn experts etc. in this situation? What could it be? What should it be? Which channels and possibilities exist to intervene?

Prevenção + remediação

Como reagir, se foram encontrados valores altos?

- aspecto específico de comunicação!
- às vezes complicado psicologicamente! → especialistas!!
- se devem elaborar respostas agora!

Consequências em regiões com nível alto:

- remediação custa \$\$\$\$
- prevenção: por leis de construção - muito menos caro!
- decisões políticas!

O verdadeiro trabalho duro começa depois do survey!

(Rn = "caixa de Pandora")

Citizen Science – Citizen Monitoring

- Citizen Monitoring: a discipline of Citizen Science (CS) - scientific research performed by citizens who are not professional scientists.
- Benefits of CS are
 - able to create large amounts of data, which is almost impossible for institutions due to necessarily limited resources
 - discovering and mapping natural or anthropogenic anomalies, not yet covered by professional measurement (**limited resources**)
 - educative: getting familiar with scientific methods
 - helping to reduce mistrust of the public towards official institutions (people can “verify” the official measurements)
 - participation increases motivation
- Problem consists in quality assurance (QA), because citizens are no trained metrologists

Literature:

https://en.wikipedia.org/wiki/Citizen_science;

Vohland K et al. (eds.) The Science of Citizen Science. 2021, DOI 10.1007/978-3-030-58278-4, www.springer.com/gp/book/9783030582777

Radon and Citizen Science

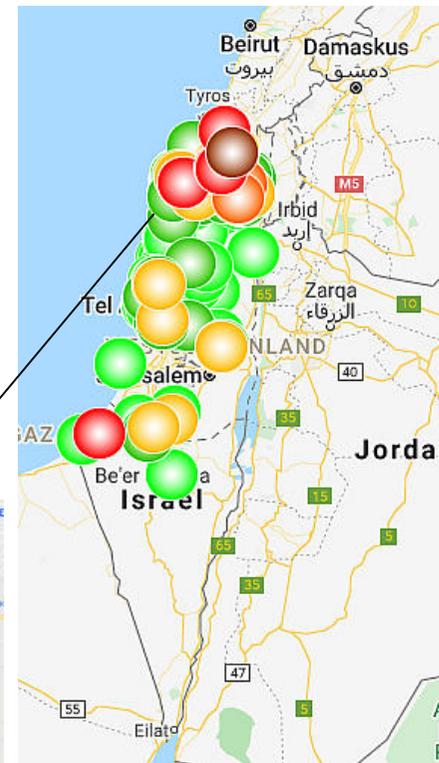
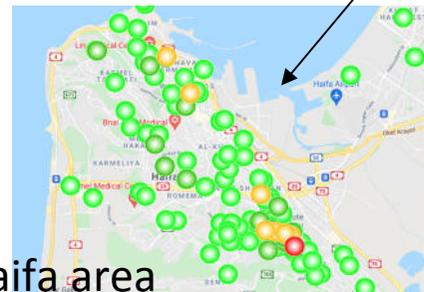
- Citizen Science (CS): scientific research conducted by citizens who are not professional scientists.
- Involvement can range between participating to different degree in defined projects and setting up entire projects.
- A great example: SAFecast (www.safecast.org).

- **Could the experience be used for Rn policy?**

Might provide more balanced weighing of stakeholder interests and lead to more transparent decision process.

- A pilot example is **RadonTest Online**, initiated in Israel, <https://radontest.online/>.

Mostly short-term scan measurements, but also long-term can be integrated. Individual data are protected.



Papel das organizações internacionais / regionais

Experiência na região Europa:

- O BSS da U.E. é motivação enorme para política de Rn, porque é uma “diretiva”, quer dizer, todos países devem transferir e implementar o BSS na legislação nacional.
- A U.E. subsidia as atividades por vários programas de pesquisa. Projetos como MetroRadon, RadonNorm, TraceRadon, LifeRespire, ...
- Associações como EURADOS, EMPIR etc.: plataformas de pesquisas pago pela U.E.
- A ERA (European Radon Association, <https://radoneurope.org/>) é eficiente em comunicar e coordenar atividades no nível mais informal. Parecido ARST nos EEUU.
- → Recentemente fundaram a AORA = Asian-Oceanic Radon Association.

D.1

Bases de dados existentes em Poços de Caldas

Preditores para construir modelo

- **Geologia**
 - muito bem estudado, mas é geologia “exótica” → generalização??
- **Taxa de dose**
 - cobertura muito fina
 - ao longo das ruas, estradas e rodovias
- **Geoquímica**
 - parcialmente muitos dados por causa da mina
 - quais elementos?
- **Rn indoor**
 - dados existem
 - metadata?
- **Rn no solo, Potencial Rn**
 - alguns dados
- **Aerogamma ?**
-

Exemplos →

Procedimento: “Data assimilation”

- Examinar qualidade dos dados
 - format;
 - missing values → possivelmente imputação;
 - erros: as vezes difícil identificar!
 - como tratar anomalias / outliers?
 - homogenizar: resolução dos mapas preditores; metodologia homogênea (muitas vezes: “top down” por modelos)
- Preparar para uso numérico na modelagem
- Por experiência: **muito trabalho!**

Harmonization bottom up // top down

bottom-up

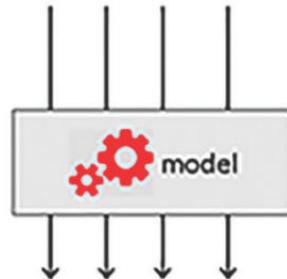
homogeneous results



same method

top-down

heterogeneous results by different methods



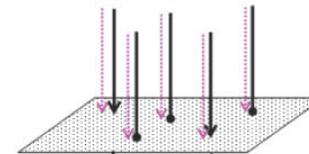
normalization through model

homogenized to common standard

Fig. 1. Concepts of bottom-up and top-down harmonization.

Filter

original data & metadata



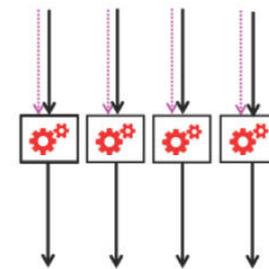
filter checks metadata

homogeneous / harmonized data

disadvantage:
fewer data

Standardize

original data & metadata



model transforms data according metadata

homogeneous / harmonized data

disadvantage:
model uncertainty in addition

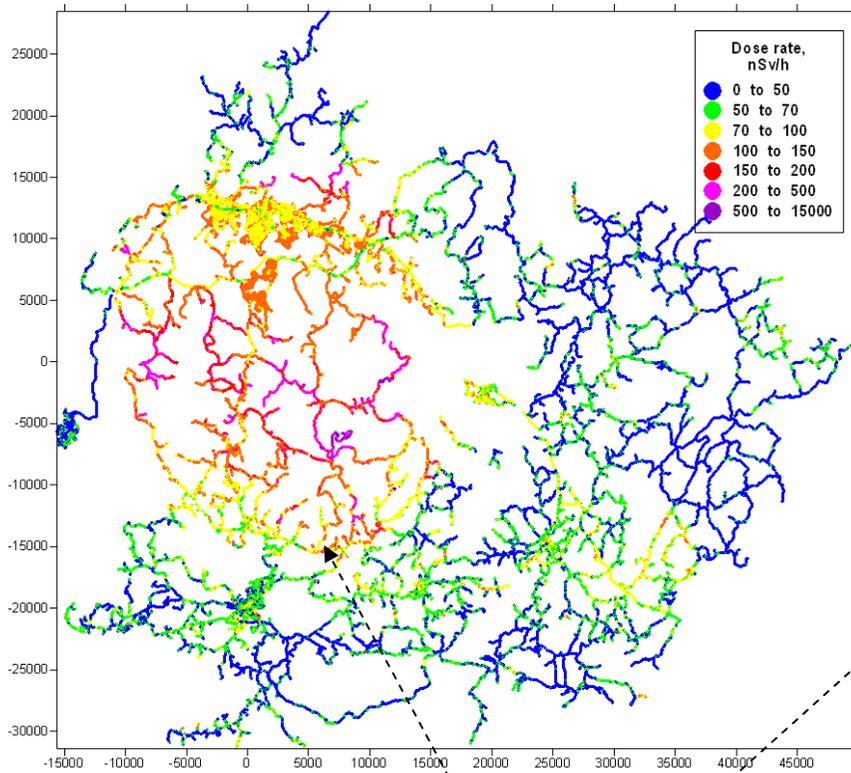
Fig. 2. Top-down harmonization by filtering and modeling.

1) Taxa de dose ambiental: metodologia

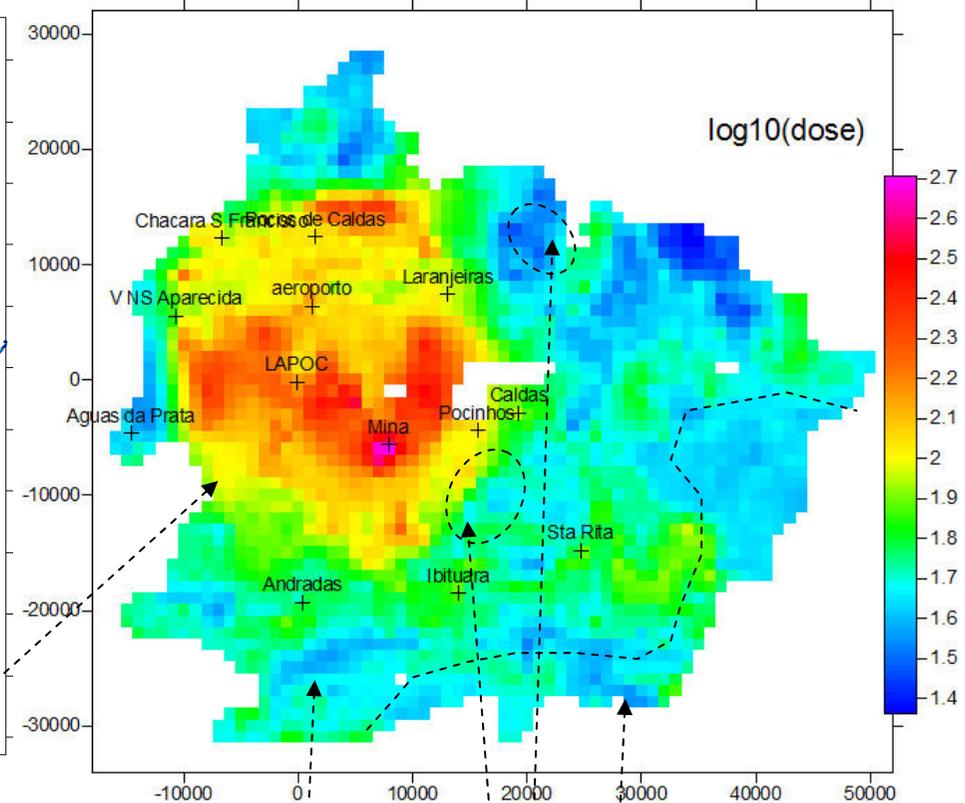
- Quantity: ambient equivalent rate, nSv/h
- Method: car-borne, explained in detail: Projeto Planalto Pocos de Caldas, Vol 1 (2009), p. 37 f.
- cosmic contribution and internal BG (ca. 10 nSv/h) not subtracted
- 522,000 values

Bossew et al. (2014): Pesquisa radiométrica Maciço Poços de Caldas: relações taxa de dose com geologia; resultados preliminares; II seminário sobre radônio no Brasil, Poços de Caldas, 19.-23.5.2014

Taxa de dose: padrão regional



Plateau Maciço Poços de Caldas



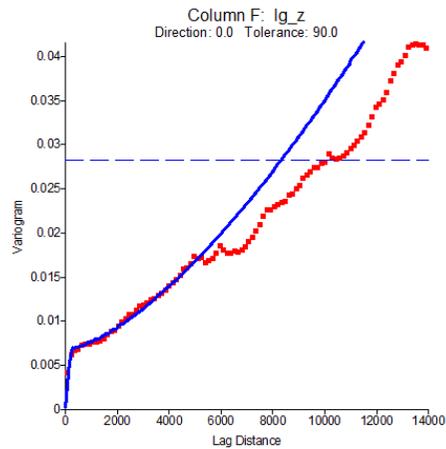
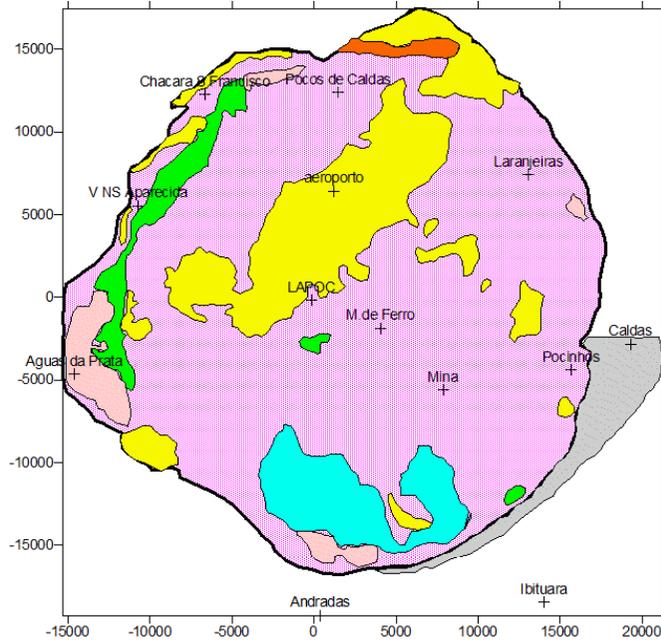
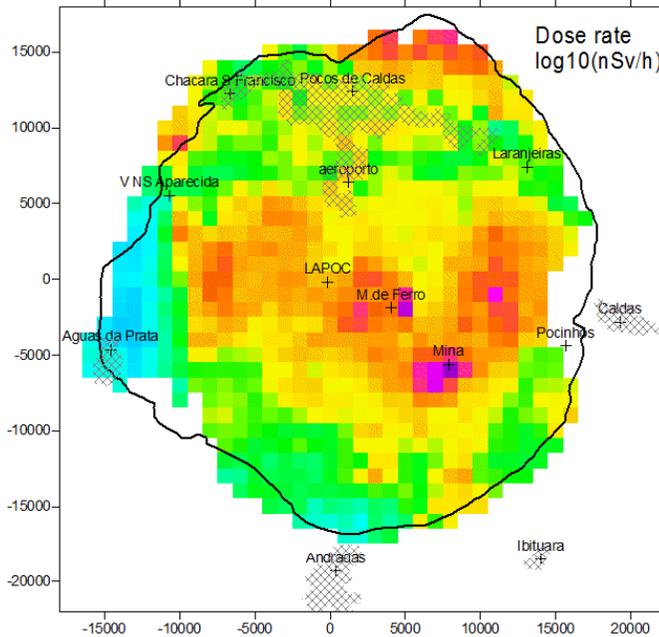
complexo Varginha

grupo Pinhal

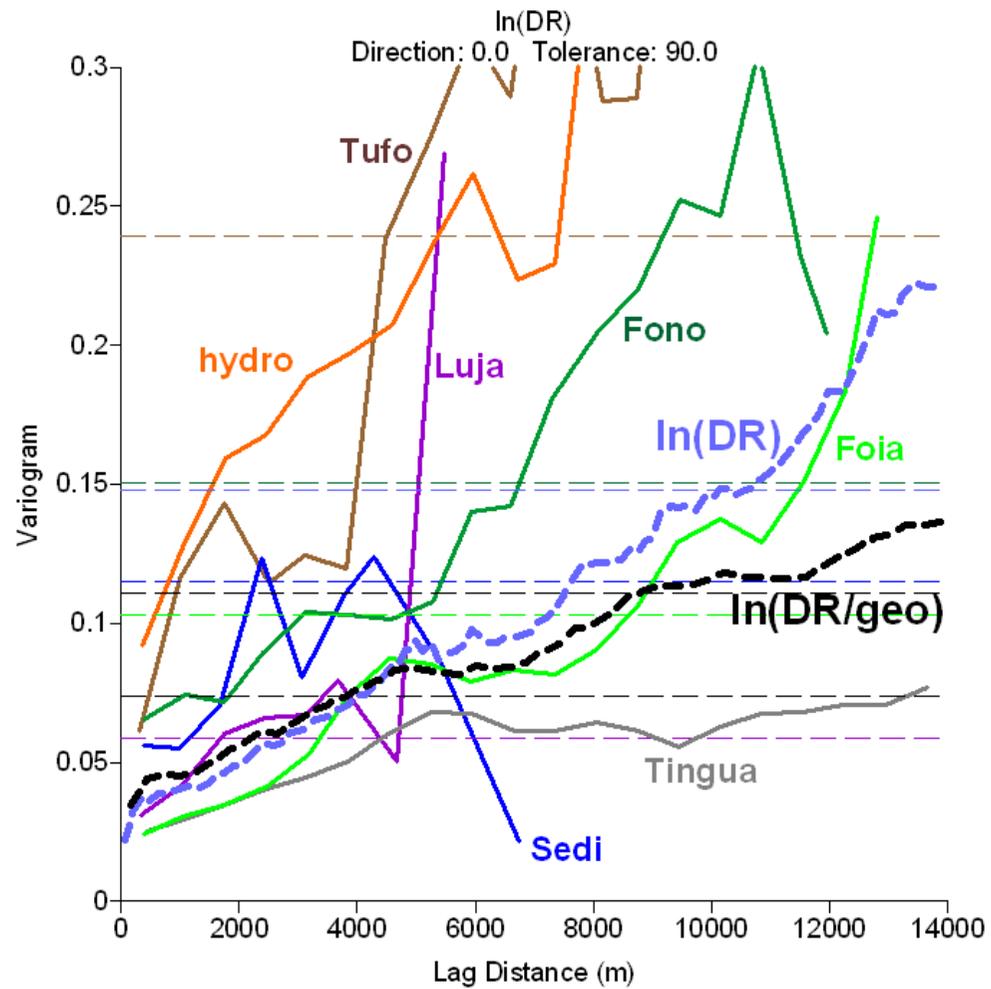
corpos sieníticos

Interpolação // geologia

resolução 1 km

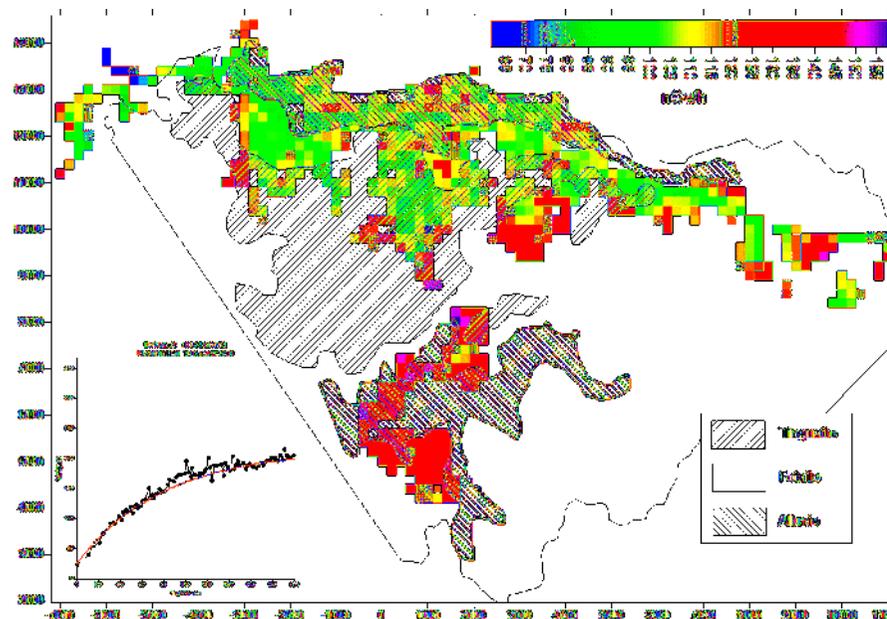
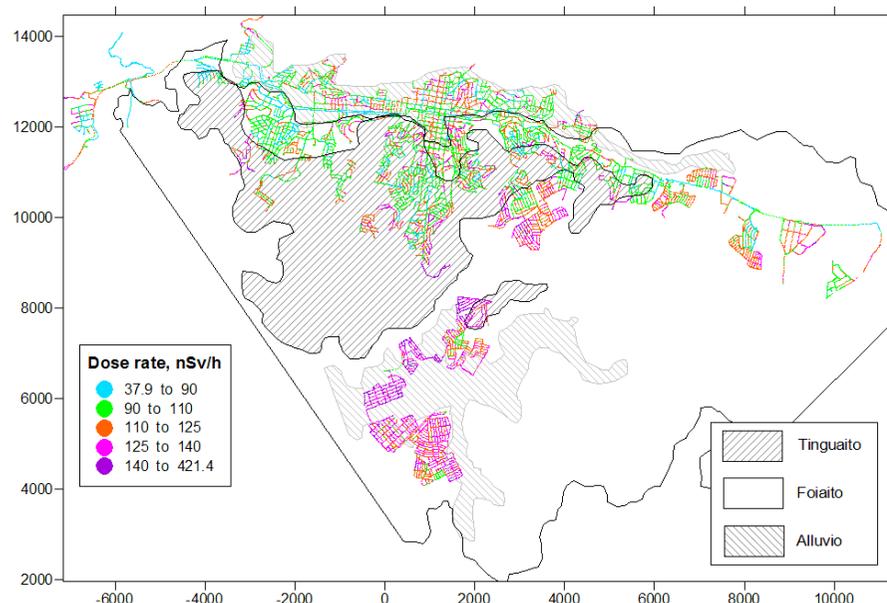


- Tinguaité
- Foioite
- Lujaurite - Khibinite
- Lava
- Volcanic
- Clastic sediments
- Fenite



variogramas por geologia:
 estrutura muito heterogénea
 → interpolação difícil!

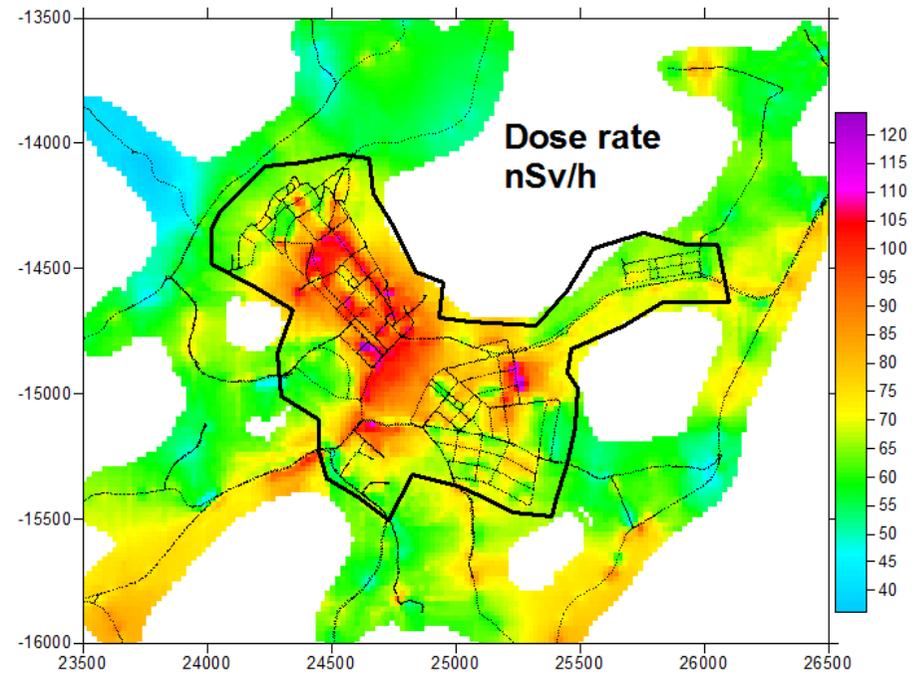
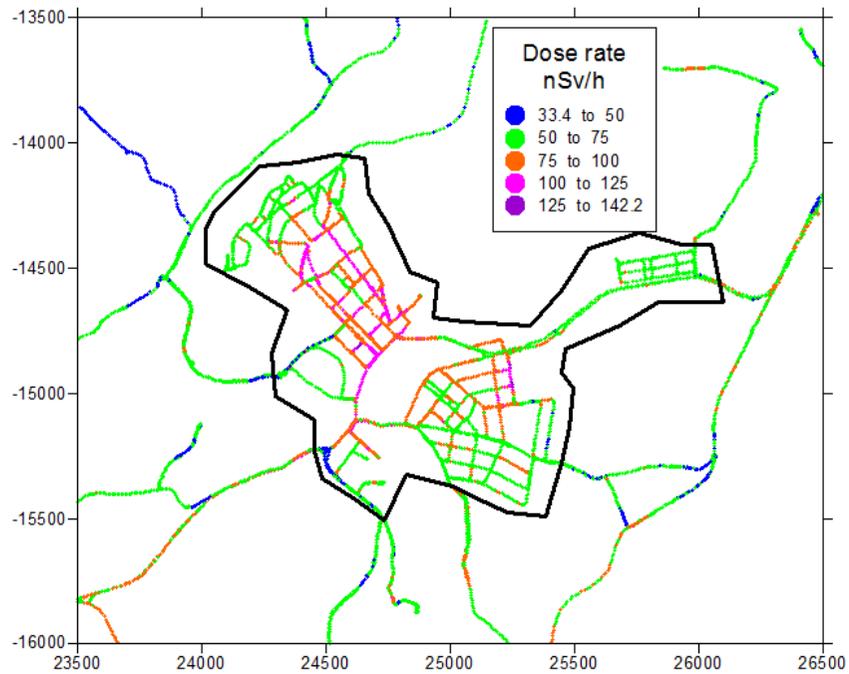
Mapeamento local



Parte urbana da municipalidade Poços de Caldas

- não existe correlação evidente da dose com geologia – pela classificação geológica acessível;
- a parte S da cidade parece com nível de dose mais alto de que o resto.

Geologia: Mapa geológico-geotécnico do município de Poços de Caldas, UNESP et al. 2008



Exemplo: Santa Rita

block OK:

AM \pm SD = 75 \pm 12 nSv/h

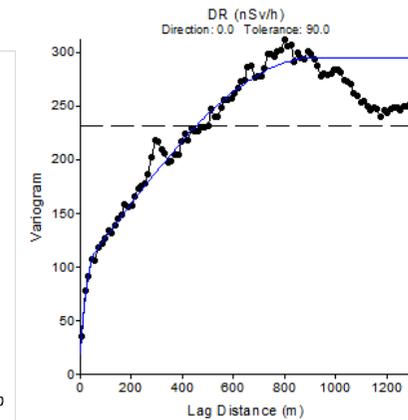
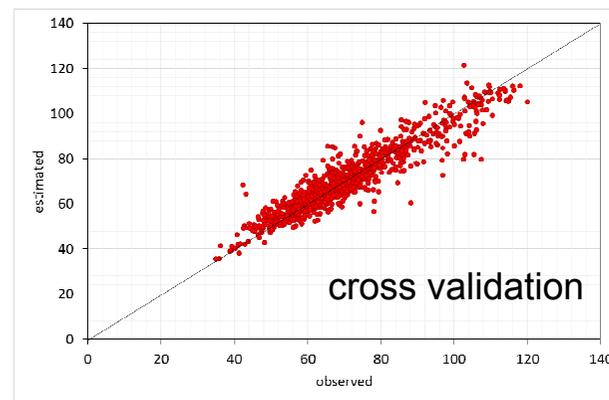
Med \pm Mad = 71 \pm 7 nSv/h

Min = 48, Max = 124

dados:

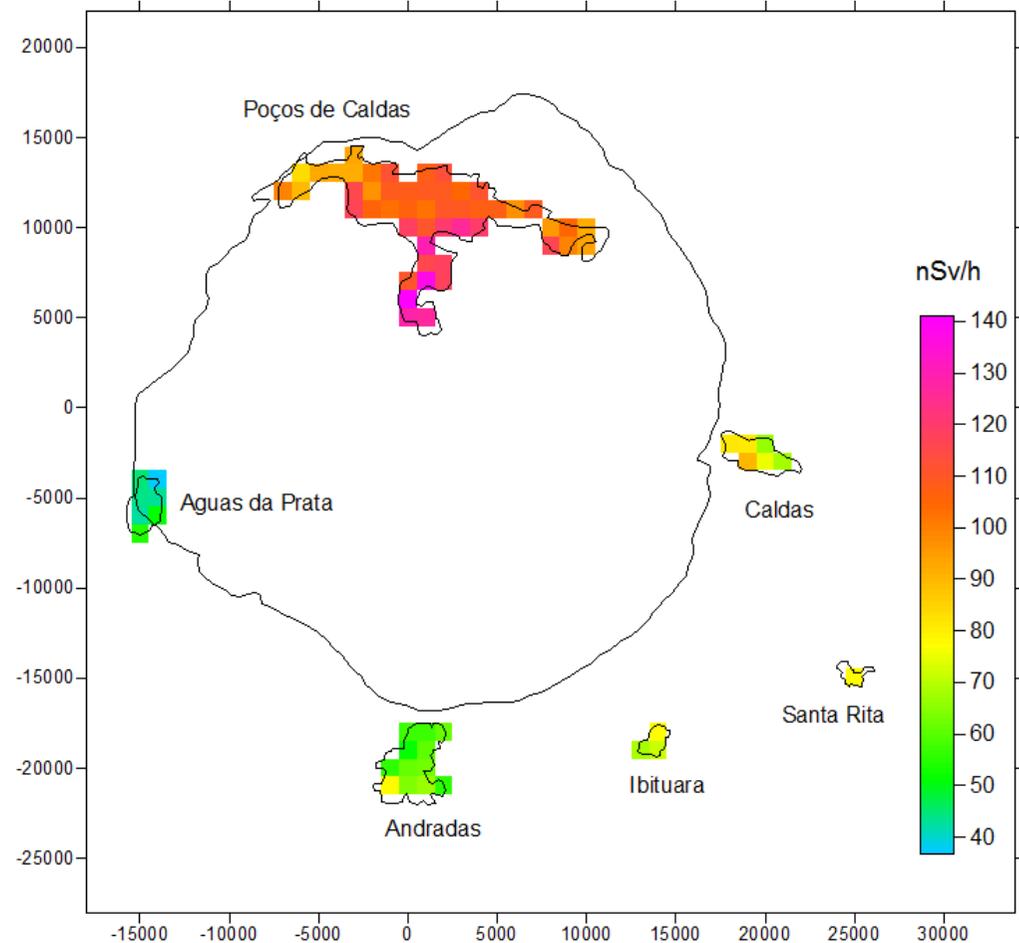
AM = 74 \pm 16

Min = 33, Max = 142



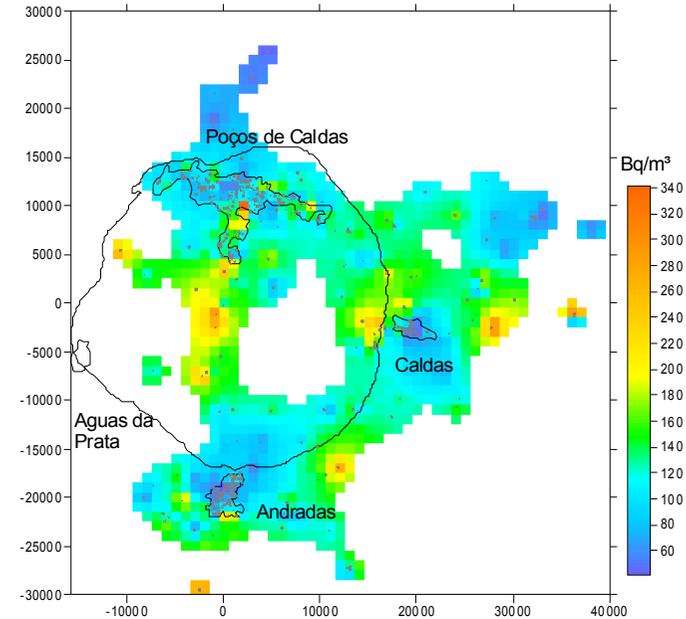
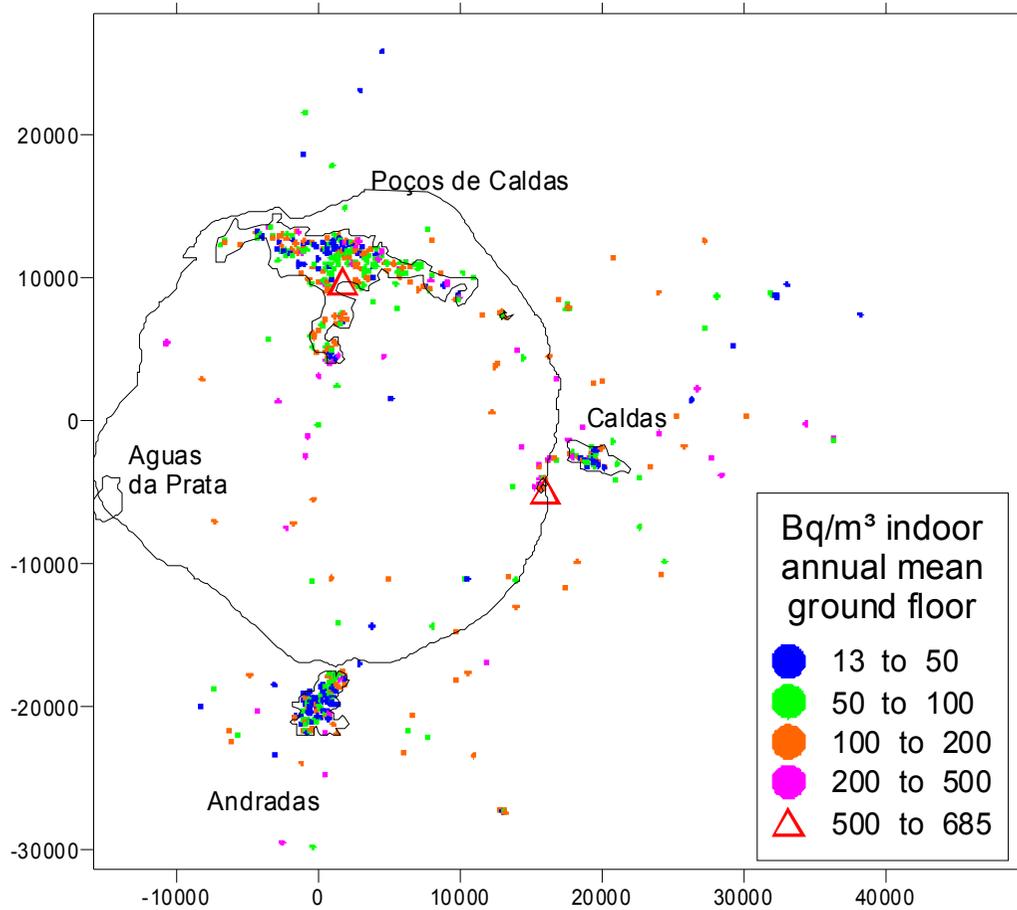
Valores médios nas áreas urbanas

cidade	AM ± SD (modelo)	dados	Projeto Planalto
Poços de Caldas Parte S (aeroporto)	108 ± 14 125 ± 14	106 ± 19 129 ± 17	118
Caldas	77 ± 8	80 ± 16	80
Andradas	61 ± 7	61 ± 16	67
Aguas da Prata	45 ± 4	47 ± 11	
Ibituara	72 ± 4	80 ± 23	63
Santa Rita	75 ± 12	74 ± 16	55
NS Aparecida	41	99 ± 11	
Laranjeiras	89	92 ± 11	
LAPOC	168	135 ± 29	
UNIFAL	47		
Pocinhos	106	82 ± 7	
Povoado X (BR)	122	126 ± 20	
Santana	51	68 ± 10	
S Pedro	42	45 ± 5	
Piau	45	57 ± 9	
S Bento	40	43 ± 11	
Campestrino	47	57 ± 15	

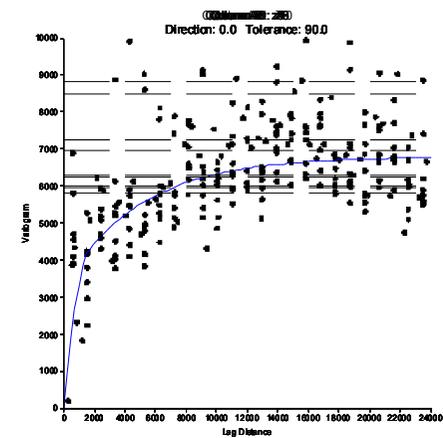


AM(cell) \approx GM * exp(SD_k²/2)
Tabela: AM_(area) [AM(cell)]

2) Rn indoor

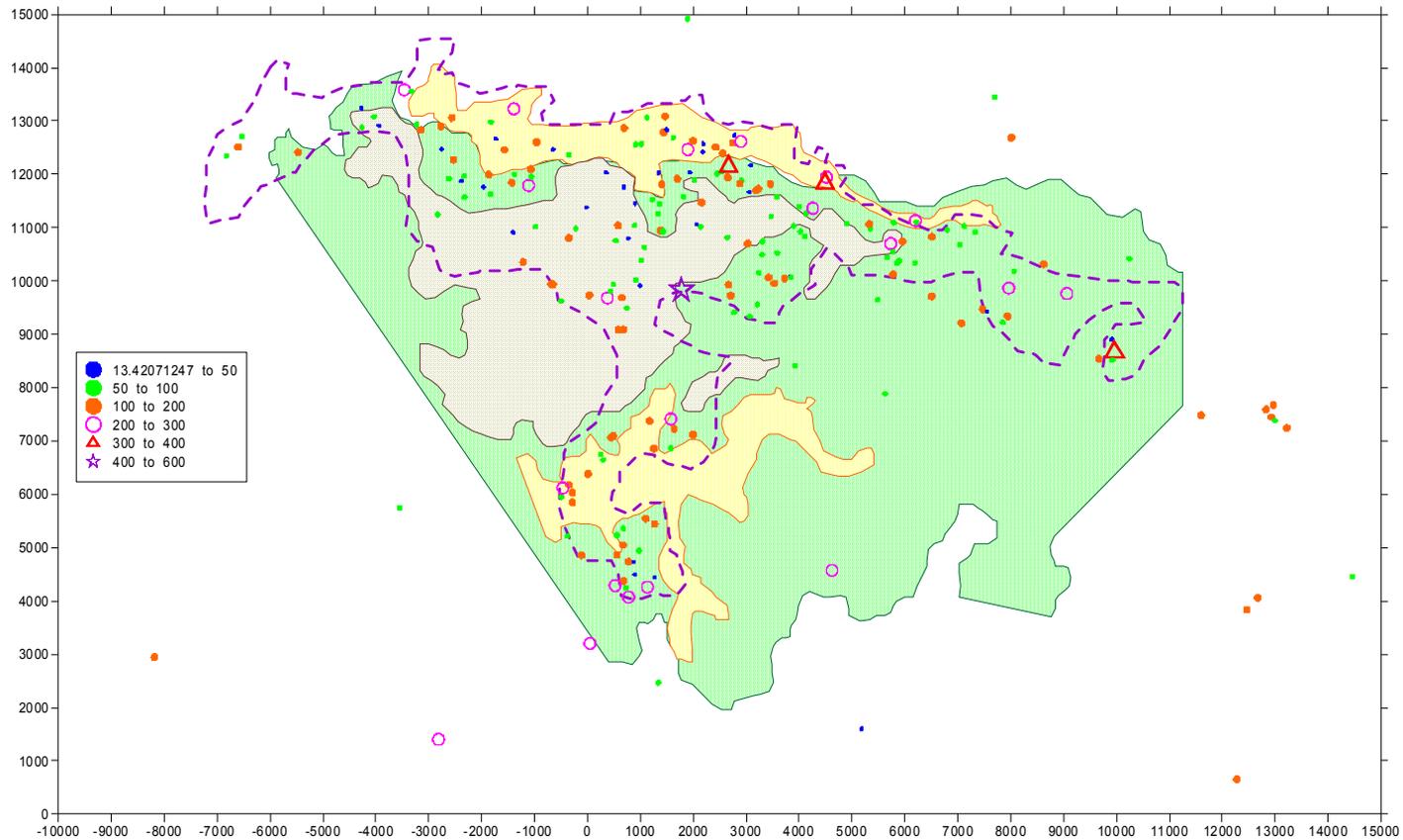


10 realiz of grid=2 km declust



Interpolação OK baseado em dados 'declustered'

Poços de Caldas:



correlação Rn indoor – taxa de dose ambiental: muito fraca de acordo com o conhecimento atual!

D.3. Upscaling

Upscaling

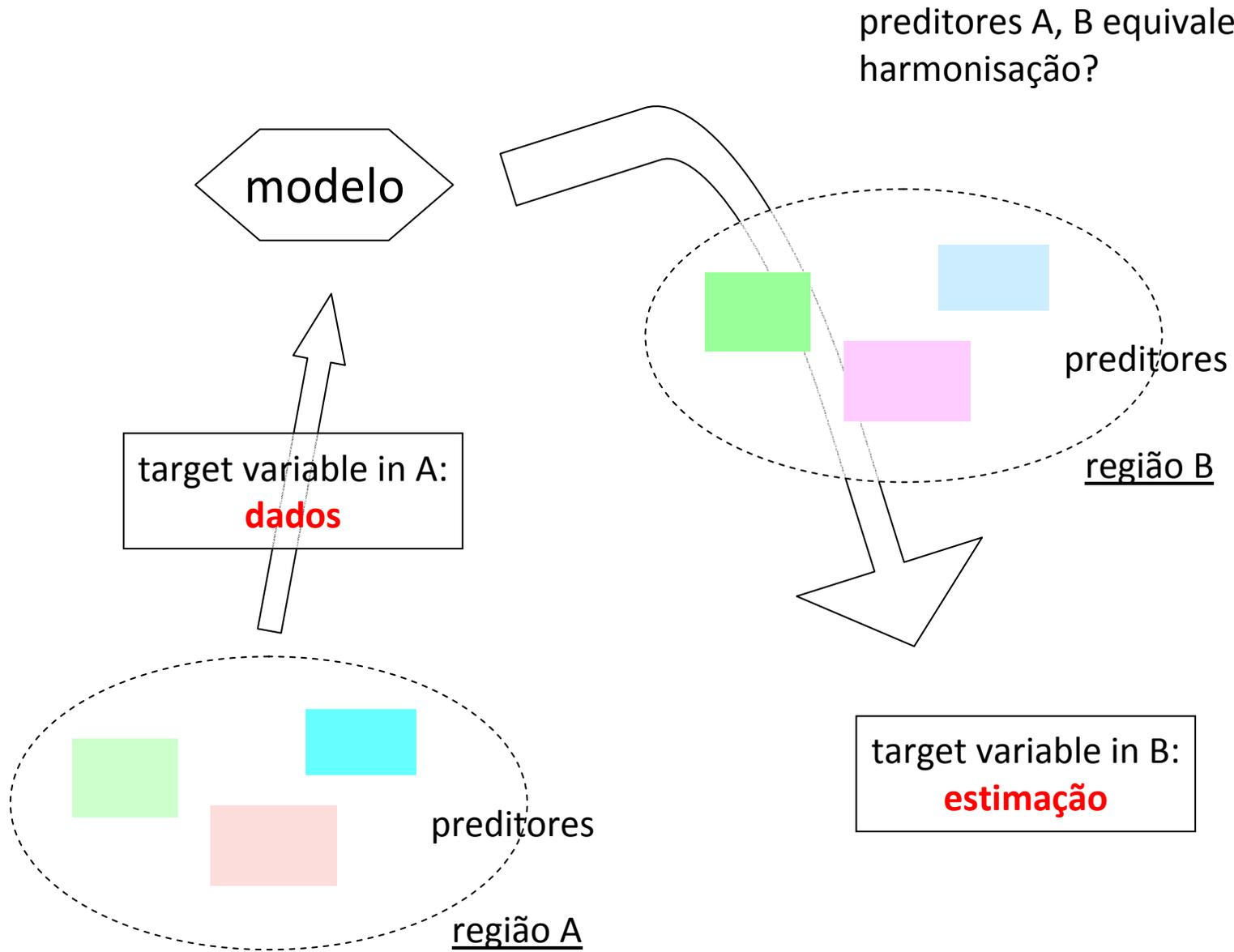
- Definição: “expanding, adapting and sustaining successful policies, programmes or projects in different places and over time to reach a greater number of people”*
- Usar valores de quantidades na escala local para estimá-los na escala regional.
- No nosso caso específico:
Será possível utilizar a experiência de Poços de Caldas (onde tem muitos dados em alta resolução espacial) para estimar Rn em escala regional – nacional?

* from a World Bank document, cited in Kern et al. 2023, <https://doi.org/10.1007/s10113-022-02020-z> .
More definitions in Guentchev et al. 2023, <https://doi.org/10.1016/j.cliser.2023.100352>

Upscaling, 2

- Método possível: por analogia
- Estabelecer modelo explicativo na escala local A (Poços de Caldas): $Rn(x) = f(\text{preditores}(x))$, $x = \text{ponto ou área pequena}$.
- Aplicar o modelo na escala regional B; condição: devem ser disponíveis os mesmos preditores na mesma escala, p.ex. mapa geológico em mesma escala. Suposição: as unidades geológicas na região B tem as mesmas propriedades que em A, quanto a Rn .
- Preditores não disponíveis na mesma escala, p.ex. mapa geológica em B = menos fino de que aquela em A: agregação de unidades em A; recalibrar o modelo.
- Pode ser complicado quando existem vários preditores em escalas diferentes; tem a ver com o MAUP.
Uncertainty budget: desafio!

construção do modelo



aplicação do modelo

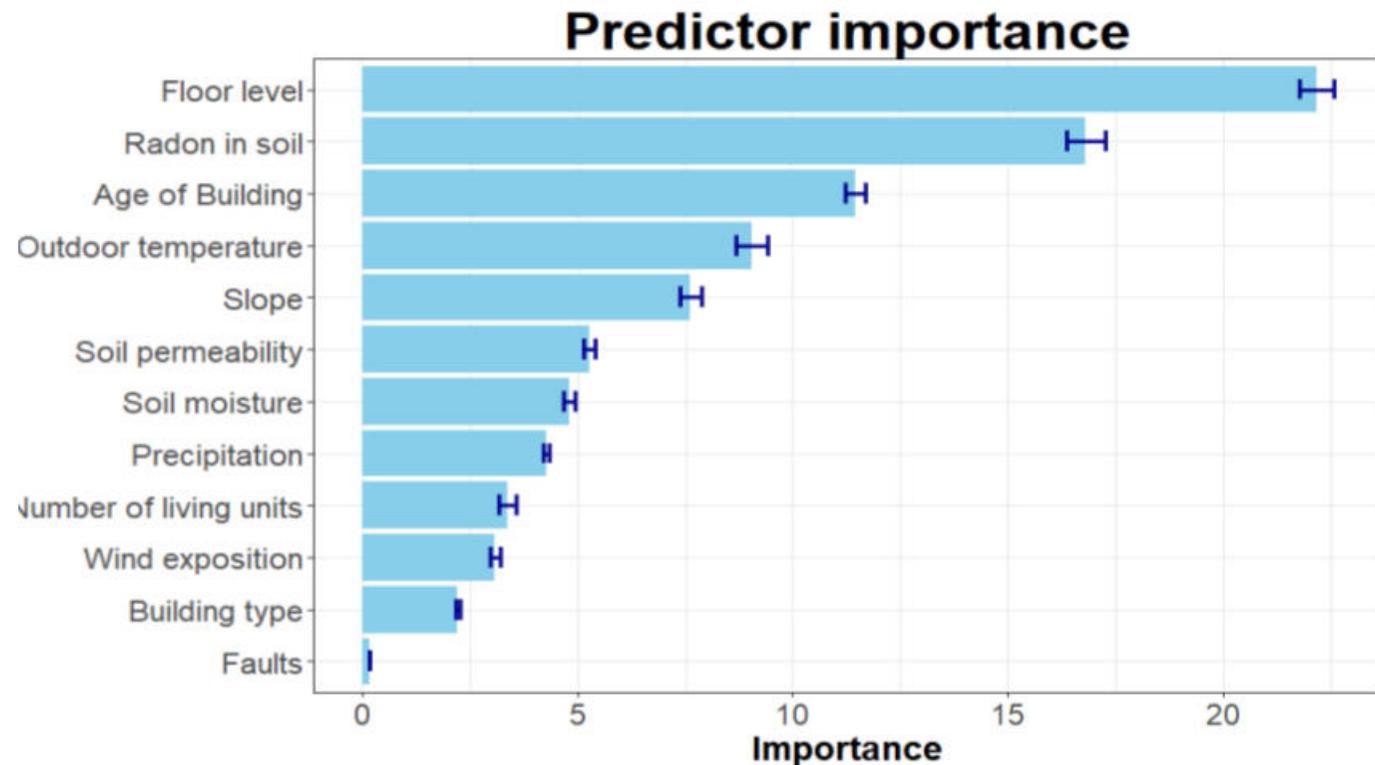
Upscaling, 3

- Quais bases de dados existem em Poços de Caldas?
- Quais bases de dados existem na escala nacional?
- Preditores candidatas:
 - Geologia
 - Pedologia, permeabilidade
 - Geoquímica
 - Soil cover, tipo de paisagem
 - Clima (p.ex. temperatura média), zona climática
 - Tipo dominante de casa
 - Fatores socio-económicos
 -
- Na área de calibração do modelo (A = Poços de Caldas), deve ter variabilidade alta dos preditores, senão não têm poder explicativo.

Importância dos preditores

- Experiencia da Alemanha – possivelmente diferente em outras regiões!
- Mapa Rn indoor por ML, Petermann et al. 2023 (<https://arxiv.org/abs/2310.11143>):

“VIP plot”



Muito obrigado!

Acknowledgement:

Agradeço as correções do Português pela Dra. Ana Paula Fleury de Macedo Soares, São Paulo.