Relationship between *Safecast* ambient dose rate and indoor radon data

Bossew P.[1], Cinelli G.[2], Petermann E.[3], Kuča P.[4], Helebrant J.[4]

1 Privatier, Vienna, Austria

2 ENEA, Palermo, Italy

3 Bfs, Berlin, Germany

4 SÚRO, Prague, Czech Republic

v.17.9.23

16 th INTERNATIONAL WORKSHOP

GARRM

(on the GEOLOGICAL ASPECTS OF RADON RISK MAPPING)

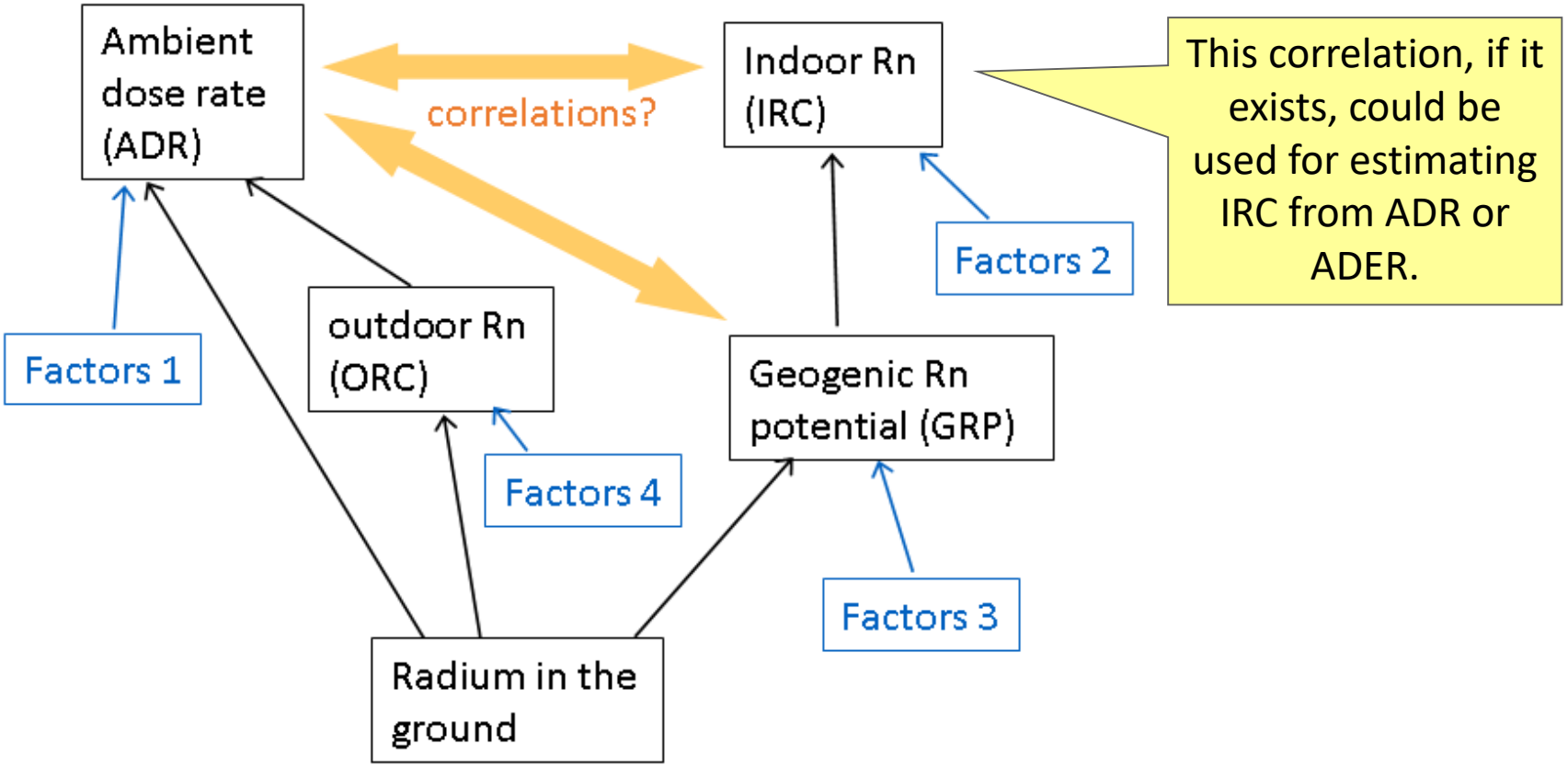
September 19 th – 21 st , 2023, Prague, Czech Republic

Research question

- Large amount of ambient dose equivalent rate (ADER) data generated by the *Safecast* project available;
- Can the data be used for predicting quantities that are relevant in radon abatement policy, such as
  - regional mean indoor Rn concentration (IRC)?
  - regional probability that IRC exceeds a reference level (RL)?
  - the status of an area as Rn priority area (RPA*)?

* Areas in which Rn abatement (prevention, mitigation, remediation) is considered necessary due to elevated Rn levels.

Rationale



Various “nuisance” factors blur the correlation: cosmic dose rate, outdoor Rn, other sources of terrestrial dose rate (^{40}K, Th series), factors which control ingression of Rn into buildings.

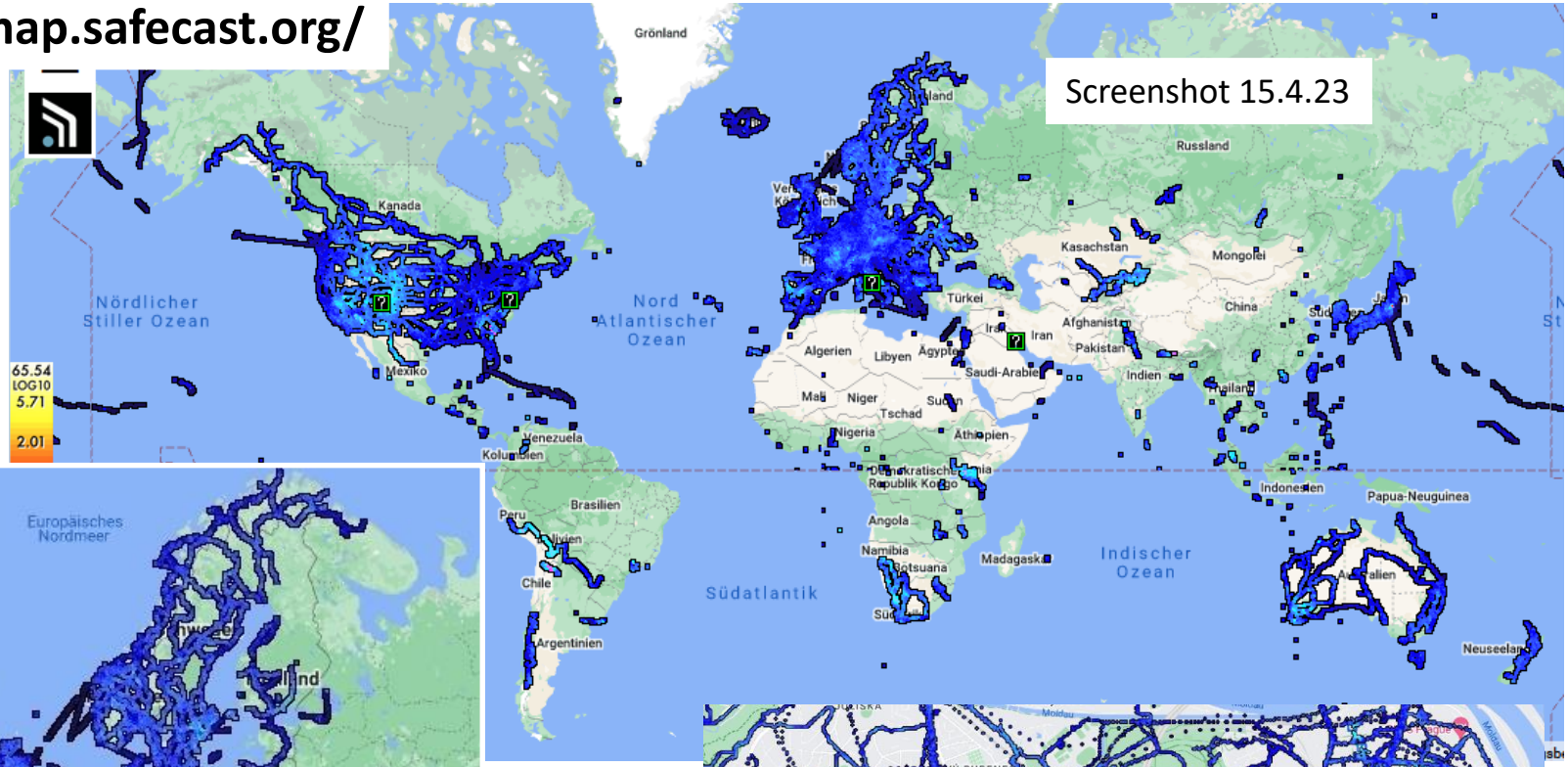
Safecast

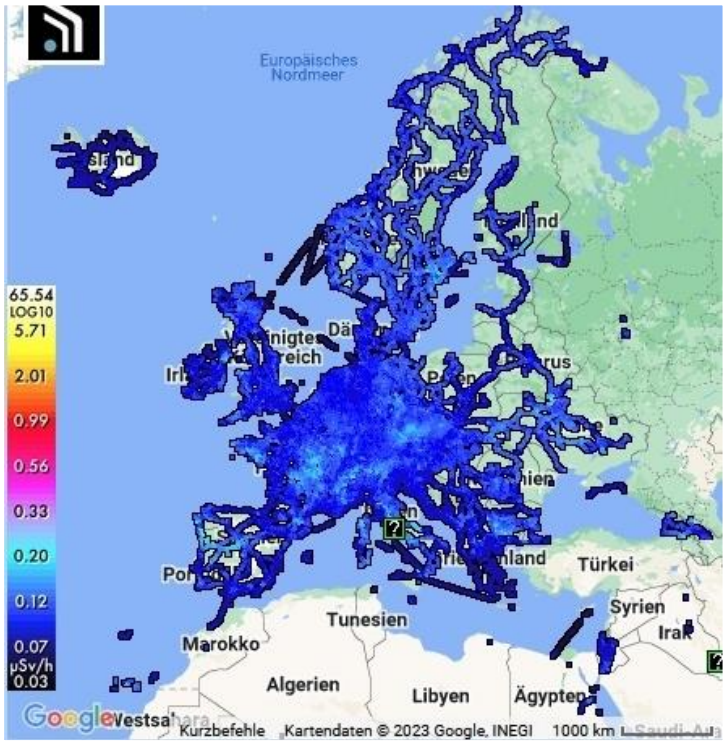
- Citizen Science project, founded in Japan after the Fukushima accident, 2011; quickly expanded world-wide;
- A monitor called "*bGeigie Nano*" used, several thousand units carried by volunteers for collecting ADER* data;
- Data can be sent to the *Safecast* team, who projects it on a publicly accessible map;
- By early 2023: about 200 mill. data in the database, about 50 mill. in Europe. Data can be downloaded.

* Ambient dose equivalent rate H*(10), nSv/h

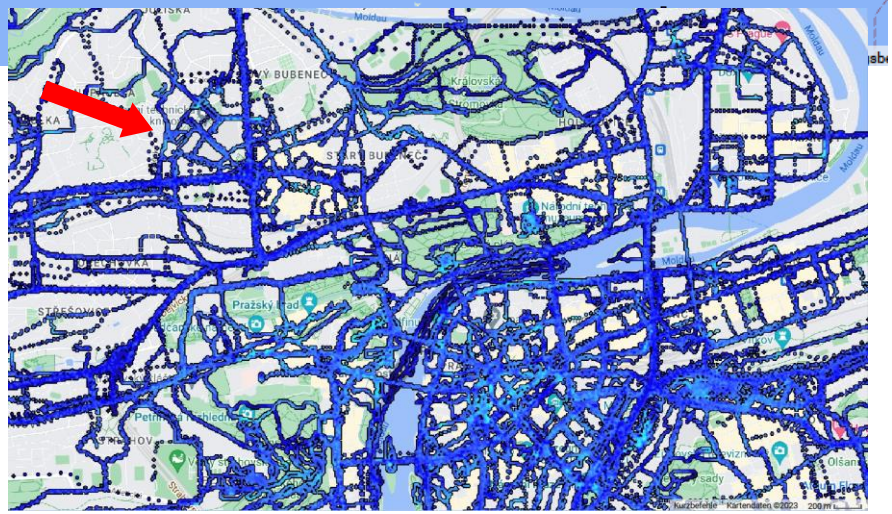
Safecast map

https://map.safecast.org/





Prague



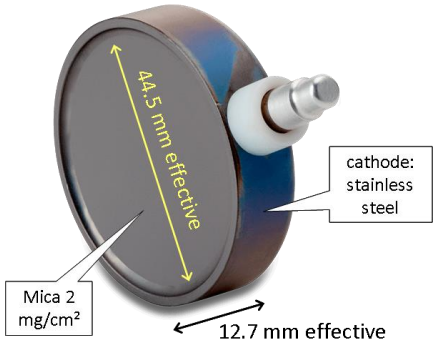
bGeigie Nano

bGeigie Nano: pancake type GM detector, coupled to GPS.

Measured quantity: number of impulses per 5 s. Count rate converted to ADER (nSv/h) with calibration factor.

Data saved on SD card in txt format: count rate, geographical coordinates and altitude, date+time, and other.

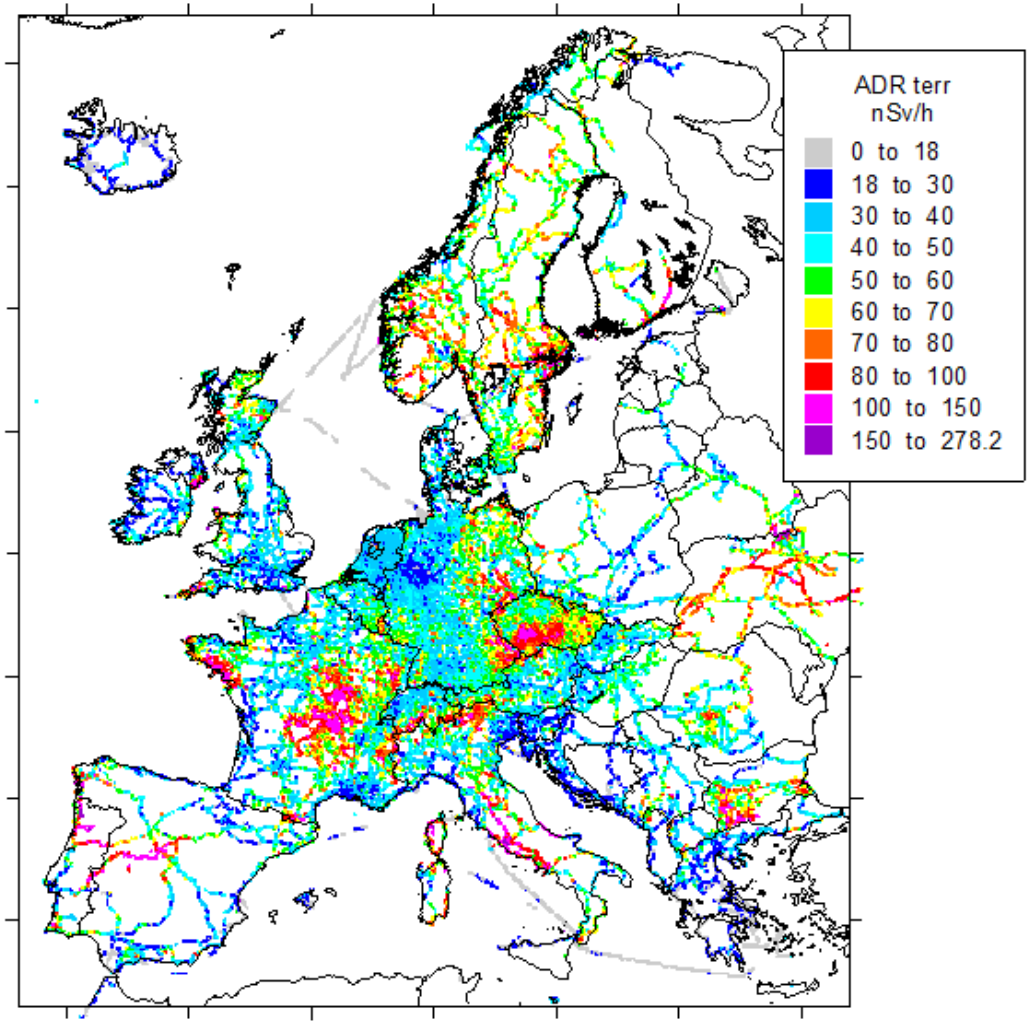
Data can be sent to Safecast and displayed on QGis using an add-in developed by SÚRO. The Safecast team projects the data on the map after plausibility check. Data available for free download.



Thin window → α, β counting possible in principle – but discouraged for practical use as it is almost impossible to interpret under field conditions.

Data 1: *Safecast*

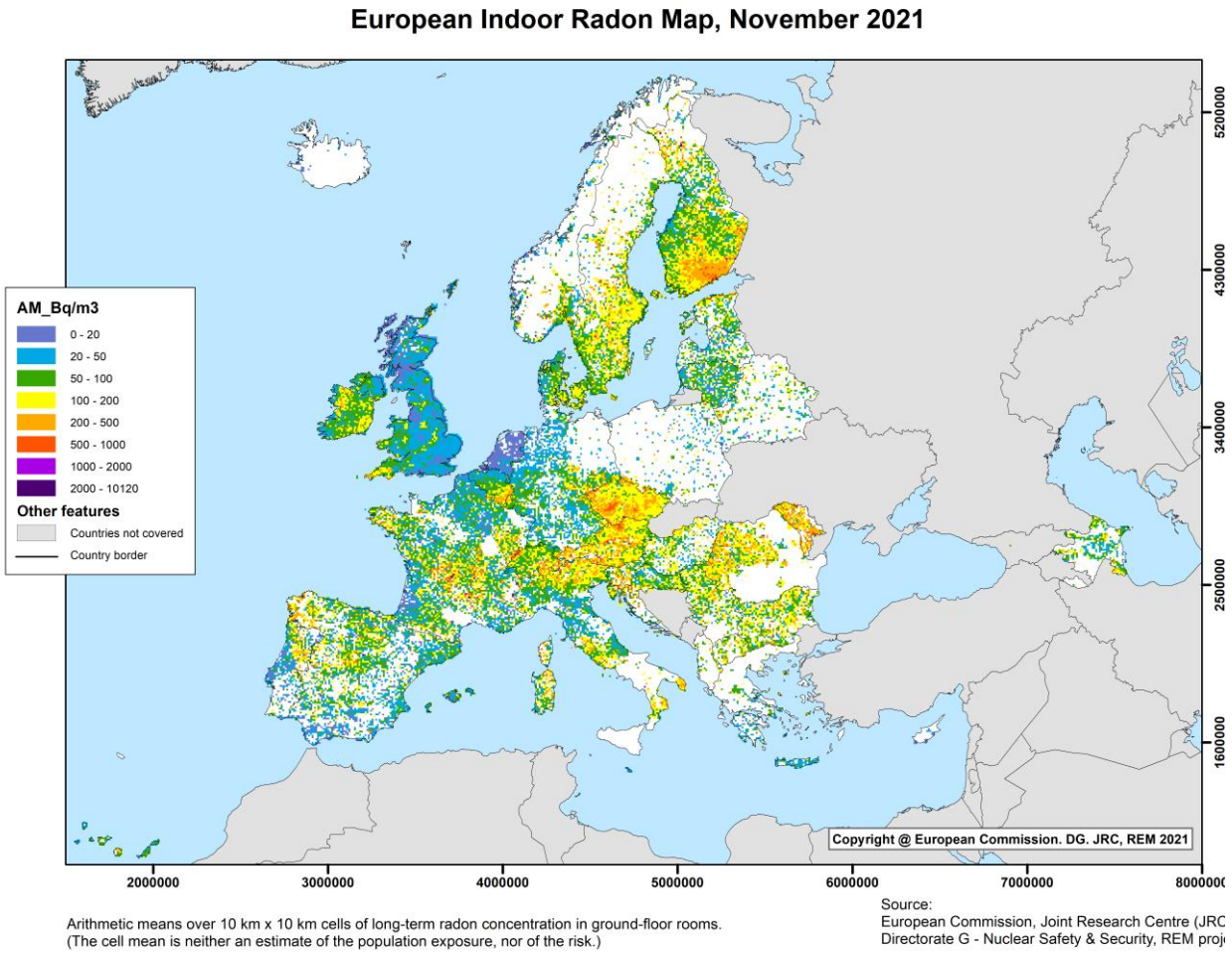
- n=52,887,234; relevant fields: coordinates, ADER (1 minute mean, µSv/h).
- Processing:
  - internal background removed (10 nSv/h);
  - cosmic dose rate subtracted, calculated from altitude, taken from DEM → terrestrial ADER,
  - conversion of geographical into European Lambert coordinates,
  - aggregation into European grid, 21,828 cells 10 km × 10 km



Data 2: Indoor radon IRC

European indoor radon database, from the European Atlas of Natural Radiation (2019) [1]:

- about 1.2 mill. measurements, ground floor rooms, aggregated into 10 km × 10 km cells.
- Statistics: AM, SD*, AM and SD of ln-transformed data, min, median, max, N (data per cell). n=29,539 cells
- Exceedance probability prob(IRC>RL)* can be calculated under log-normal assumption in cells.



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

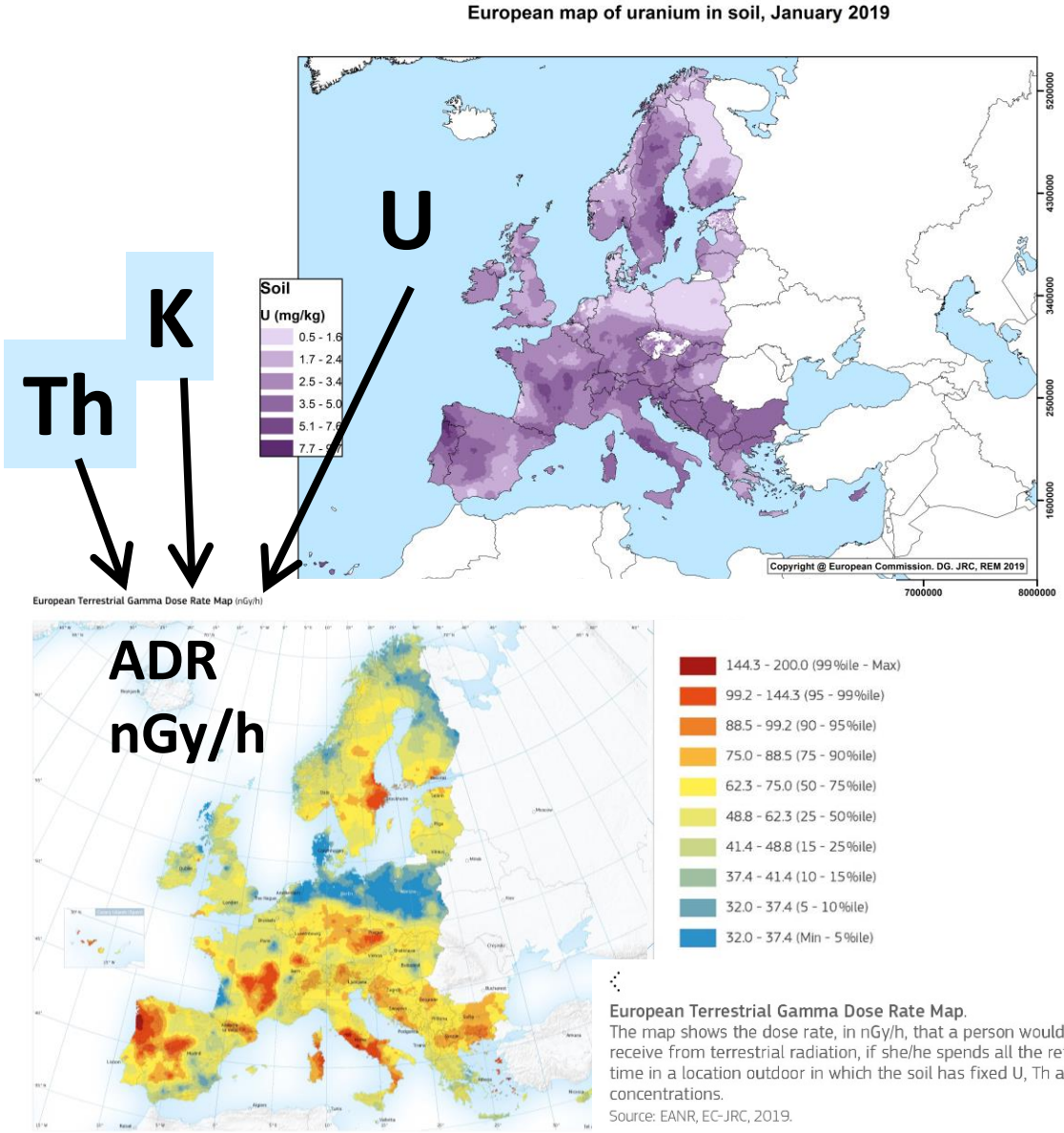
* AM, SD – arithmetical mean, standard deviation; RL – reference level

[1] https://remon.jrc.ec.europa.eu/About/Atlas-of-Natural-Radiation

Data 3, Geochemistry → ADR

Geochemical database from the Atlas:

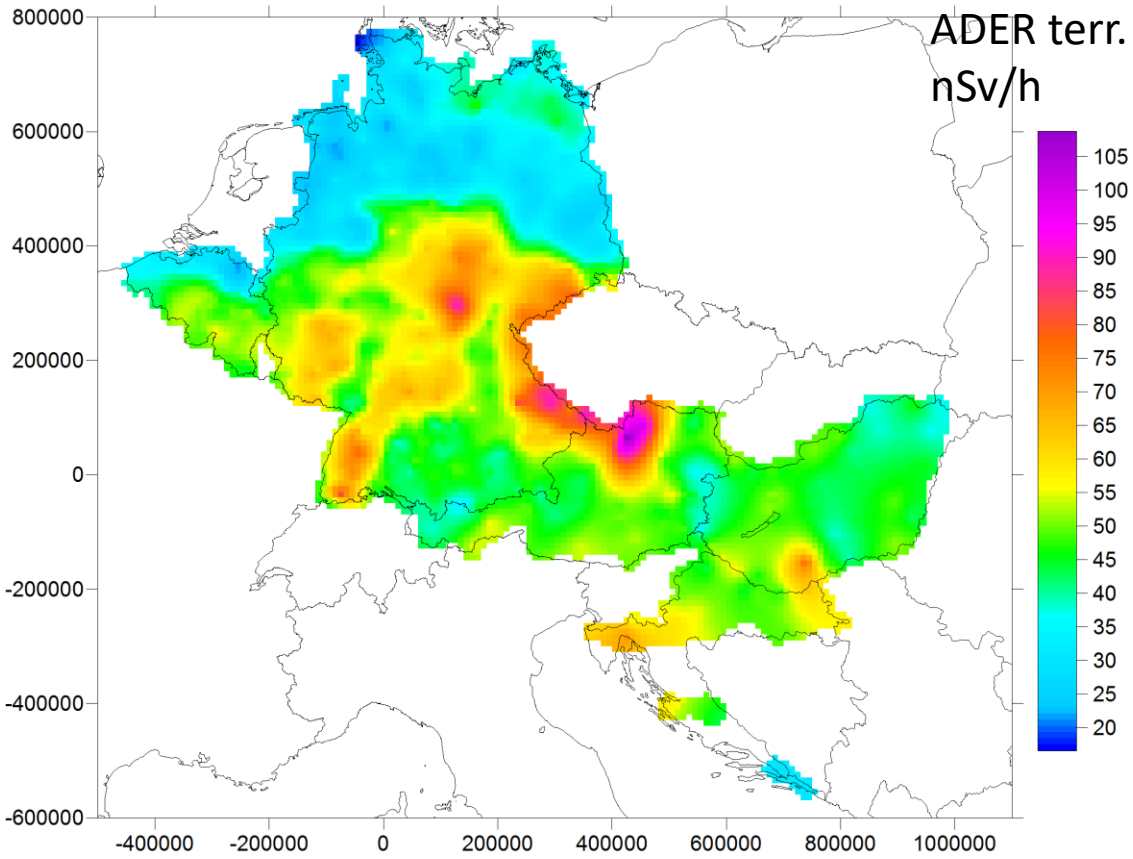
- from U, Th, K concentrations terrestrial ADR (nGy/h) calculated.
- 80% dry matter assumed
- Converted to ADER (nSv/h), 0.7 Sv/Gy.
- Same grid as indoor Rn map. n=12,101 cells



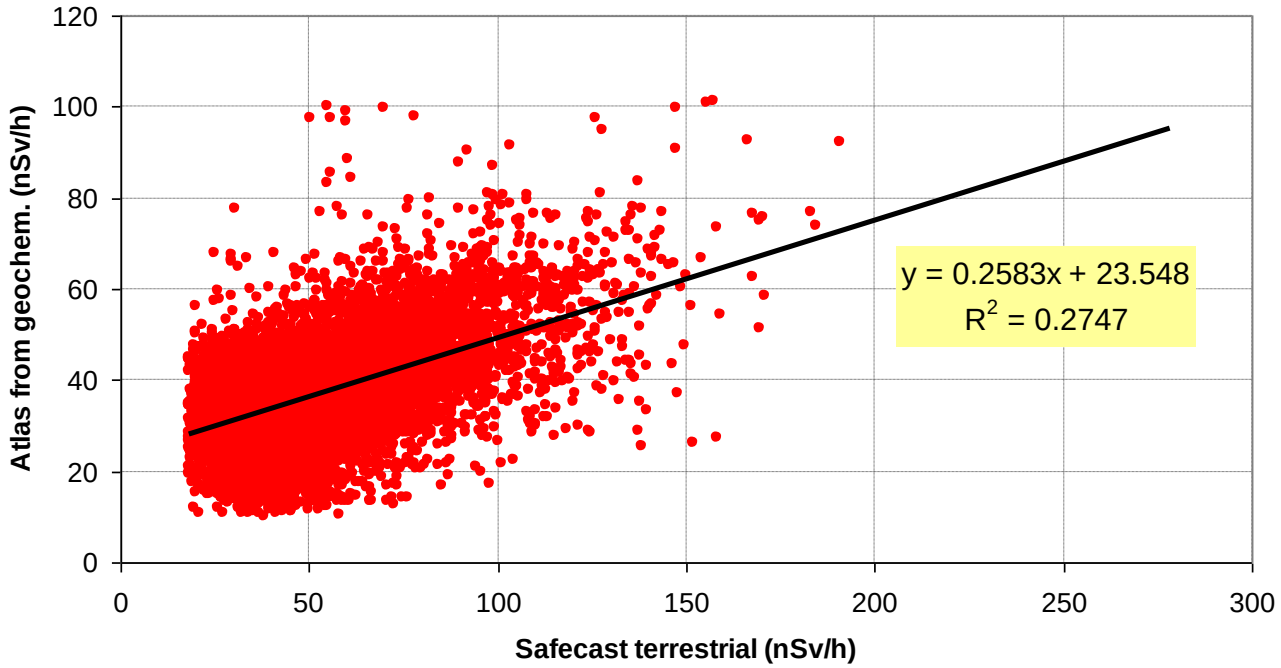
https://remon.jrc.ec.europa.eu/About/Atlas-of-Natural-Radiation/Digital-Atlas/Uranium-in-soil/Uranium-concentration-in-soil-

Data 4: ADER EURDEP

Terrestrial background
ADER extracted from EURDEP data
By analysis of ADER time series, discarding Rn peaks.
Ref: Bossew et al.(2017): Estimating the terrestrial gamma dose rate by decomposition of the ambient dose equivalent rate.
http://dx.doi.org/10.1016/j.jenvrad.2016.02.013



Comparison ADER Safecast / Atlas



slope should theoretically be =1, if both ADER were correct.

Problem: Both have uncertainty.

Uncertainty of independent variable (x) leads to “regression dilution”: reduces slope.

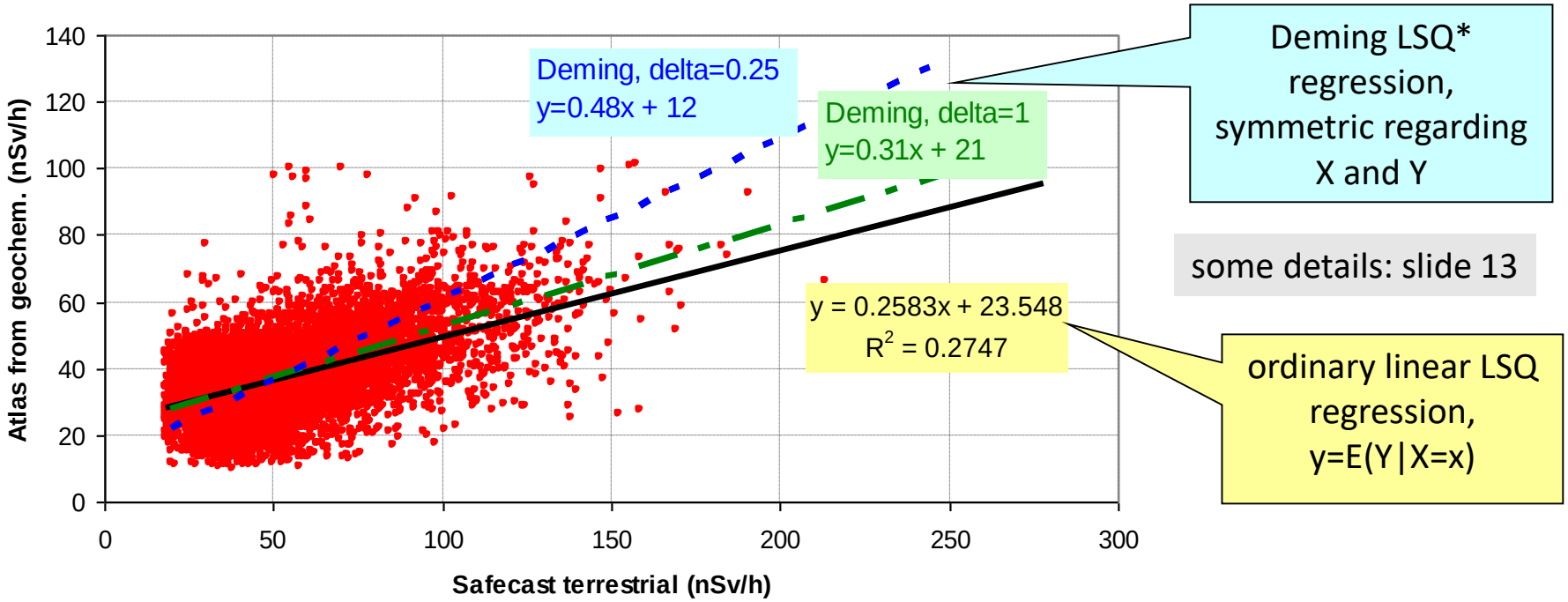
Uncertainty of Safecast data:

- Measurement locations not representative for a cell;
- Incorrect measurement method

Uncertainty of Atlas data:

- Computation from U, Th, K;
- Interpolation

Comparison ADER Safecast / Atlas

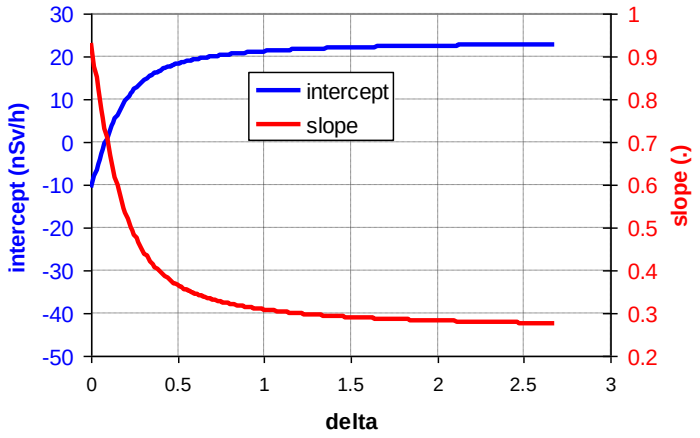


Deming LSQ* regression, symmetric regarding X and Y

some details: slide 13

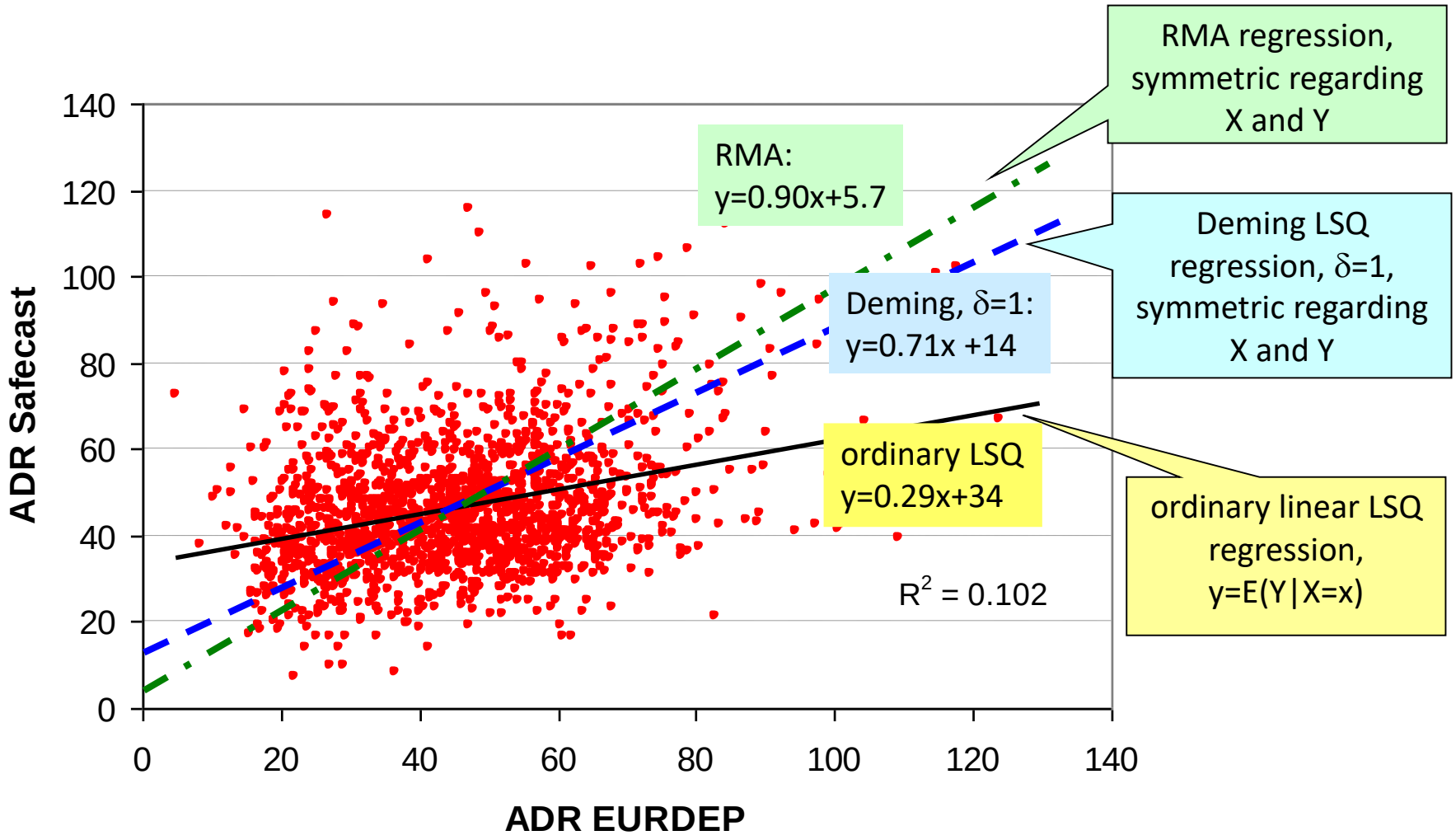
ordinary linear LSQ regression, y=E(Y|X=x)

Problem: Both variables have uncertainty (in ordinary regression only Y) → Deming (orthogonal) regression. Result is very sensitive against choice of $\delta := unc_y^2/unc_x^2$ Here tentatively $\delta=1$ and 0.25 chosen: both about plausible, but requires further research! For $\delta \to \infty$: ordinary regression



* LSQ – least square

Comparison ADER Safecast / EURDEP



EURDEP Station data associated to the 10 km × 10 km cell in which they are located – probably no optimal method of association.

Footnote: Linear regression

- **Common Least square regression:**
  estimates a value y of Y, given a value x of the independent variable X.

$$\hat{y} = E[Y \mid X = x]$$

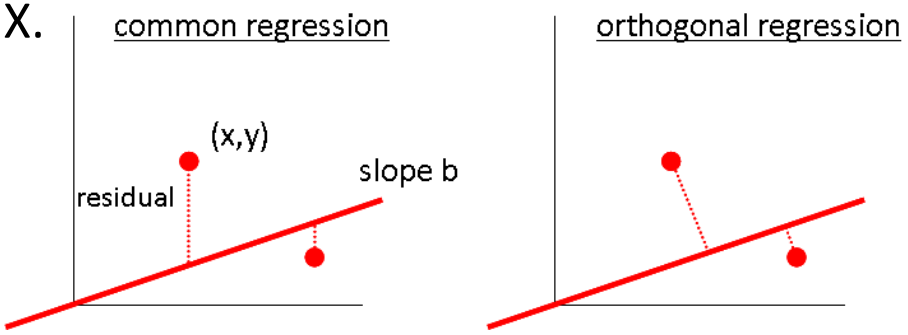
  Or: y=b*x + a (intercept) + ε (error)

- b = slope(Y|X=x) ≠ 1/slope(X|Y=y) Not symmetric!! The slope is no measure of association! X is assumed without uncertainty.
- **RMA regression**:
  b =GM(slope(Y|X), 1/slope(X|Y)) = √(Var(Y)/Var(X))
- **Orthogonal / Deming regression**:

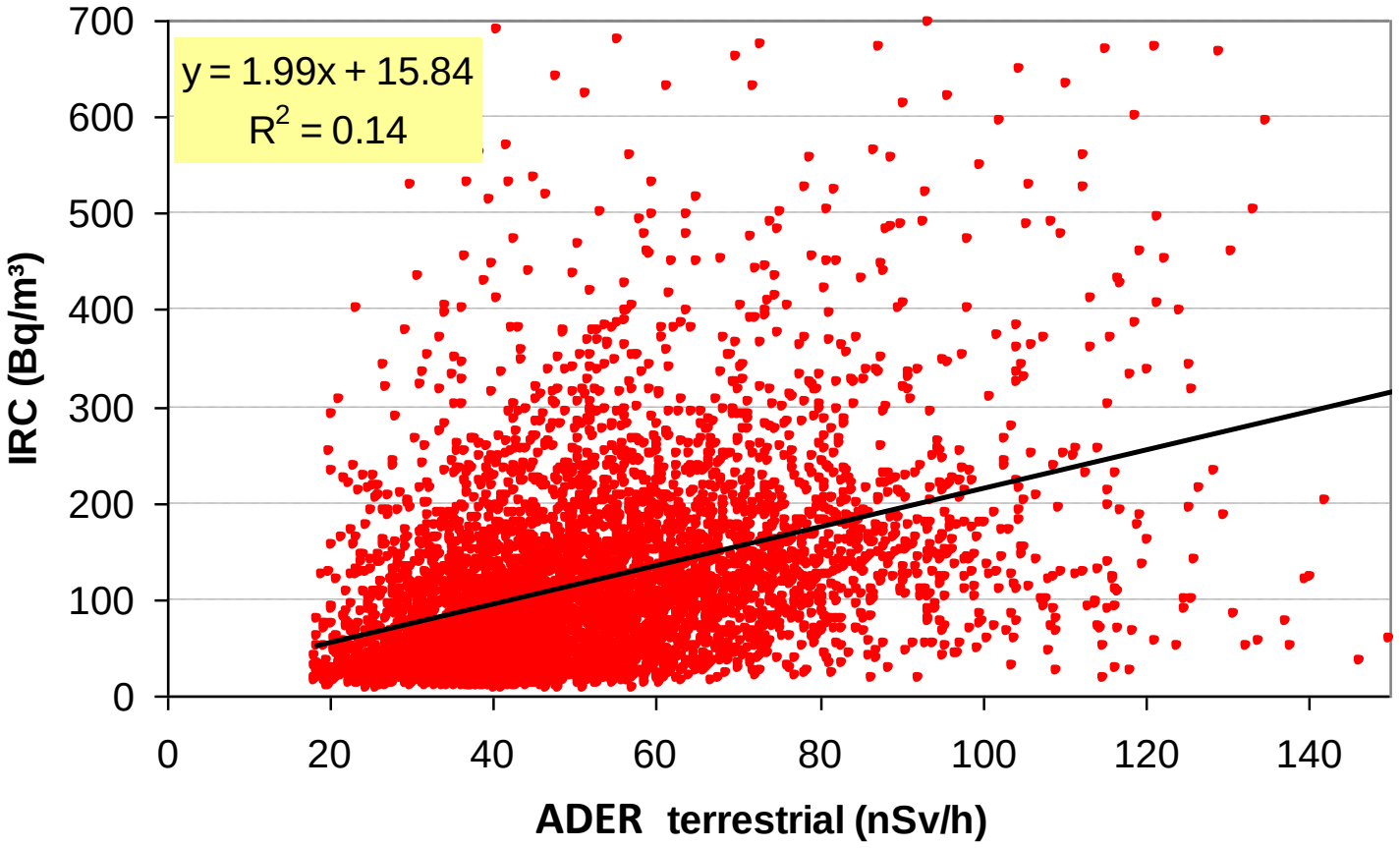
$$b(Deming) = \frac{VarY - \delta\, VarX + \sqrt{(VarY - \delta\, VarX)^2 + 4\delta\, CovarXY^2}}{2\, CovarXY}$$

  orthogonal regression: δ=1
  δ accounts for uncertainty of Y and X.

RMA and Orth. regression:
Symmetric, b(Y|X) = 1/b(X|Y)
slope is here an association measure



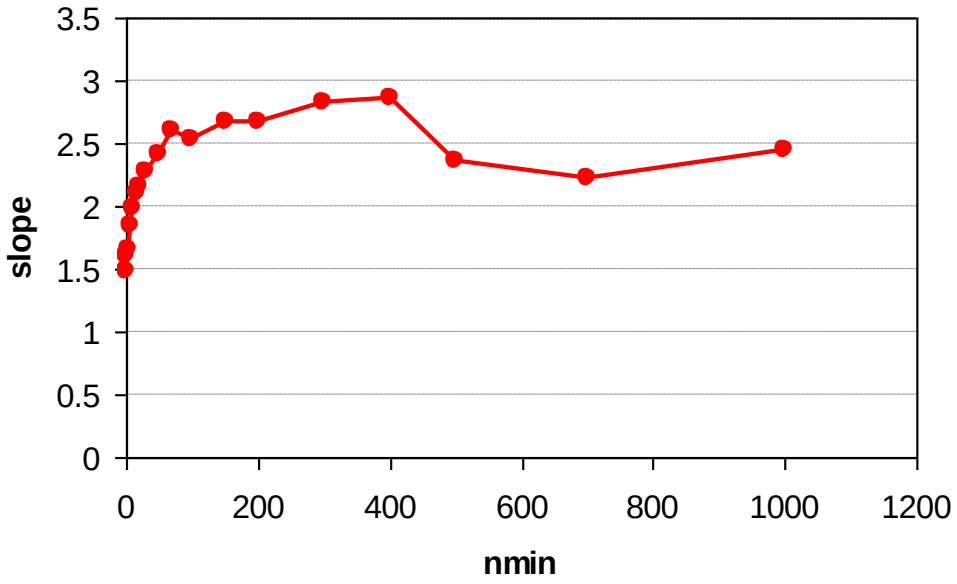
ADER Safecast – IRC mean per cell



for IRC, only cells with n>10 used
uncertainty not yet considered

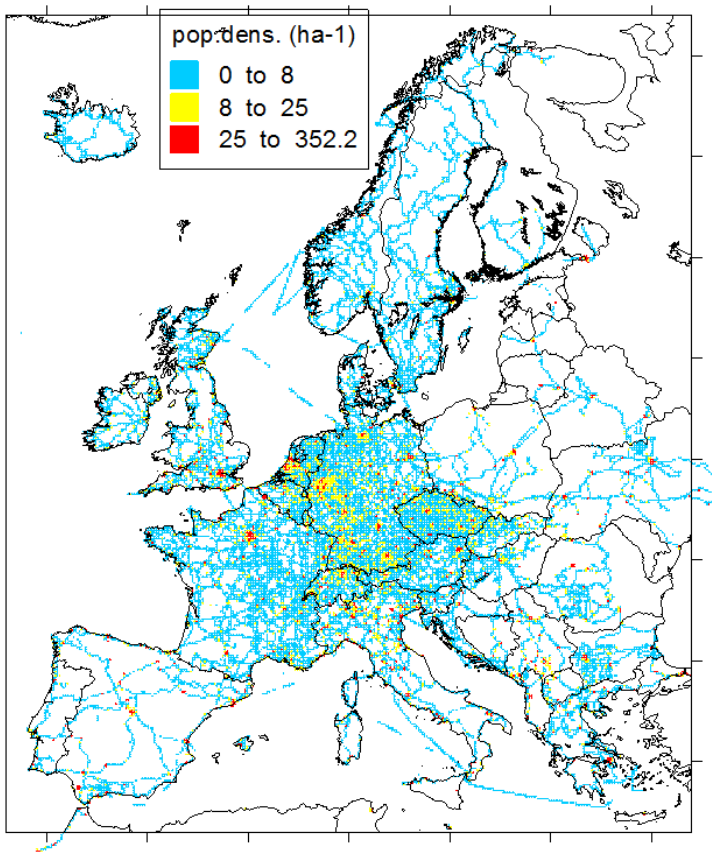
dependence on number of data

- The slope (previous slide) depends on the chosen minimal number of IRC data (n) per cell.
- Reason? – Perhaps because mean IRC per cell has uncertainty, which is the lower, the higher is the number of data per cell.



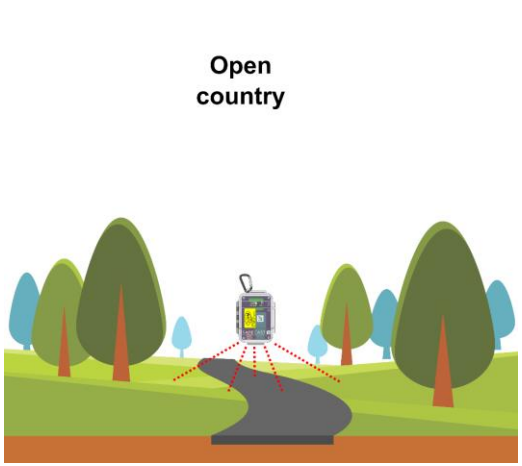
To be investigated further!

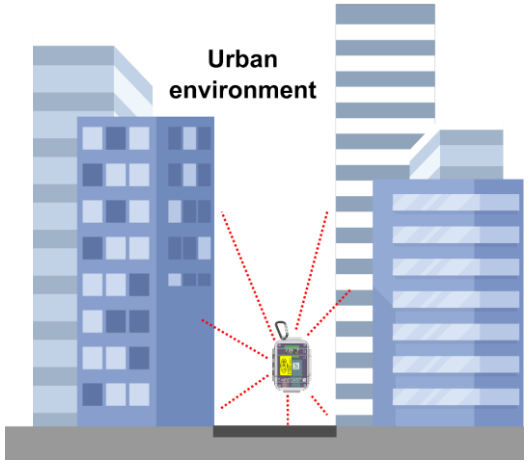
influence of urbanization? - 1



Idea:

The association between ADER and radon may be different in open country and cities

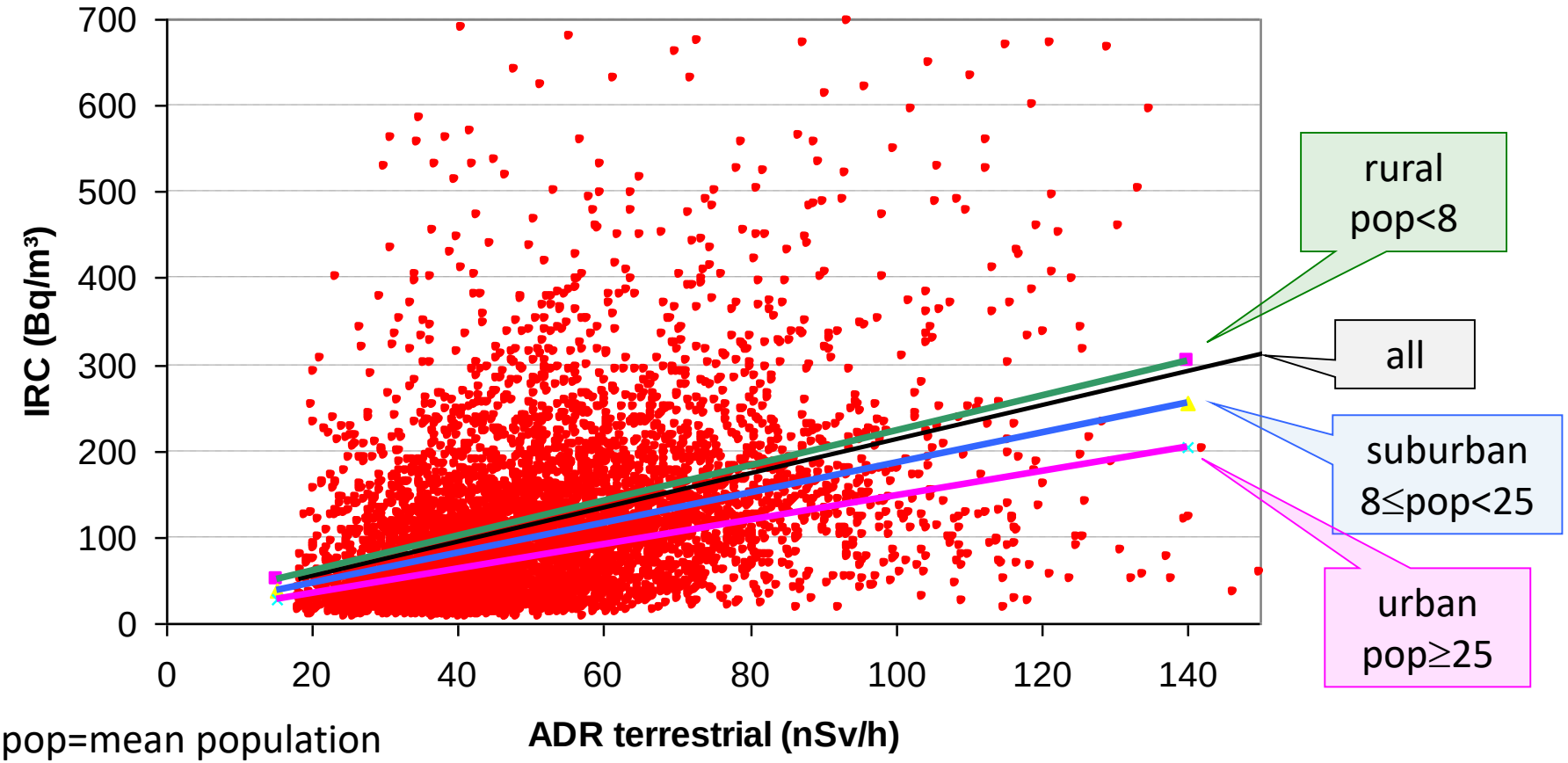




proxy of degree of urbanization: population density, overlaid on *Safecast* traces

other proxies available in Europe: built surface, built volume – but not yet evaluated

Influence of urbanization? - 2

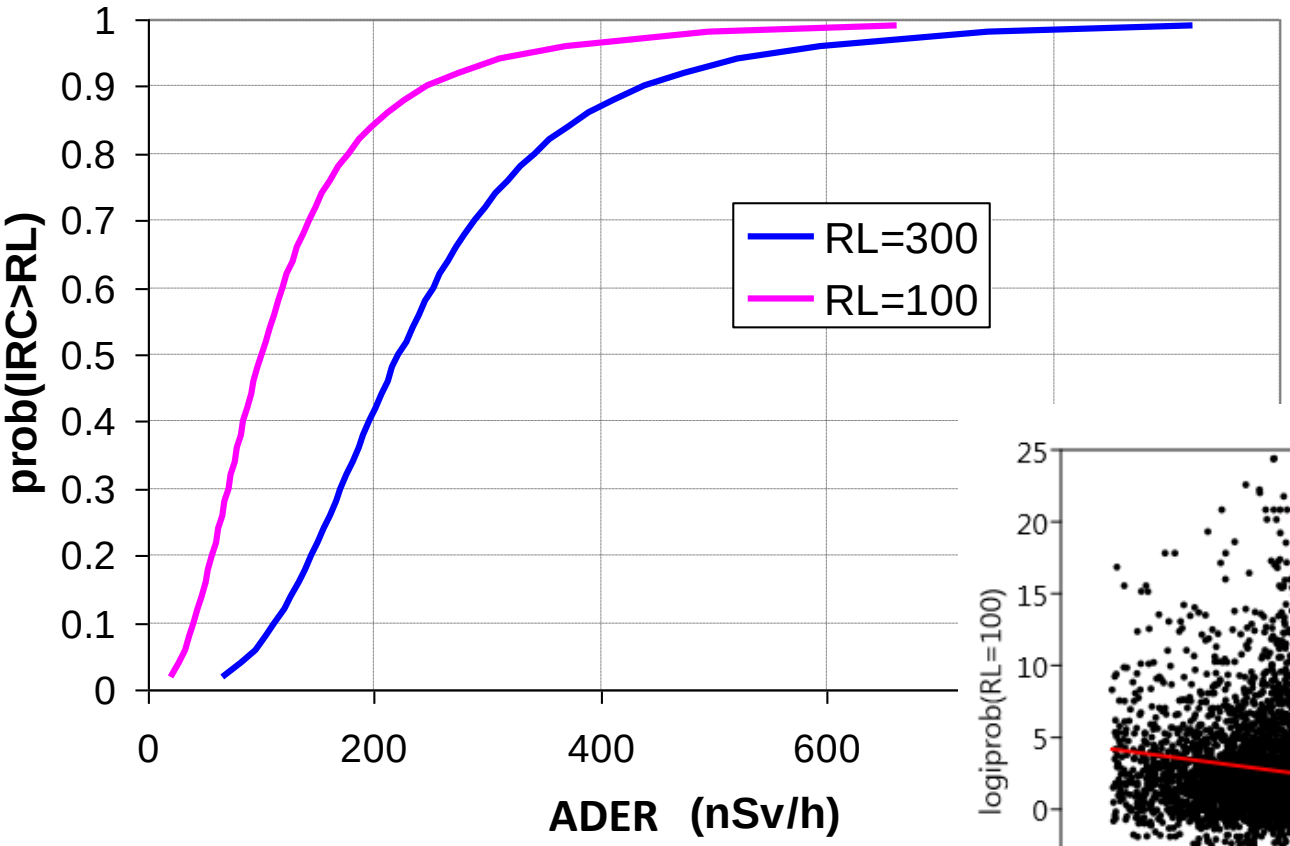


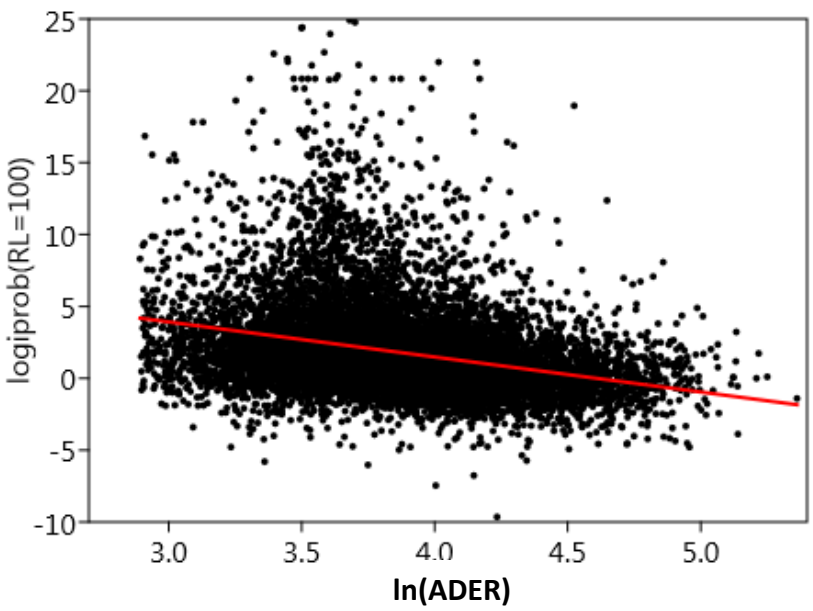
pop=mean population density in cell (ha^{-1})

Tentative interpretation: Association between IRC and ADER is weaker in cities than in the open country – could be expected. More detailed interpretation still necessary!

ADER Safecast – IRC exceedance probability

log-logistic relationship: $p=1/(1+b\,x^{-a})$

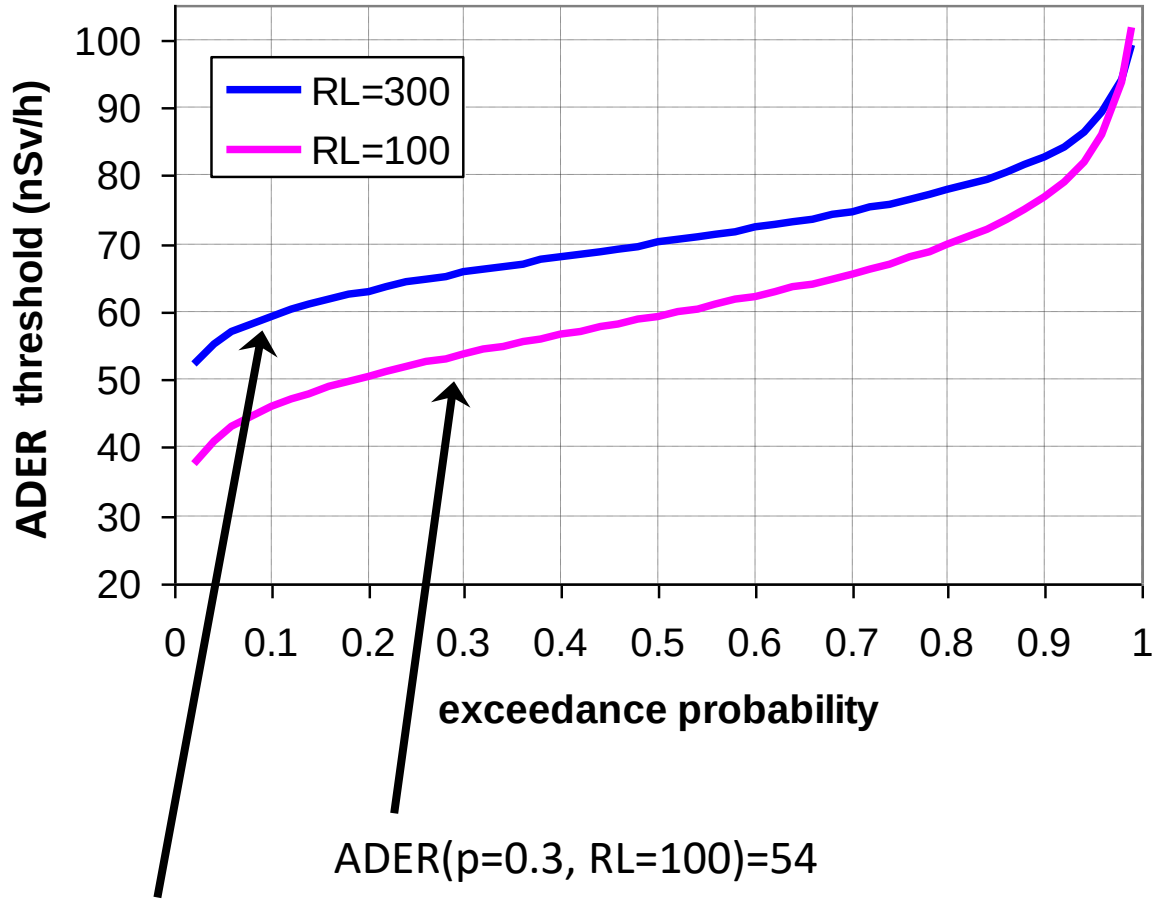




log-logistic relationship transformed to linear, y=a+bx,
y=logiprob(p):=ln((1/p)-1), x=ln(ADER)

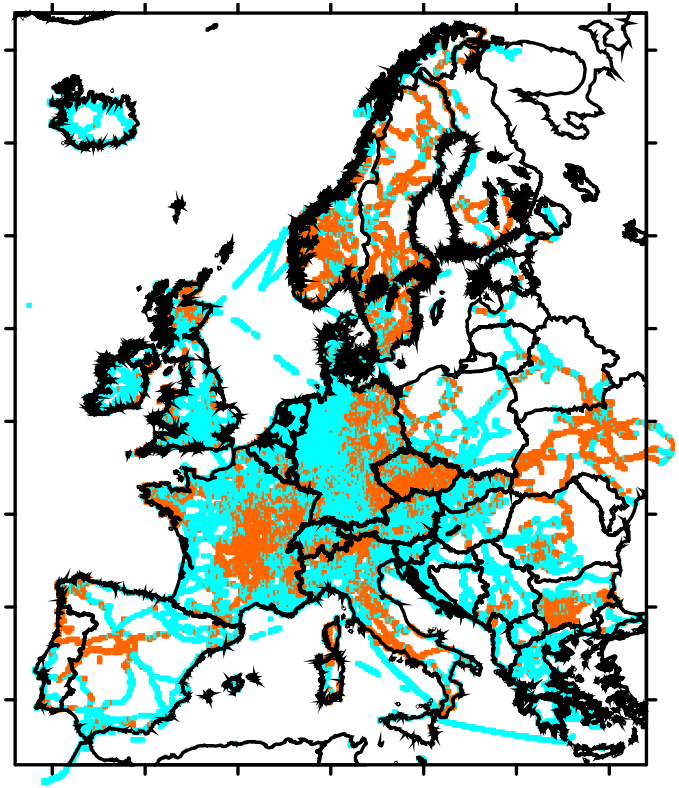
RPA: ADER thresholds by logistic regression

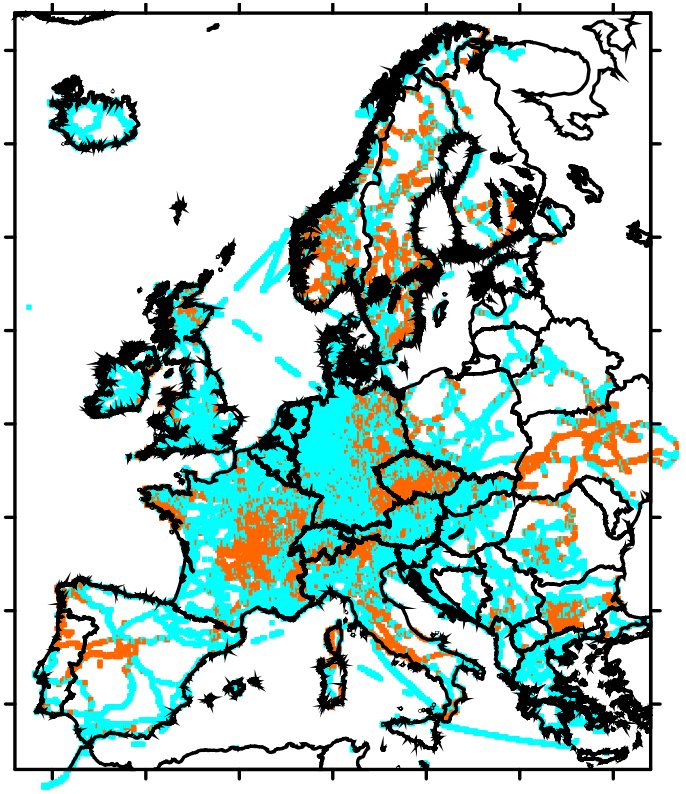
inverse log-logistic relationship: $y=e^{A}((1/p)-1)^{B}$



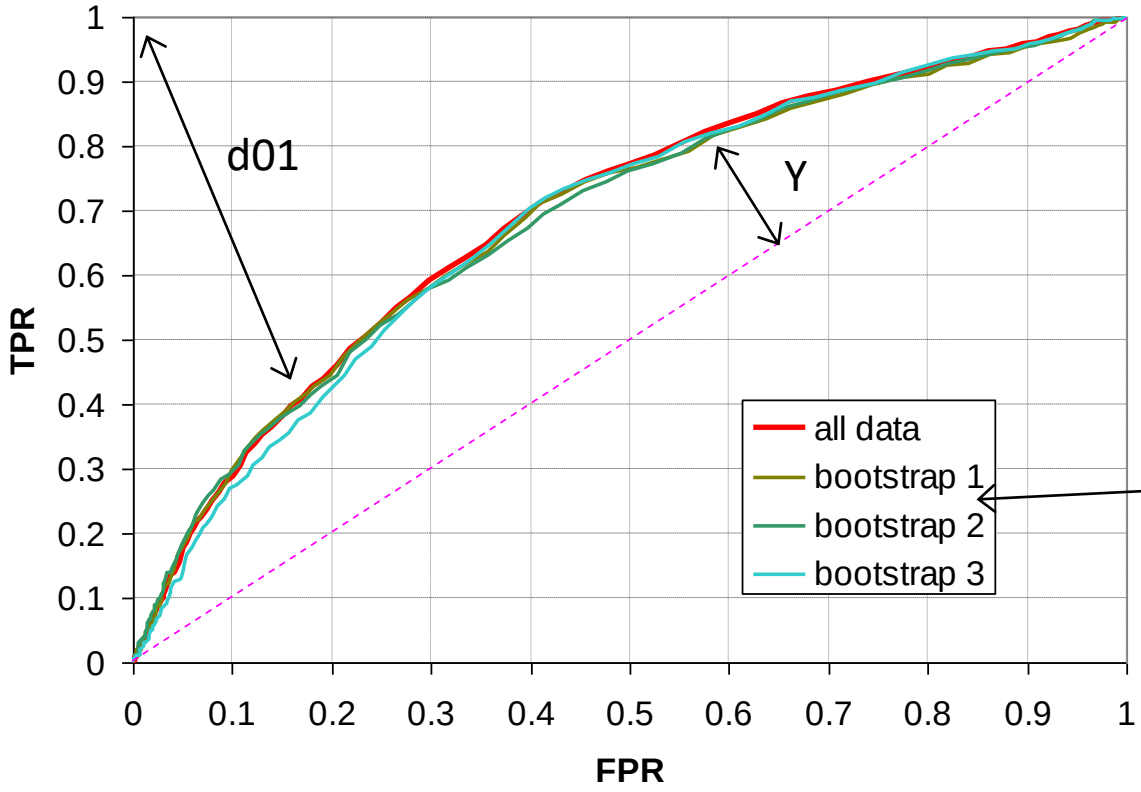
Estimating radon priority areas? (1)

- Method 1: Logistic regression.
- RPA: prob(IRC>100)>0.3 and prob(IRC>300)>0.1
  (red areas in the maps)





Estimating radon priority areas? (2)



Method 2: ROC, cross-classification, nonparametric

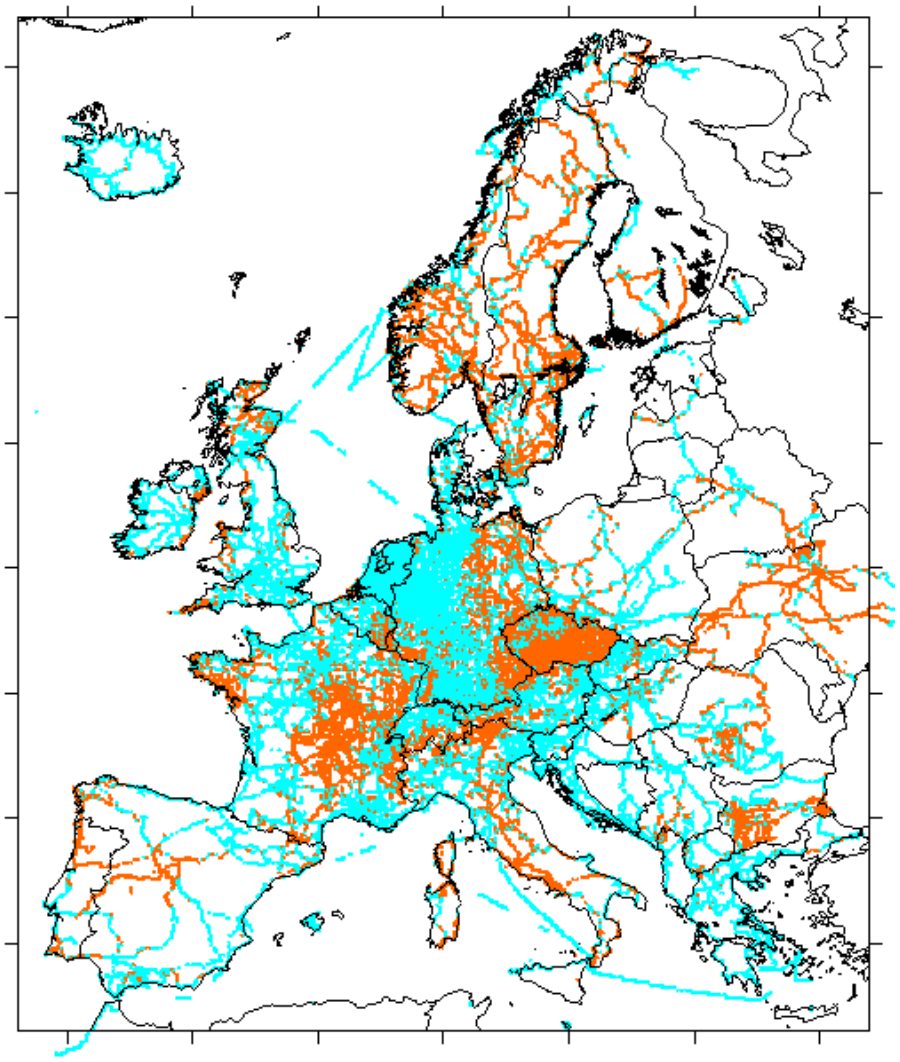
in order to assess the uncertainty of the ROC curve. Appears very reliable. Total 100 bootstraps

thresholds of ADER, corresponding to the crit

TPR, FPR – true, false positive ratios
Optimal point in the ROC curve by optimizing statistics, e.g., finding the point on the curve which maximizes Y- or minimizes d01-statistic
AUC – area under curve, measure of strength of relationship. / min IRC data per cell = 20.

crit	Y	d01	AUC
prob(IRC>300)>0.1	**51**	**53**	0.71
prob(IRC>100)>0.3	**48**	**48**	0.73

Estimating radon priority areas? (2)



prob(IRC>300)>0.1
threshold:
ADER terr.= 52 nSv/h
1.kind error prob≈ 40%
2.kind error prob≈ 30%
quite high!

RPA too large!
Result of method 1 appear more plausible, although method 2 is usually more robust.
Pattern essentially correct.

Conclusions

- Gridded *Safecast* ADER data appear plausible.
  Problem: *Safecast* data with unknown uncertainty, possibly serially correlated.
- *Safecast,* Atlas ADER and EURDEP ADER are significantly correlated.
- Relationship between *Safecast* ADER and IRC related quantities exists, but not very strong.
- Results refer to means in 10 km × 10 km cells. This cannot be extrapolated to local values! (I.e., local IRC or RPA status prediction by local ADER measurement is not possible!) How the relationships look like for other aggregation sizes is still to be investigated.
- RPA over-estimated; high classification error probability

To do

- Better consideration of data uncertainty!
- More detailed classification by type of environment, in which *Safecast* ADER has been measured: rural, sub-urban, urban
- Influence of number of measurements per Rn cell to be further investigated... why?

Acknowledgement

The authors wish to acknowledge the JRC (Joint Research Centre of the European Commission) G.I.4 Unit for allowing usage of the datasets underlying the European Atlas of Natural Radiation.

Thank you!

Dr. phil. Peter Bossew
Privatier
Physics & metaphysics en gros & en detail

Bundesamt
für Strahlenschutz

Agenzia nazionale per le nuove tecnologie,
l'energia e lo sviluppo economico sostenibile

Státní ústav radiační ochrany, v. v. i.
National Radiation Protection Institute