Can citizen science generated ambient dose rate data (*Safecast*) be used for predicting indoor radon?

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.21.4.23

EGU General Assembly 2023

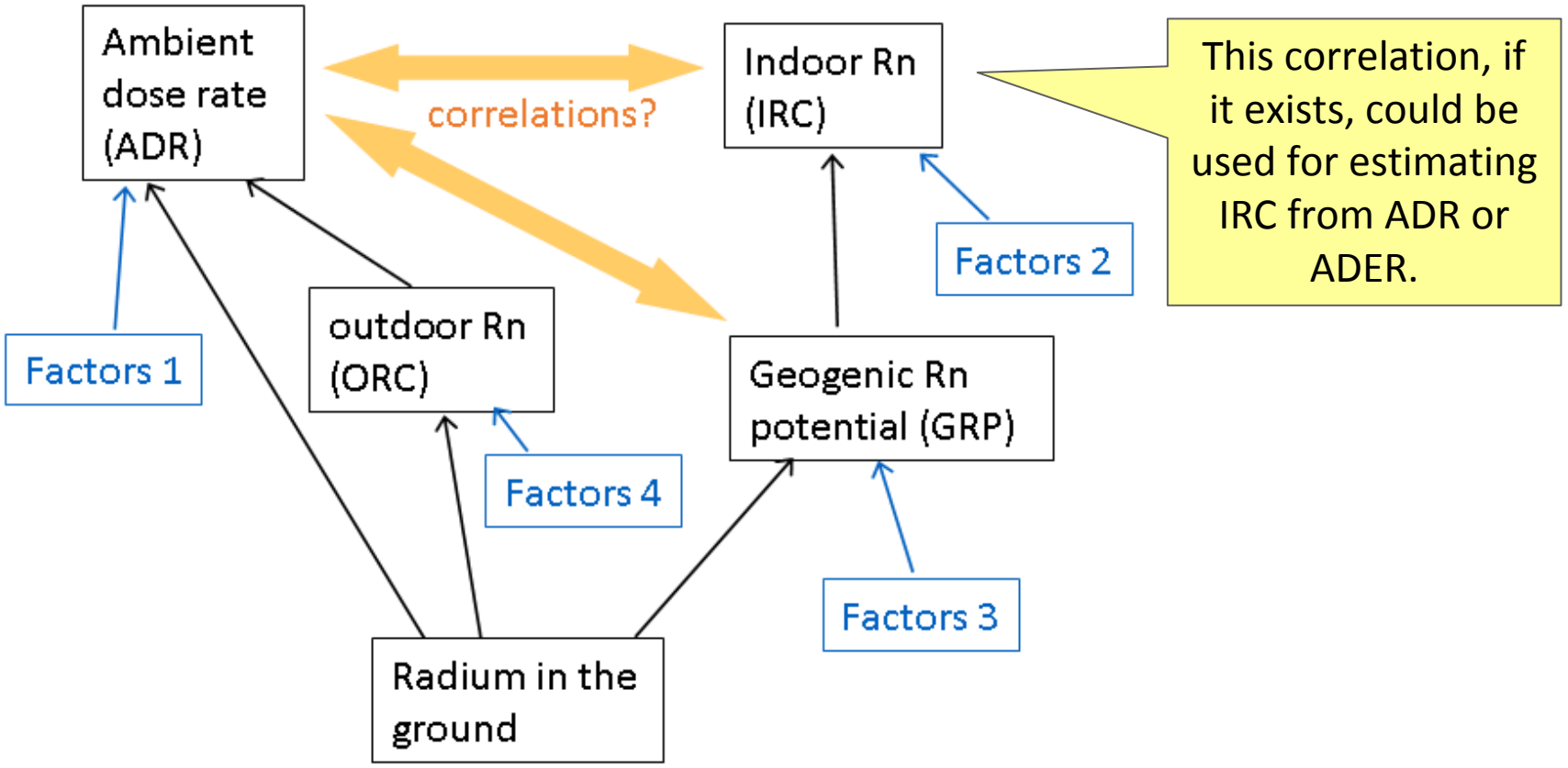
Vienna, Austria , 23–28 April 2023

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

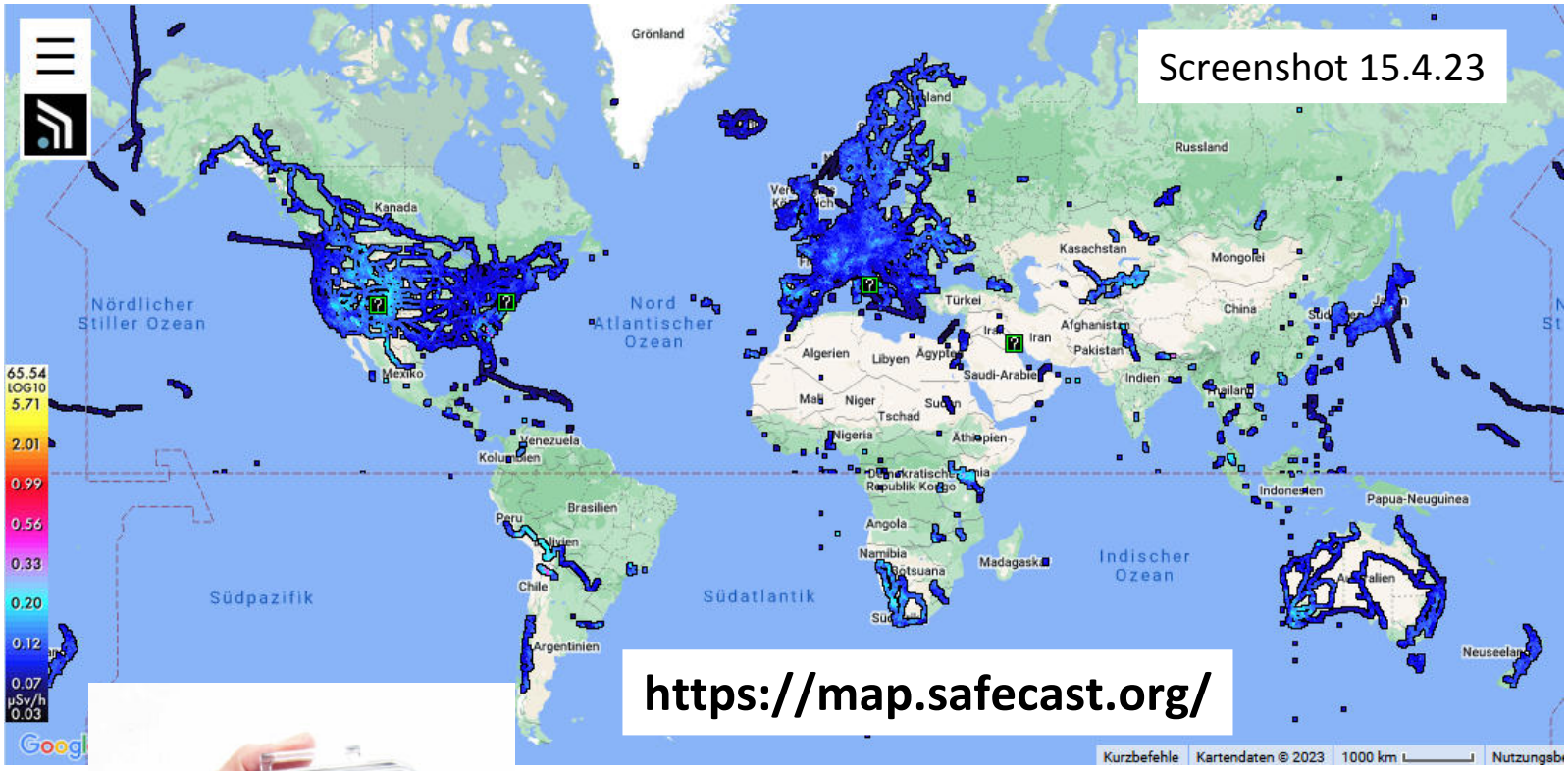


Various “nuisance” factors blur the correlation: cosmic dose rate, outdoor Rn, other sources of terrestrial dose rate (^{40}K, Th series), factors which control ingression of Rn into buildings.

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.

Safecast map



bGeigie Nano: GM detector, coupled to GPS. Output: ADER, nSv/h

Data saved on SD card in txt format.

Data

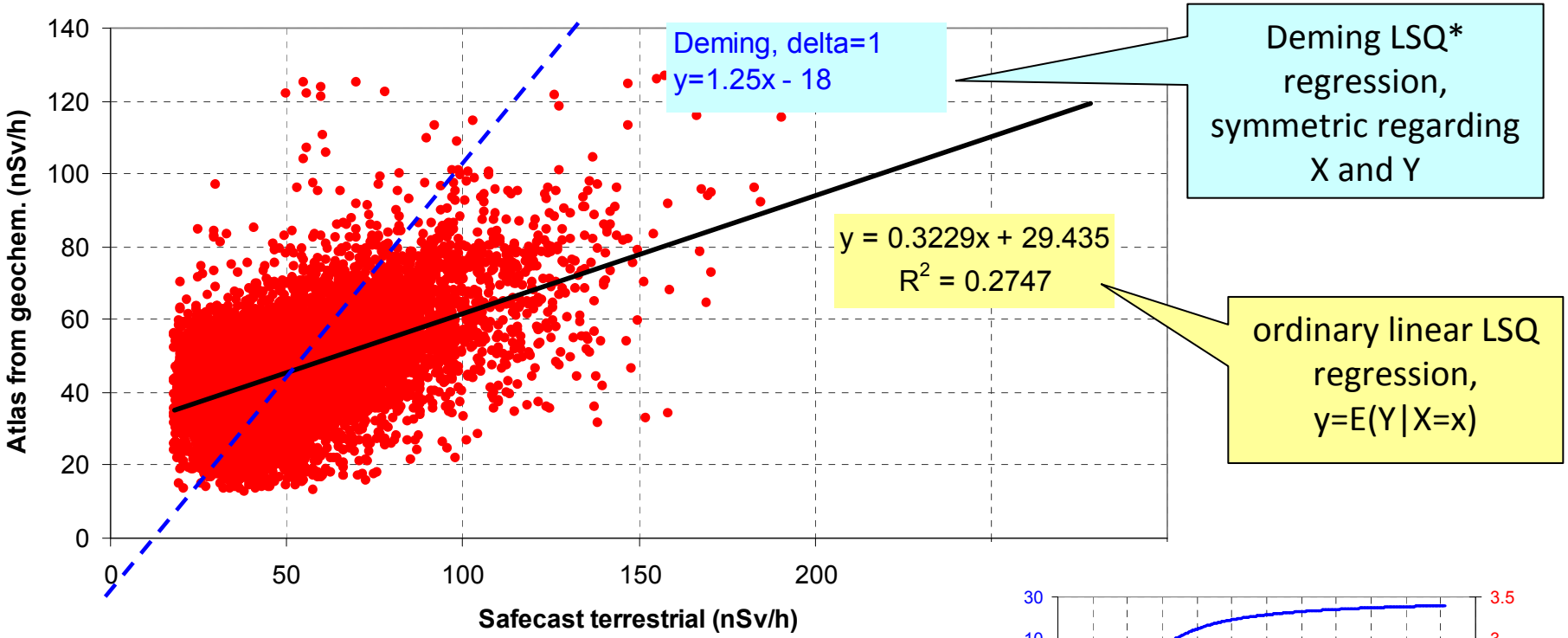
- *Safecast* data
  - n=52,887,234; relevant fields: coordinates, ADER (1 minute mean, µSv/h).
  - Processing: internal background removed; cosmic dose rate subtracted, calculated from altitude, taken from DEM → terrestrial ADR, conversion of geographical into European Lambert coordinates, aggregation into European grid, 21,828 cells.
- European indoor radon database, from the European Atlas of Natural Radiation (2019) [1]:
  - about 1.2 mill. measurements, 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.
- Geochemical database from the Atlas:

  from U, Th, K concentrations terrestrial ADR (nGy/h) calculated. Converted into ADER (nSv/h). Same grid. n=12,101 cells

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

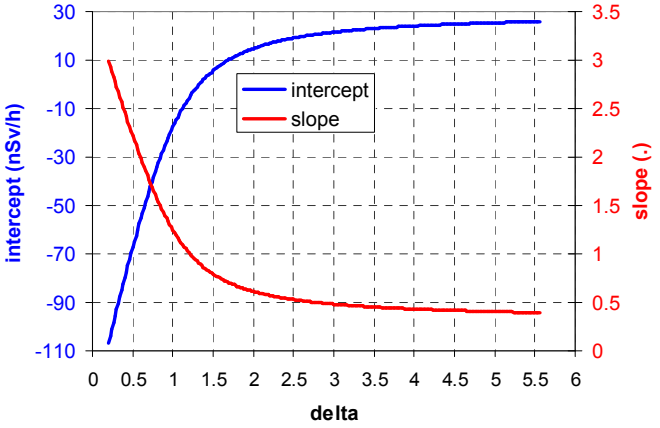
* AM, SD – arithmetical mean, standard deviation; RL – reference level

Comparison ADER Safecast / Atlas



Deming LSQ* regression, symmetric regarding X and Y

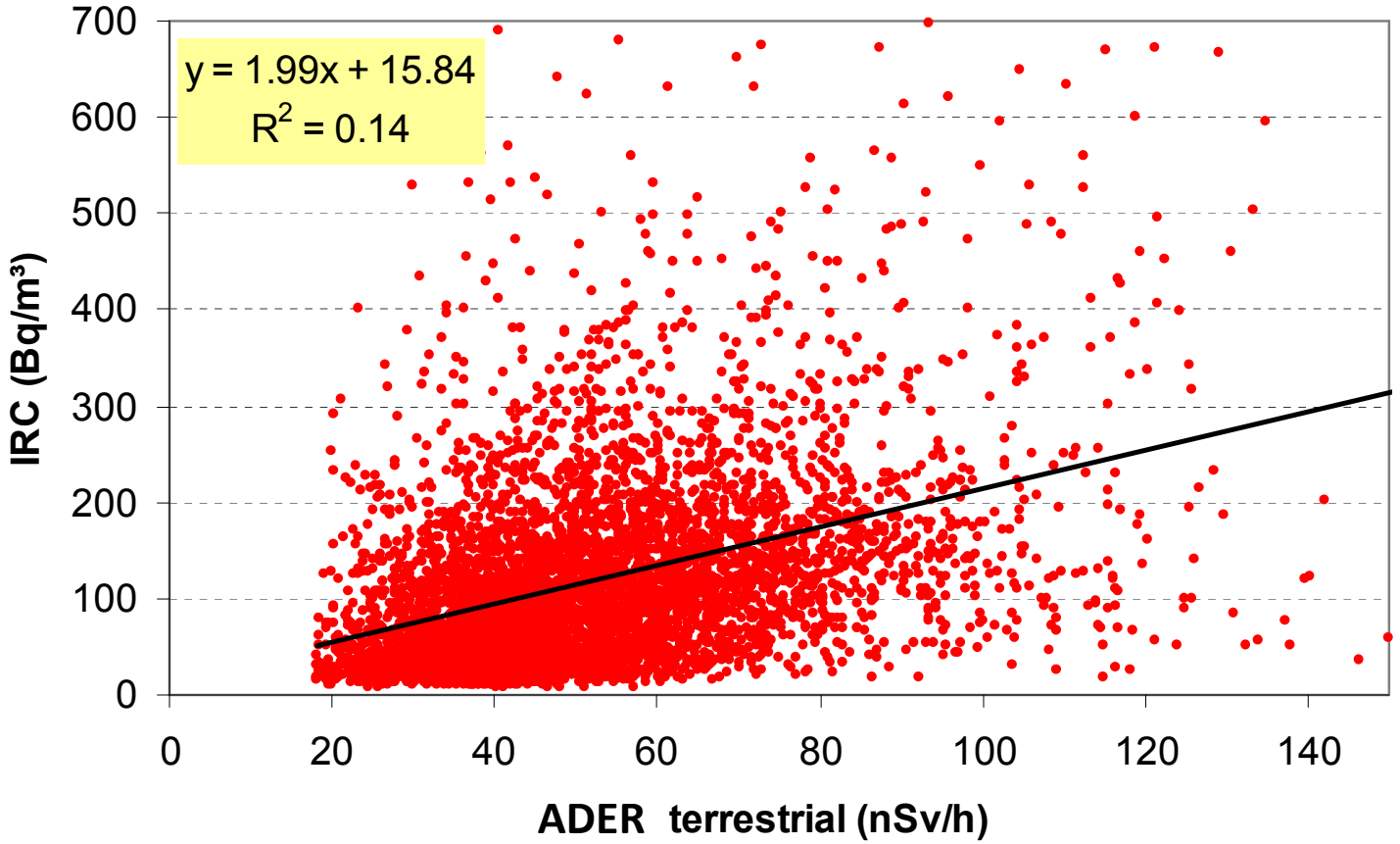
ordinary linear LSQ regression, y=E(Y|X=x)



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 $\delta=1$ chosen: result not convincing, requires further research! For $\delta \to \infty$: ordinary regression

* LSQ – least square

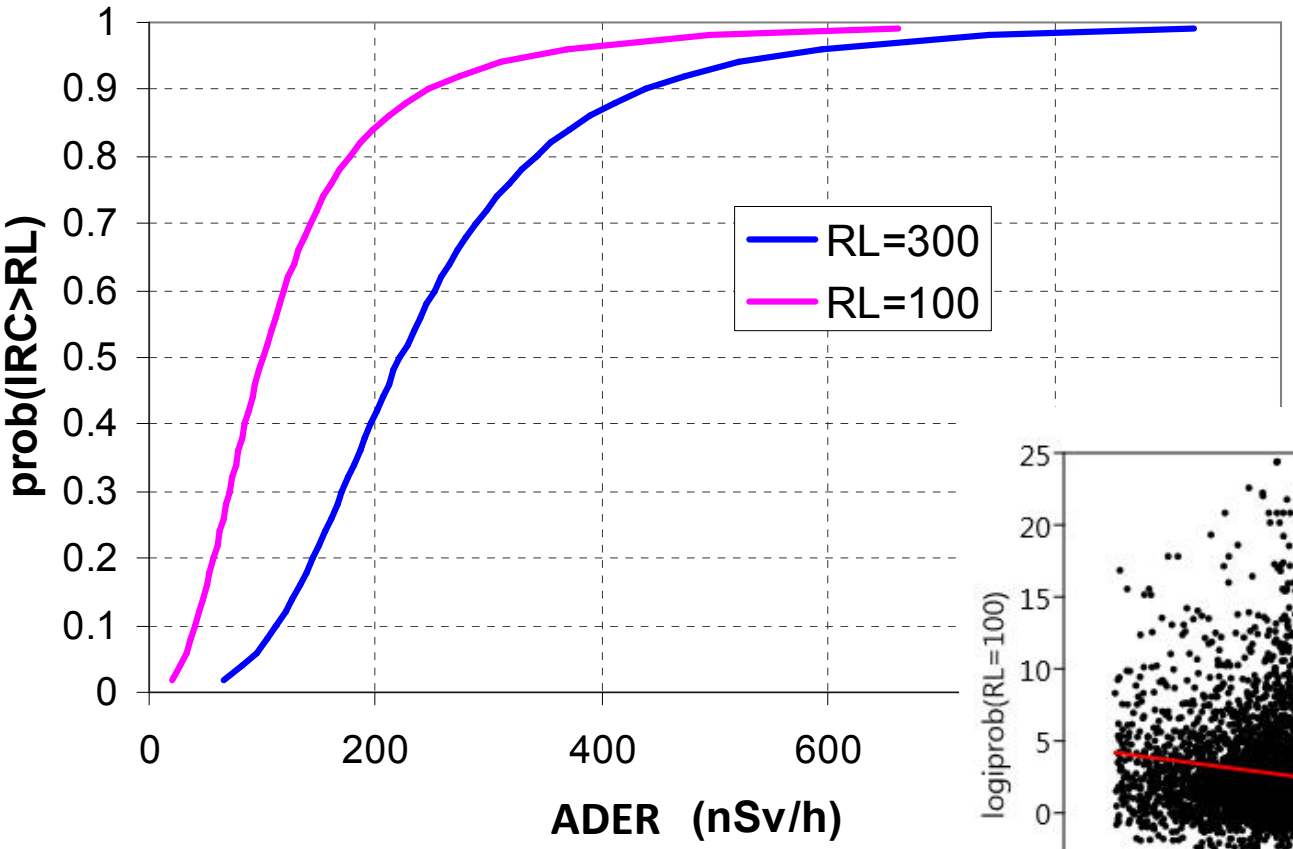
ADER Safecast – IRC mean per cell

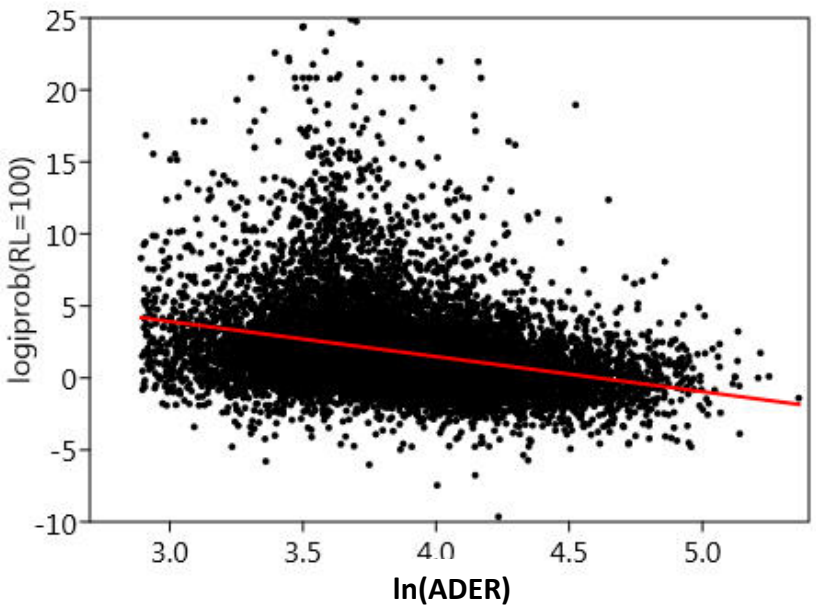


for IRC, only cells with n>10 used
uncertainty not yet considered

ADR Safecast – IRC exceedance probability

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

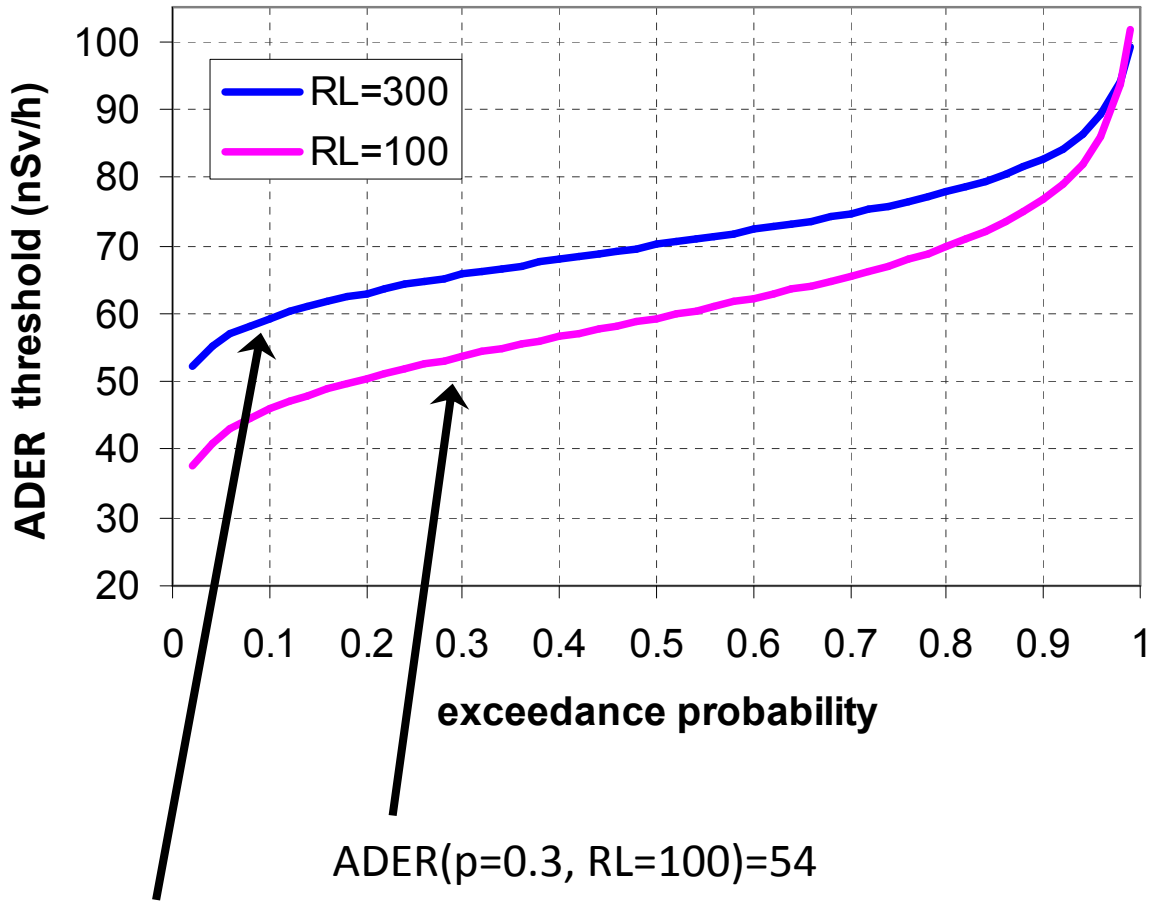




log-logistic relationship transformed to linear, y=a+bx,
y=logiprob(p):=ln((1/p)-1), x=ln(ADER)

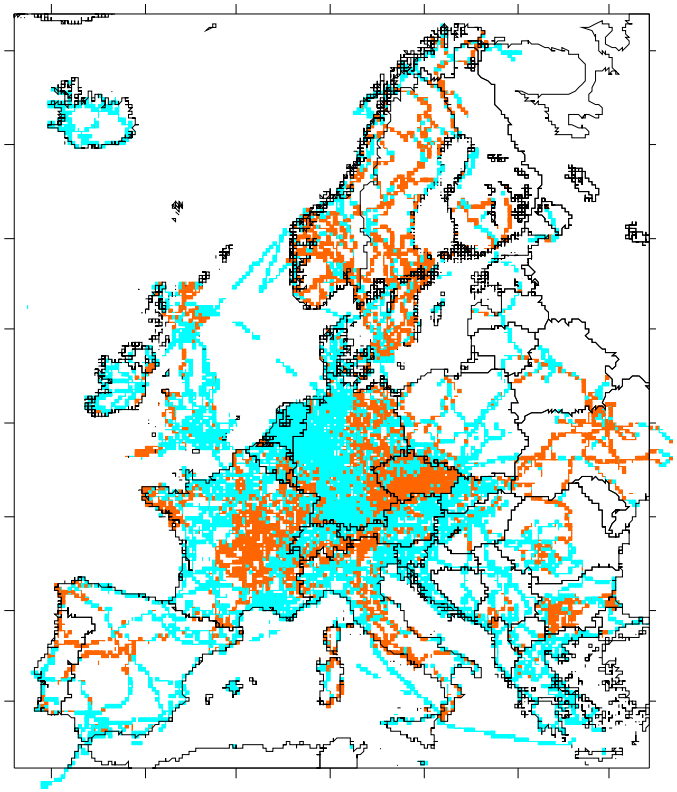
RPA: ADER thresholds by logistic regression

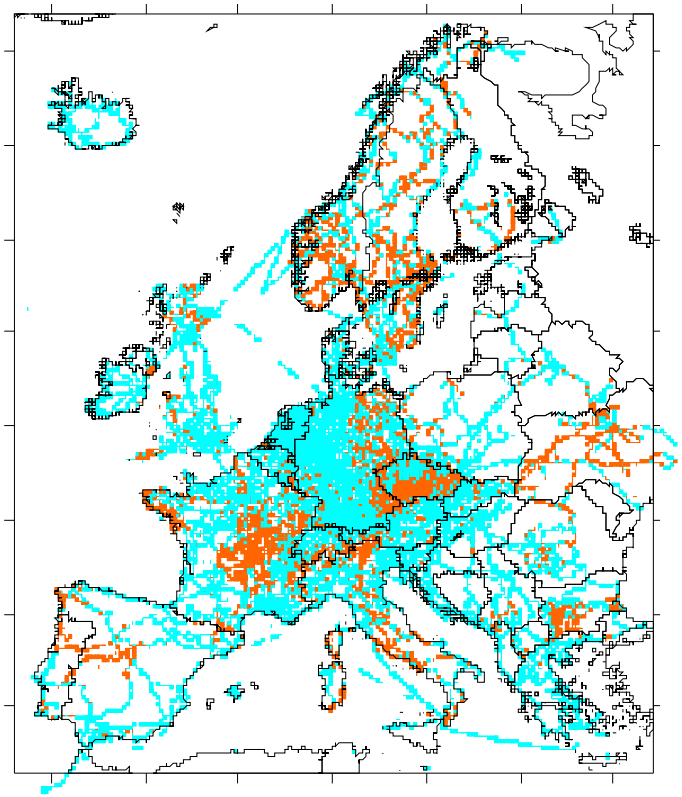
inverse log-logistic relationship: $y=e^{A}((1/p)-1)^{B}$



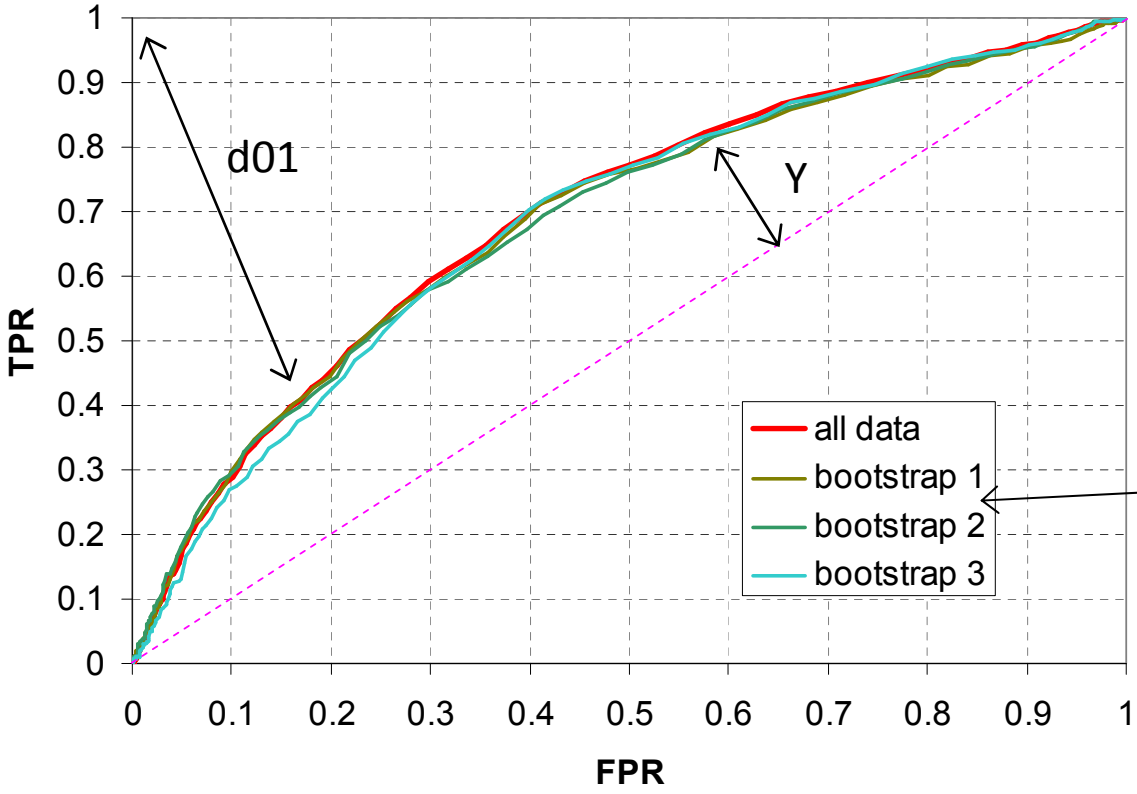
Estimating radon priority areas? (1)

- <u>Method 1:</u> 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

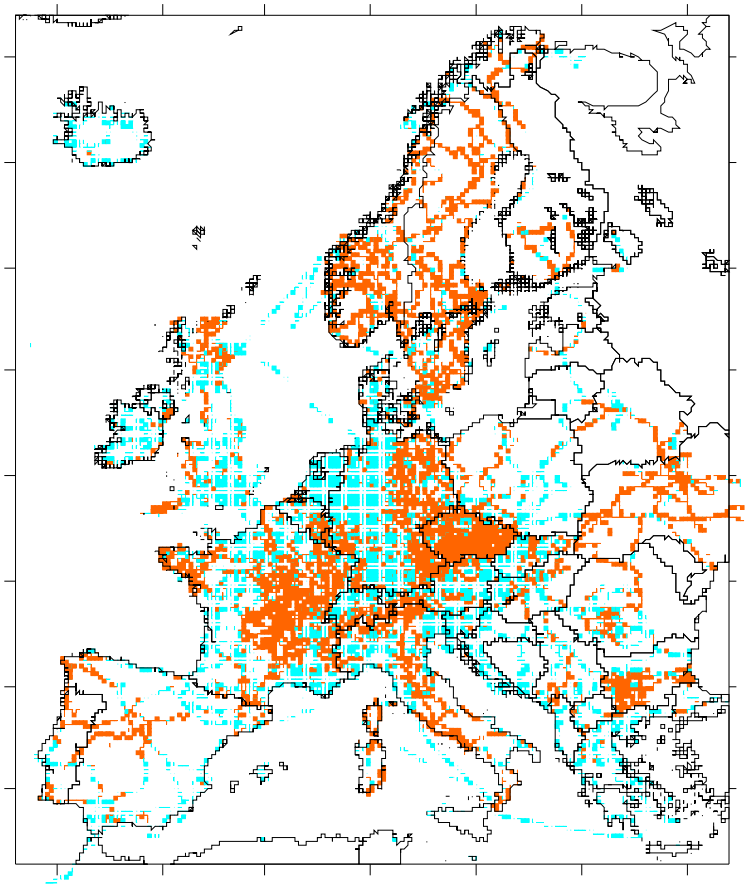
in order to assess the uncertainty of the ROC curve. Appears very reliable

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.

crit	Y	d01	AUC
prob(IRC>300)>0.1	**49**	**51**	0.69
prob(IRC>100)>0.3	**48**	**48**	0.70

Estimating radon priority areas? (2)



prob(IRC>300)>0.1
threshold:
ADER terr.= 50 nSv/h
1.kind error prob≈ 40%
2.kind error prob≈ 30%
quite high!

RPA too large!
Method 1 appears more plausible, although method 2 is usually more robust.
Pattern essentially correct.

Conclusion

- Gridded *Safecast* ADER data appear plausible.
  Problem: *Safecast* data with unknown uncertainty, possibly serially correlated.
- *Safecast* and Atlas ADER are significantly correlated.
- Relationship between *Safecast* ADER and IRC related quantities exists, but not very strong.
- RPA over-estimated; high classification error probability

To do

- Better consideration of data uncertainty!
- 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 investigated.

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