Development of a geogenic radon hazard index GRHI

Bossew P.[1], Cinelli G.[2], Ciotoli GC.[3], Crowley Q.[4], De Cort M.[2],
Elío J.[4], Gruber V.[5], Petermann E.[1], Tollefsen T.[2]

1 German Federal Office for Radiation Protection, Berlin
2 European Commission, Joint Research Centre (JRC), Ispra, Italy
3 Inst. of Environmental Geology and Geoengineering, National Research Council, Rome, Italy
4 Trinity College Dublin, School of Natural Sciences, Geology, Dublin, Ireland
5 Austrian Agency for Health and Food Safety, Department for Radon and Radioecology, Linz

v.24.5.19b

3rd International Conference Radon in the Environment
Krakow, 27-31 May 2019

Content

- **Motivation, objective**
- **Role of Metro Radon**
- **Different approaches**
- **Previous attempts**
- **Examples**

Motivation, concept, objective 1

- A quantity which quantifies the contribution of geogenic factors to the potential risk that exposure to indoor Rn causes.
- A quantity which measures the availability of geogenic Rn at surface level.
- Measure of "Rn proneness" or "Rn priorityness" of an area due to geogenic factors.

It is known that many geogenic factors contribute. Much literature available!

Motivation, concept, objective 2

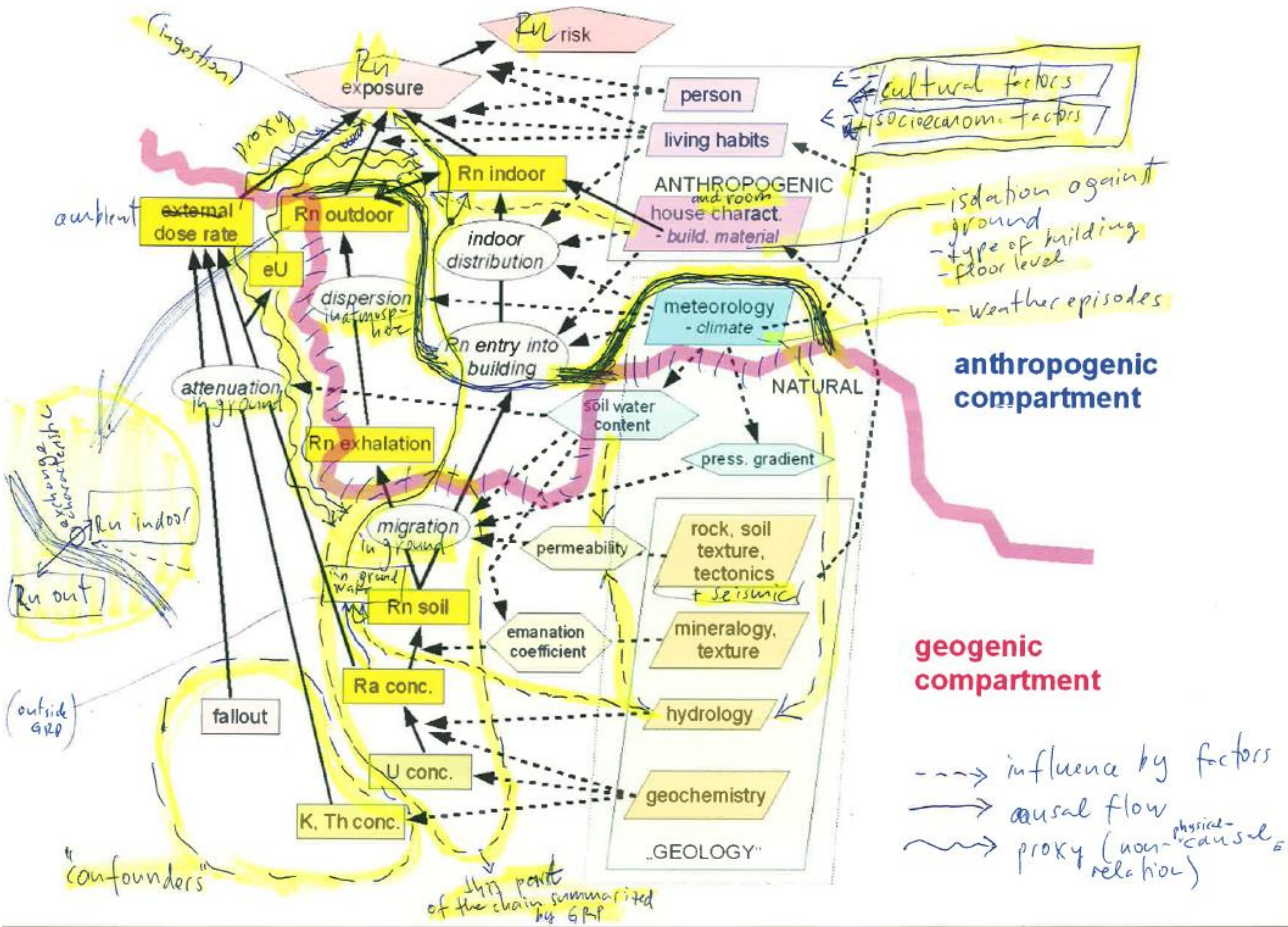
- One quantity whose regional variability represents as much as possible the variability of the geogenic controls of Rn hazard. In other words, these factors shall be squeezed appropriately into one quantity “GRHI”;
- A measure of geogenic Rn hazard which is defined homogeneously across Europe. This means, determine a value of GRHI everywhere in Europe, irrespective of regionally available geogenic databases, but still comparable between any locations. Such GRHI would be the base of a European map of geogenic Rn and a European wide determination of Rn priority areas.

Role of MetroRadon

- Development of the GRHI is one of the objectives of MetroRn! (WP 4.3.4)
  Text being generated in parallel to this presentation.
- Harmonization of geogenic Rn quantification across Europe (~ WP 3.2)
- Possibly harmonized Rn priority areas (delicate subject!) (WP 4.4)

The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States

From rock to risk



Improved version! --- We wanted to make it a bit clearer.

(Draft for the Atlas – still under graphics processing.)

Predictors and proxies

<u>Geogenic quantities of interest:</u>

- Rn concentration in soil gas
- gas permeability
- U concentration,
- ambient dose rate ADR,
- geological units / lithology,
- fault density,
- groundwater recharge coefficient,
- soil properties (texture, humidity,...),
- karstification,
- standardized indoor Rn concentration,
- various geochemical concentrations,
- climate

Geogenic Radon Potential GRP (e.g. Neznal definition); but: available only regionally - CZ, DE, BE, (IT), (ES), (AT), ?

Initial idea (Cinelli et al. 2015)

European Geogenic Radon Map: Multivariate classification approach

European Commission

Grid 10 km x 10 km

Input Variable (i)	SOIL GAS Rn (1)	INDOOR Rn (2)	GEOLOGY (3)	U_{soil} (4) U_{rock} (5)	SOIL Properties (5)	TGDR (7)	....
Classification A B C D	s_1	s_2	s_3	s_4 , s_5	s_6	s_7	.
Weight	$\omega_1(n)$	$\omega_2(n)$	ω_3	$\omega_4(n), \omega_5(n)$	$\omega_6(n)$	$\omega_7(n)$	

Geogenic Radon Risk Index

Low *High*

n – number of samples per grid cell

Joint Research Centre

Desired properties of the GRHI

I. Flexible while consistent: see next slide
  - flexible, i.e. to be applied to as many different situations as possible;
  - consistent: across borders between regions in which different databases are used for estimation; this implies independence of actual database used

II. should reflect as much information as possible

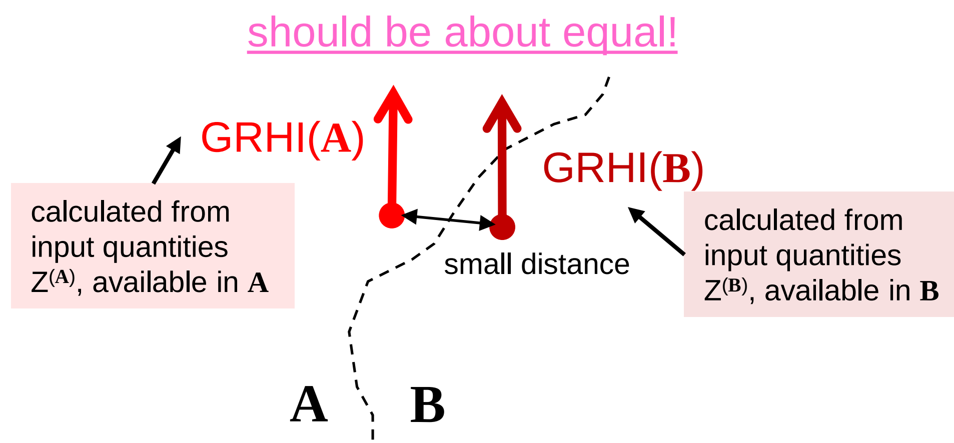
III. should be simple to calculate!

IV. optimal predictor of the geogenic contribution to indoor Rn

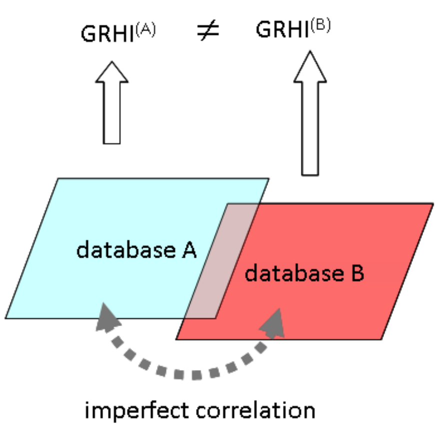
(I) consistency

Its value at a location must be independent on which quantities it has been estimated from.

I.e., GRHI calculated from U concentration in soil should have approximately the same value as if calculated from dose rate or GRP, etc.

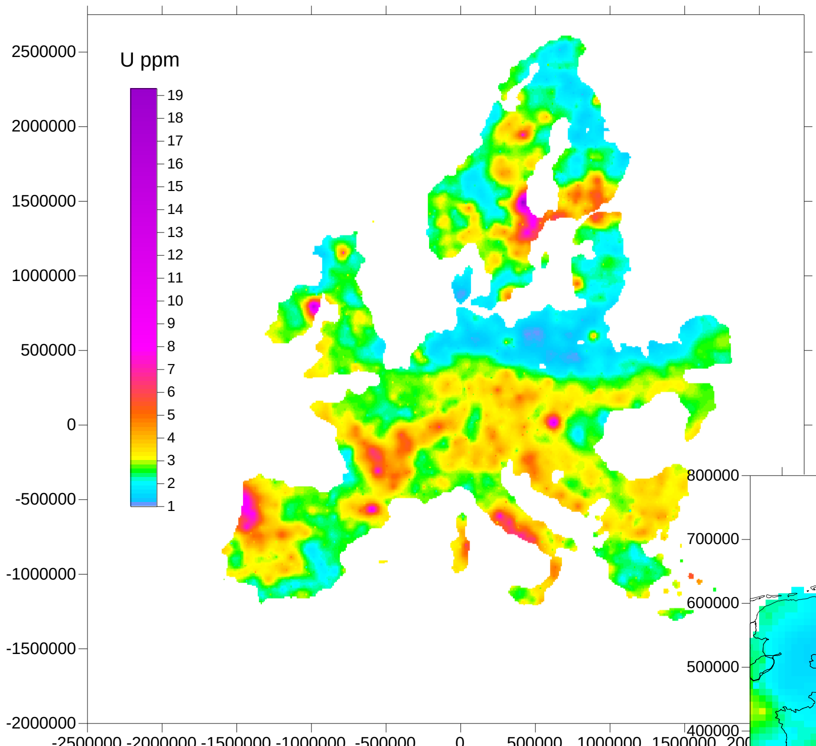


This follows from the requirement to be consistent cross borders, or regions in which different input quantities are available.

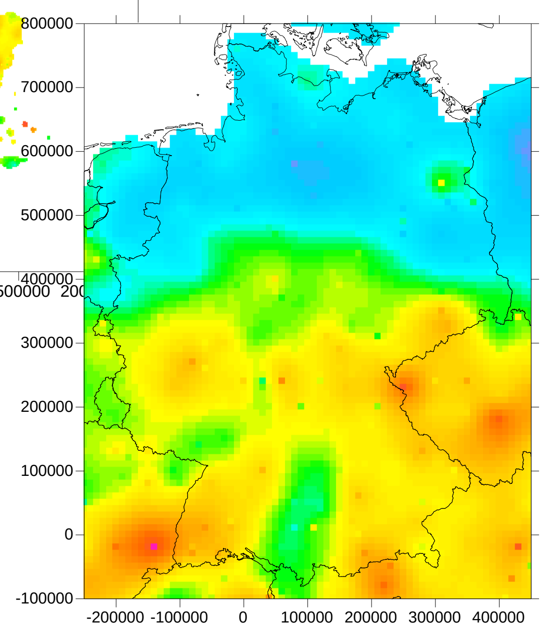


Reason for the lack of consistency between GRHI calculated from different databases.

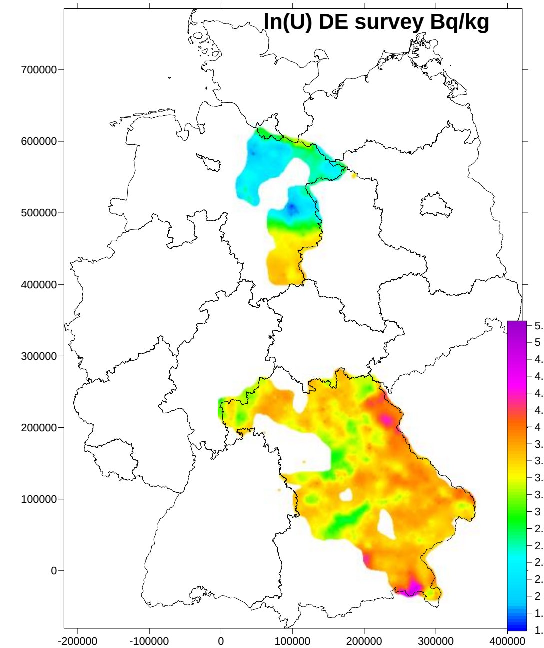
(II) maximal information



FOREGS+GEMAS
(XRF data)



German
incomplete
U survey,
n=2194
(Bq/kg)



regionally less
complete, but
much higher
resolution →
more
information

Approaches, 1

A. “global”

- databases of relevant predictors X_i are available for $\pm$ the entire domain (Europe). From these, a model GRHI = f(X_i) is derived. Option: use regional subsets for calibration / training and validation.
- A1: Machine Learning (ML): Find optimal combination of X_i that best explains IRC. → presentation Eric.
- A2: Dimensional reduction: Point- or cell-wise construction of GRHI by “condensing” the multivariate $\underline{X}$; Either by selection of relevant covariates of by combining them (e.g. as PC) such that they best explain IRC, as in A1. (= Desired property IV !)

Approaches, 2

B. “regional”

- Use regional databases, which ever available in a region. Apply transfer model tailored such that the results match across region borders.
- Methods: Regression, regional dimensional reduction, regional ML.
- Big challenge: consistency (property I.)

This presentation: only A (“global”) discussed!

all approaches

- Models tailored such that they best explain IRC (or derived quantity, such as prob(IRC>RL) etc. = Property IV.
- IRC (or derived) = f(X) + residual ("error");
- Define GRHI := f(X). (Possibly rescale to [0,1] etc.)
- The residuals have essentially 3 sources:
  - data uncertainty;
  - inadequacy of the model (inadequate functional dependence, insufficient predictors)
  - non-geogenic factors! which of course contribute to IRC importantly!

Desired properties fulfilled?

	(I) consistent	**(I) flexible**	**(II) as much inf. as possible**	**(III) easy**	**(IV) optimal pred. of C**
A1 **A2**	by definition	no, because global model	No, because regional data not included; Yes, insofar as structural properties of a large area accounted for	Not really; depending on how done in detail	to be checked by trying!
B	problematic	yes	potentially yes	more or less	

pros - cons

	pro	con
A	• consistent by default → no harmonization problem • Structural properties of a large region accounted for (can learn on a wide variety of situations)	• regionally resolution lower than with regional model; • does not exploit all regionally available data; • not flexible: models has to be calculated from entire dataset; • technically quite complicated
B	• regionally higher resolution; • makes use of more data; • relatively easy	• consistency & harmonization between regions difficult; • regionally valid transfer models must be found.

Option: classification

If predictor Z in class j $\Rightarrow$ then GRHI in class k
Not easy for multivariate, multinomial

- One predictor, GRHI classed
  $\Rightarrow$ logistic regression, ROC;
- Several numerical predictors, GRHI classed
  $\Rightarrow$ discriminant analysis.
- Machine learning:
  Multivariate classification e.g. by random forest technique, support vector machine etc.

Previous attempts

- “Long Way” 2011, sec. 5.4.3, H. Friedmann:
  - Definition of RH from soil Rn and perm; transfer models to estimate soil Rn from U or ADR
  - Classify RH into 3 classes.

- Kraków 2015: Cinelli et al. -- see slide 9!

- TREICEP-5, Veszprém 2016: Bossew et al.
  - transformed variables
  - options: GRHI constructed such that
    (a) covariates considered as proxies or predictors of GRP; or
    (b) covariates should best predict indoor Rn
  - weights:
    (1) through correlations between variables;
    (2) loadings of 1. principal component
  - performance of GRHI assessed as RPA predictor, DE data

- GARRM-13, Prague 2016: Bossew et al.
  - 3 “families” of methods:
    ‘F’: GRHI=mean of distribution functions of covariates;
    ‘R’: GRHI=mean of GRP predicted by covariates through regression;
    ‘P’: 1.PC, as above.
  - performance of RHI assessed as predictor of indoor Rn exceedance probability, DE data;
    no convincing advantage of any method

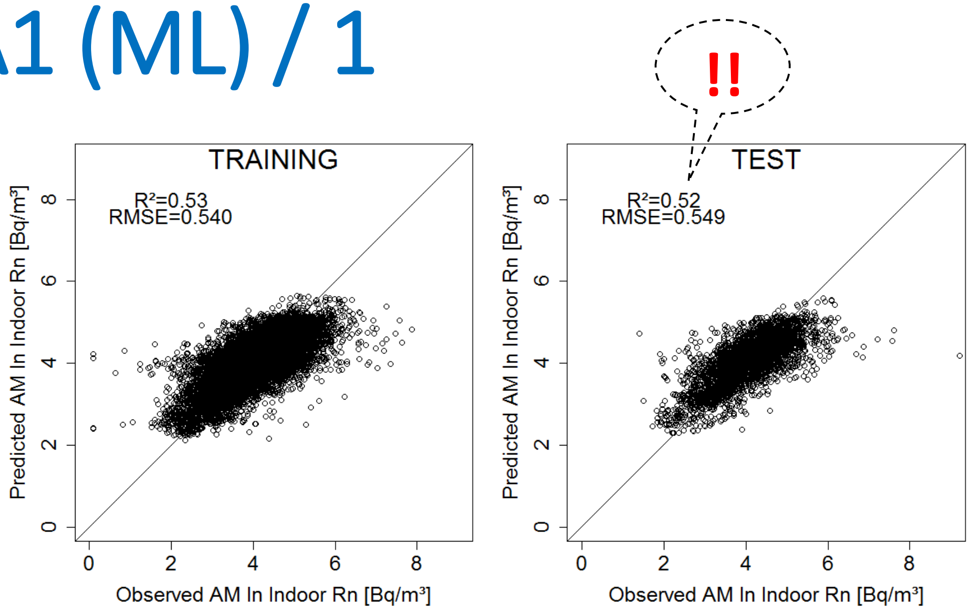
- TEERAS, Sofia 2017: Cinelli et al.
  - Case study Cantabria:
    covariates: soil Rn, GDR, fault density, U in soil, lithology, permeability, karstification
  - weights: correlation with indoor Rn; GDR and U excluded
  - 3 “hazard classes”: if prob(C>300), estimated from GRHI, >0.1 → high; if prob(C>100)<0.1 → low; otherwise medium.
  - Performance through underestimation rate (2.kind error): 7%

- IWEANR, Verbania 2017: Ciotoli et al.; Bossew et al.
  - Case study DE: GRHI constructed from European (Atlas) and DE data. Comparison with GRHI derived
    from GRP shows error with regional trend → not desirable
  - RHI map constructed of PCA(U, Th, K, fault density, heat flow, seismic density, soil fine fraction)

Example A1 (ML) / 1

Method: MARS regression

- Target variable Y: arithmetic mean of ln (Indoor Rn conc.)
- Only cells with n>30 were used
- 17,018 cells with data remaining



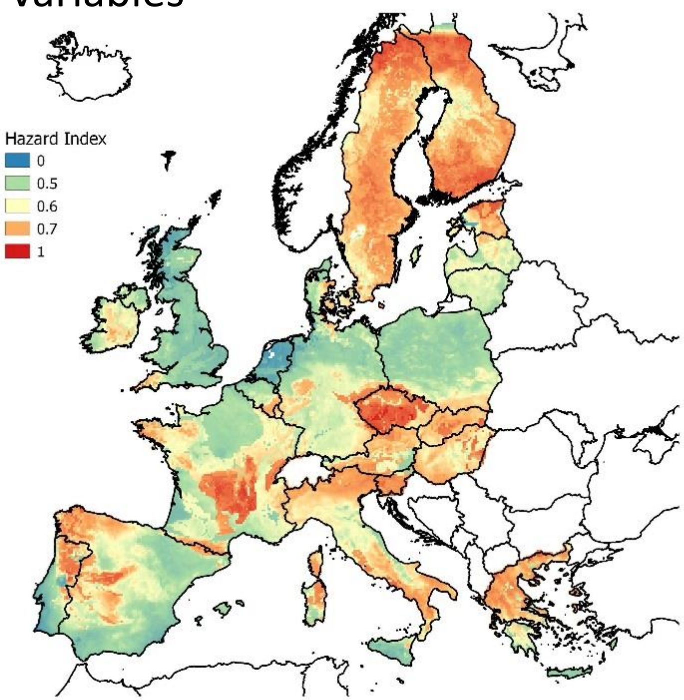
Input data/predictors

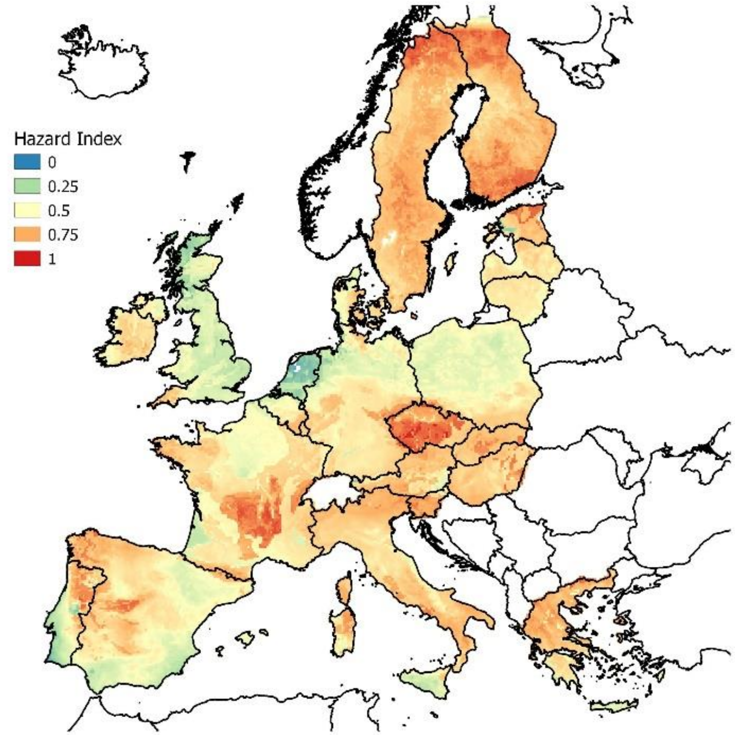
Data source	Informative	Not informative
Geology (IGME 1:5M)	**Petrography**	Distance and type of fault
Hydrogeology (IHME 1:1.5M)	**Hydrogeological classes**	Karstification
Soil type	**Soil types**	-
LUCAS topsoil survey	**Silt, Available water capacity,** Clay, coarse fraction, Bulk Density	Sand, Texture class,
Coordinates	**X, Y**	-
Soil hydraulic properties	Hydraulic conductivity	-

Example A1 / 2

Tentative GRHI maps by ML

- grid: 10 km × 10 km
- Cells in prediction grid represent
  - dominant categorical variable (covering the most of the cells area; mode)
  - Arithmetic mean of numerical variables





- target = AML
- linear rescaling to values between 0 (min) and 1 (max)

← different color ramp to emphasize regional differences

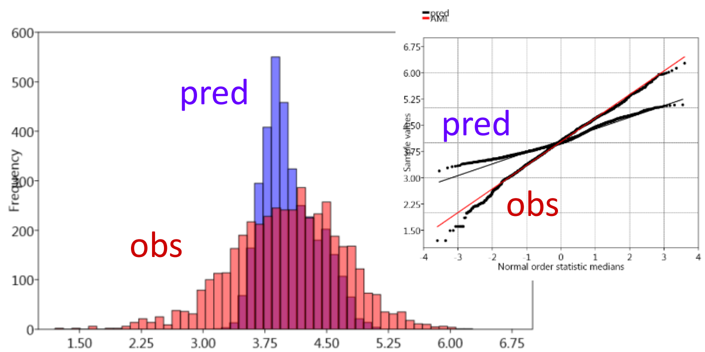
Example A2 (simple regression) / 1

- **Multivariate input data X:**

- starting with 100 covariates:
  -- Geochemistry: FOREGS + GEMAS, 59 elements ;
  missing U estimated by La and Ce; 4982 data points
  -- soil properties: from LUCAS, projected to geochem. data points
  -- FF:=(clay+silt+.05*sand)/(100+CF) (Whether this is a reasonable def. as perm proxy -- ?)
  -- geology: IGME 5000
- Used for further analysis:
  pH, TOC, FF, CF, bulk dens, AWC, ln(U), K_2O, Al_2O_3, SiO_2, Fe_2O_3, CaO, geo1
- geo1:={Carbonate, Meta-sediments, Siliciclastics, Tert/Quart sediments, Igneous basic, Ign intermediate, Ign acid pre-Variscan, Ign acid Var., Ign acid post-Var.} … strongly simplified from IGME 5000. (Perhaps too much simplified!)

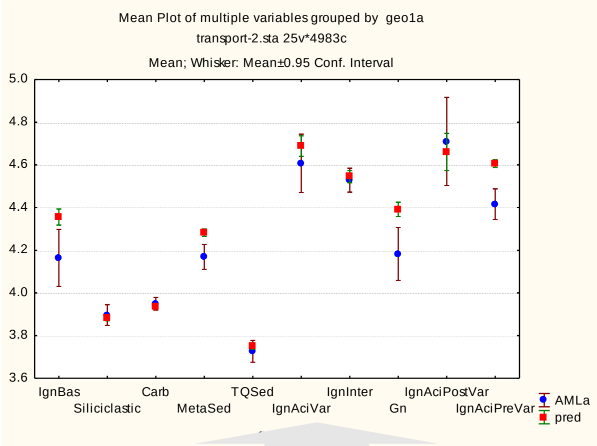
- **Target variable** Y=AML (Atlas cells); interpolated from Atlas to geochem. locations, i.e. hypothetical AML of 10x10 km² squares on geochem. data locations.
- Dim. reduction:
  finding combination of X by PCA which best predicts Y was not successful;
- By trying GLM $\rightarrow$ geo1, FF, pH, dens., K_2O, ln(U): best predictor of AML; r^2=26% only
  Including "annual mean temperature" as proxy of anthrop. factors would lead to 29%.
- f(X) $\rightarrow$ OK (Atlas cells, i.e. 10 km resolution) $\rightarrow$ rescale [0,1] …. this defines the GRHI.

Example A2 / 2

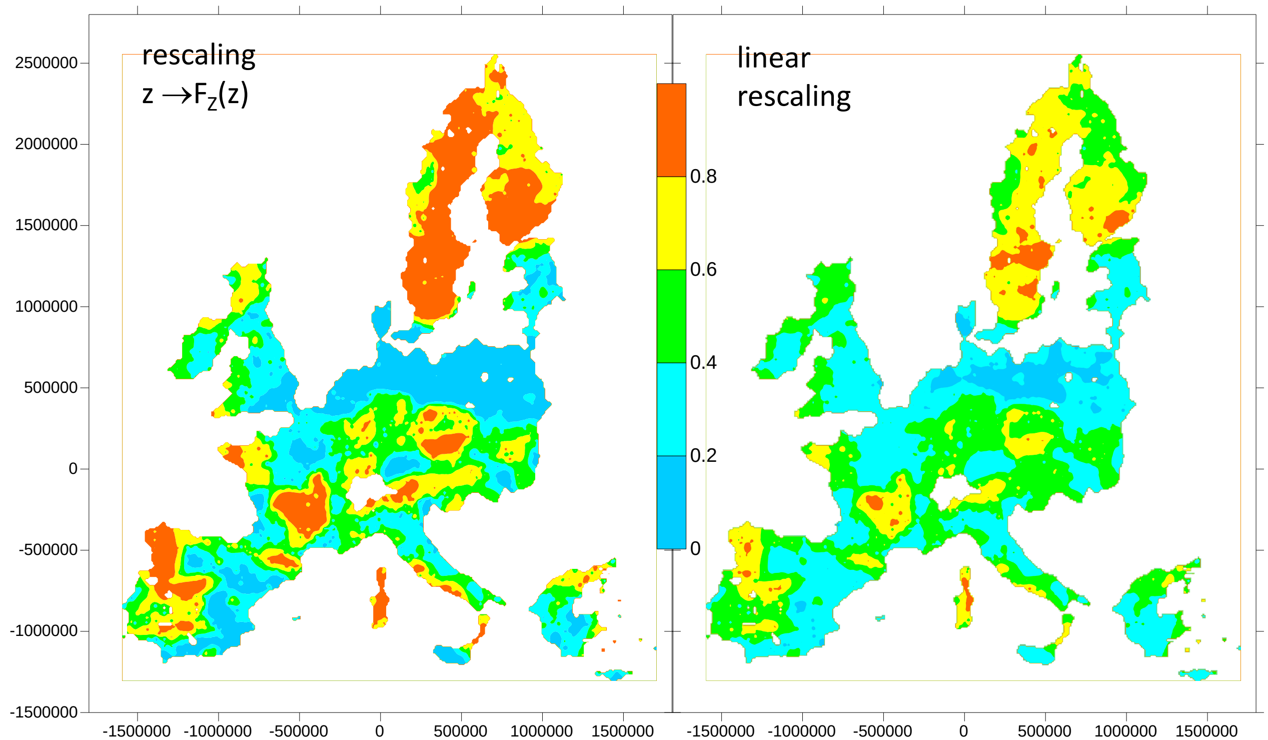


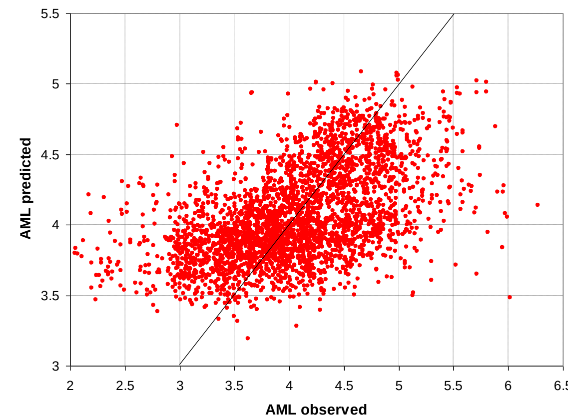
both about normal; can be rectified by rescaling

tentative GRHI maps based on simplified geology, U, K, soil parameters which to some degree substitute permeability;
Made simply by GLM



geological grouping reasonably well preserved





prediction of AML by geogenic predictors moderately good

definition of rescaling is relevant!

Conclusions

- Global approaches (A1 and A2) seem promising in spite of disadvantages.
- ML approach (A1) leads to very high r^2;
  simple GLM approach (A2) easy to perform, but lower r^2.
  Currently ML seems better choice.
- Still badly missing:
  Gas permeability data, European scale:
  We have data on texture and other soil properties, but one essential parameter missing: mean soil humidity.
- Unclear: improvement by using CoDa for geochem. data? Theoretically sounder, but more complicated.
- Variable construction by PCA: theoretically more satisfactory, but not successful here.

To-do

- Try different ML methods;
- For methods based on dim. red.: further explore methodology, selection of input variables
- Establish criteria for model performance: r^2 etc, but also compliance to desired properties
- Explore classification approaches → GRPA maps!
- Compare with regional approach (B)

Thank you!

Bundesamt für Strahlenschutz

European Commission

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

AGES

Consiglio Nazionale delle Ricerche

Institute of Environmental
Geology and Geoengineering

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