kernel regression based mapping of indoor radon concentrations in switzerland

G. Kropat, F. Bochud, M. Gruson, C. Murith, J.-P. Laedermann, S. Baechler

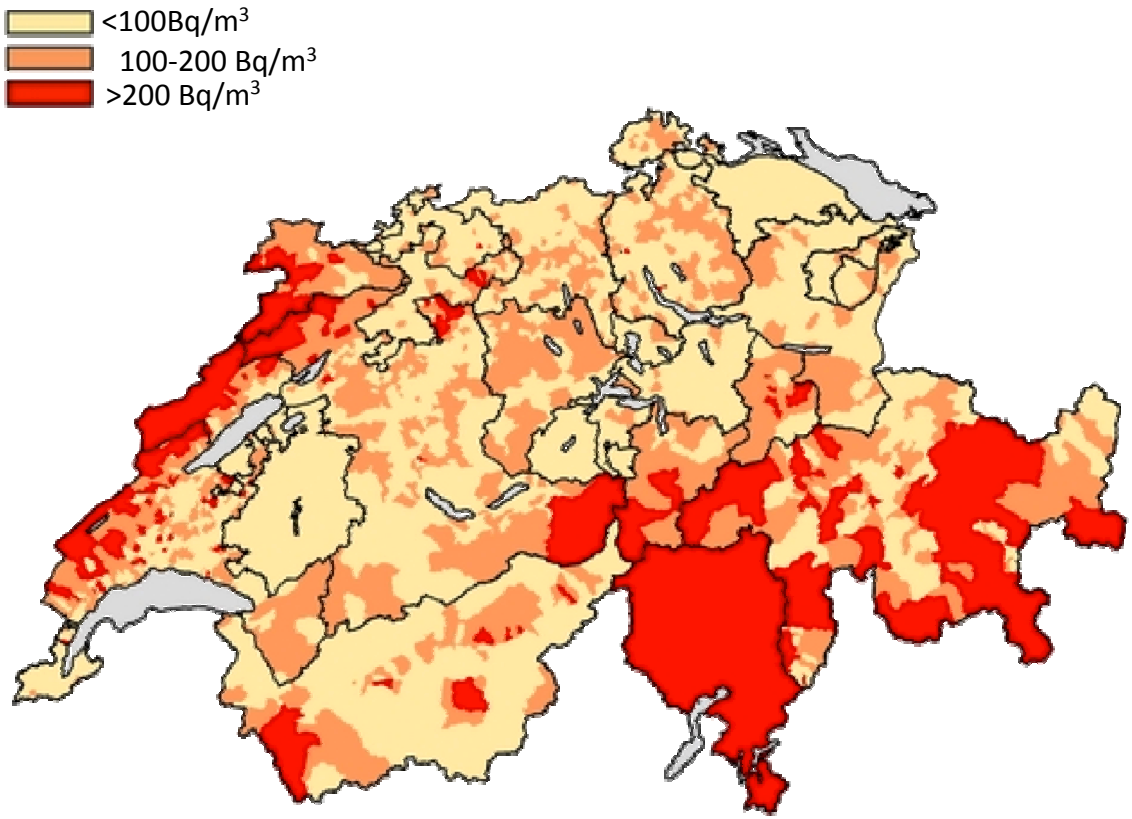
Institut de radiophysique

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de l'intérieur DFI
Office fédéral de la santé publique OFSP

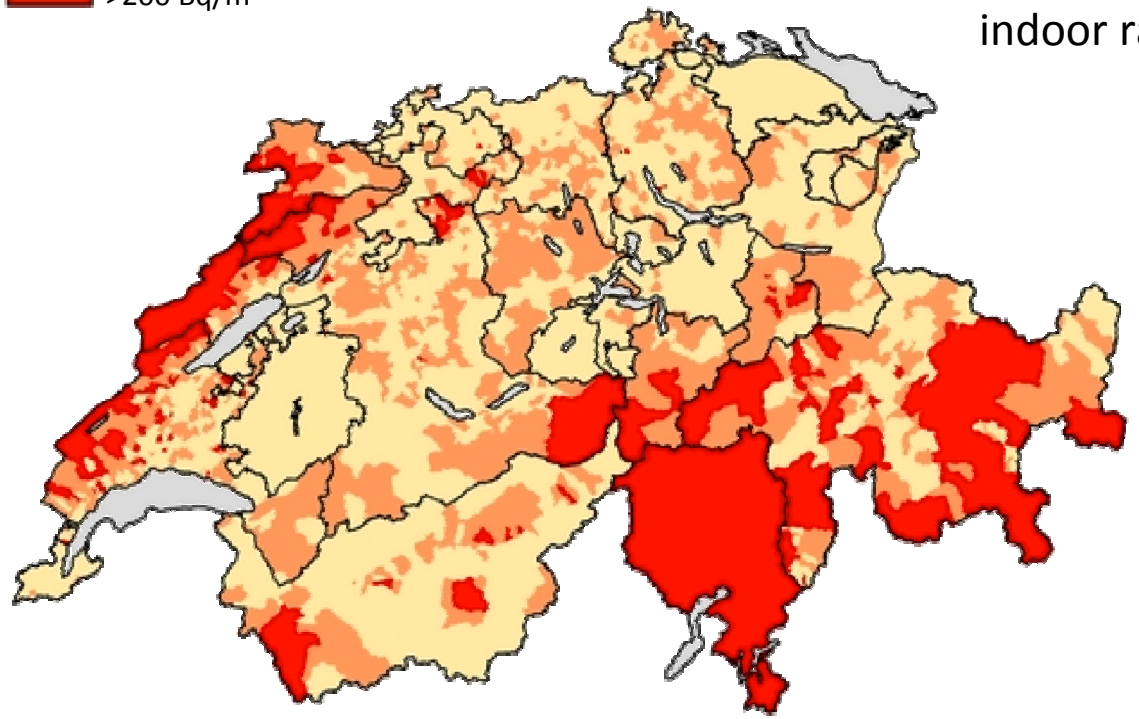
current map in switzerland

intro



current map in switzerland

intro

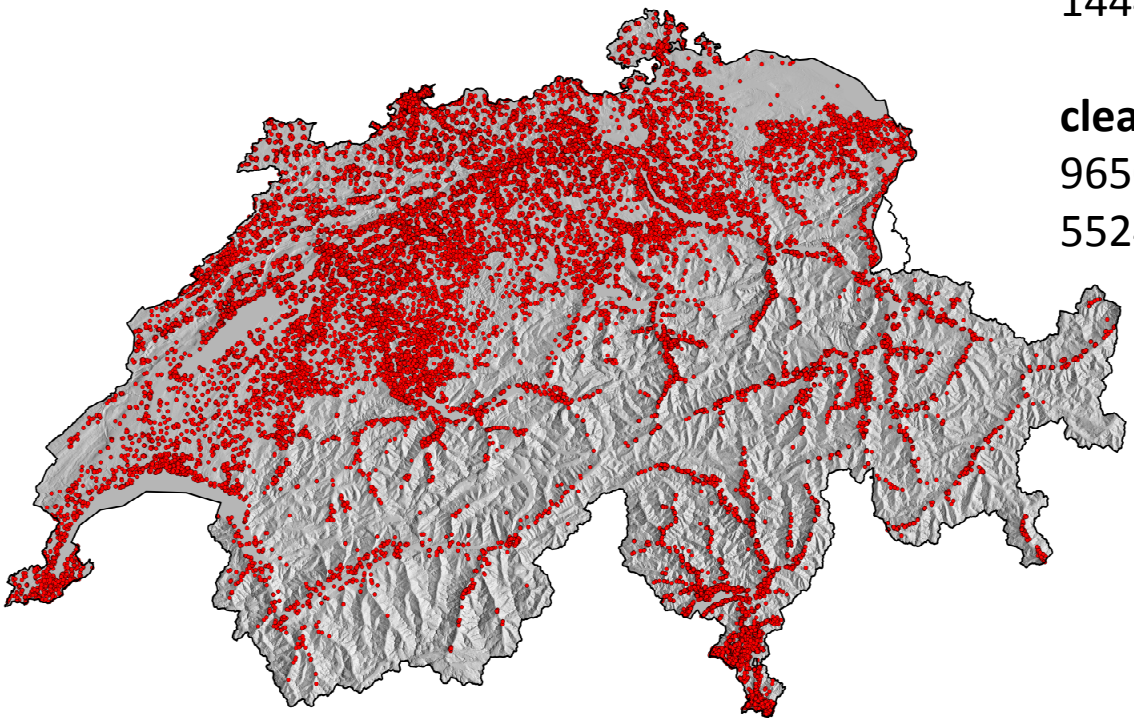


aim of project:
find best prediction for
indoor radon concentrations

measurements

data

- radon measurement sites

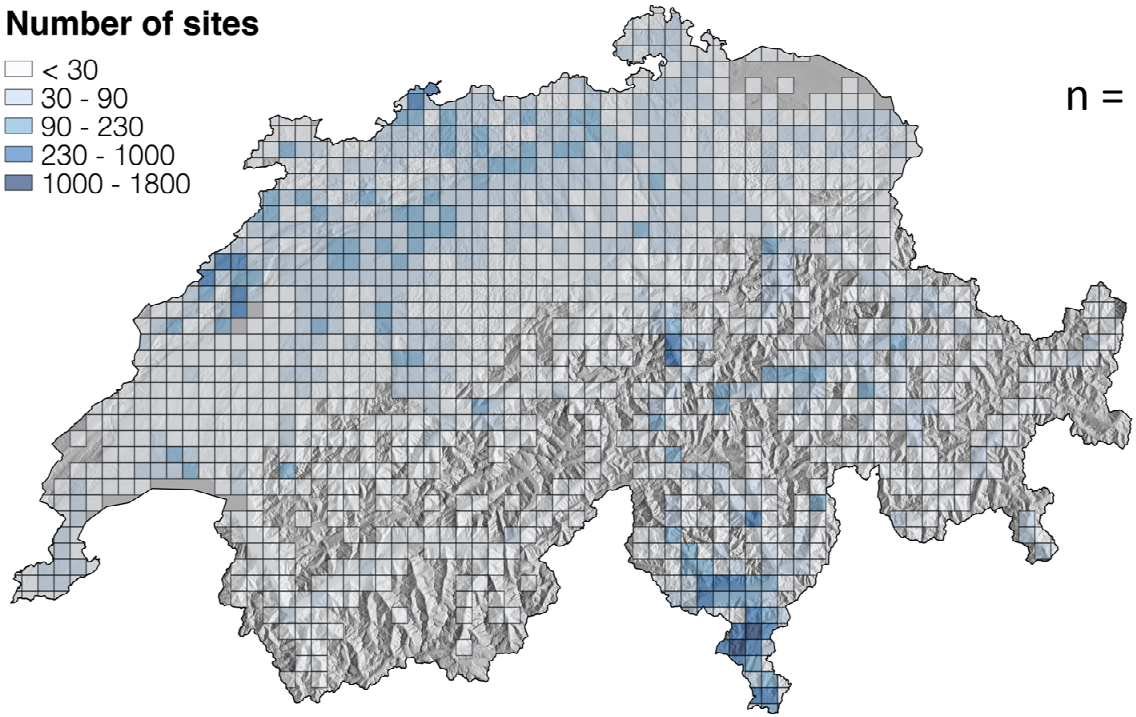


raw database:
229215 measurements
144407 houses

cleaned database:
96508 measurements
55248 houses

sampling density

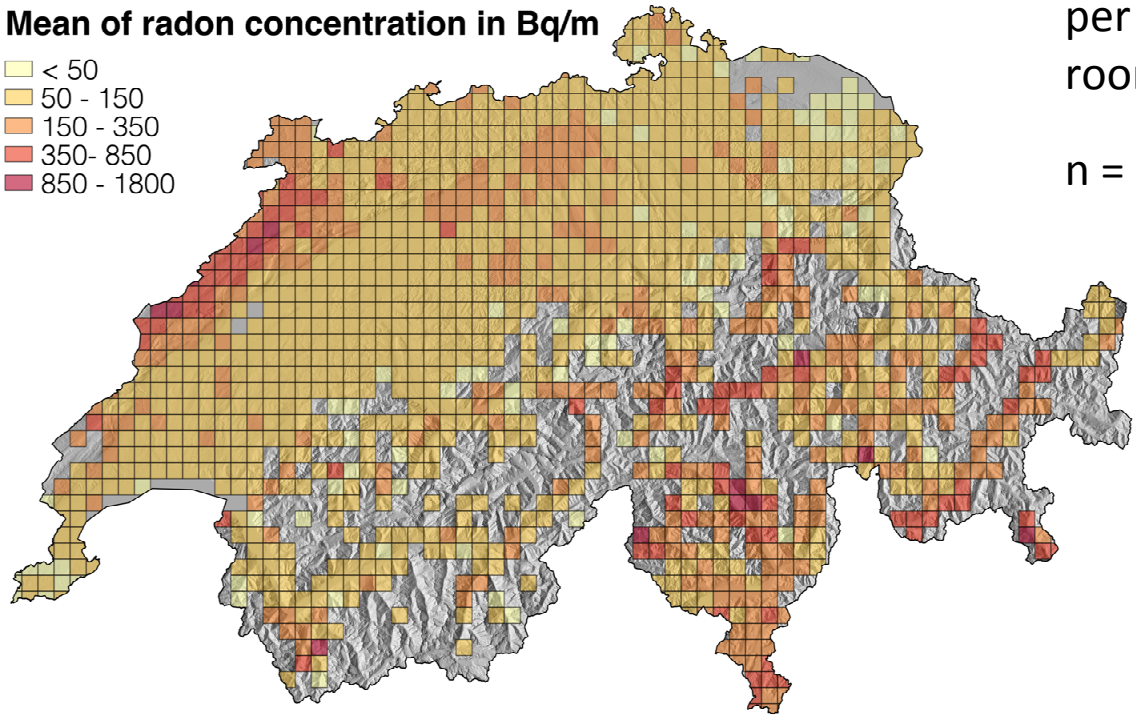
data



n = 52211

local sample means

data

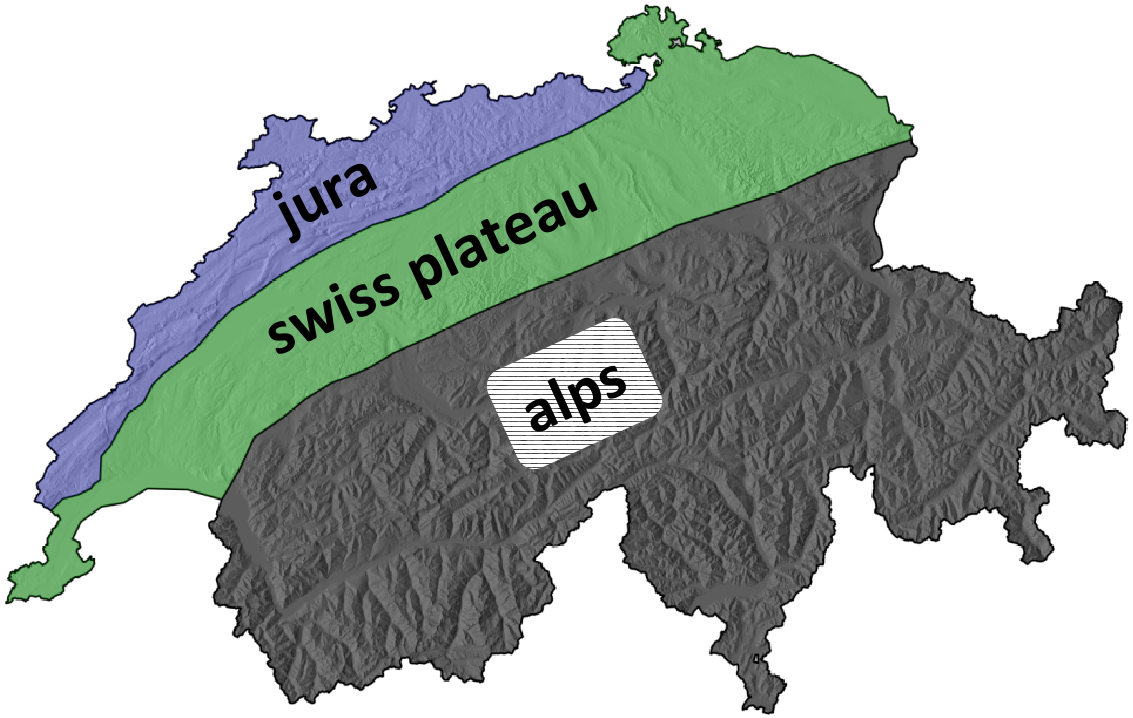


maximum measurement per house in inhabited rooms

n = 52211

geological regions in switzerland

data



available radon influencing variables

data

continuous variables

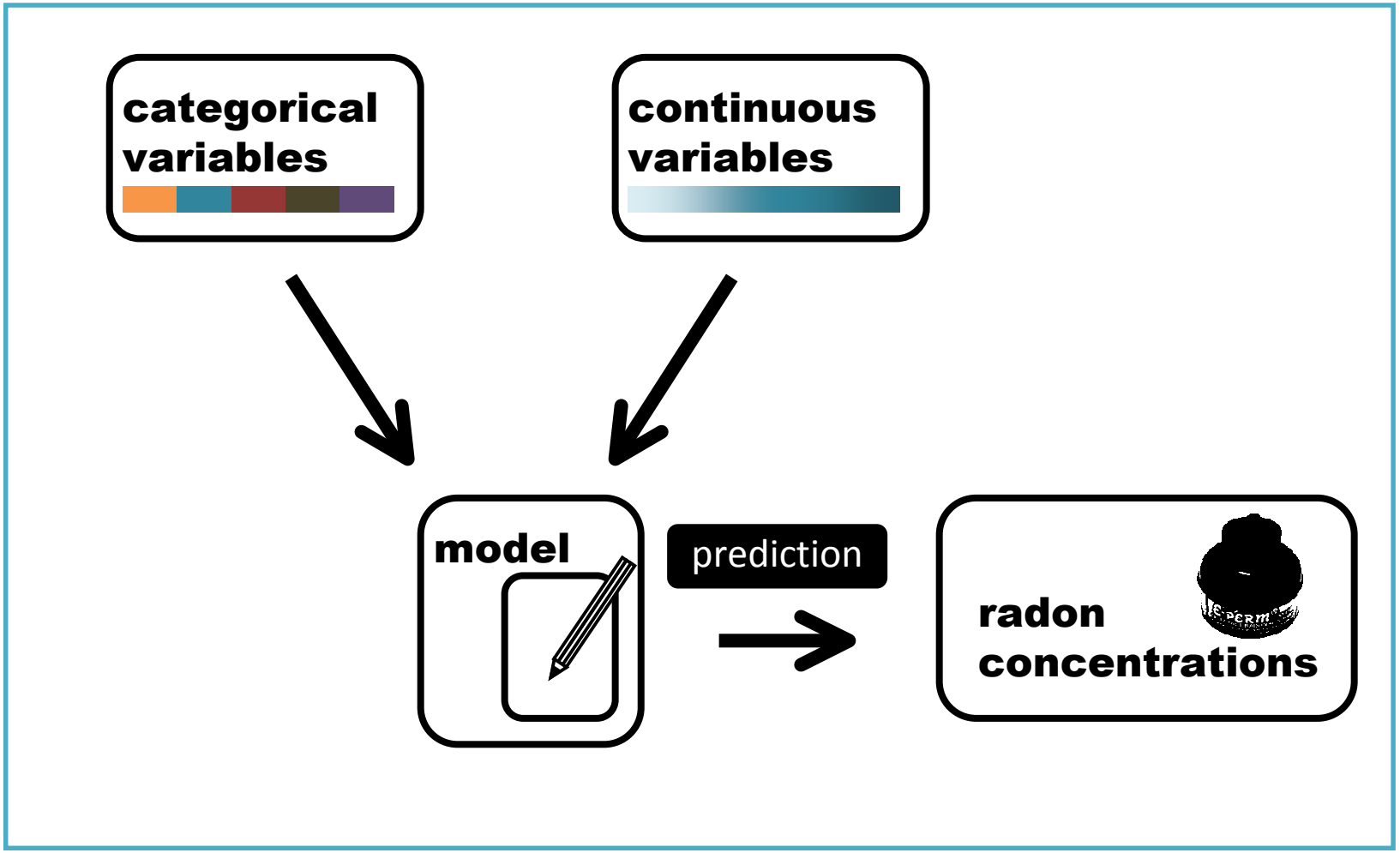
- coordinates 
- altitude 

categorical variables

- housetype 
- floorlevel 
- foundation 
- roomtype 
- geology 
- inhabited or not 

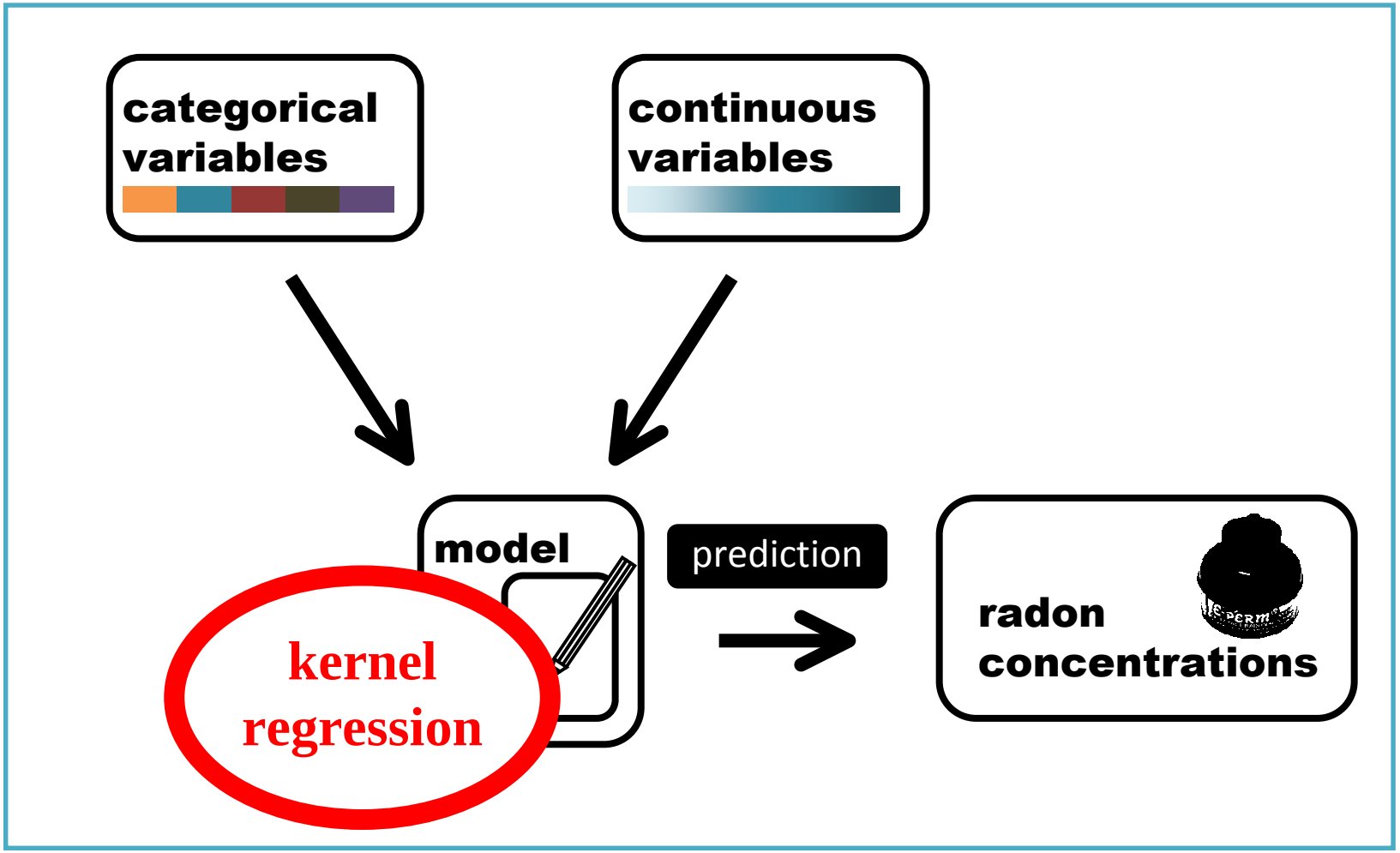
find model to predict radon

method



find model to predict radon

method



kernel regression

method

idea is to find conditional expectation value $E\left(Rn|x_1,\ldots,x_n\right)$

$$x_1 = coordinateWE$$
$$x_2 = coordinateSN$$
$$x_3 = floorlevel$$
$$\vdots$$

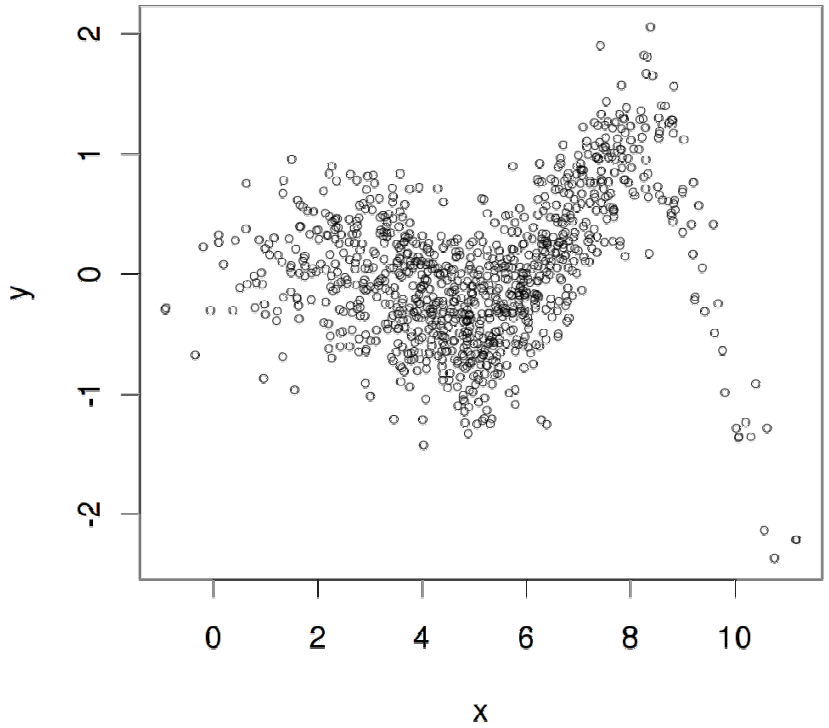
$$E\left(Rn|\vec{x}\right) = \int_{-\infty}^{\infty} Rn f\left(Rn|\vec{x}\right) dRn \qquad f\left(y\,|\,x\right) = \frac{f\left(y,x\right)}{f\left(x\right)}$$

$$= \int_{-\infty}^{\infty} Rn \frac{f\left(Rn,\vec{x}\right)}{f\left(\vec{x}\right)} dRn \qquad f\left(\vec{x}\right) = \int_{-\infty}^{\infty} f\left(Rn,\vec{x}\right) dRn$$

kernel regression example

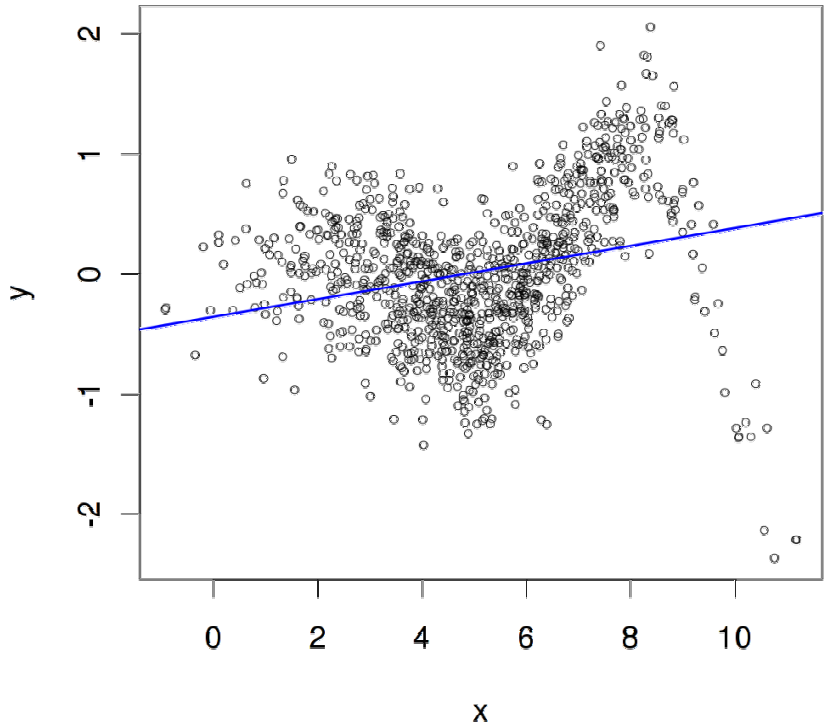
method

- “unknown” process



kernel regression example

method

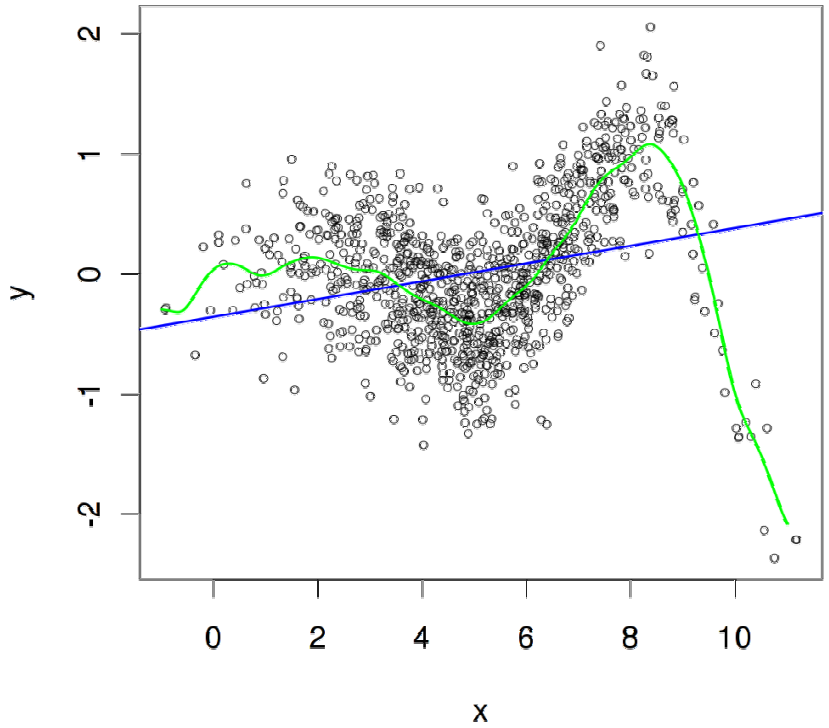


“unknown” process

linear regression

kernel regression example

method



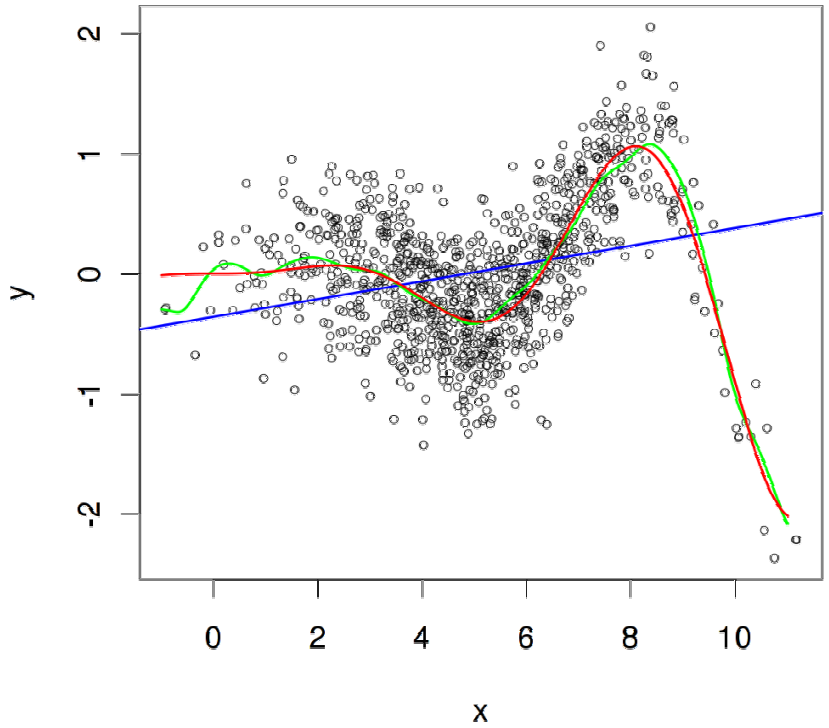
“unknown” process

linear regression

kernel regression

kernel regression example

method



“unknown” process

linear regression

kernel regression

real curve

Advantages of kernel regression:

- non linearity
- multivariate
- mixed models with categorical and continuous variables

geological features

results

geologie scale 1 : 500 000

lithology scale 1 : 500 000

pedology (soil texture) 1 : 1 000 000

stratification to earth foundation

five fold cross validation

variables included into model or not						validation
longitude	**latitude**	**altitude**	**geology**	**lithology**	**soil texture**	R^2 (%)
yes	yes	yes	no	no	no	**24**
yes	yes	yes	yes	no	no	**25**
yes	yes	yes	no	yes	no	**25**
yes	yes	yes	no	no	yes	**24**
yes	yes	yes	yes	yes	yes	**25**

different foundation types

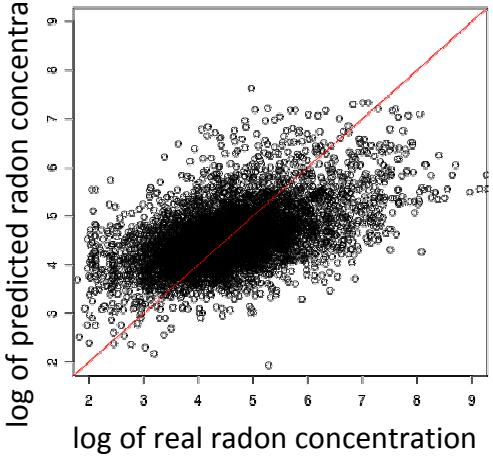
results

variables:

coordinates	year of construction	geology	foundation
altitude	roomtype	lithology	
housetype	dosimetertype	soil texture	

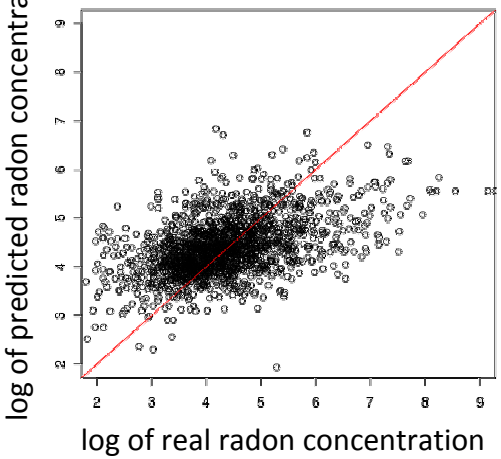
different foundations

R^2=28% N_{test}=6983



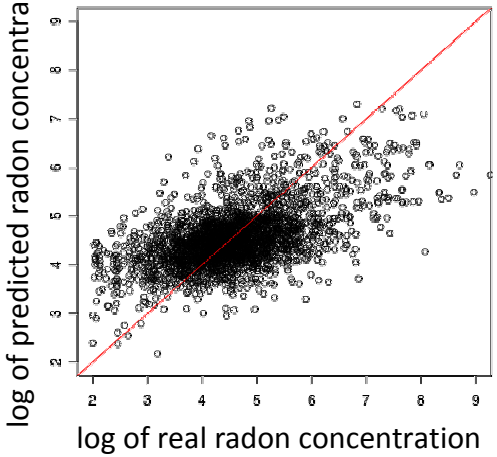
concrete foundation

R^2=23% N_{test}=1993



earth foundation

R^2=31% N_{test}=3098



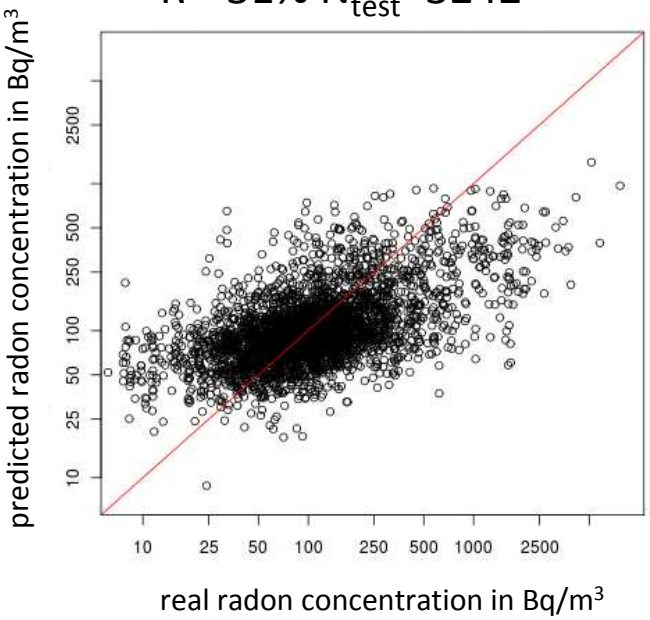
different housetypes

results

model learned on measurements only in houses with earth foundation

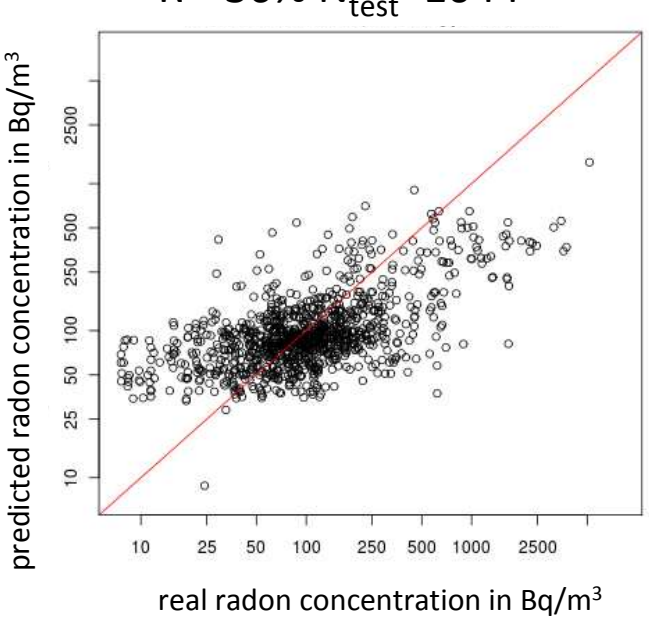
several housetypes

R^2=31% N_{test}=3242



farms

R^2=36% N_{test}=1044



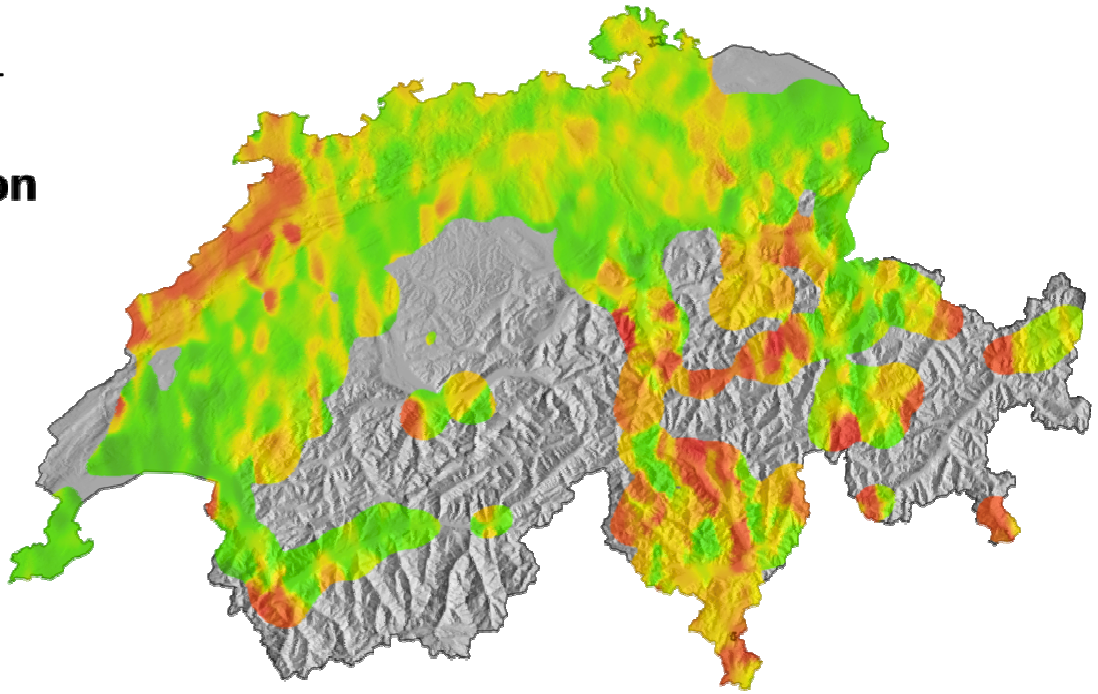
mapping

results

prediction for living rooms
in groundfloor for detached house
built between 1900 – 1970
with earth foundation

training data N=14084

no prediction carried out at points
where mean distance to next 20
nearest neighbours is more than 8km



conclusions

- kernel regression performs interpolation of indoor radon data by accounting for categorical variables
- in case of data on coarse scale geology, lithology and soil texture don't seem to add information to model
- radon concentrations are better predictable in houses with earth foundations than with concrete foundations

conclusions

- kernel regression performs interpolation of indoor radon data by accounting for categorical variables
- in case of data on coarse scale geology, lithology and soil texture don't seem to add information to model
- radon concentrations are better predictable in houses with earth foundations than with concrete foundations

Thanks for your attention!