Examples and considerations of spatial prediction modeling of geogenic groundwater contamination

Joel Podgorski, Dahyann Araya, Michael Berg

18 May 2022

eawag
aquatic research
Swiss Federal Institute of Aquatic Science and Technology

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development
and Cooperation SDC

GAP
Groundwater
Assessment
Platform

Geogenic (natural) groundwater contamination

- Up to 10% of the global groundwater is contaminated by high levels of arsenic (As) or fluoride (F)
- Over 400 million people globally are exposed

Arsenic (As)
causes skin lesions,
pulmonary disorders,
internal cancers

WHO: 10 μg/L

~220 million people

Fluoride (F)
causes dental mottling &
decay,
crippling skeletal deformation

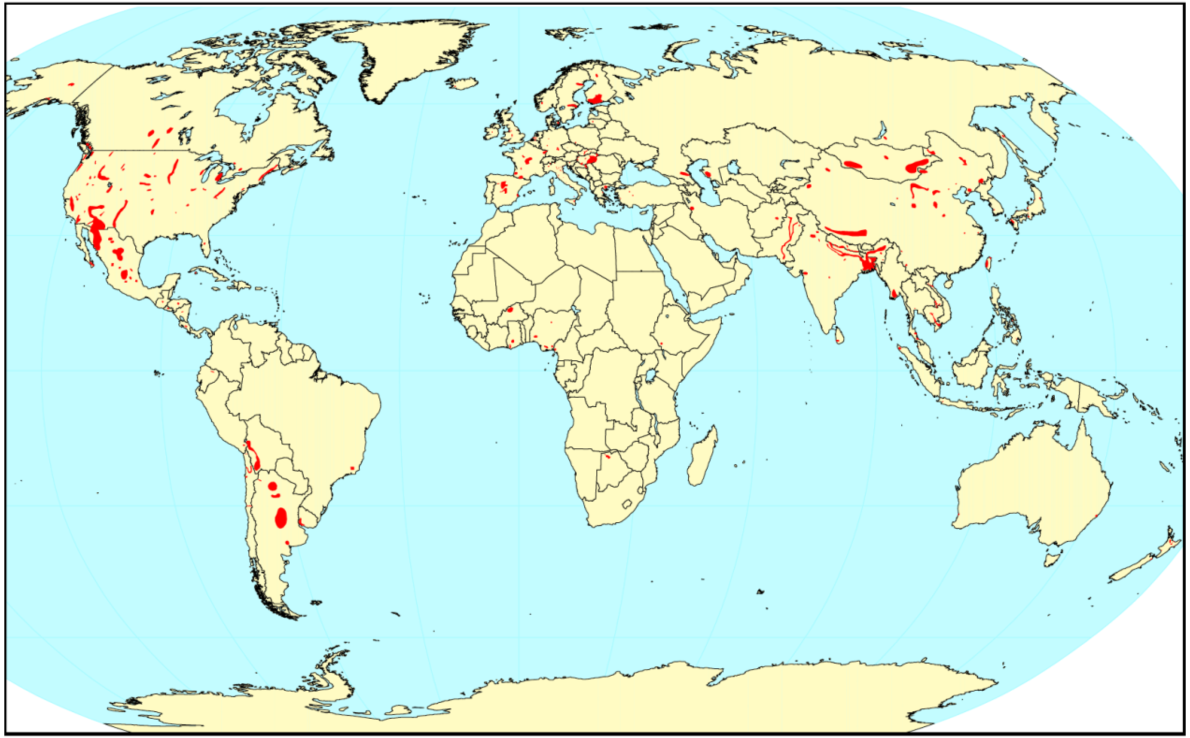
WHO: 1.5 mg/L

~200 million people

→ Geochemistry, occurrence and distribution, impact on population, mitigation measures

Geogenic arsenic in groundwater resources

- Naturally occurring around the world
- Exposure via drinking water & crops
- Difficult to detect/measure
- New affected areas continually found



Ravenscroft et al., 2008

Common conditions

Overarching factors influencing As groundwater contamination

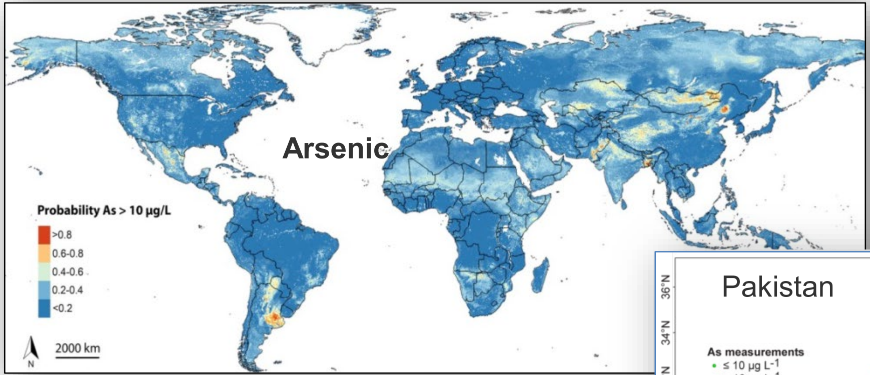
- Low-oxygen (reducing) aquifers
- Organic-rich sediments
- Young sediments (Holocene)
- Low-lying areas (predominantly young river deltas)

➡ Is it possible to forecast water quality and delineate safe and unsafe areas ??

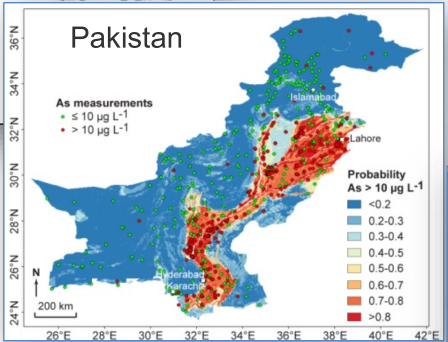
Southeast Asia delta plains

Hug et al., ES&T **2008**

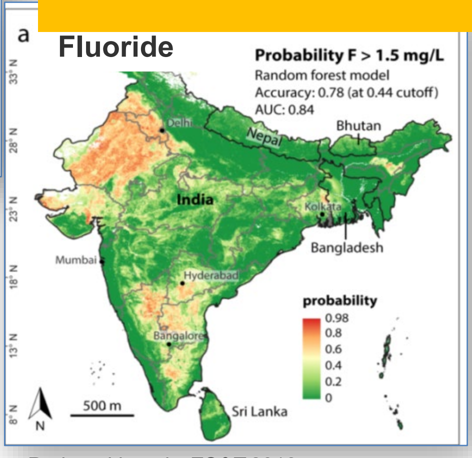
Prediction map examples



Podgorski et al., *Science* 2020



Podgorski et al., *Science Advances* 2017



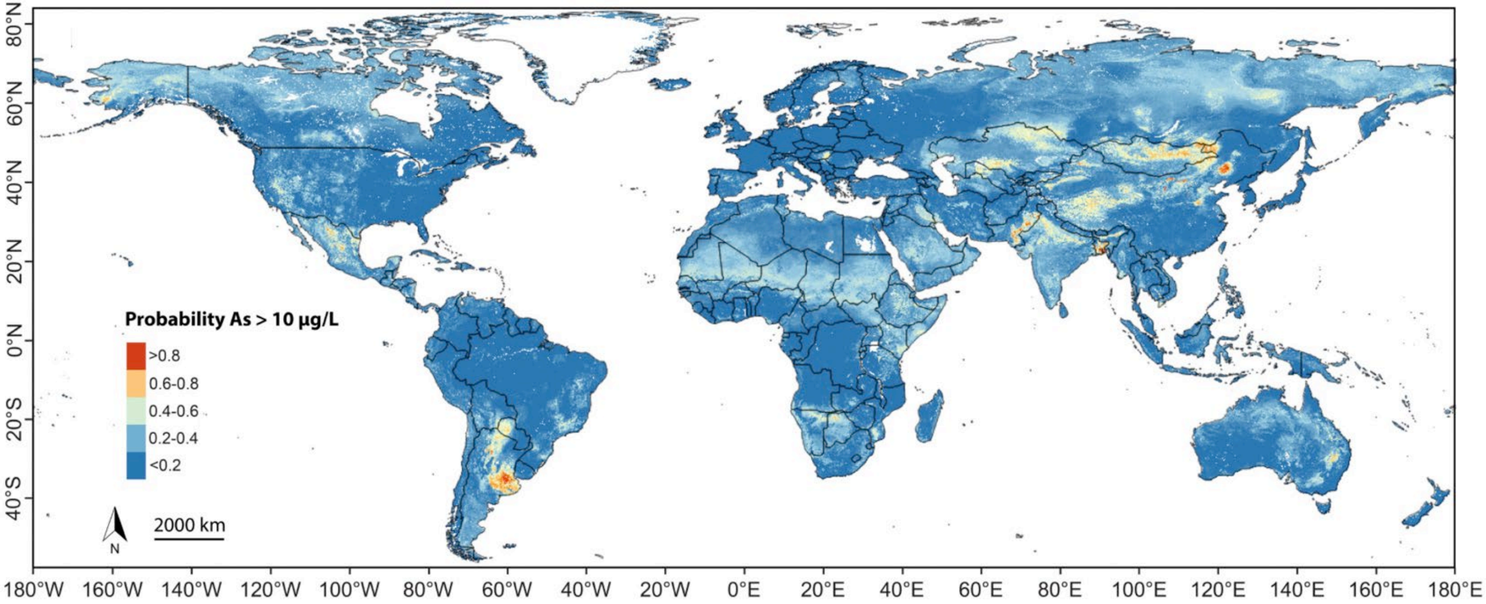
Podgorski et al., *ES&T* 2018

Impacts:
- Increased awareness
- Political discussions
- Follow-up studies
- Targeted sampling campaigns

Predictor variables:
- Geology
- Soil pH
- Soil texture
- Soil type
- Aridity
- Slope
- Cropland
- Irrigation

Groundwater arsenic model (10 µg/L)

Created with machine learning (random forest)



Podgorski and Berg, *Science* **2020**

11 independent variables:

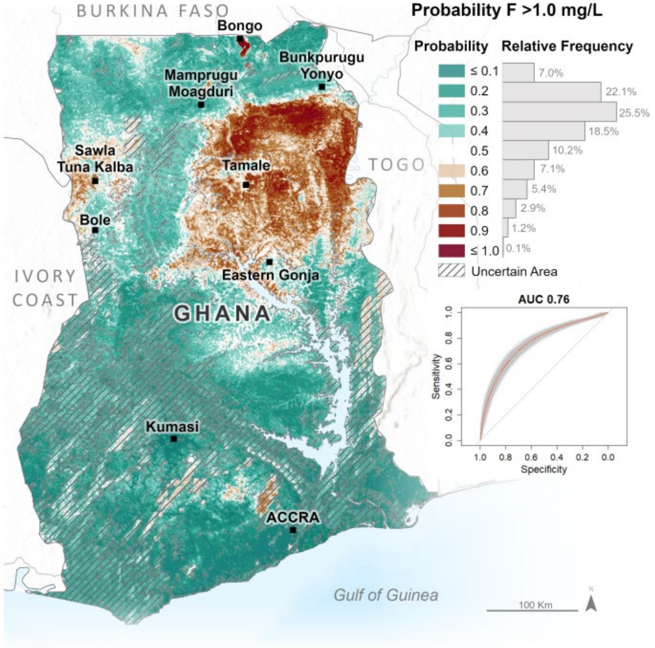
Climate: actual evapotranspiration, aridity, potential evapotranspiration, precipitation, Priestley-Taylor coefficient, temperature

Soil: clay content, fluvisols, soil pH, sand content

Topography: topographic wetness index

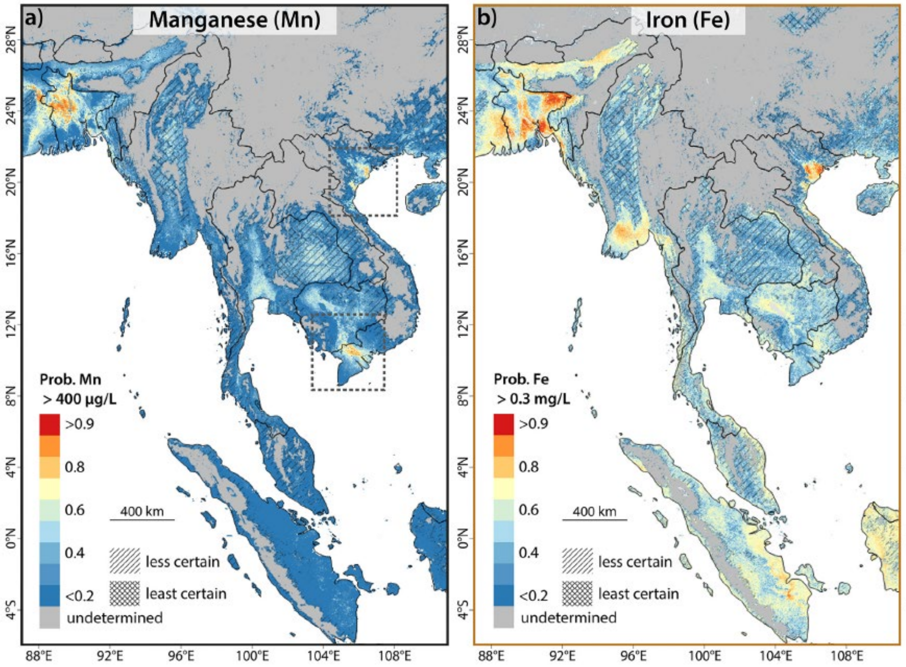
Recent work

Fluoride in Ghana



Araya et al., *Water Research* 2022

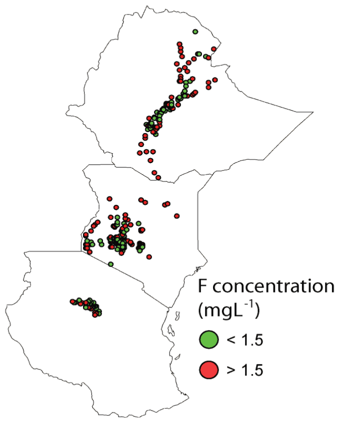
Manganese and iron in Southeast Asia



Podgorski et al, *STOTEN* 2022

Concept of geostatistical prediction modelling

Concentration data → Geo-statistical modeling → Identifying risk areas → Mitigation



Groundwater data

GIS data

Combining geochemical and statistical expertise

Risk maps

Relevant geospatial data
- Increasingly available in digital GIS format
- Often free of charge
- Resolution and coverage increasing

General modeling process

1. Collect data points of contaminant concentrations
   - Cover area of map to be developed
   - Possibly convert to binary format
   - Should cover range of values
2. Assemble predictor variables
   - Based on known or suspected relationships with contaminant
   - Spatially continuous
   - Should contain wide range of values
3. Statistical modeling (e.g. R, python)
   - Free, many different packages available
4. Apply model to predictors -> create map

Statistical modeling algorithms

Find relationships between predictor variables and contaminant concentration

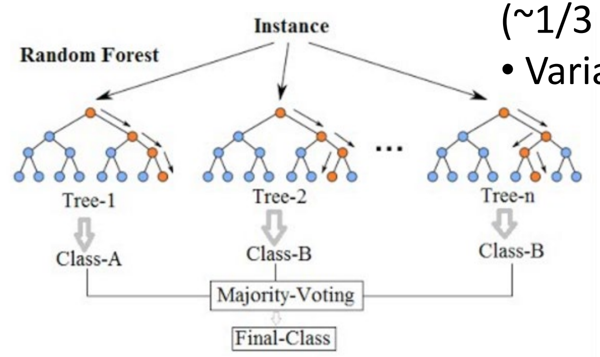
→Logistic regression

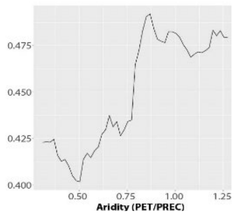
$$P = \frac{e^{(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}}$$

P: probability
$x_1, \ldots, x_k$: predictor variables
$\beta_1, \ldots, \beta_k$: regression coefficients

→Random forest

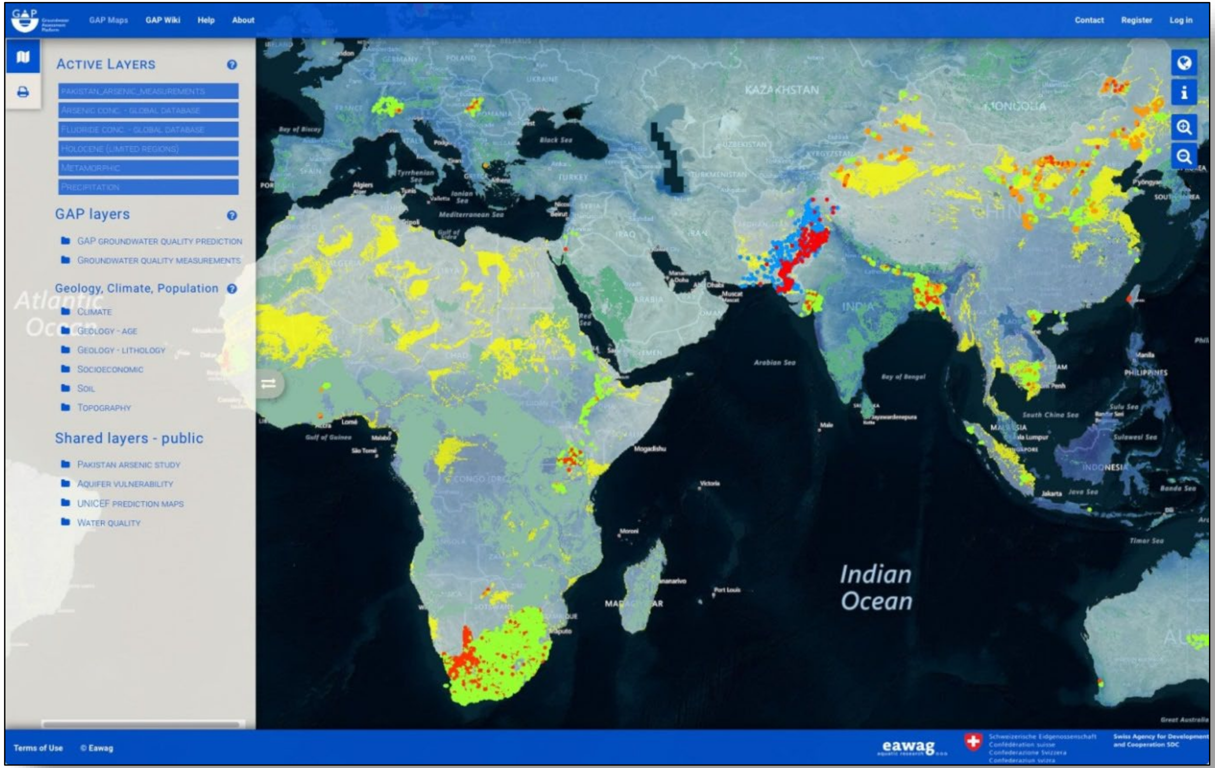
- Ensemble (average) of decision trees
- Randomness introduced to avoid overfitting:
  - random subset of possible predictor variables at each node
  - random selection with replacement of data rows used in each tree (~1/3 of data row not present in a given tree)
  - Variable importance, partial dependence plots





Also: generalized boosted regression models, neural networks, support vector machines eawag aquatic research

GAP Groundwater Assessment Platform



- Visualize groundwater data
- Upload and view own data
- Create own predictive maps
- Share data and models with other users or publicly

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development and Cooperation SDC

www.gapmaps.org

Thank you

joel.podgorski@eawag.ch

eawag aquatic research

GAP Groundwater Assessment Platform