SIPTA The Society for Imprecise Probabilities: Theories and Applications

Application of uncertainty theory in the field of environmental risks

Dominique Guyonnet, BRGM

September 27th 2023

Geoscience for a sustainable Earth
brgm

1. Introduction
2. Communicating on information theories
3. Examples of applications of IP to environmental problems
    - 3.1 Imprecise soil quality objectives
    - 3.2 Leakage from a heap-leaching operation
    - 3.3 Interpolation of SIC data (Sparse, Imprecise, Clustered)
4. Communicating on IPs
5. Conclusions

1. Introduction

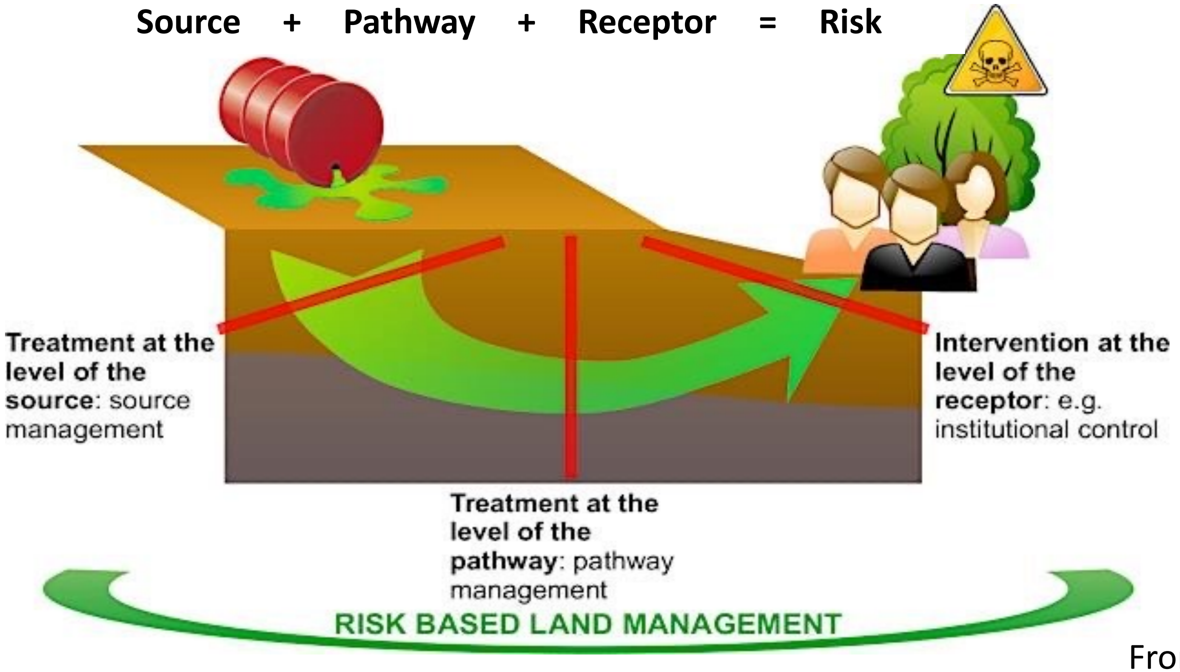
- This material is based primarily on methods developed in the nineties.
- My first encounter with epistemic uncertainty methods:
- General framework for the assessment of risk for health of humans or ecosystems:

Ensembles flous, nombres flous, et incertitudes dans les sciences de la terre

B. CÔME
Commission des Communautés Européennes Rue de la loi, 200 – B-1049 BRUXELLES

Conference: GEOPROBA 90

Source + Pathway + Receptor = Risk



- In the field of environmental risks, the common practice in the nineties was to use subjective; i.e., typically postulated, single probability distributions and to apply the Monte Carlo method

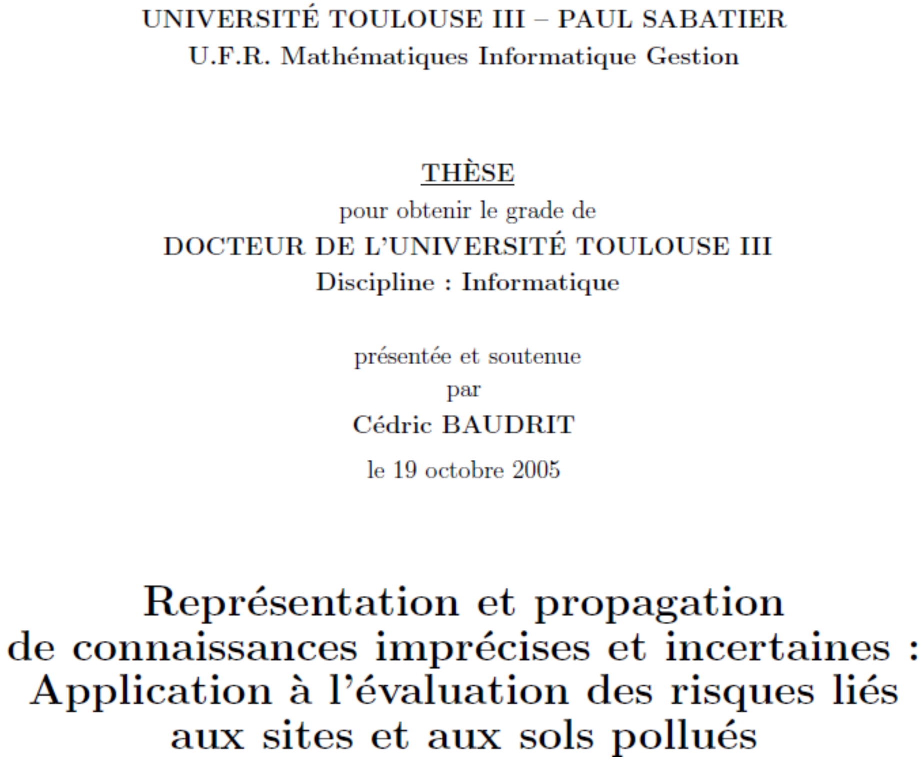
From: Tack, F.M.G., & Bardos, P. "Soil and groundwater remediation technologies - A practical guide", CRC Press.

- But the postulated pdf approach didn’t seem satisfactory (« *pulling probability distributions out of a hat* »)
- Scientific question addressed with IRIT: how to combine both stochastic and epistemic uncertainty in the assessment of risk?
- PhD of Cédric Baudrit
- A notable result over that period: the « hybrid » or « joint propagation » method:

IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 14, NO. 5, OCTOBER 2006 593

Joint Propagation and Exploitation of Probabilistic and Possibilistic Information in Risk Assessment

Cédric Baudrit, Didier Dubois, *Senior Member, IEEE*, and Dominique Guyonnet



UNIVERSITÉ TOULOUSE III – PAUL SABATIER
U.F.R. Mathématiques Informatique Gestion

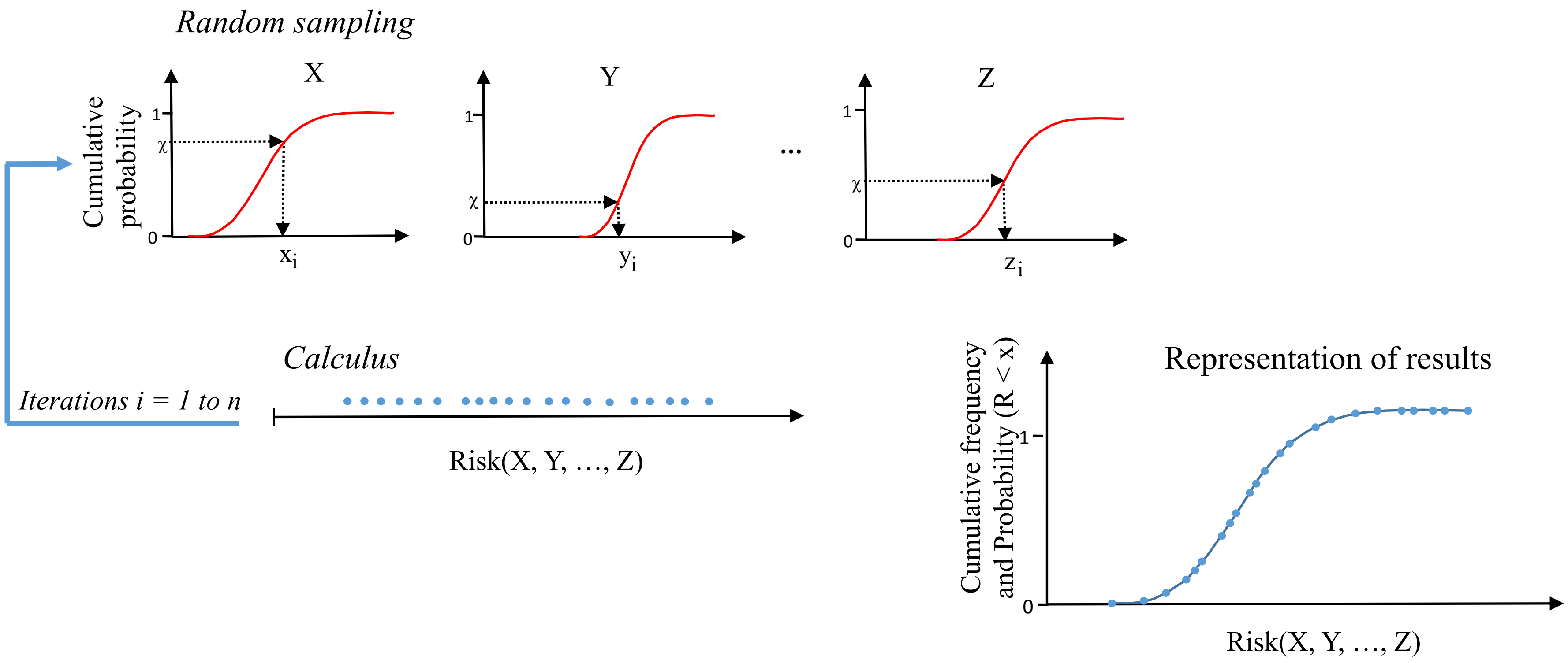
THÈSE
pour obtenir le grade de
DOCTEUR DE L’UNIVERSITÉ TOULOUSE III
Discipline : Informatique

présentée et soutenue
par
Cédric BAUDRIT
le 19 octobre 2005

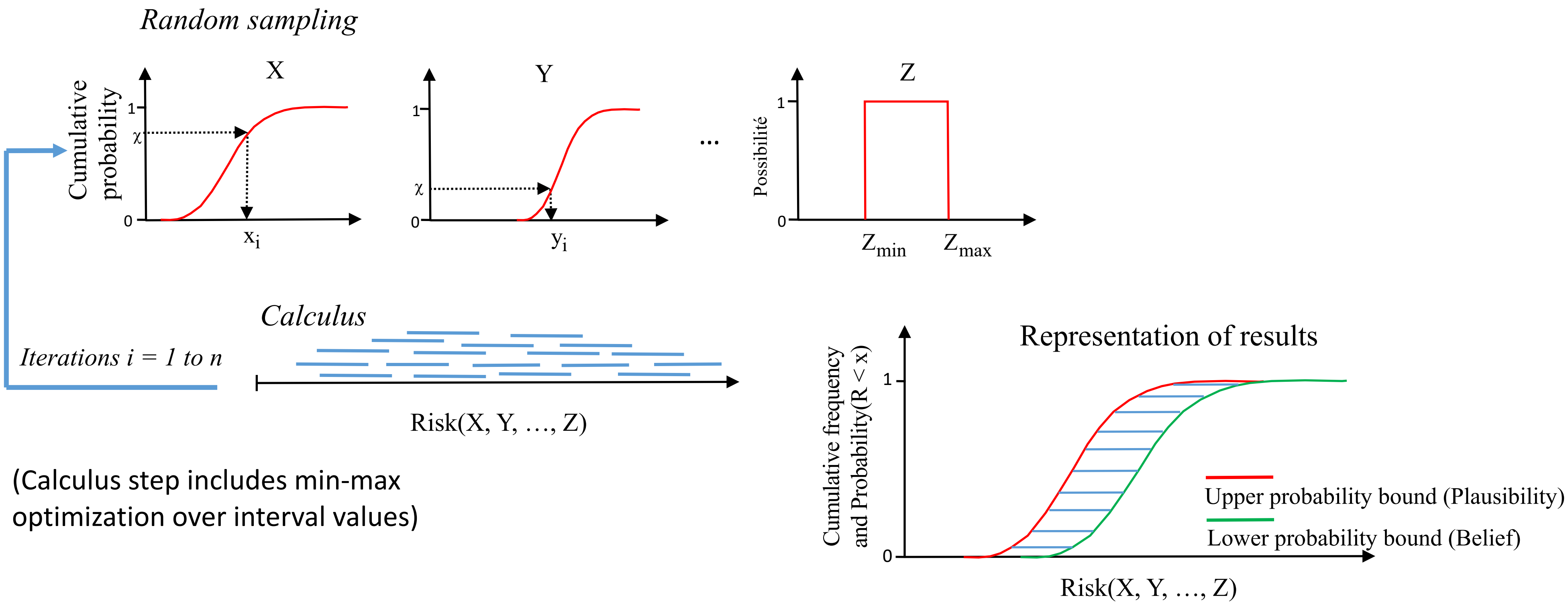
Représentation et propagation
de connaissances imprécises et incertaines :
Application à l’évaluation des risques liés
aux sites et aux sols pollués

Geoscience for a sustainable Earth
brgm

- Reminder: the classical « Monte Carlo » method for propagation of stochastic uncertainty through a model of risk. Risk is function of uncertain parameters X, Y, … Z



- The hybrid (or joint) propagation method: associating stochastic and epistemic uncertainty in the propagation



(Calculus step includes min-max optimization over interval values)

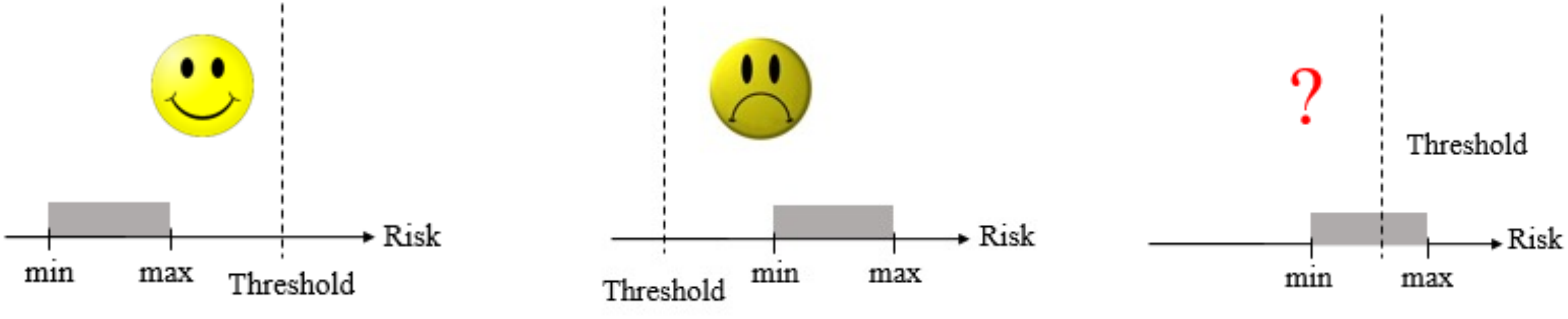
A free application with « R »:

https://cran.r-project.org/web/packages/HYRISK/index.html

2. Communicating on information theories

- In the field of environmental risks, a quantitative approach to uncertainty is uncommon. The notion of probability is barely adopted in most cases. So imprecise probability still has a long way to go…

- Typical approach to uncertainty : evaluator has a risk « model » that is a function of certain parameters (e.g., source concentration, vector velocity, receptor vulnerability, …). Approach to uncertainty consists in varying parameters (usually one by one) within expert judgment-based intervals and observing effect on estimated risk in comparison with a « risk threshold »



No discrimination
of evidence within
the interval

- Uncertainties of epistemic origin: a first step is to explain the difference between knowing and not knowing
- Illustration
  - ✓ Scenario : gas leak in the auditorium
  - ✓ I am in charge of performing a health risk assessment
  - ✓ Among the required parameters: body weight
  - ✓ Assume 2 situations :

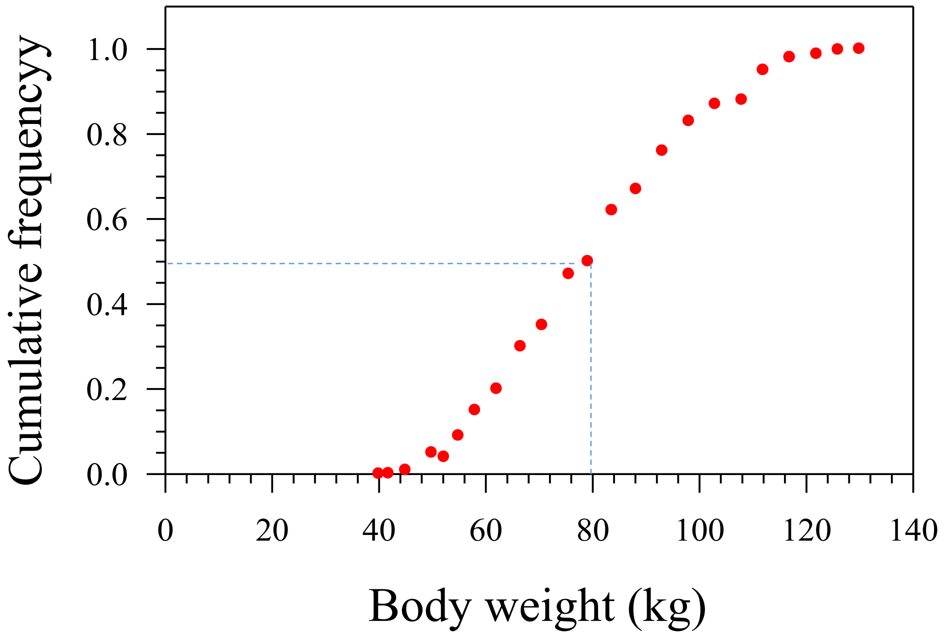
    1) I have a (precise) scale and I can weigh each person in the auditorium

    2) I only have my expert judgment and the opportunity to rapidly glance into the room

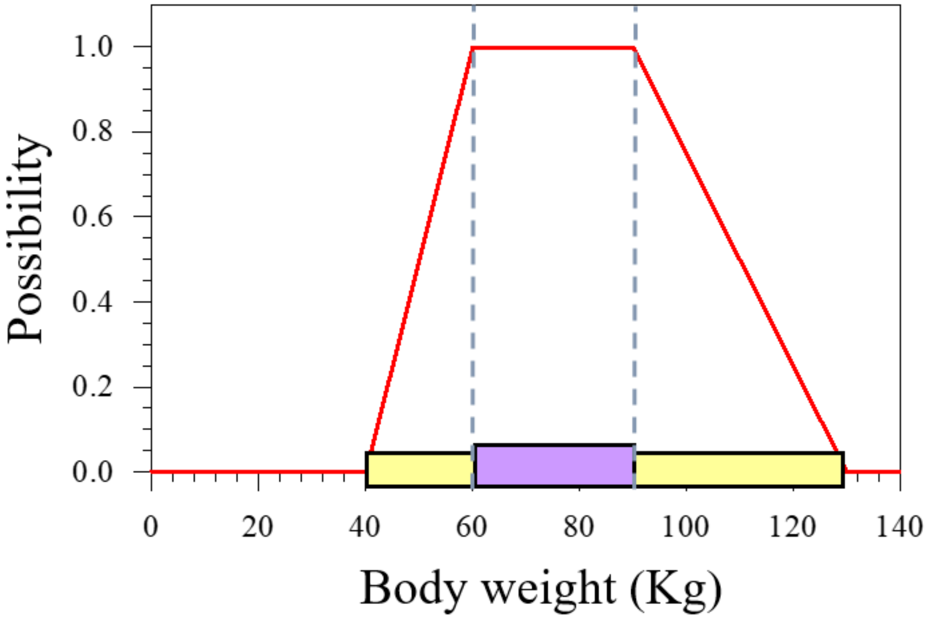
Geoscience for a sustainable Earth

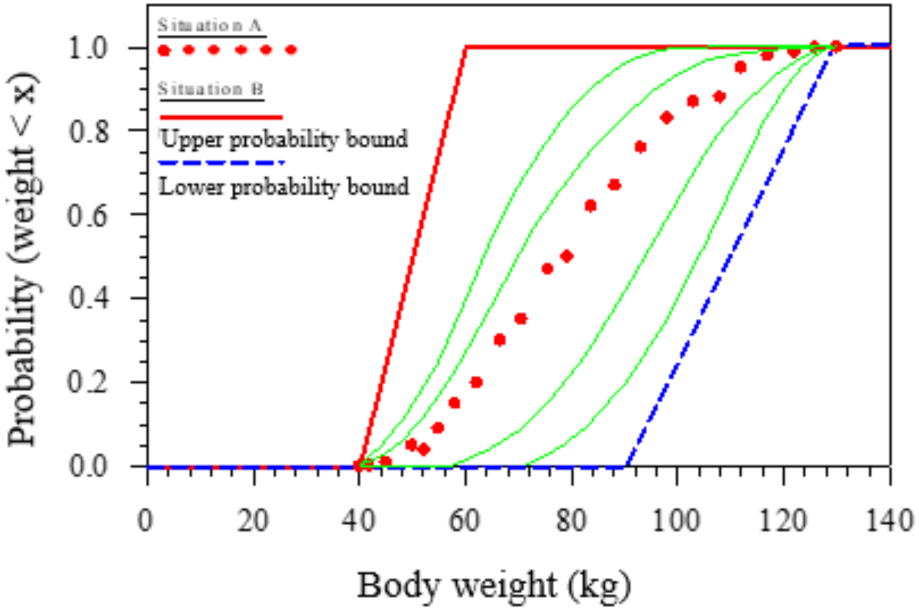
brgm

- Situation A: I can report results as a diagram of cumulative frequencies
- For example: if we were to pick a person in the auditorium randomly, there would be a 50% chance that the person’s weight would be lower than 80 Kg
- We can fit a probability distribution to the data, or propagate it directly in the risk model



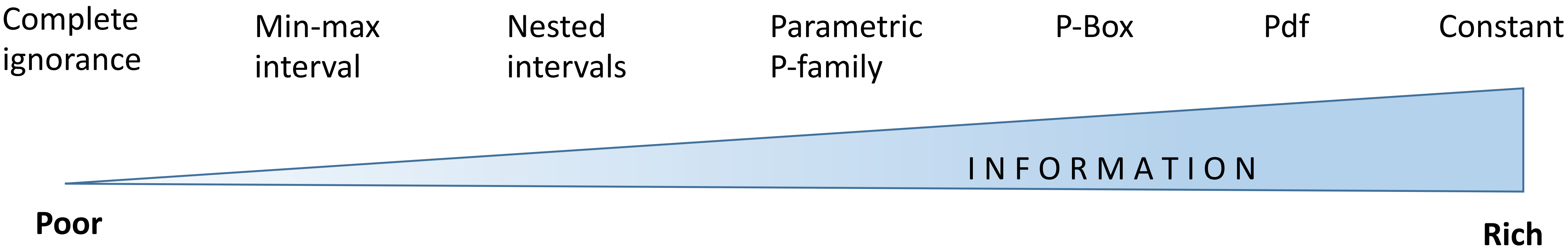
- Situation B - type of information collected is:
  - ✓ An interval of weight values outside which I would consider values are very unlikely: [40 – 120 Kg]
  - ✓ An interval of values that appear most likely (notion of preference): [60 – 90 Kg]
- This information can be represented as a nested interval or fuzzy set
- This type of representation is particularly well suited to expert judgment





Geoscience for a sustainable Earth
brgm

- Guiding principle: **Represent information in a manner that is consistent with the nature of that information**



Complete ignorance

Min-max interval

Nested intervals

Parametric P-family

P-Box

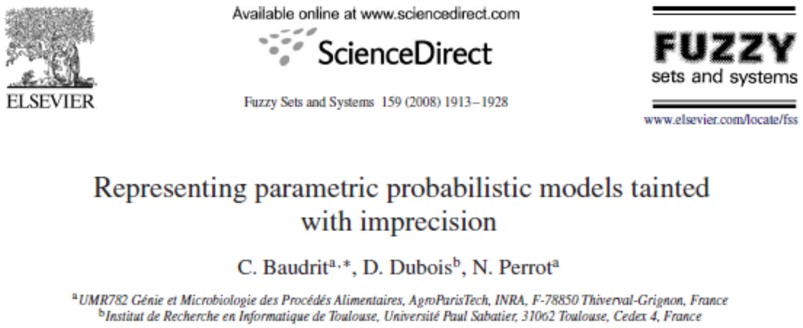
Pdf

Constant

INFORMATION

Poor

Rich



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 159 (2008) 1913–1928

FUZZY sets and systems

www.elsevier.com/locate/fss

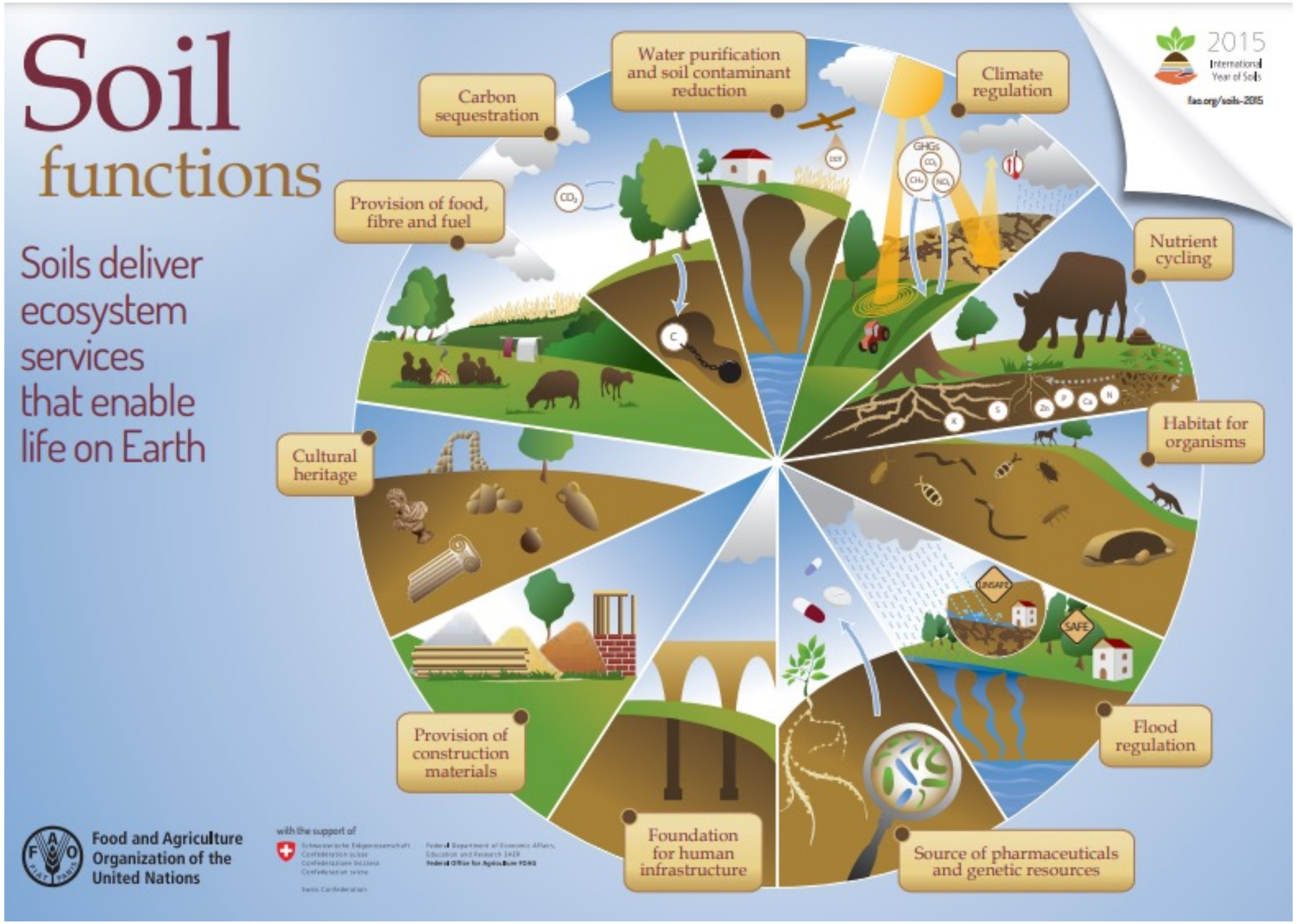
Representing parametric probabilistic models tainted with imprecision

C. Baudrit[a,*], D. Dubois[b], N. Perrot[a]

[a]*UMR782 Génie et Microbiologie des Procédés Alimentaires, AgroParisTech, INRA, F-78850 Thiverval-Grignon, France*
[b]*Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, 31062 Toulouse, Cedex 4, France*

3. Examples of applications of IP to environmental problems

3.1 Soil health and human health




Soil contamination

Soil erosion

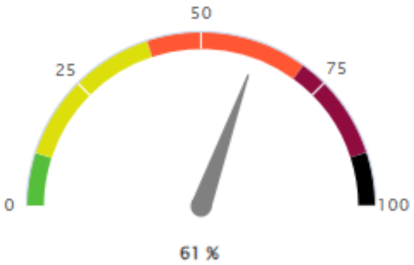

Soil salinization

Soil sealing

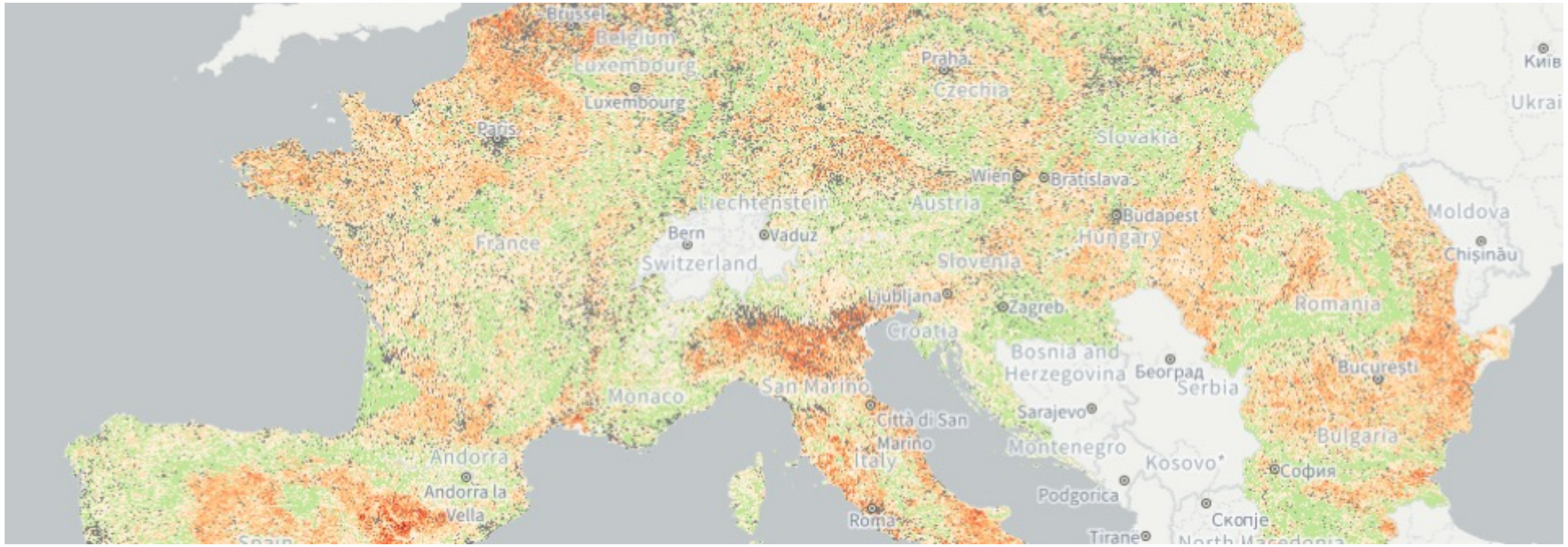
Etc.

https://esdac.jrc.ec.europa.eu/esdacviewer/euso-dashboard/

Proportion of land affected by soil degradation in the EU



The speedometer indicates the proportion of land likely to be affected by one or more soil degradation processes or by soil sealing in the EU. It is based on the convergence of evidence approach described below. This figure is subject to a degree of uncertainty. It is likely an under-estimate as it is based on soil degradation indicators for which data is available; however, scientific evidence is lacking for many other soil degradation processes which are not reflected in the current figure.

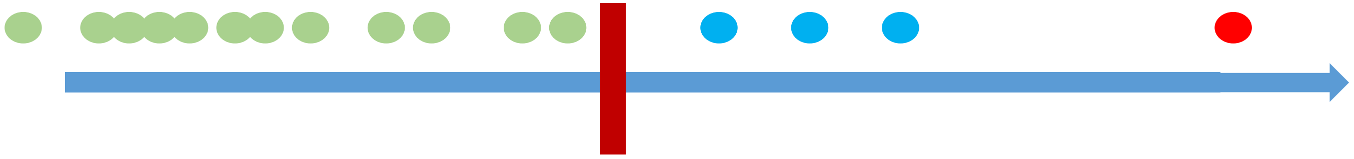


Geoscience for a sustainable Earth
brgm

- In some cases soils require remediation (treatment). The environmental authority defines an objective (a threshold value) for soil concentrations after treatment
- In practice, threshold values are « crisp » and defined on a health-risk basis

For example: arsenic concentrations in a given soil should be lower than X mg/Kg

- Uncertainty is typically taken into account by stating that « a certain proportion » of soil concentrations after treatment may lie above the threshold



Soil values below the threshold

A proportion of values are allowed to lie above the threshold

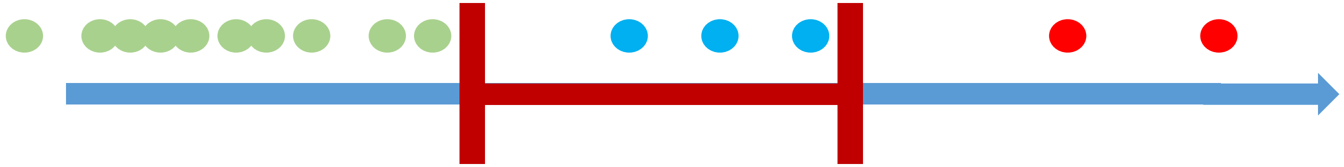
🟢 Acceptable values
🔵 Admissible values
🔴 Unacceptable values

Pollutant concentration in soil

Risk-based
precise threshold

- But how far above? By a factor 2, 3, … And on which basis? This approach is not conservative

- If we define post-remediation soil quality objectives as risk-based **intervals**, rather than precise values, then we have an upper safeguard



Upper safeguard constrains monitoring values

Pollutant concentration in soil

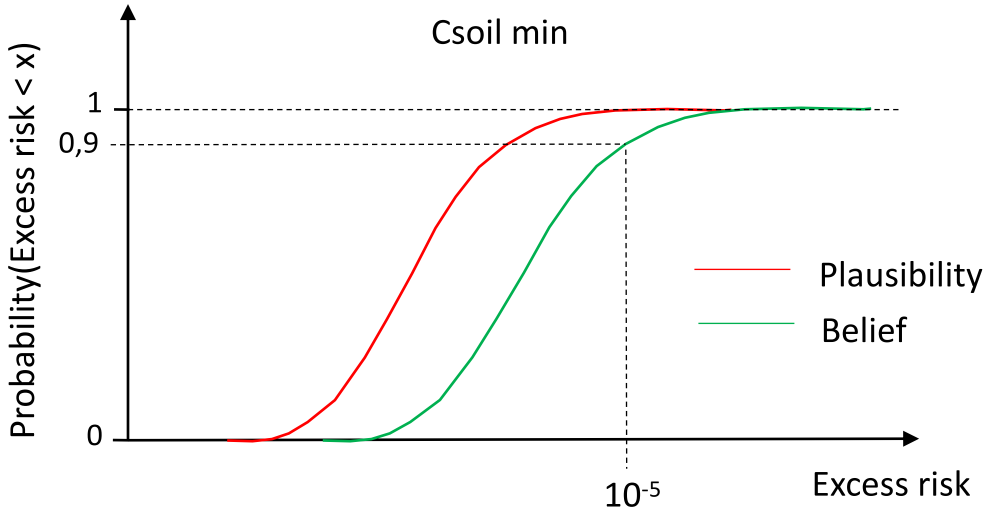
Risk-based **imprecise** threshold

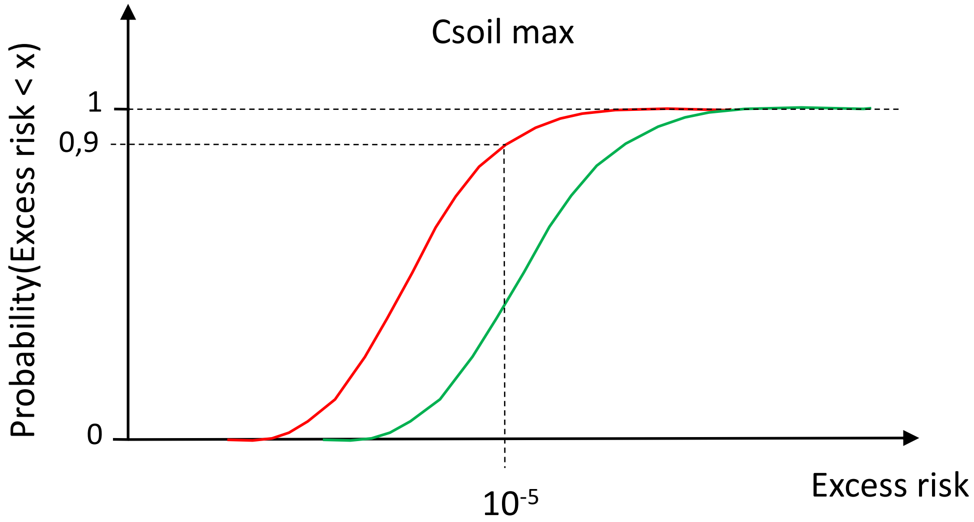
● Acceptable values

● Admissible values

● Unacceptable values

- Proposed approach for defining imprecise post-remediation soil quality objectives:
  - ✓ Estimate **lower** Csoil limit such that we are « certain » (Belief) that Risk < Threshold (10^{-5}) with 90% confidence, despite all unfavourable parameter value combinations
  - ✓ Estimate **upper** Csoil limit such that it remains « Plausible » that Risk < Threshold (10^{-5}) with 90% confidence, considering all favourable parameter value combinations





- ✓ $C_{soil\ min}$ is a conservative limit, as it considers the most unfavourable (pessimistic) parameter value combinations. With $C_{soil\ min}$ there is high **certainty** (Bel = 90%) that risk is lower than the threshold (10^{-5})
- ✓ $C_{soil\ max}$ is a non-conservative limit, as it considers the most favourable (optimistic) parameter value combinations. But it still remains highly **plausible** (Pl = 90%) that risk is lower than the threshold (10^{-5})

Geoscience for a sustainable Earth
brgm

Application to Arsenic-contaminated soils

- Slags left over by steel industry in the North-Est of France
- Several million cubic metres of slag (Photo)
- BRGM was in charge of assessing risks associated with these materials
- Assessment included:
  - ✓ Selection and characterisation of representative samples
  - ✓ Mineralogy and speciation of As and Pb in the samples
  - ✓ Characterisation of bioaccessibility of As and Pb and links with speciation
  - ✓ Quantitative evaluation of health risks considering bioaccessibility and risk parameter uncertainties

Geoscience for a sustainable Earth

brgm

- **Risk model**

$$IER = D \times UER$$

IER = Individual Excess Risk (expected excess cancers resulting from dose D)

D = Dose absorbed ($mg/Kg\ d^{-1}$)

UER = Unit Excess Risk (expected excess cancer per unit dose; $(mg/Kg\text{-}d)^{-1}$)

$$D = \frac{SI \times CS \times BA \times EF \times ED}{BW \times AT}$$

SI = Soil Ingestion (Kg/d)

CS = As Concentration in Soil (mg/Kg)

BA = As Bioaccessibility (unitless)

EF = Exposure Frequency (days/yr)

ED = Exposure Duration (yrs)

BW = Body Weight (Kg)

AT = Averaging Time (yrs)

Geoscience for a sustainable Earth

brgm

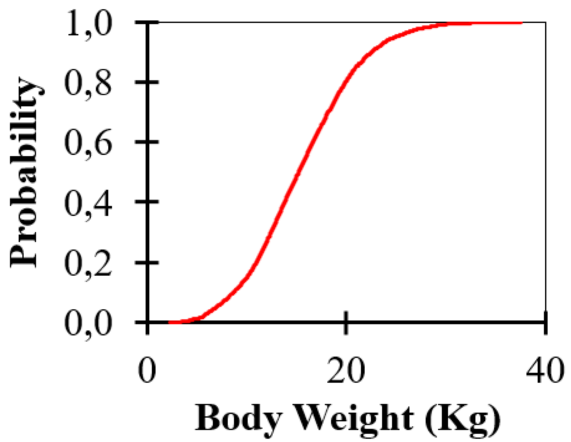
- **Uncertainty representation**

✓ Constants

- Csoil (in the procedure Csoil is varied such that the risk threshold is respected either by the lower probability indicator; Belief, or the upper probability indicator; Plausibility)
- UER = 1.5 (mg/Kg d)$^{-1}$ (US EPA 2009 / OEHHA 1998, recommended by INERIS 2010)
- ED = 6 yrs (child scenario)
- AT = 70 yrs (standard health risk procedure)

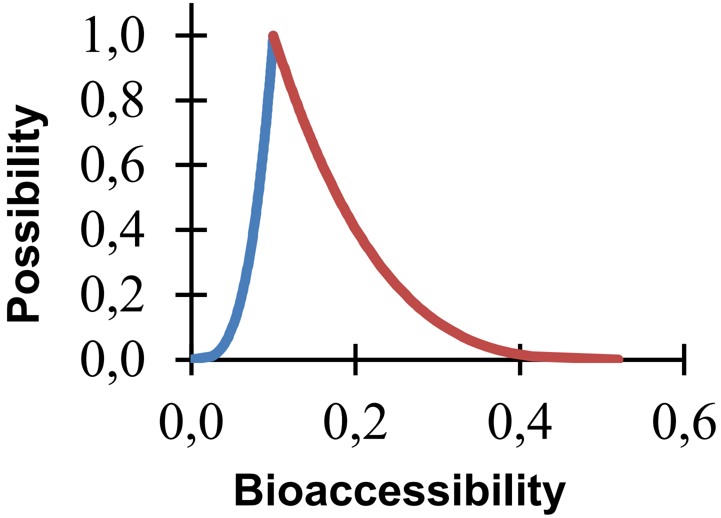
✓ Probability distribution

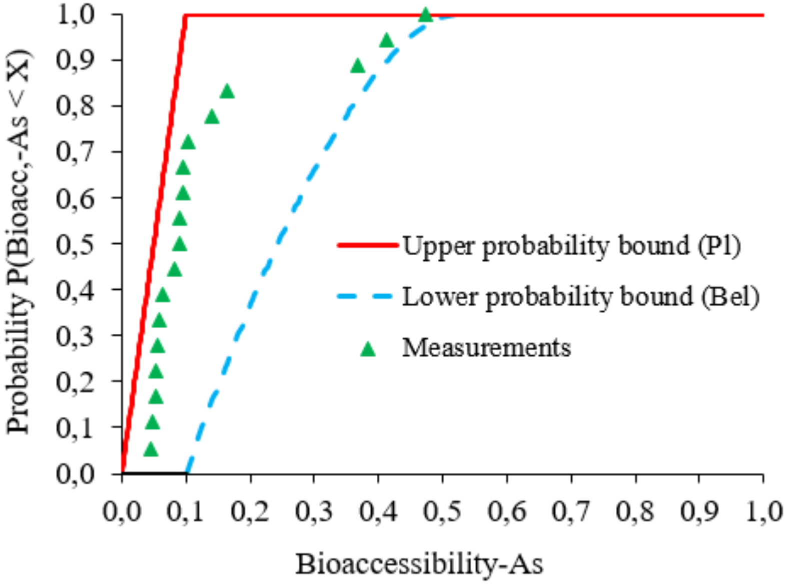
- BW: based on statistical data, average = 15.5 Kg, standard deviation = 5.4 Kg



(derived from Dereumeaux et al., 2012)
Dereumeaux C, Kairo C, Zeghnoun A. Synthèse des travaux du Département santé environnement de l'Institut de veille sanitaire sur les variables humaines d'exposition. Saint-Maurice: Institut de veille sanitaire; 2012. 29 p.

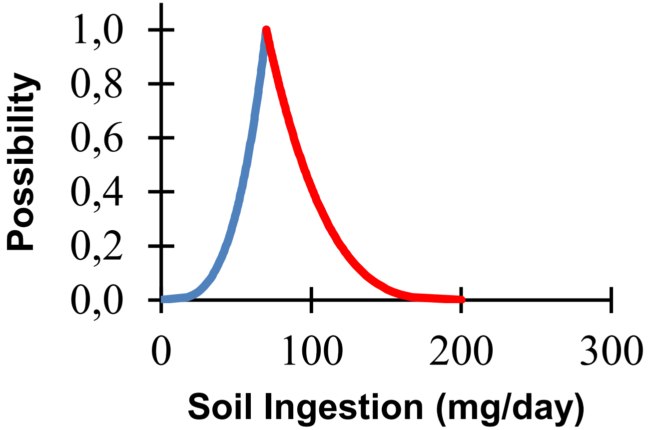
- ✓ Possibility distributions
  - ○ Defined based on scarce data or on expert opinion
  - ○ These possibility distributions define families of probability distributions (instead of just one)
  - ○ Bioaccessibility ( )
    - ▪ Preferred value: 10%
    - ▪ Support: 0 – 52%





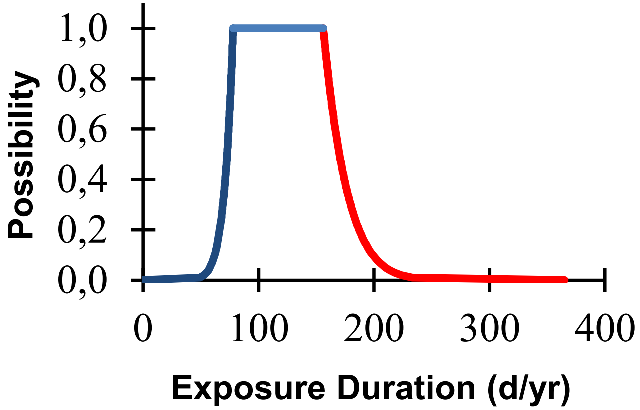
Geoscience for a sustainable Earth
brgm

- Soil Ingestion (mg/d)
  - Preferred value: 70
  - Support: 0 – 200



(intermediate choice between Bonnard, 2017 and US-EPA, 2017)

- Exposure Frequency (d/yr)
  - Preferred values: 78-156
  - Support: 0 – 365

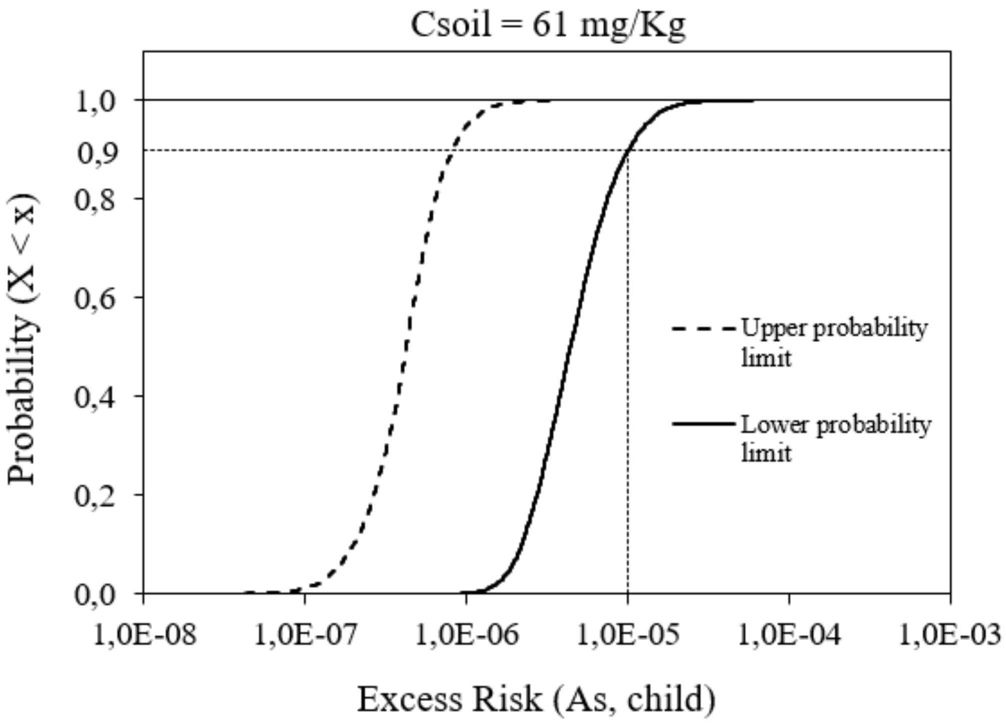


(based on expert opinion consistent with Roy et al., 1993)

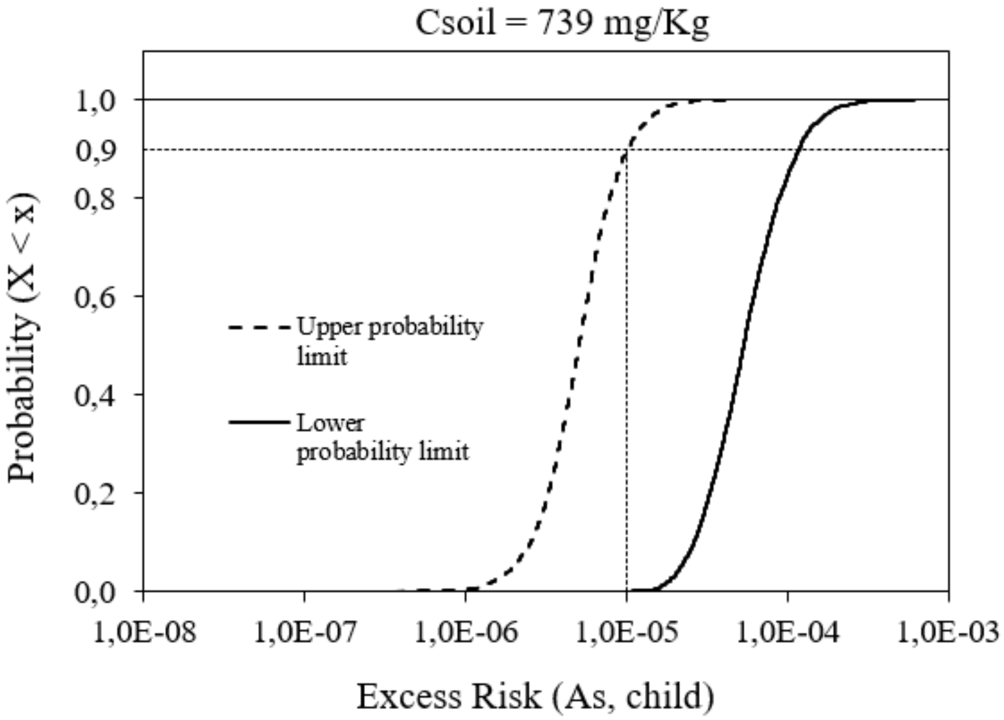
Geoscience for a sustainable Earth
brgm

- **Results**

Lower soil concentration bound



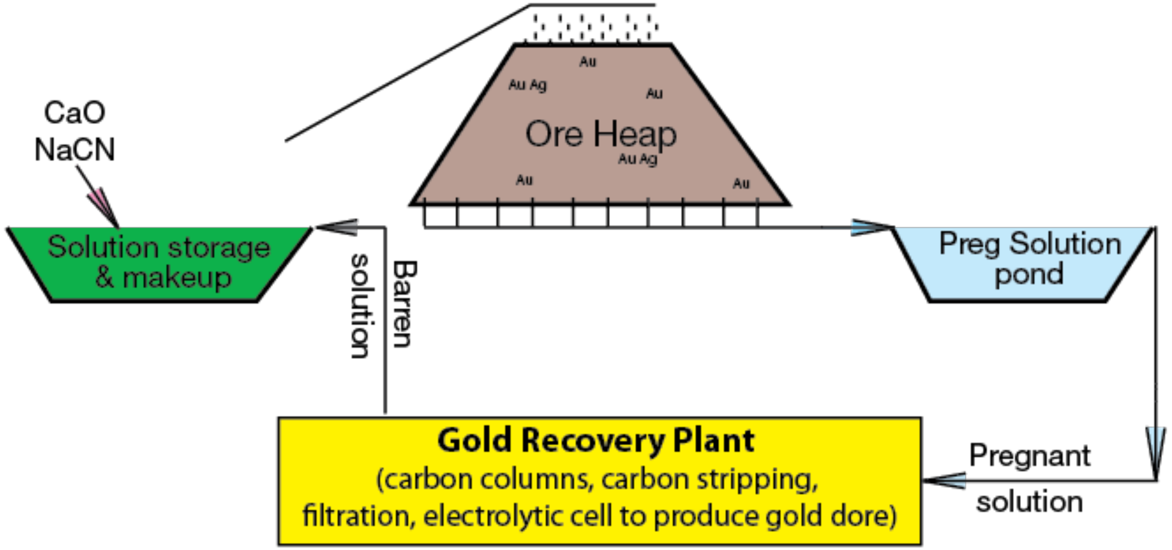
Upper soil concentration bound

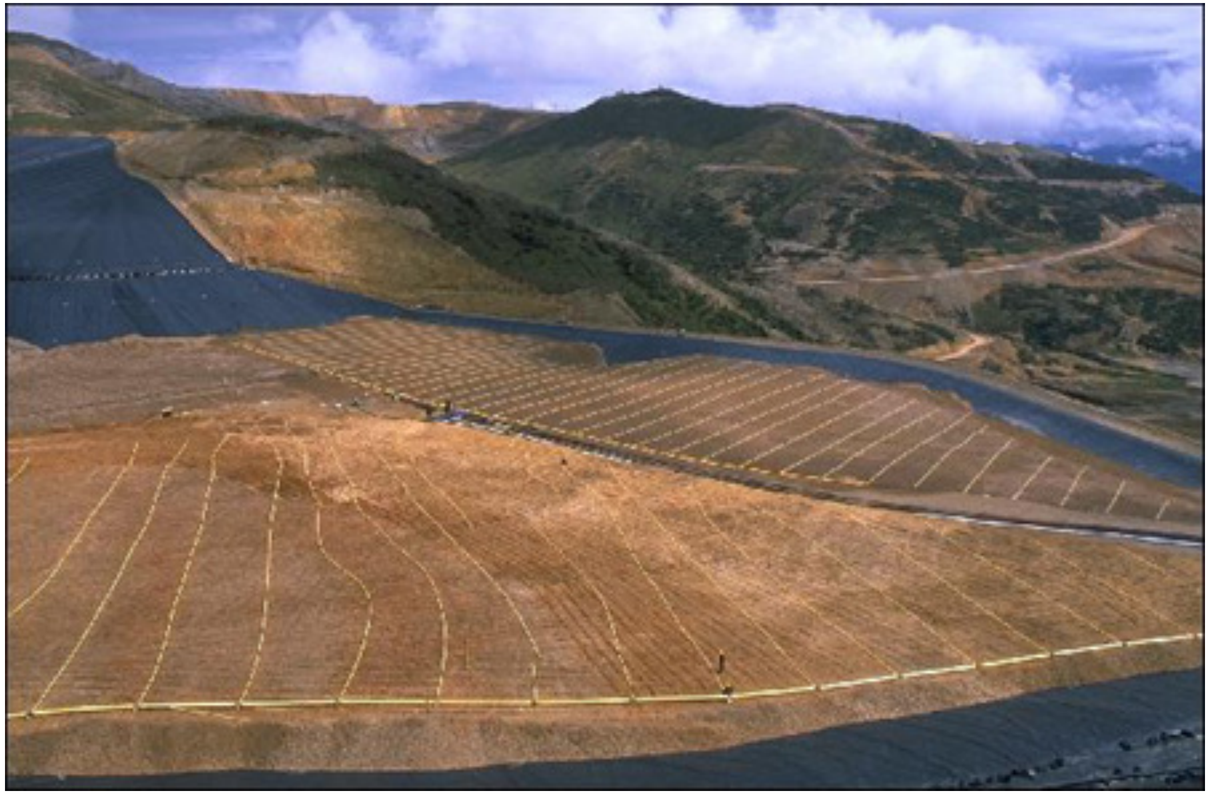


- In terms of a **post-remediation soil concentration objective**: post-remediation soil monitoring should yield values **≤ 61 mg/Kg**, while X% (to be defined by decision-maker) could exceed this value but should always remain **≤ 739 mg/Kg**

3.2 Risk of leakage from a heap-leaching facility

- Heap leaching is a common method in the mining industry for treating ore





- For ex. cyanurated solutions for precious metals such as gold:
  $Au^{+}(s) + 2\ CN^{-}(aq) \rightarrow Au(CN)^{-}_{2}(aq)$

- Acid solutions for copper, nickel, etc.

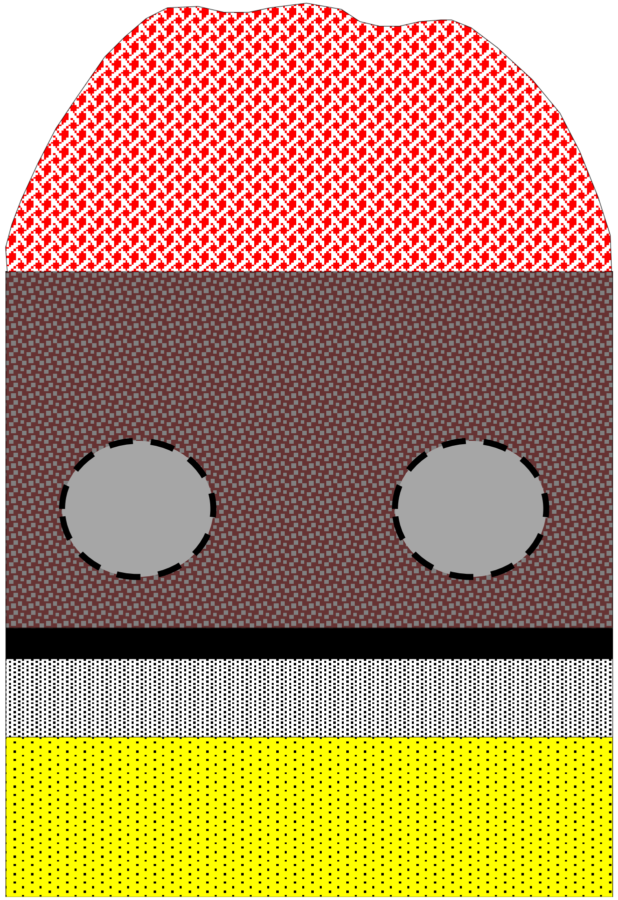
- The ore is piled on top of a drainage system and the fluids are percolated through the ore

- The drainage system is underlain by barriers to avoid oil and water contamination

- Risk assessment is performed at an early stage for design purposes

Geoscience for a sustainable Earth
brgm

Schematic of heap leaching



Ore

Drainage and protection layers

High-density geomembrane

Low-permeability mineral layer

Foundation layer

Geomembranes in storage

Geoscience for a sustainable Earth
brgm

- The scale of such operation is huge:

Andean Valley Fill (Thiel et Smith)
65 hectares, 100s of meters ore height

Installing the geomembrane

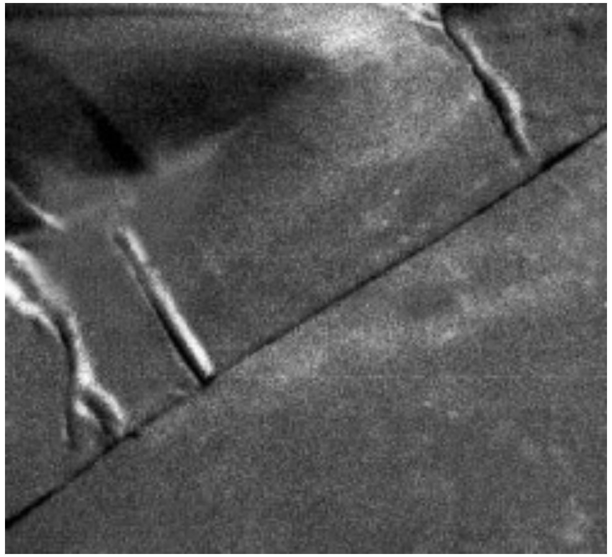
High-density polyethelene drains at the bottom, to collect fluids (before putting the ore)

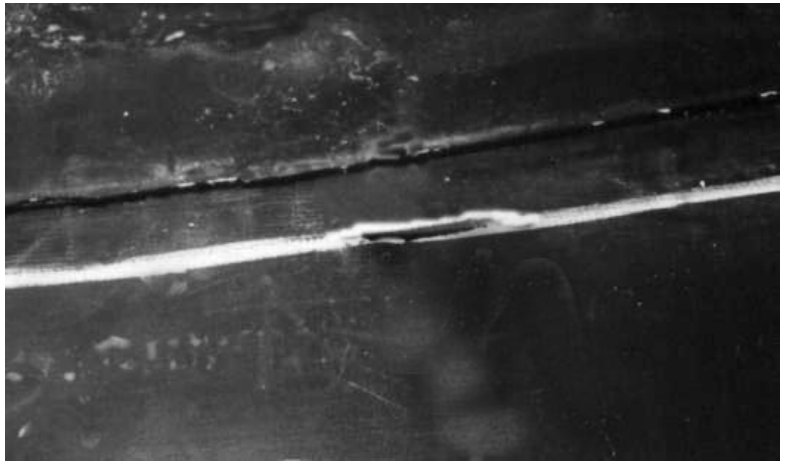
- Tubes to infliltrate leaching fluids into the ore

Geoscience for a sustainable Earth

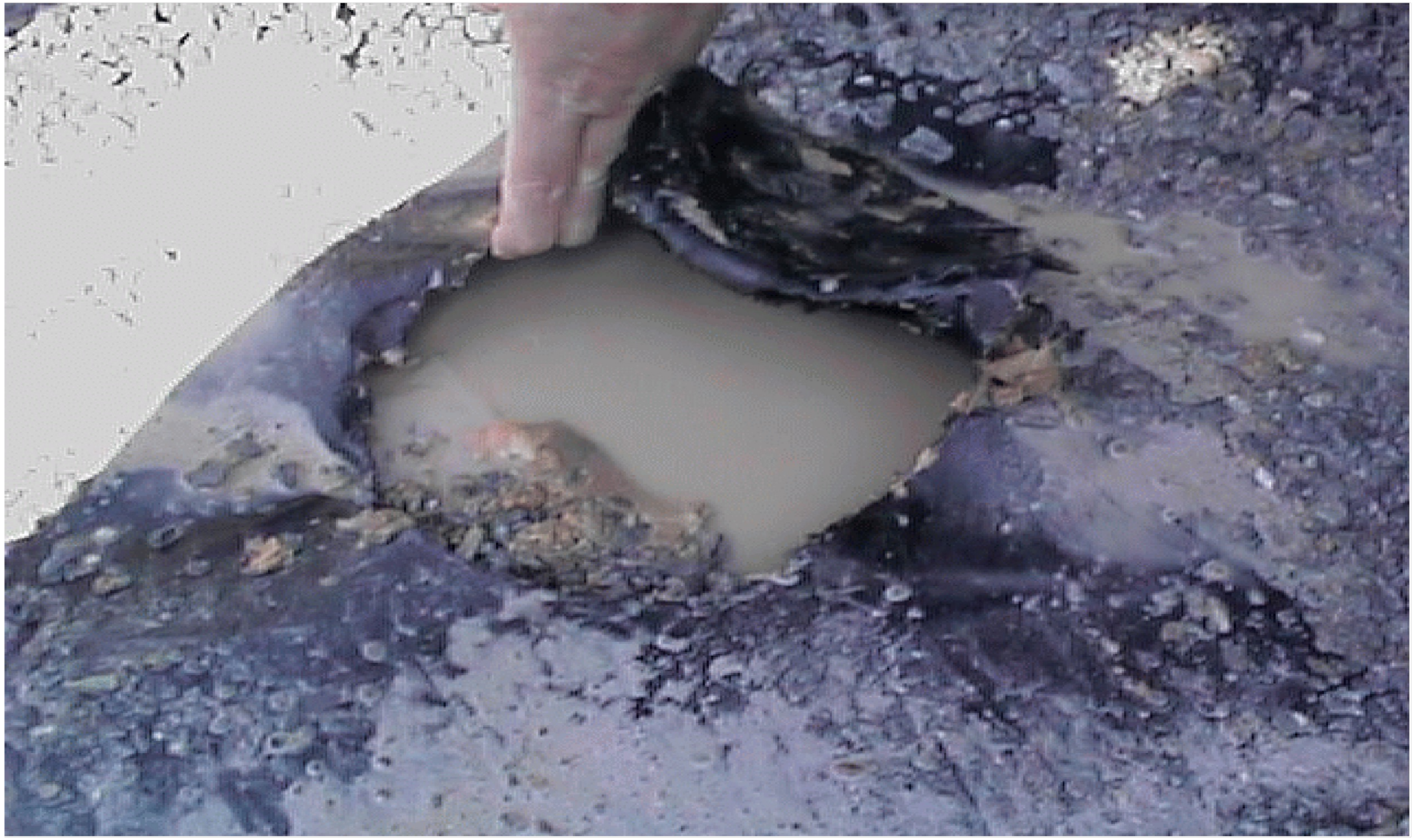
brgm

There can be defaults in the geomembrane





- The geomembrane can be damaged by elements in the foundation layer



Objectives

- Estimate the imprecise probability of leakage rate through the bottom barrier
- Take into account uncertainties relative to controlling parameters

Leakage model

$$Q = n\,0{,}21\,h_w^{0{,}9}\,a^{0{,}1}\,K_s^{0{,}74}\left(1 + 0{,}1\left(\frac{h_w}{H_s}\right)^{0{,}95}\right)$$

Giroud, 1997;
Touze-Foltz et al., 2008

Q = leakage rate (m^3/s),

n = number of defaults per hectare

h_w = hydraulic head above the geomembrane (m),

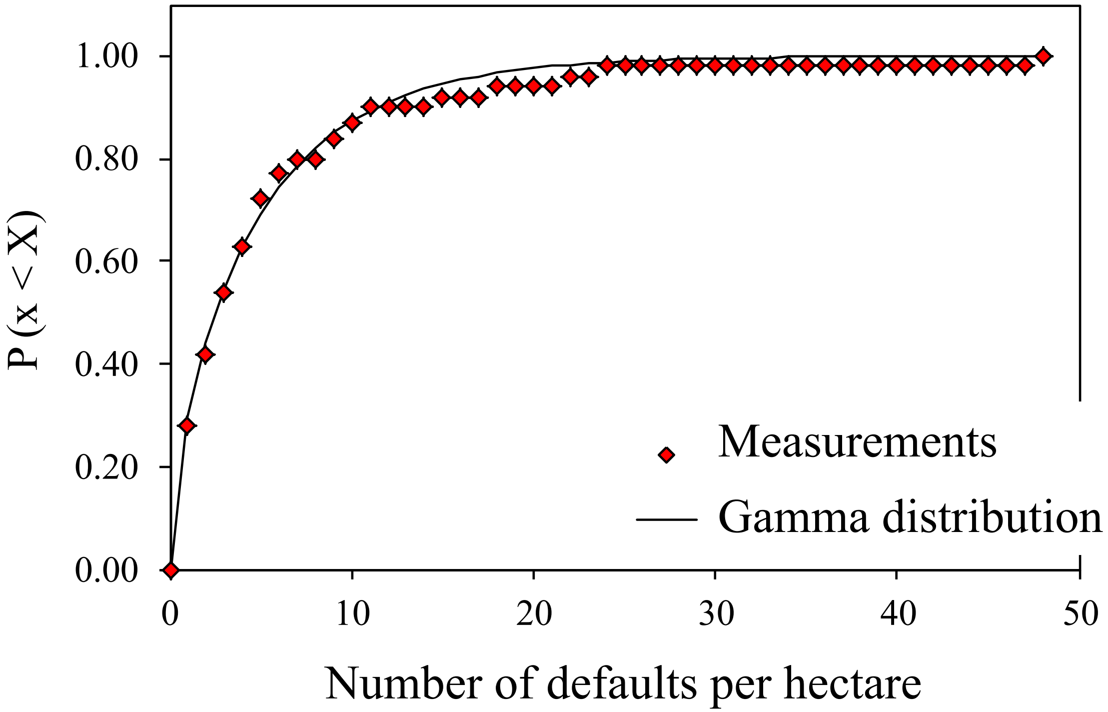
a = default surface area (m^2),

K_s = hydraulic conductivity of the mineral layer (m/s),

H_s = thickness of the mineral layer (m).

Model parameters

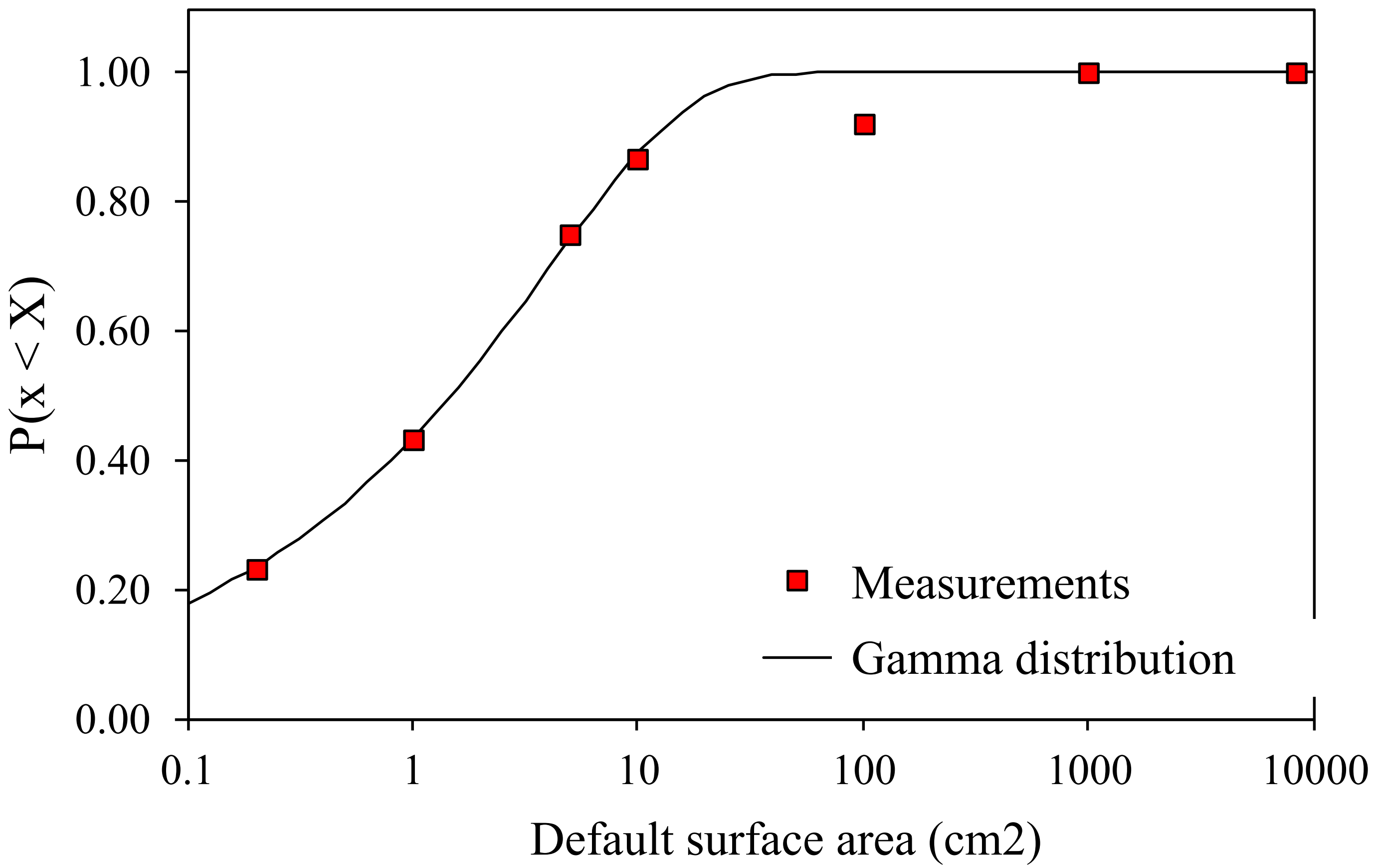
- Number of defaults per hectare



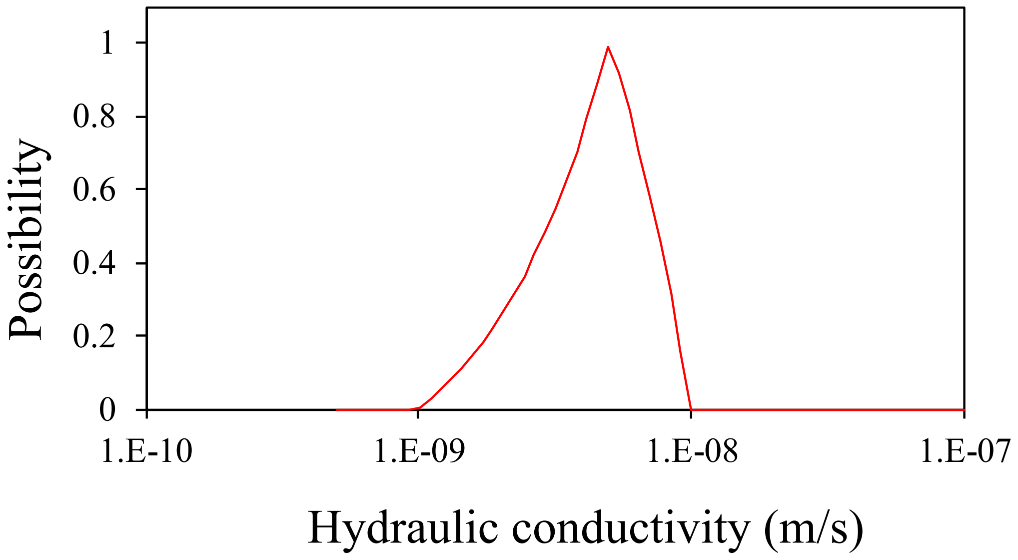
Data from Forget et al., 2005

- Default surface area

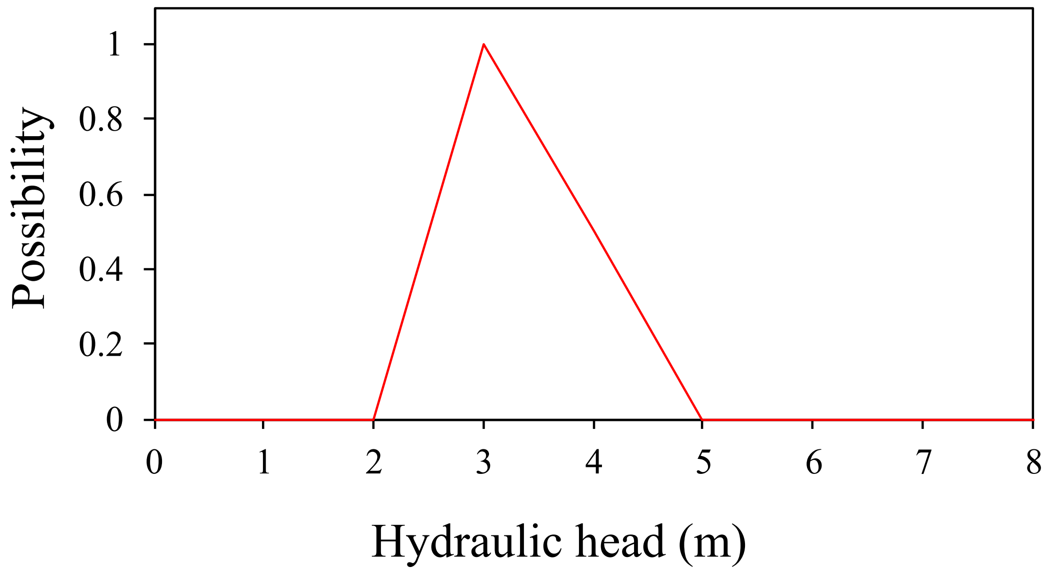
Data from Colucci & Lavagnolo, 1995



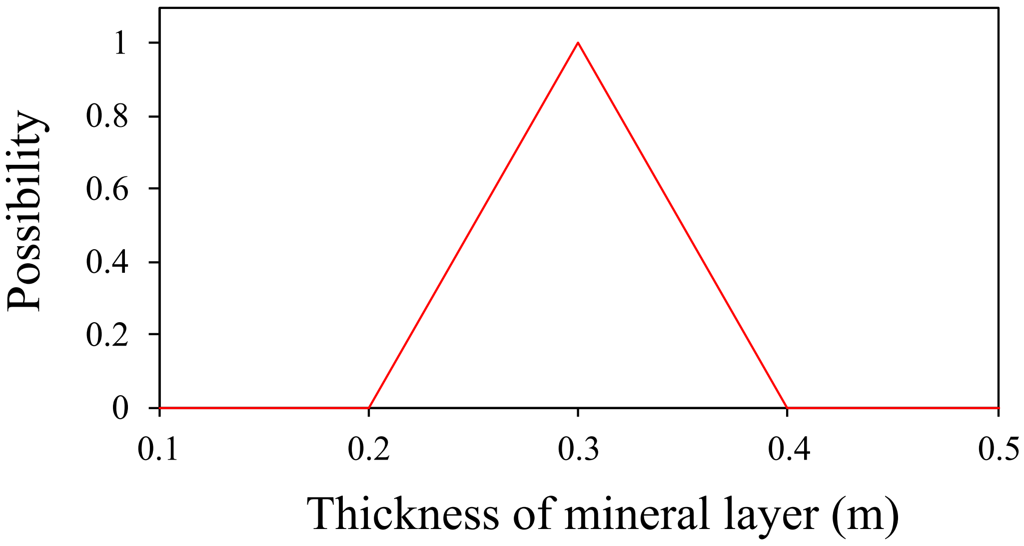
- Hydraulic conductivity of the low-permeability mineral layer



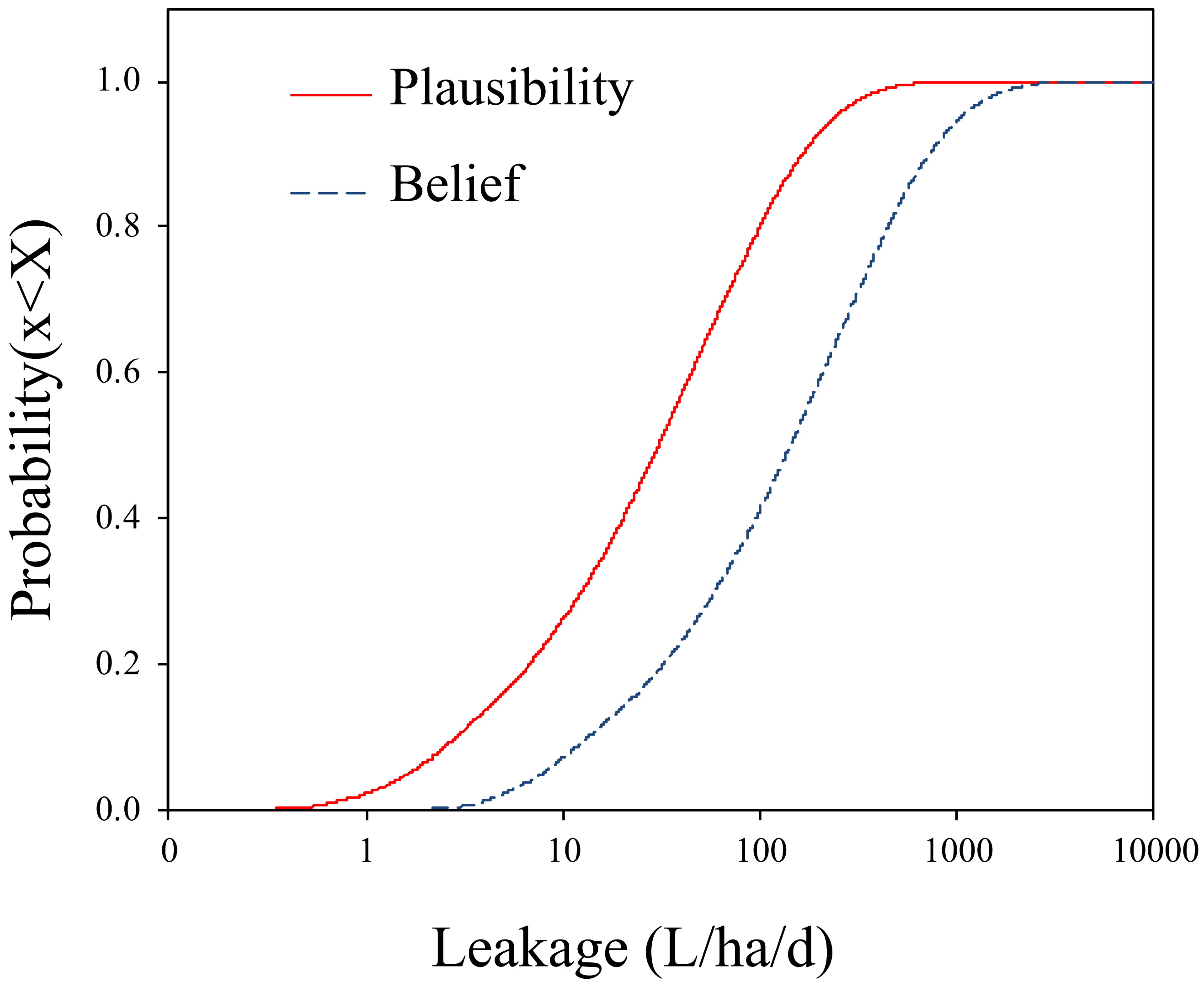
- Hydraulic head above the geomembrane



- Thickness of the low-permeability mineral layer

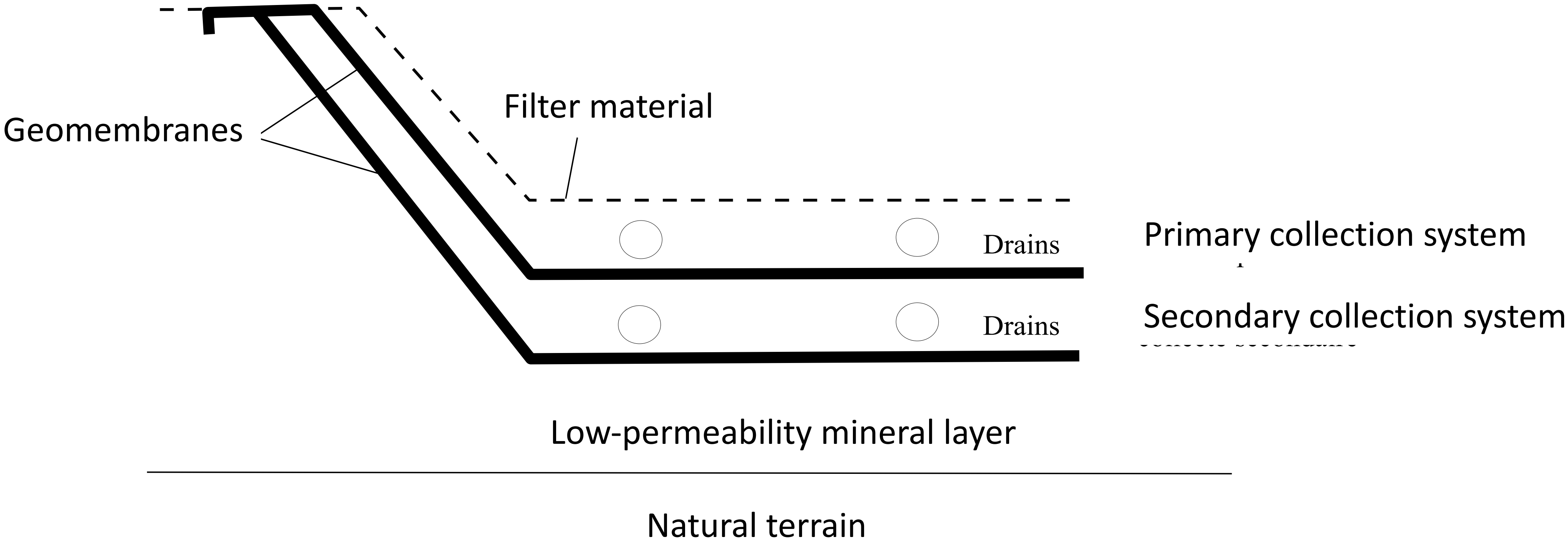


Results of propagation



Comparison with measurements

- Measurements performed in double barrier systems



Geomembranes

Filter material

Drains

Primary collection system

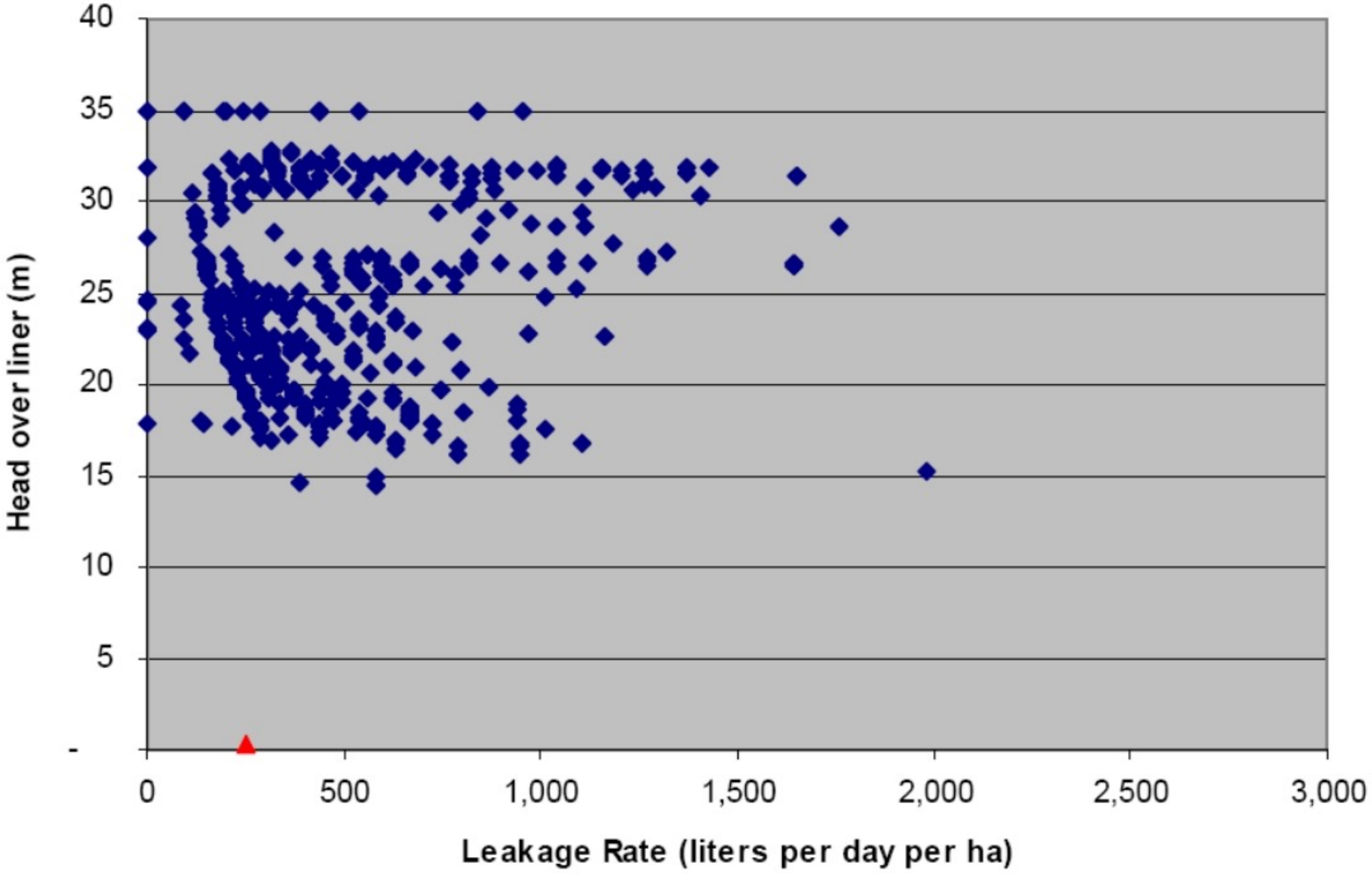
Drains

Secondary collection system

Low-permeability mineral layer

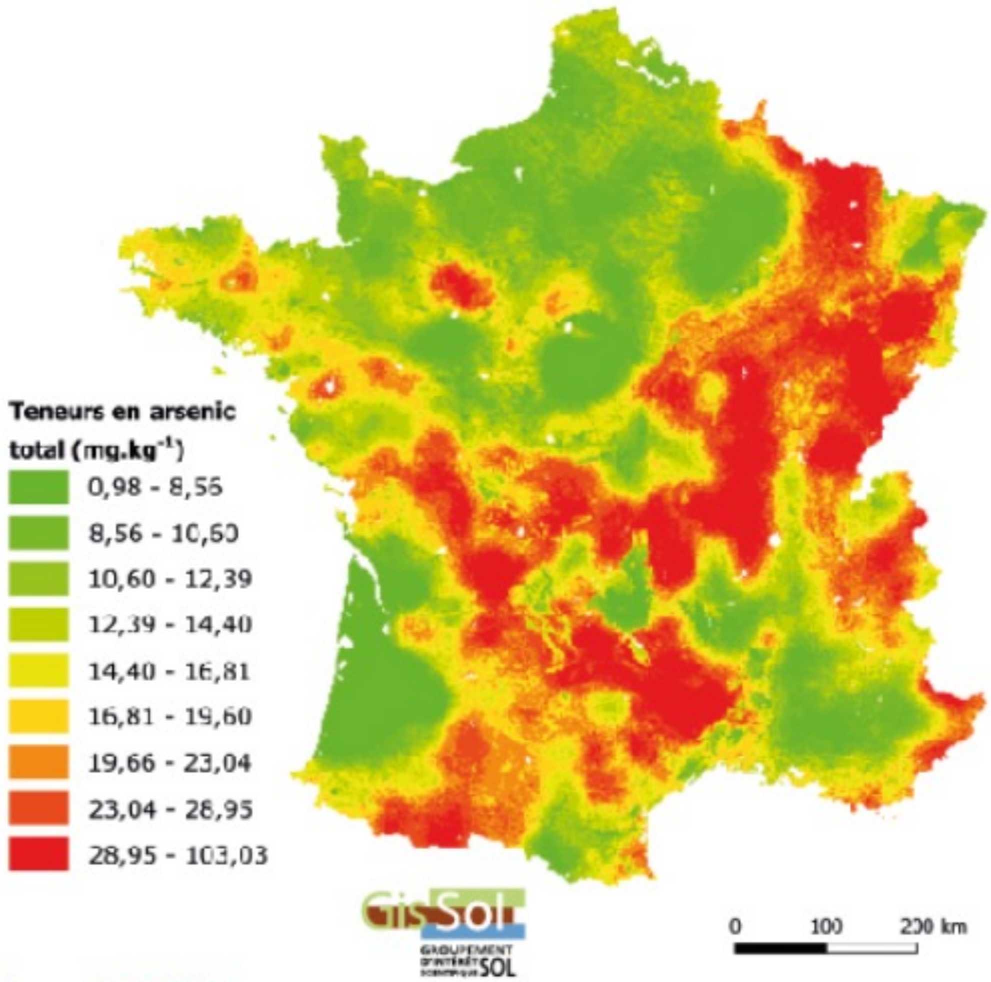
Natural terrain

Measurements of Thiel et Smith (2003)



3.3 Interpolation of SIC data

- Creating maps is very common in environmental work
- Example: map of arsenic concentrations in French soils
- Making such maps usually involves some sort of interpolation method
- If there is enough data, geostatistical methods are used (kriging)
- But if data are SIC (Sparse, Incomplete, Clustered), geostatistical methods reach their limits



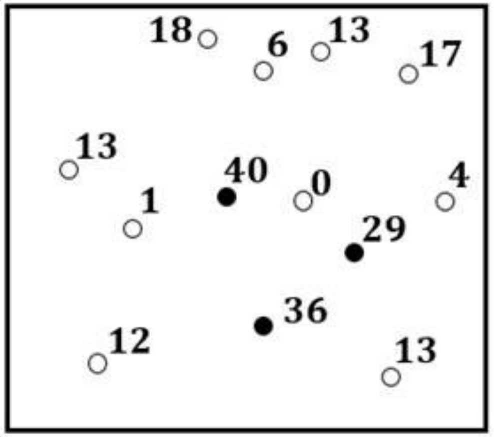
Source : Gis Sol, 2019

- Example

E.C. Dahlberg, 1975
Relative effectiveness of geologists and computers in mapping potential hydrocarbon exploration targets
Mathematical Geology

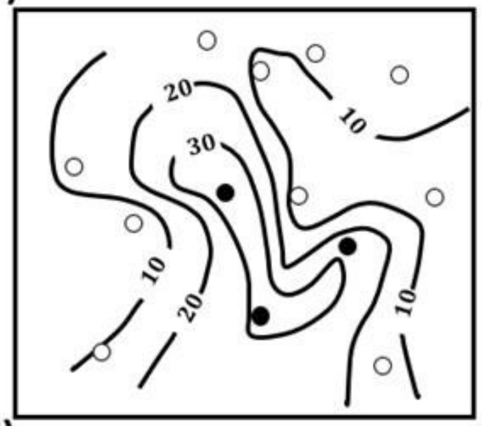
Original data: 13 values of sand thicknesses in borehole cores

a)



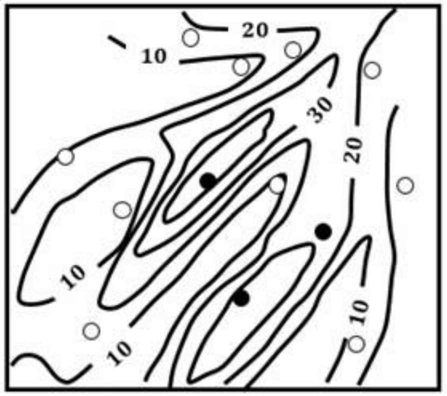
Result of manual triangulation

b)



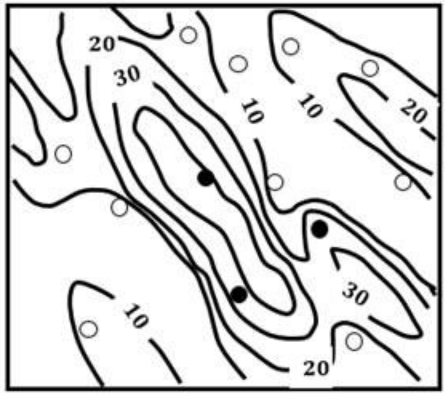
Interpolation by Geologist 1 who assumed fluviatile deposits

c)

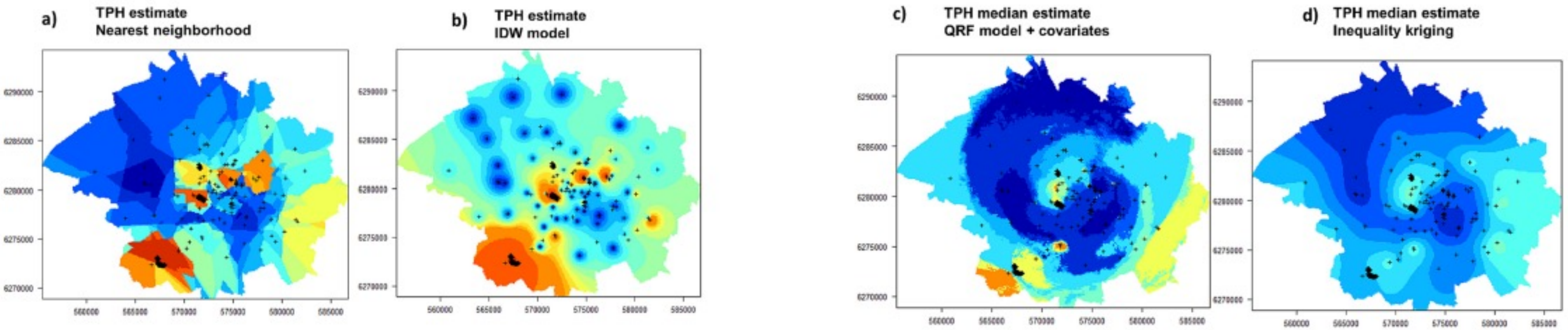


Interpolation by Geologist 2 who assumed channel deposits

d)

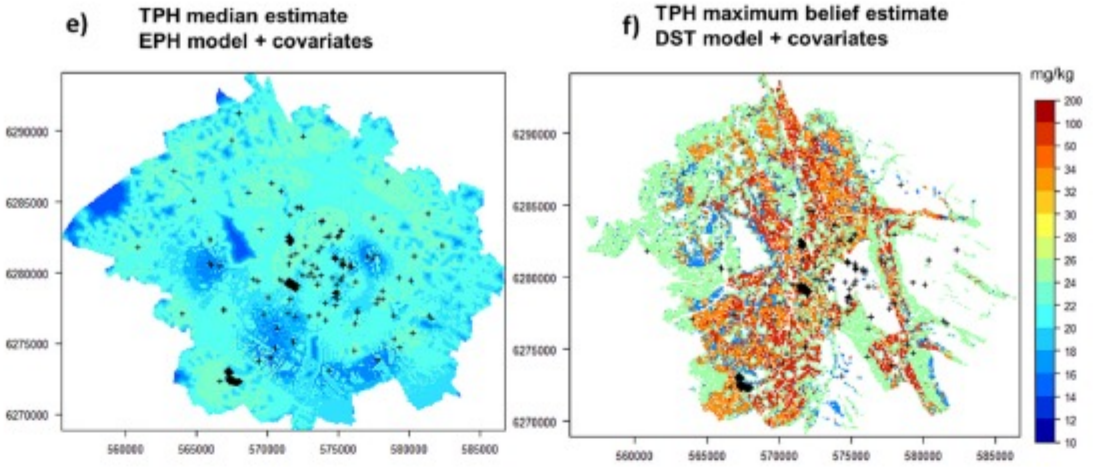


- Ongoing efforts as part of the HOUSES project (ANR) with BRGM, HEUDIASYC, IRIT, Paris School of Mines



Six interpolation for mapping Total Petroleum Hydrocarbon (TPH) in the city of Toulouze:

a) Nearest neighbour
b) Inverse distance weighting
c) QRFF quantile random forest
d) Inequality kriging
e) EPH experimental probabilistic hypersurface
f) DST-Belief map




Journal of Geochemical Exploration 254 (2023) 107298

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

Defining urban soil geochemical backgrounds: A review for application to the French context

Stéphane Belbèze, Jérémy Rohmer, Philippe Négrel*, Dominique Guyonnet

BRGM, 45060 Orléans, France

Geoscience for a sustainable Earth

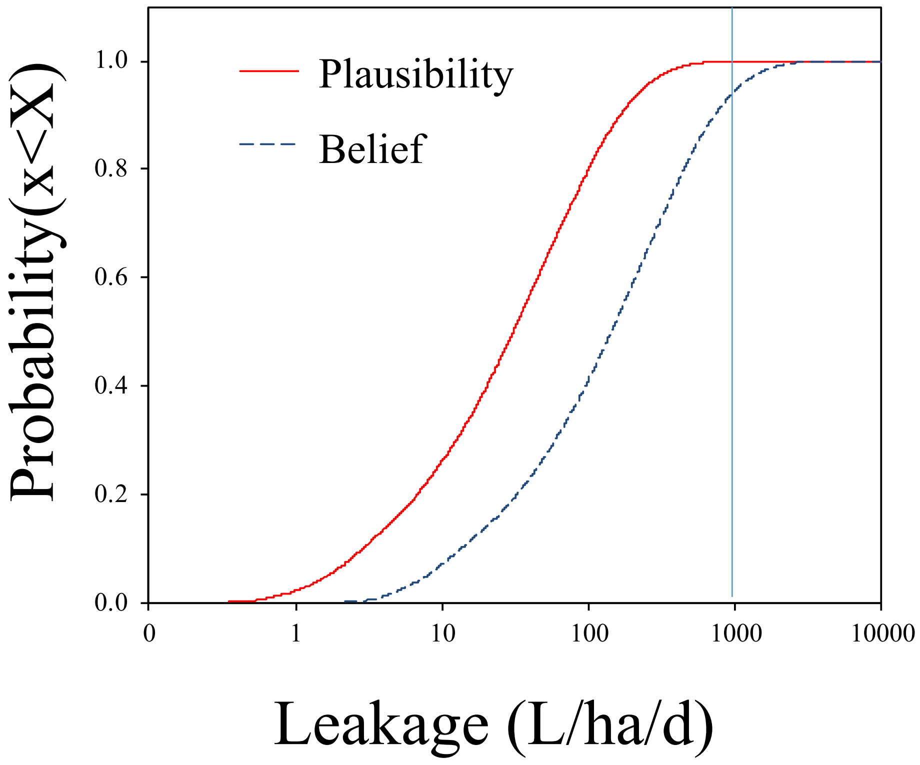
brgm

4. Communicating on IPs

- In the field of environmental risk engineering in France, it is uncommon to see crisp probabilities of exceeding an acceptance threshold, let alone imprecise probabilities

- For example, in this result, probability that leakage < 1000 L/ha/d is between 0,95 and 1

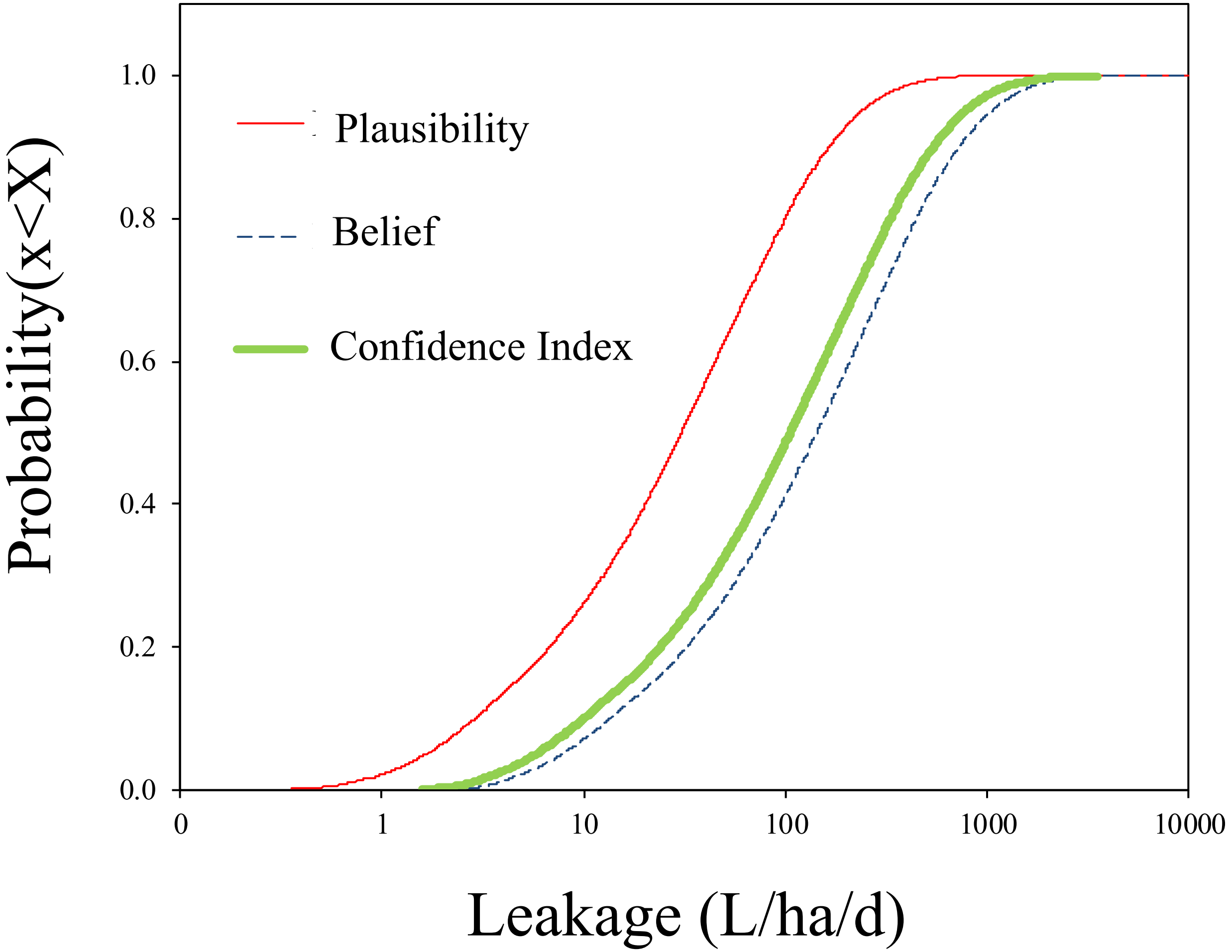
- Alternative: based on work by Hurwicz (1951), define a confidence index as a weighted average of Pl and Bel



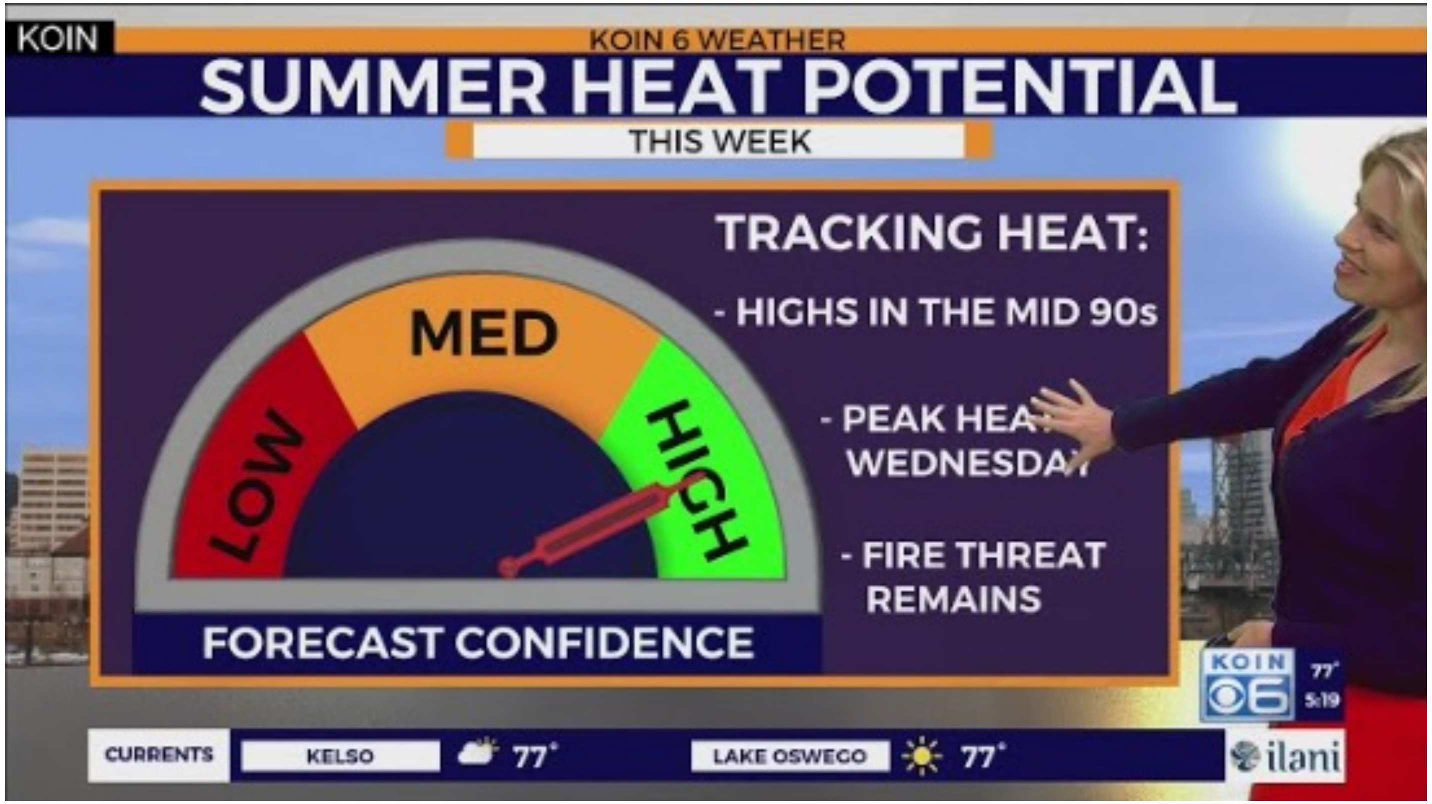
- At each level of probability:
  Confidence Index = α Pl + (1- α) Bel

- In a context of risk aversion, give more weight to Bel than to Pl

- Result for α = 0,33



- Could be aggregated into simple indicators of confidence for communication purposes
- Beneficial to draw from experience of weather forecasting community



Geoscience for a sustainable Earth
brgm

5. Conclusions

- Uncertainty is an unavoidable aspect of risk-based contaminated land and soil remediation
- In addressing uncertainty, it is important to first look at the information, then choose a mathematical framework for representing and propagating this information
- Taking into account uncertainty of epistemic origin is important to convey to decision-makers the range of alternative outcomes
- It is also important for highlighting the need for additional data collection
- When probability distributions are postulated, there is no way of distinguishing, in the variance of computed output, the actual variability resulting from true stochastic randomness from apparent variability due to subjective probability judgements
- There remains a long way to go before imprecise probabilities are part of the decision-making process in the field of environmental risks

Thank you for your attention