ANR HOUSES

Harmonized Operation of Uncertainties in Spatialized Environmental Systems

Jeremy Rohmer (1), Abel Henriot (1), Stephane Belbeze (1), Dominique Guyonnet (1), Chantal de Fouquet (2), Thomas Romary (2), Sebastien Destercke (3), Benjamin Quost (3), Jean-François Leger (3), Helene Fargier (4), Romain Guillaume (4), Didier Dubois (4), Emmanuel Cazeneuve (5), Camille Chabrol (5)

2 déc. 2022

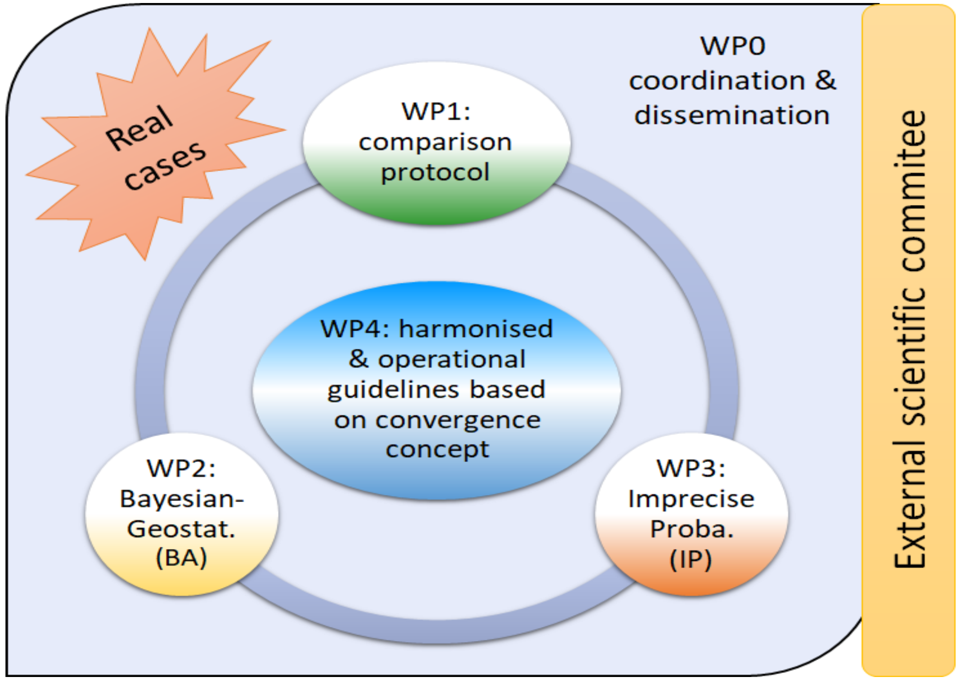
(1) Géosciences pour une Terre durable brgm
(2) ARMINES
(3) heudiasyc
(4) IRIT
(5) HESUS

anr agence nationale de la recherche

In one slide…

Objective:

Define a **harmonized framework** to exhaustively and transparently reflect **all uncertainties** along the **modelling chain** of **spatial data** while keeping **track of their origins** (knowledge imperfection and/or random variability)



Budget (ANR grant): 582 keuros; (total): 1.23 Meuros

Duration: 42 months (expected starting date April 2023)

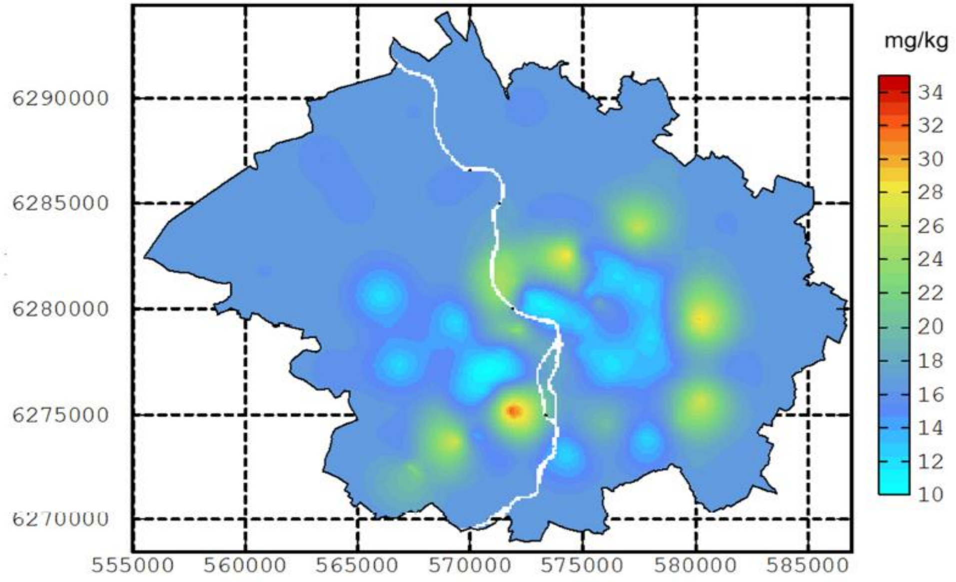
Early career scientists:

- 1 18-month post-doc (WP2)
- 1 12-month post-doc (WP4)
- 1 Phd (WP3, 1/2 salary)
- 1 research engineer (WP3)

	(1)	(2)	(3)	(4)	(5)
Statistics for environments	■	■			
Geostatistics	■	■			
Bayesian analysis		■	■	■	
Imprecise probability			■	■	
Decison making under uncertainty	■		■	■	■
Operational use	■	■			■

(1) Géosciences pour une Terre durable brgm (2) ARMINES (3) heudiasyc (4) IRIT (5) HESUS

Motivation: maps as a support for decision making

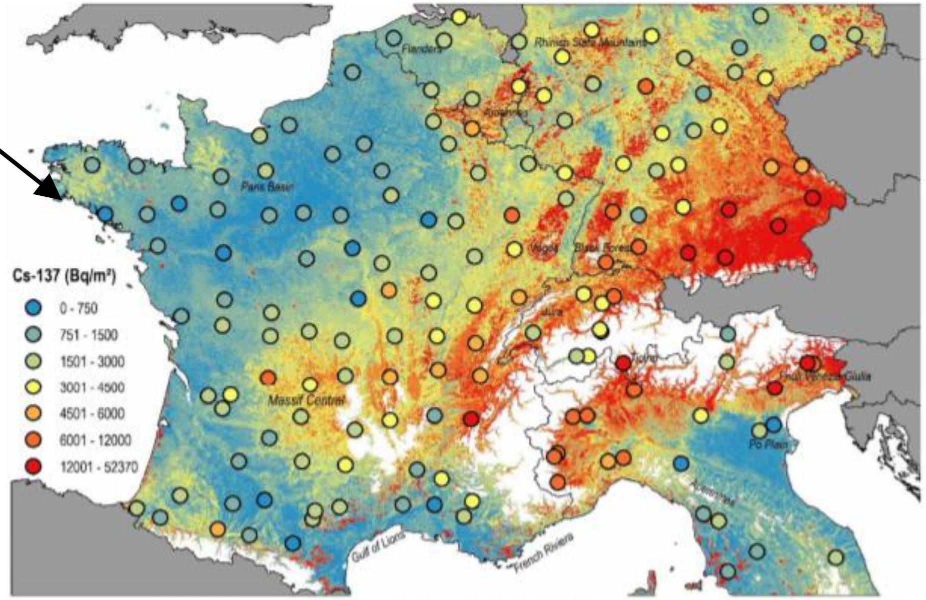


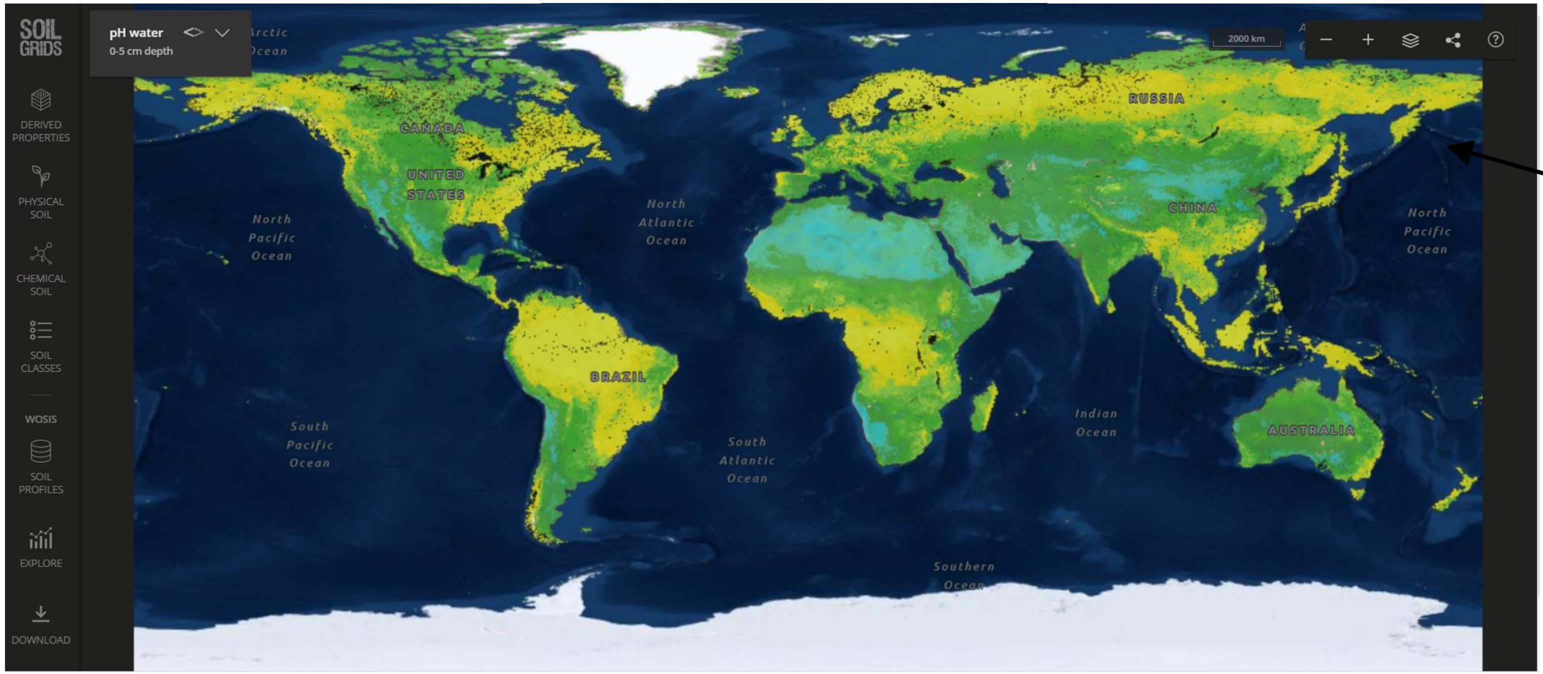
Total Petroleum Hydrocarbon (TPH) - Toulouse

Belbeze et al. (2019)

Concentration Cs

LUCAS database

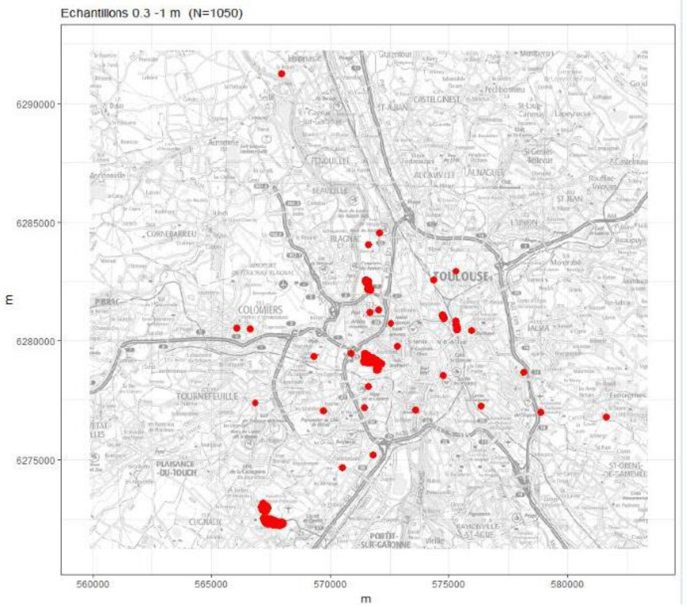


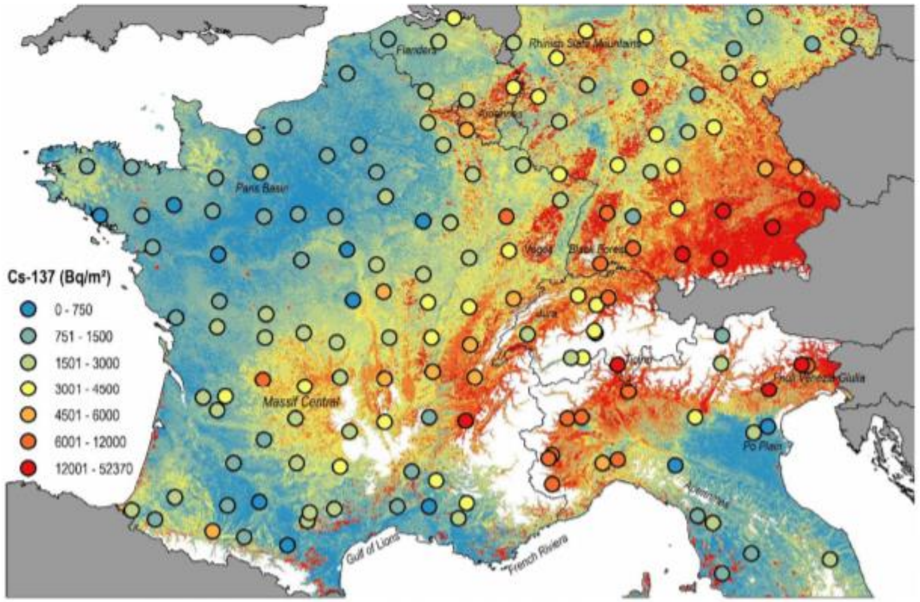


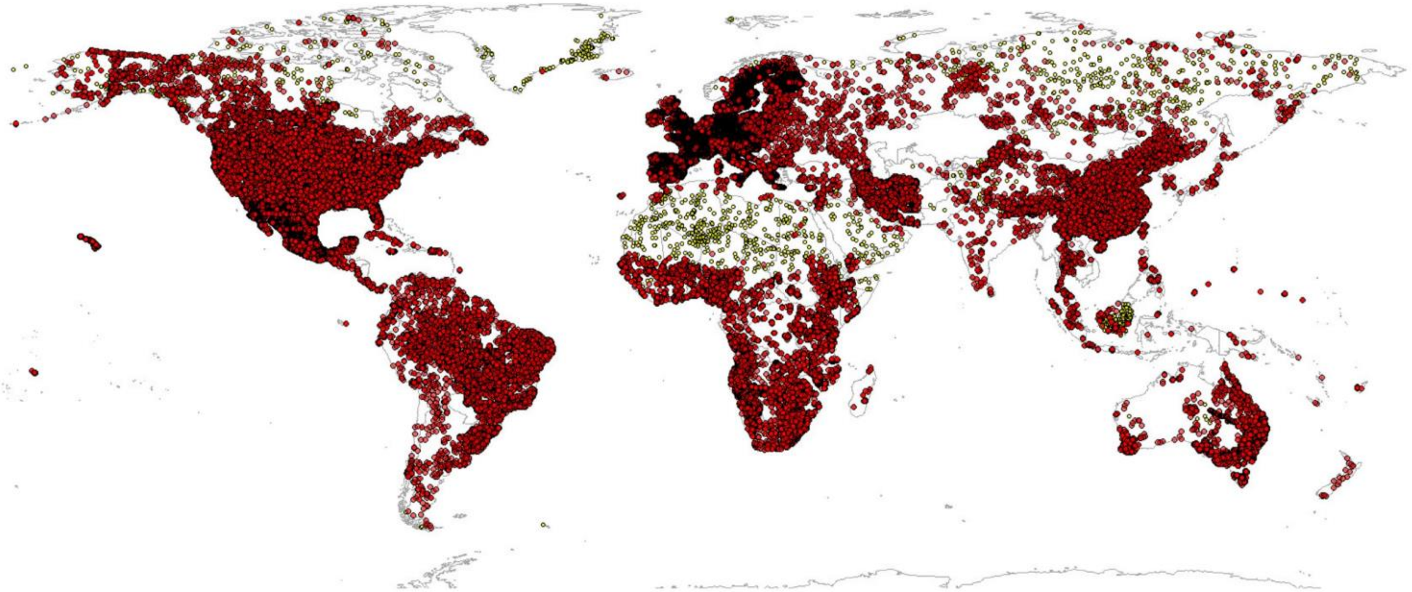
pH water

soilgrid.org

Problem: we only have limited (point) information about the environment

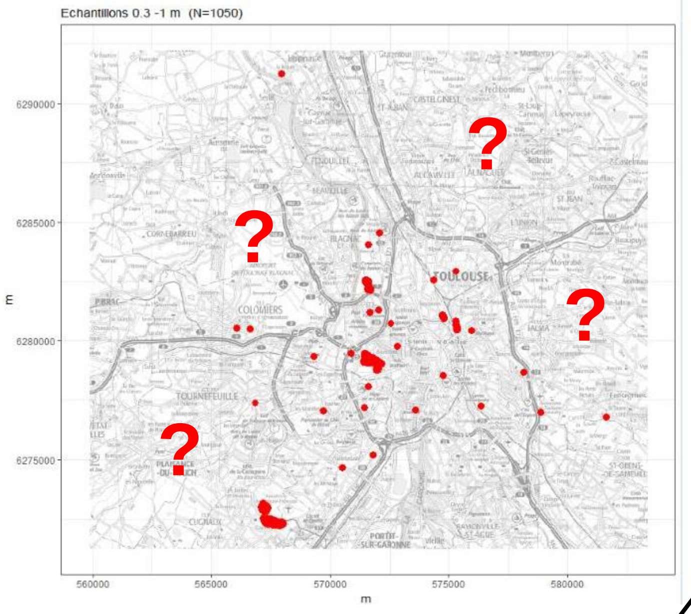




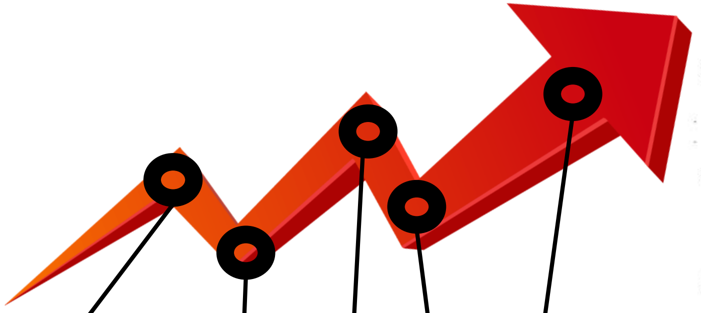


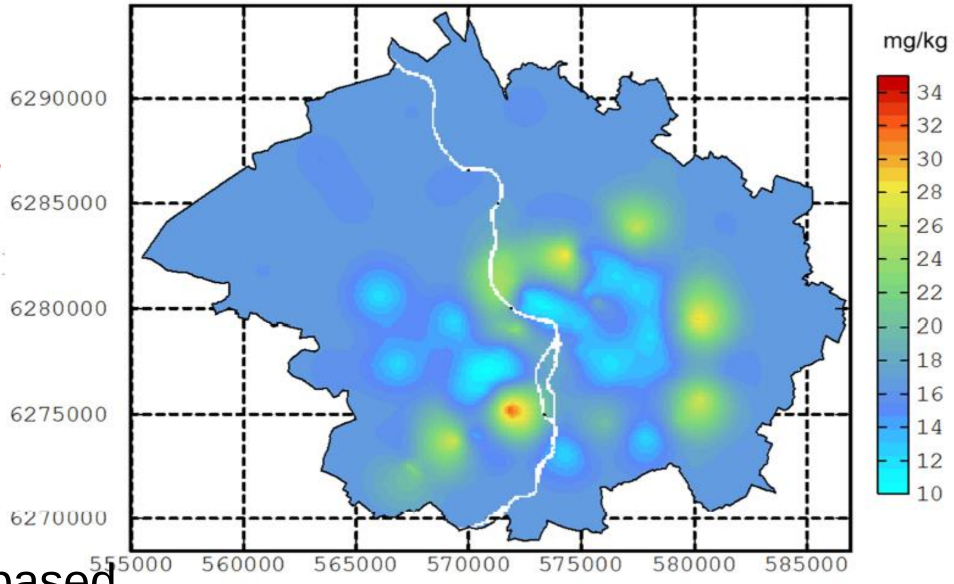
Problem: cascade of uncertainties

[Belbeze et al. (2019)]



Spatial modelling chain*





Model-based predictions

Model learning

Spatial model set up

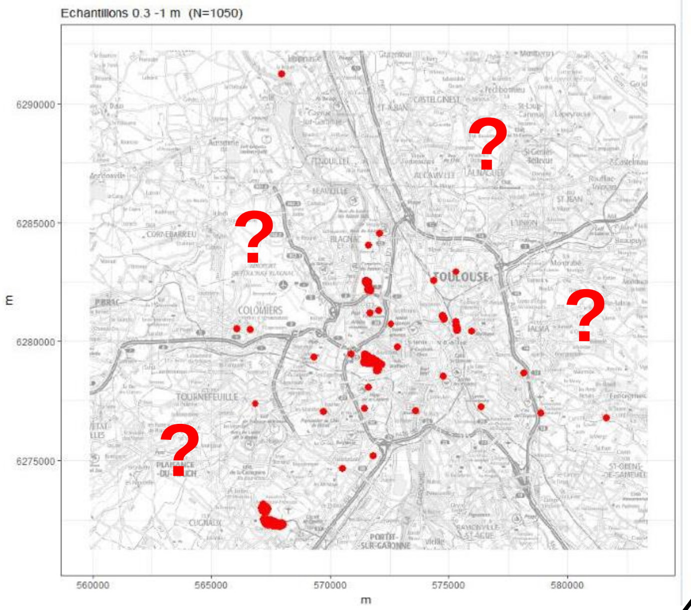
Information on spatial dependence structure

Data measurements & collection

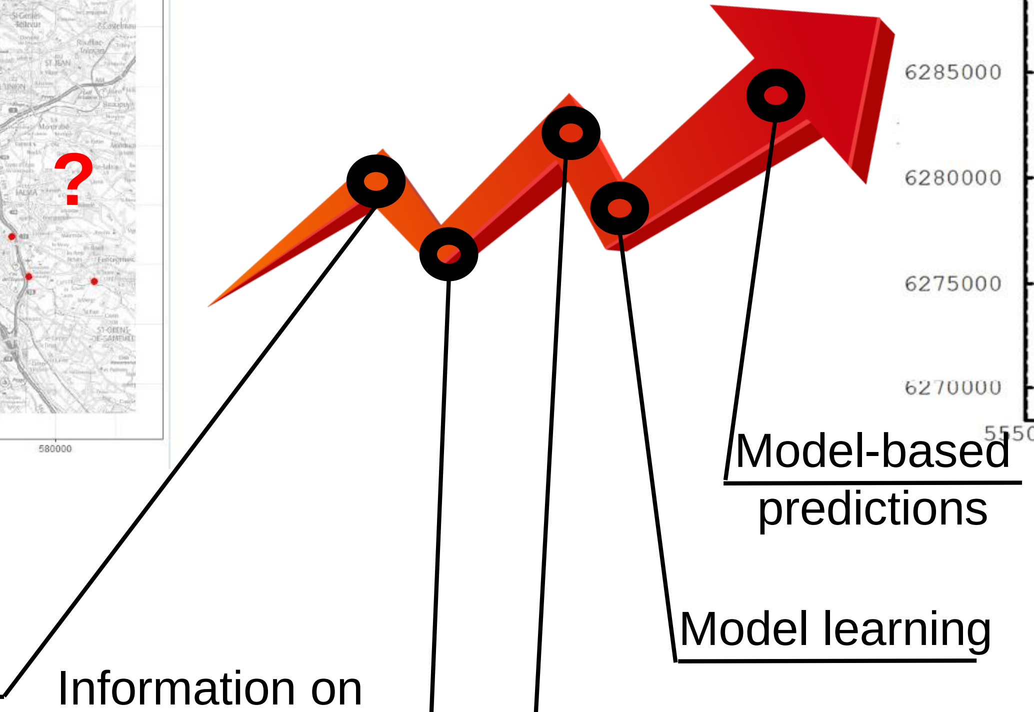
*HOUSES focuses on Geostatistics

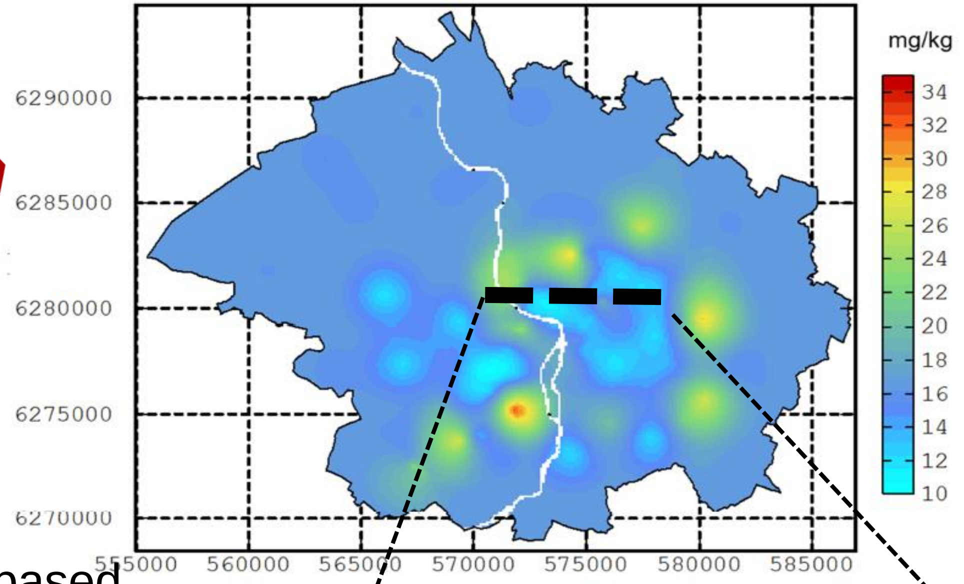
Problem: cascade of uncertainties

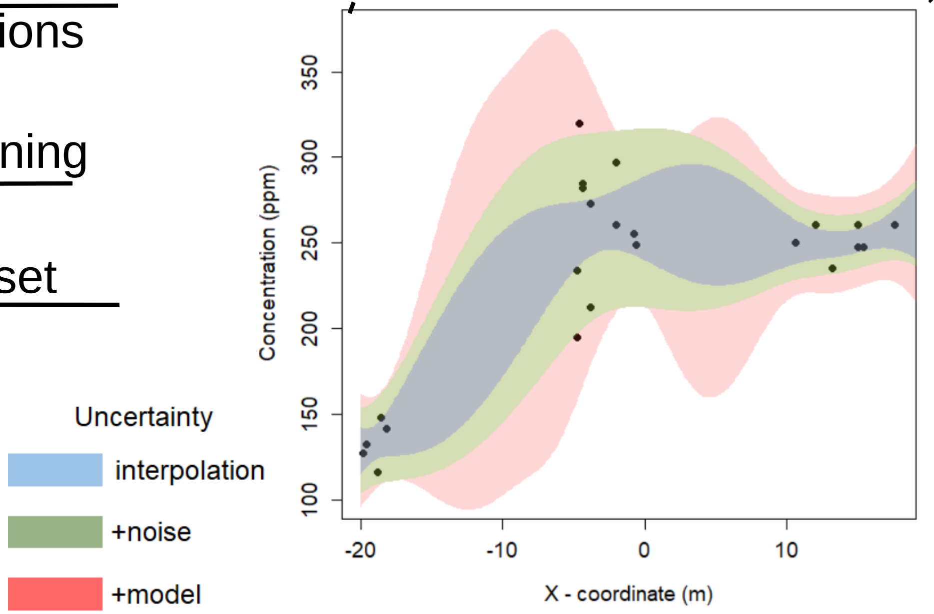
[Belbeze et al. (2019)]



Spatial modelling chain*





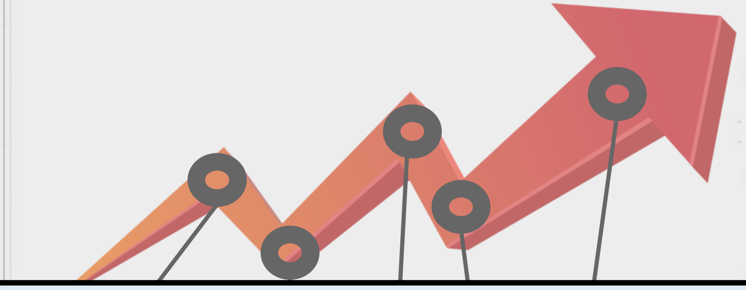


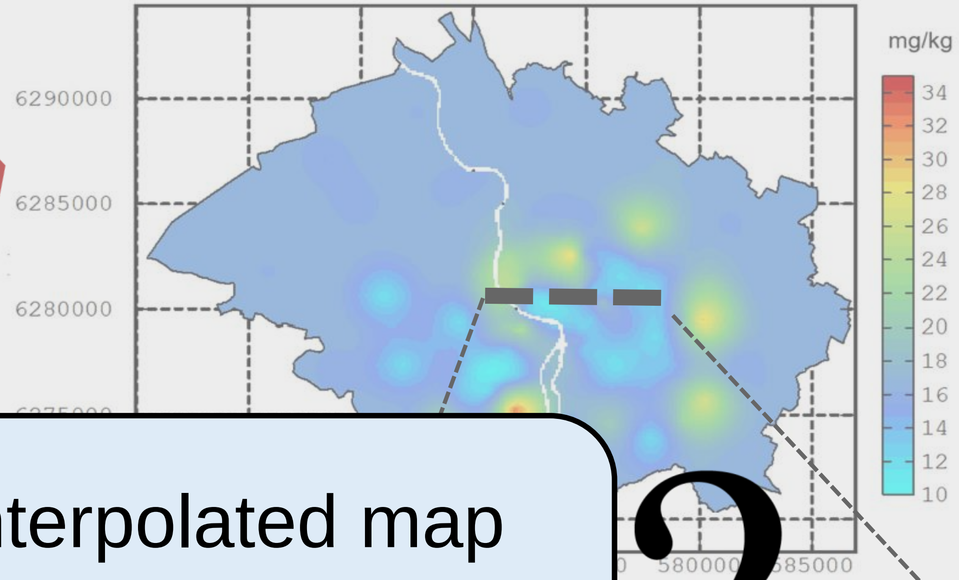
*HOUSES focuses on Geostatistics

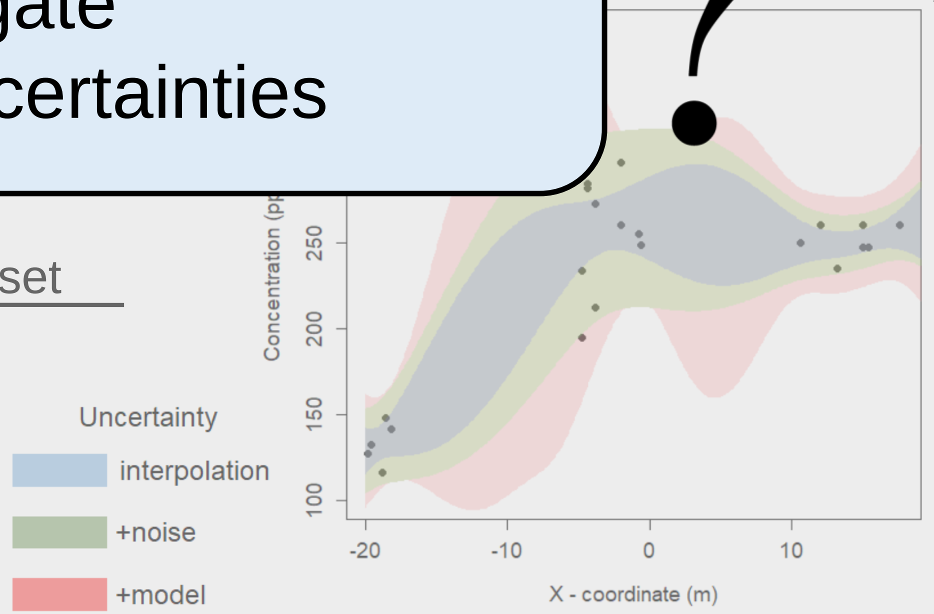
Problem: cascade of uncertainties

[Belbeze et al. (2019)]

Spatial modelling chain*







1. What **confidence** in the interpolated map
2. How **uncertainties** propagate
3. What **methods** to treat uncertainties

Data measurements & collection

Information on spatial dependence structure

Spatial model set up

*HOUSES focuses on Geostatistics

Key gaps in the current practices


Earth-Science Reviews 210 (2020) 103359

Contents lists available at ScienceDirect

Earth-Science Reviews

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

ELSEVIER

Review Article

Machine learning for digital soil mapping: Applications, challenges and suggested solutions

Alexandre M.J.-C. Wadoux*, Budiman Minasny, Alex B. McBratney

~50% with uncertainty estimates


Projet Exploratoire sur les Fonds Pédo-Géochimiques Urbains

Rapport final

BRGM/RP-70282-FR

Octobre 2020

<5% with uncertainty analysis

Key gaps in the current practices


Earth-Science Reviews 210 (2020) 103359

Contents lists available at ScienceDirect

Earth-Science Reviews

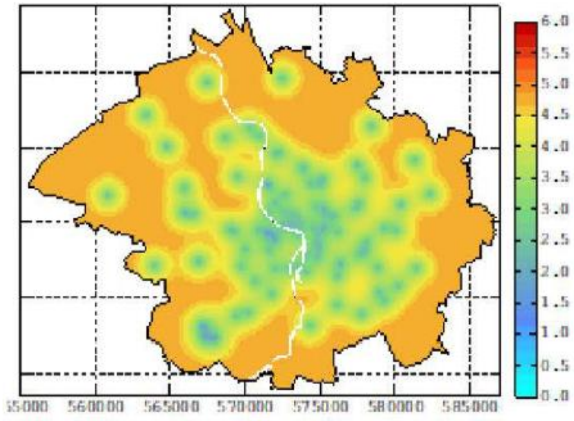
journal homepage: www.elsevier.com/locate/earscirev

Review Article

Machine learning for digital soil mapping: Applications, challenges and suggested solutions

Alexandre M.J.-C. Wadoux*, Budiman Minasny, Alex B. McBratney

~50% with uncertainty estimates



<5% with uncertainty analysis

Kriging variance?
Only partly reflects the uncertainties (interpolation error related to the spatial distribution of observations)

Key gaps in the current practices


Earth-Science Reviews 210 (2020) 103359

Contents lists available at ScienceDirect

Earth-Science Reviews

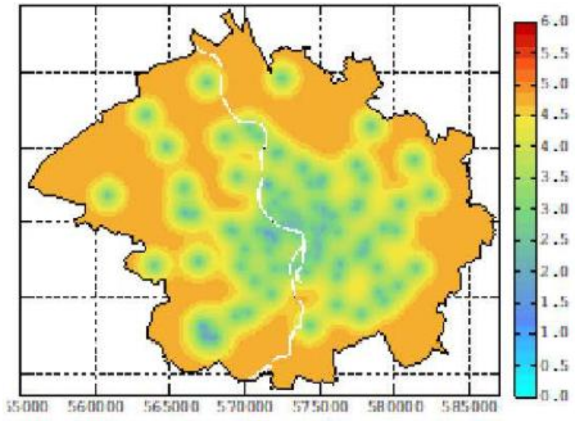
journal homepage: www.elsevier.com/locate/earscirev

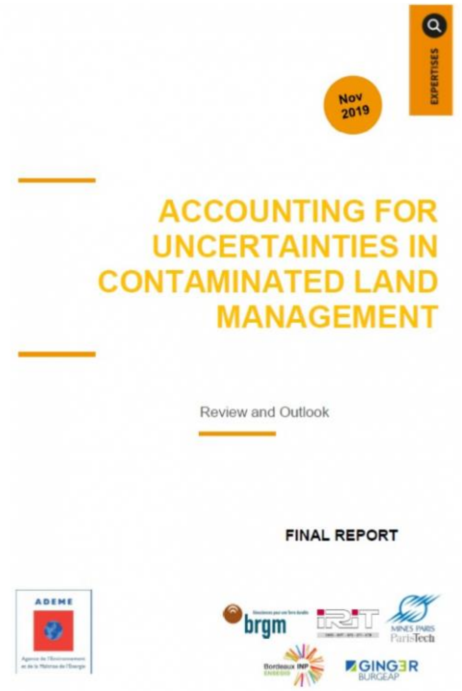
Review Article

Machine learning for digital soil mapping: Applications, challenges and suggested solutions

Alexandre M.J.-C. Wadoux*, Budiman Minasny, Alex B. McBratney

~50% with uncertainty estimates





<5% with uncertainty analysis

Kriging variance?
Only partly reflects the uncertainties (interpolation error related to the spatial distribution of observations)

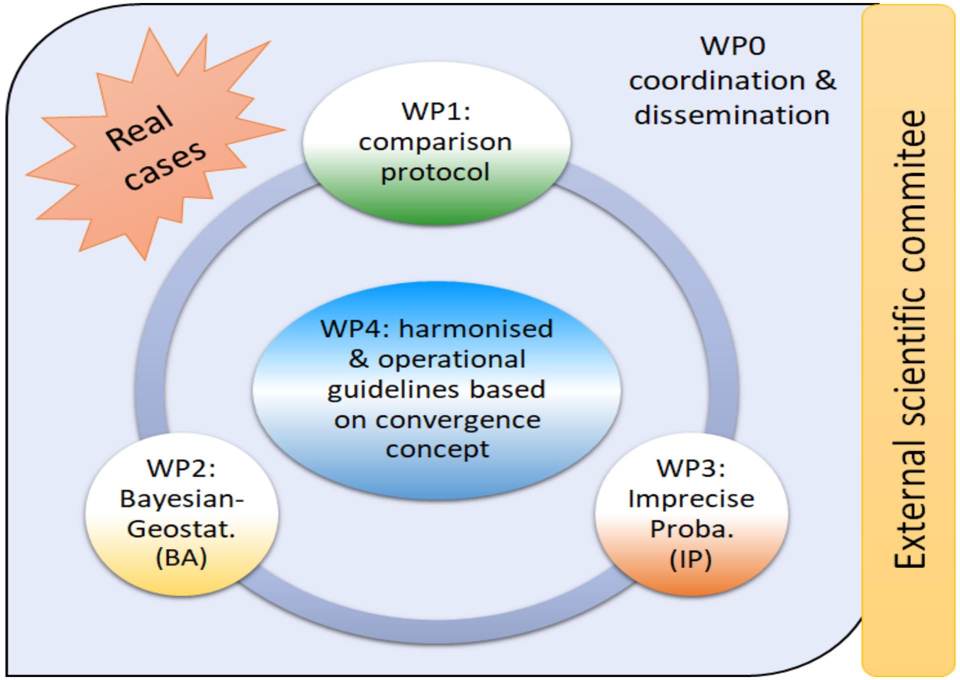
[…] importance of "*uncertainty sources*".

[…] the random character of these phenomena and which can be described (uncertainty of so-called "*stochastic*" origin),
[…] the incomplete and/or imprecise nature of our knowledge regarding these phenomena (uncertainty of "*epistemic*" origin)?

…Motivation for HOUSES…

Objective:

Define a **harmonized framework** to exhaustively and transparently reflect **all uncertainties** along the **modelling chain** of **spatial data** while keeping **track of their origins** (knowledge imperfection and/or random variability)



Budget (ANR grant): 582 keuros; (total): 1.23 Meuros

Duration: 42 months (expected starting date April 2023)

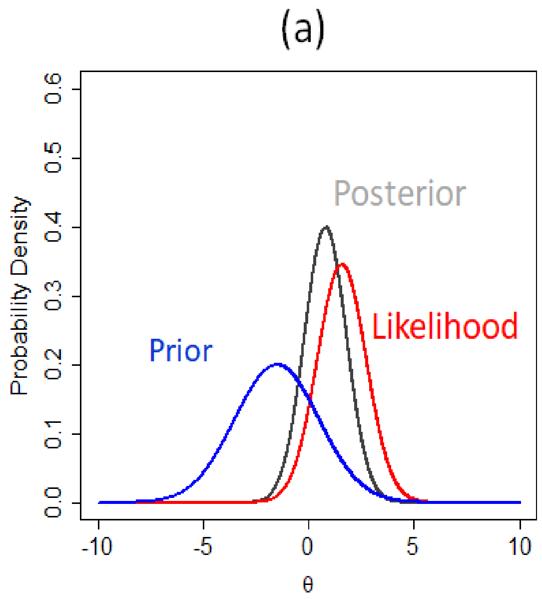
Early career scientists:

- 1 18-month post-doc (WP2)
- 1 12-month post-doc (WP4)
- 1 Phd (WP3, 1/2 salary)
- 1 research engineer (WP3)

	(1)	(2)	(3)	(4)	(5)
Statistics for environments	■	■			
Geostatistics	■	■			
Bayesian analysis		■	■	■	
Imprecise probability			■	■	
Decison making under uncertainty	■		■	■	■
Operational use	■	■			■

(1) Géosciences pour une Terre durable brgm (2) ARMINES (3) heudiasyc (4) IRIT (5) HESUS

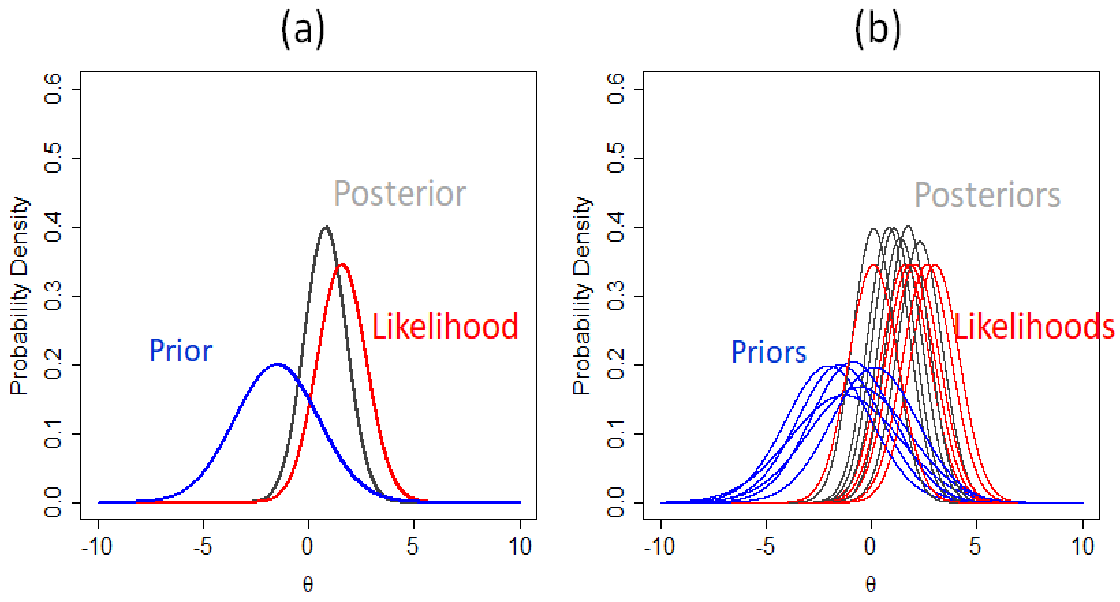
Key ingredient 1 – diverse uncertainty management frameworks



Bayesian approach
(BA)

Key ingredient 1 – diverse uncertainty management frameworks

Sensitivity to priors

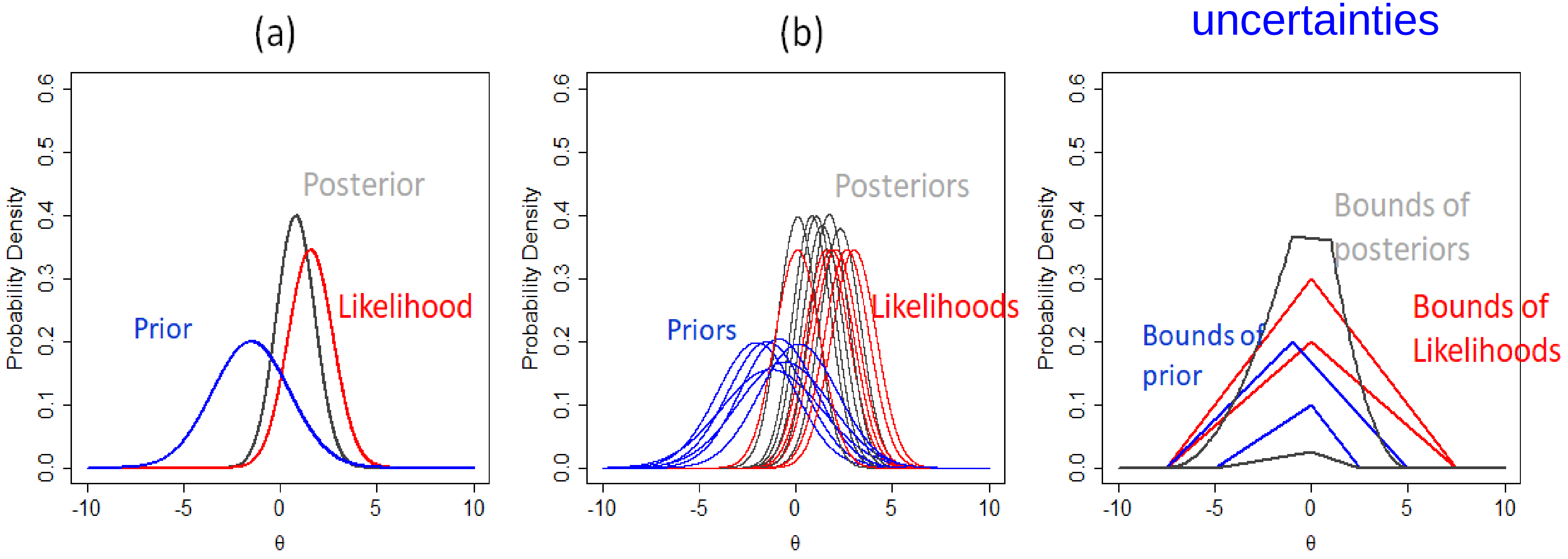


Bayesian approach (BA)

Robust Bayesian approach

Key ingredient – diverse uncertainty management frameworks

Probabilities to represent epistemic uncertainties

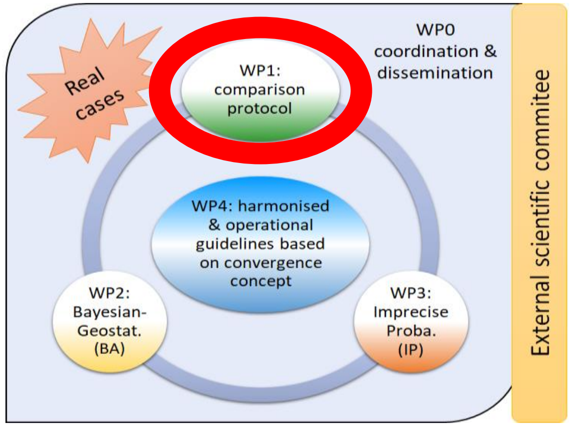


Bayesian approach (BA)

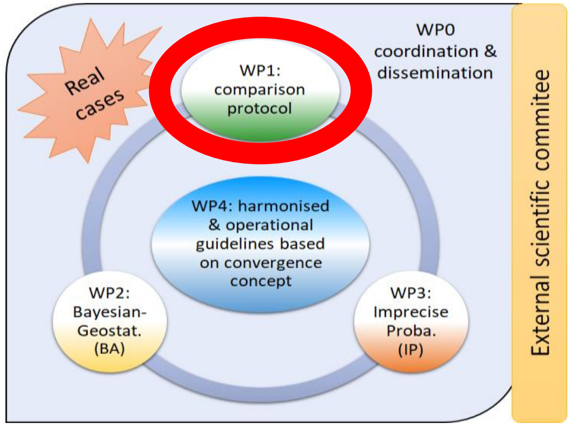
Robust Bayesian approach

Imprecise Probability (IP)

WP1 Setting up a common framework of comparison (co-lead. BRGM/HEUDIASYC)

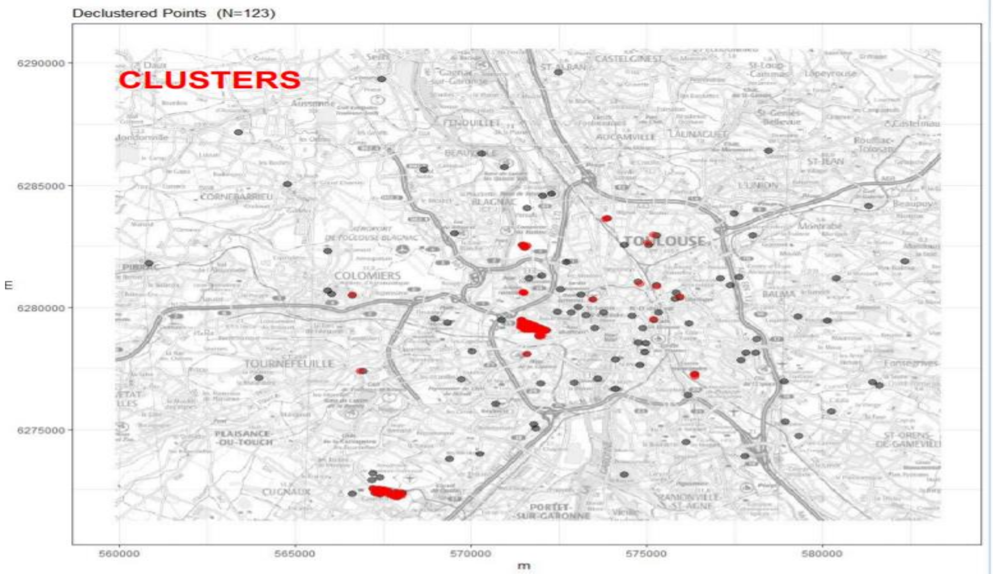


WP1 Setting up a common framework of comparison (co-lead. BRGM/HEUDIASYC)

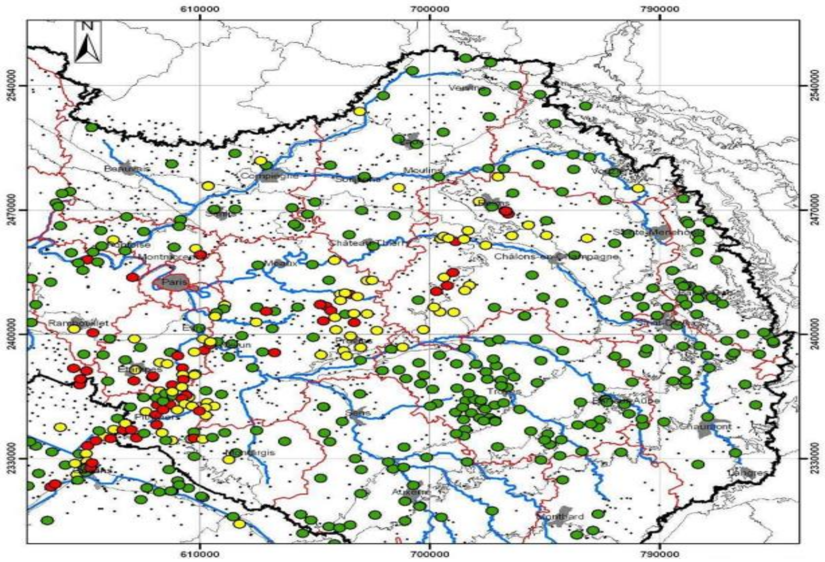


Task 1.1 – Design of experiments

➡ Experiments based on **real** Sparse Imprecise Clustered cases

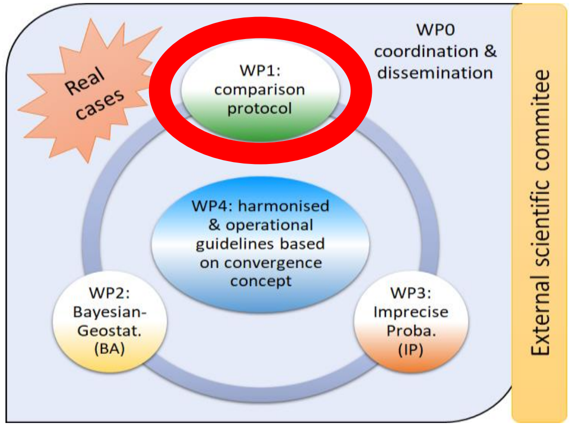


Sparse and clustered data for geochemical background mapping in Toulouse city ***[Belbèze et al. 2019]***



Clustered data for Trace elements' concentrations over a very large area in Paris basin ***[Gourcy et al. 2011]***

WP1 Setting up a common framework of comparison (co-lead. BRGM/HEUDIASYC)



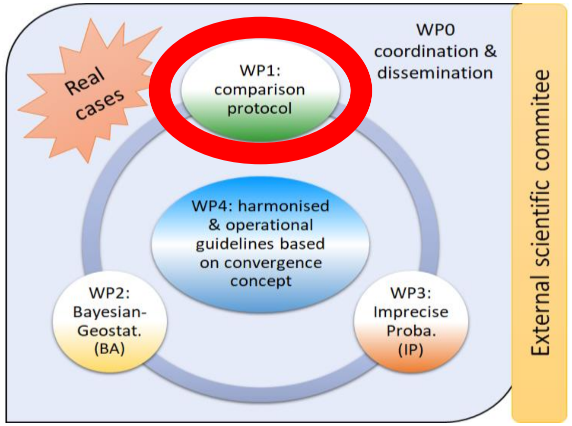
Task 1.1 – Design of experiments

Random experiments based on large datasets

Fig. 2. Examples of studied spatial samples. (a-b) simple random samples; (c) systematic random sample; (d-e) moderately clustered samples; (f-g) strongly clustered samples; (h-i) strongly clustered, gapped samples. Except for the systematic sample (c), the sample size always amounted to 5000. The systematic sample had an expected size of 5000 but realized samples varied in size between 4998 and 5056.

WP1 Setting up a common framework of comparison (co-lead. BRGM/HEUDIASYC).

Task 1.2 – Protocol of comparison



What is a **'good'** uncertainty model?

Envisioned criteria to compare methods

Reliability: coverage and width of uncertainty intervals.

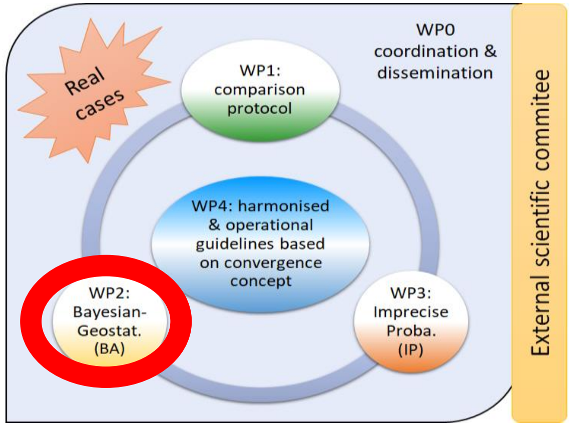
Computability: computational burden, "simplicity" of implementation, degree of expertise required for the implementation, the interpretability, the simplicity for communicating

Relation to knowledge context. Capability to reflect the whole cascade of uncertainties, flexibility and adaptation to the knowledge context (following the convergence concept

WP2 Uncertainty analysis within geostatistical & Bayesian framework (ARMINES).

Task 2.1 – Developments (2 year post-doc)

Focus on Bayesian hierarchical modelling [Gelfand et al. 2010].



Complexiy of environmental variables

Appropriate parametrisation

Computational burden

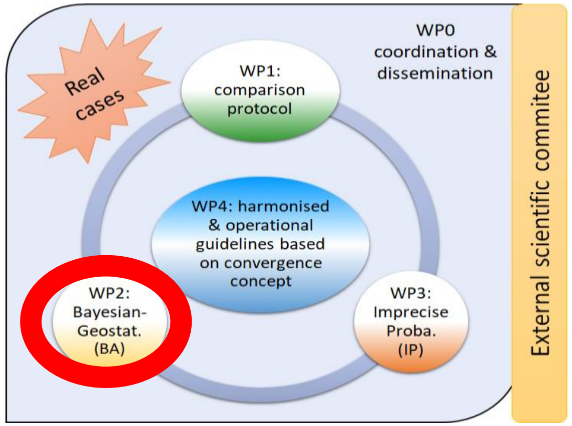
Added value w.r.t. classical approaches

- **Trans-Gaussian processes**
  - Necessitate a preliminary transformation
  - **Penalised complexity priors** to define the parameters' transformations
- Bayesian inference → **MCMC algorithms** = computationally intensive.
  - variational approaches like INLA **[Rue et al. 2017]**
  - Vecchia's approximation **[Katzfuss & Guiness 2021]**
- **Comparison with classical approaches**

WP2 Uncertainty analysis within geostatistical & Bayesian framework (ARMINES).

Task 2.1 – Developments (2 year post-doc)

Focus on Bayesian hierarchical modelling [Gelfand et al. 2010].



Complexiy of environmental variables

Appropriate parametrisation

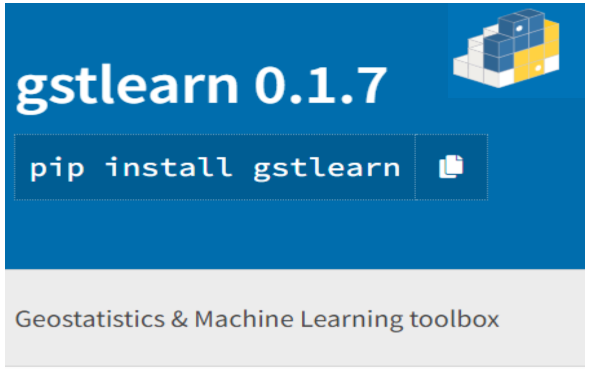
Computational burden

Added value w.r.t. classical approaches

- **Trans-Gaussian processes**
  - Necessitate a preliminary transformation
  - **Penalised complexity priors** to define the parameters' transformations
- Bayesian inference → **MCMC algorithms** = computationally intensive.
  - variational approaches like INLA **[Rue et al. 2017]**
  - Vecchia's approximation **[Katzfuss & Guiness 2021]**
- **Comparison with classical approaches**

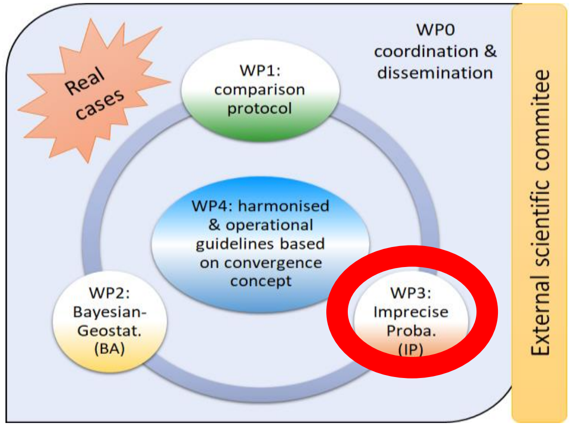
Task 2.2 – Experiments

- **Application to comparison protocol (link to WP1)**
- **Implementation in gstlearn**
- **Hosting short period visits** of HOUSES researchers at the Geostatistics group of the centre de Géosciences of ARMINES



WP3 Uncertainty analysis within IP framework (UTC/IRIT).

Task 3.1 – Developments (1 PhD)



Two **IP approaches will be investigated**

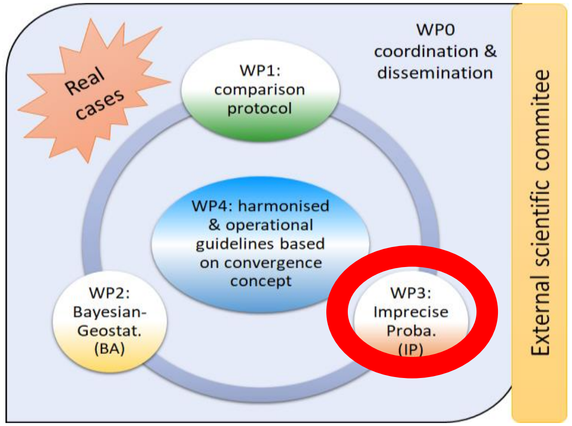
Boosting existing approaches

Or

Change the viewpoint?

- **Same underlying statistical hypothesis.**
  - multiple priors (in the form of sets) instead of one, such as in **[Mangili 2016]** extending Gaussian processes,
  - a plug-in to the classical approach, such as in conformal prediction **[Mao et al. 2020],**
- **Departing from the traditional statistical hypothesis**
  - a constrained (fuzzy) optimisation problem **[Dubois et al. 2014]**
  - information fusion problem **[Shinde et al. 2021]**

WP3 Uncertainty analysis within IP framework (UTC/IRIT).



Task 3.1 – Developments (1 PhD)

Two **IP approaches will be investigated**

Boosting existing approaches

Or

Change the viewpoint?

- **Same underlying statistical hypothesis.**
  - multiple priors (in the form of sets) instead of one, such as in **[Mangili 2016]** extending Gaussian processes,
  - a plug-in to the classical approach, such as in conformal prediction **[Mao et al. 2020],**
- **Departing from the traditional statistical hypothesis**
  - a constrained (fuzzy) optimisation problem **[Dubois et al. 2014]**
  - information fusion problem **[Shinde et al. 2021]**

Task 3.2 – Experiments (1 Research Engineer)

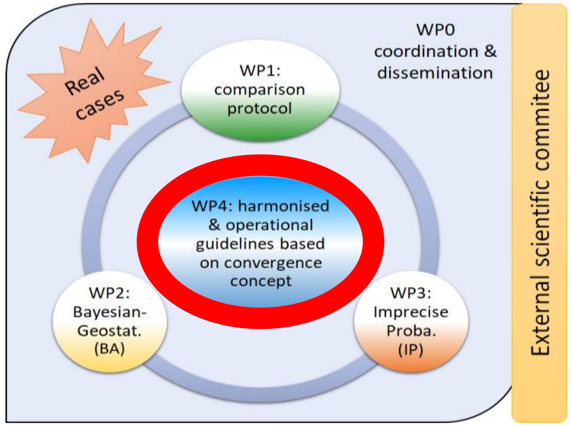
- **Application to comparison protocol (link to WP1)**
- **Implementation by a 12-month Research Engineer e.g. HYRISK**
- **Visiting periods** of researchers at ARMINES – center of Geostatistics **(link to WP2)**

HYRISK

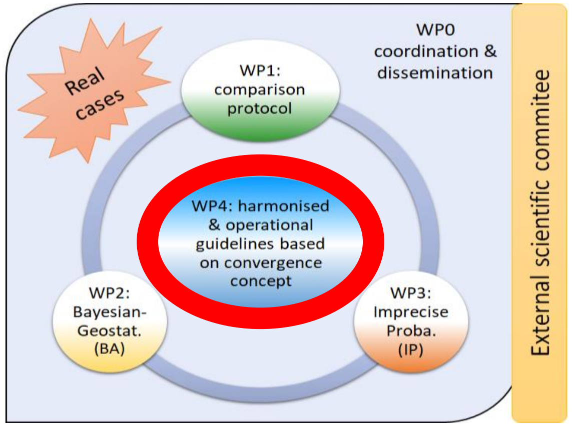
WP4 Toward harmonized and operational guidelines (HESUS/BRGM, incl. 1 postdoc)

Task 4.1 – Inter-comparison & recommendations (12 month post-doc)

- **Based on WP2.2 and WP3.2**
- **Data competition** (Hackathon)
- **Supplement or Complement** vision?



WP4 Toward harmonized and operational guidelines (HESUS/BRGM, incl. 1 postdoc)

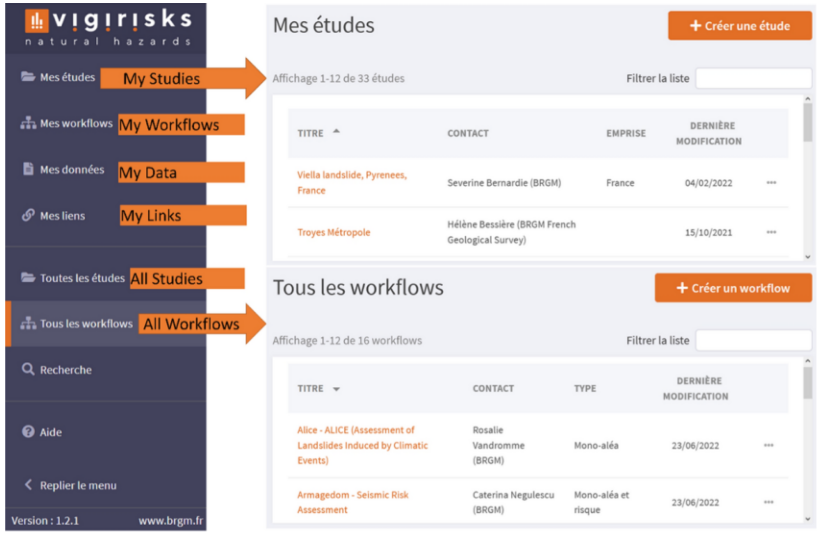


Task 4.1 – Inter-comparison & recommendations (12 month post-doc)

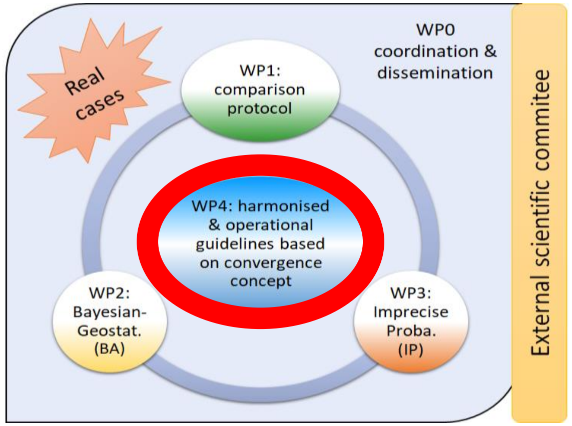
- **Based on WP2.2 and WP3.2**
- **Data competition** (Hackathon)
- **Supplement or Complement** vision?

Task 4.2 – Support to operational activities (HESUS, BRGM)

- Connection and **transfer feasibility** to already existing operational facilities
- Set the **basis to operationalization**



WP4 Toward harmonized and operational guidelines (HESUS/BRGM, incl. 1 postdoc)



Task 4.1 – Inter-comparison & recommendations (12 month post-doc)

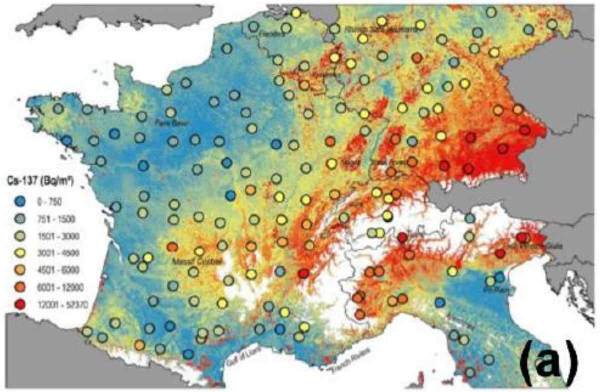
- **Based on WP2.2 and WP3.2**
- **Data competition** (Hackathon)
- **Supplement or Complement** vision?

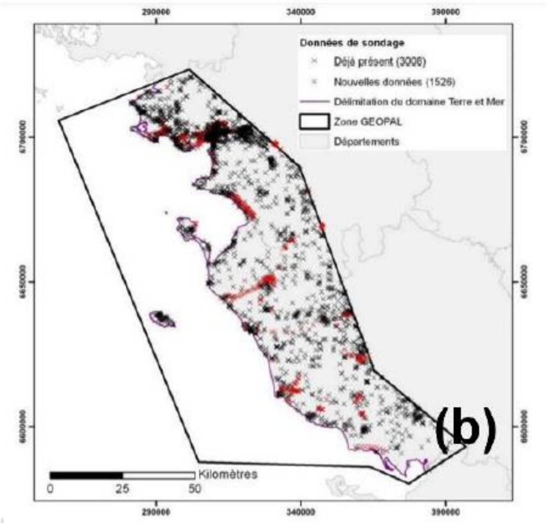
Task 4.2 – Support to operational activities (HESUS, BRGM)

- Connection and **transfer feasibility** to already existing operational facilities
- Set the **basis to operationalization**

Task 4.3 – Guidelines & Transfer to other domains

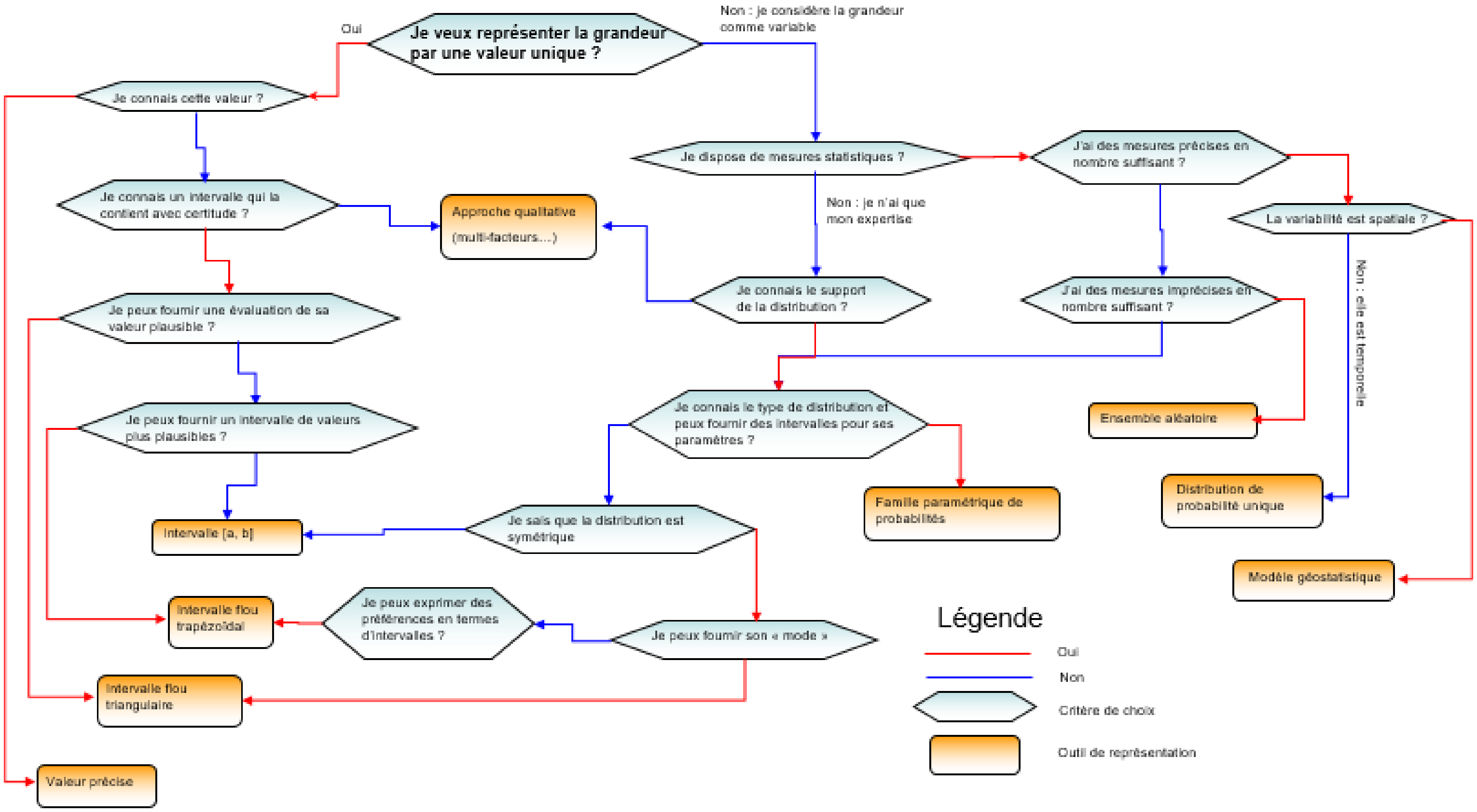
- Harmonized **guidelines**
- **(Stress) testing** using new real cases
- 2-day **final workshop**





An example of guidelines for non-spatial data

[Dubois & Guyonnet (2011)]



HOUSES Expected results

1. **Operational. Bring together different communities**
   - Geostatistics, uncertainty quantification (UQ), environmental and natural risk analysis
   - (1) explore the added values of **uncertainty-aware practices** to improve decision-making;
   - (2) converge towards **harmonized guidelines**.

Useful for any geo-scientists working with maps

HOUSES Expected results

1. **Operational. Bring together different communities**
   - Geostatistics, uncertainty quantification (UQ), environmental and natural risk analysis
   - (1) explore the added values of **uncertainty-aware practices** to improve decision-making;
   - (2) converge towards **harmonized guidelines**.

2. **Methodological. In-depth feasibility and inter-comparison analysis**
   - Major frameworks for uncertainty management (**geostatistical BA and IP**)
   - (1) potentialities for **capturing types of information/knowledge** important to the decision-maker;
   - (2) the **validity domains** w.r.t. the context;
   - (3) Unlock **key implementation limitations** for making these frameworks usable and operational.

Links to decision-making under uncertainty

HOUSES Expected results

1. **Operational. Bring together different communities**
   - Geostatistics, uncertainty quantification (UQ), environmental and natural risk analysis
   - (1) explore the added values of **uncertainty-aware practices** to improve decision-making;
   - (2) converge towards **harmonized guidelines**.

2. **Methodological. In-depth feasibility and inter-comparison analysis**
   - Major frameworks for uncertainty management (**geostatistical BA and IP**)
   - (1) potentialities for **capturing types of information/knowledge** important to the decision-maker;
   - (2) the **validity domains** w.r.t. the context;
   - (3) Unlock **key implementation limitations** for making these frameworks usable and operational.

3. **Formal. Define the mathematical setting for knowledge representation of spatial data**
   - probabilities, intervals, Fuzzy sets, hybrid, new?
   - (1) sufficient **flexibility** for modelling the different information at all stages of the spatial modelling chain;
   - (2) **continuity/transition** from quasi-total ignorance to data-rich situations.

Links to AI/machine learning

Thank you for your attention!