RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Geosciences pour une Terre durable
brgm

Explainable machine learning to help the prediction of Geoscience processes: introduction with a focus on the challenges

Jeremy Rohmer

25 May 2023

With H. Breuillard, S. Belbeze, R. Chassagne, A. Henriot

Géosciences pour une Terre durable

brgm

THE FRENCH GEOLOGICAL SURVEY

The BRGM is France’s public reference institution for **Earth Science** applications for the management of surface and subsurface resources and risks.

Its activities are geared to scientific research, support to public policy development and international cooperation.

Geology and knowledge of the subsurface

Groundwater management

Risks and spatial planning

Mineral resources and the circular economy

Subsurface potential for the energy transition

Data, digital services and infrastructure

BRGM SERVICE GÉOLOGIQUE NATIONAL WWW.BRGM.FR

Outline

- Context of ‘prediction’ at BRGM
- Current practices based on Uncertainty Quantification tools
- Towards explainable machine learning and open questions

Outline

- Context of ‘prediction’ at BRGM
- Current practices based on Uncertainty Quantification tools
- Towards explainable machine learning and open questions

General setting [1,2]

$X \longrightarrow$ **Natural system** $\longrightarrow Y$

"Predictor variables" "Response variables"

Mathematically

$$y = f(X)$$

- **Science:** Extract information about the law of nature—the function *f.*
- **Prediction:** Predict what the response variables Y are going to be with the predictor variables X revealed to us.
- **Numerical simulators or Machine Learning (ML) tools (denoted g)** try to quantify the relationship under "nature" creating an input output mapping:

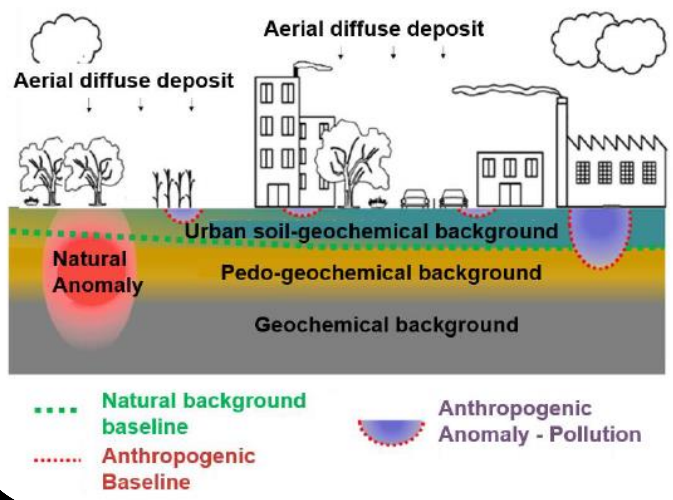
$$y = f(X) \approx g(X)$$

[1] Zhao & Hastie (2021); [2] Borgonovo (2021)

Soil & water pollution

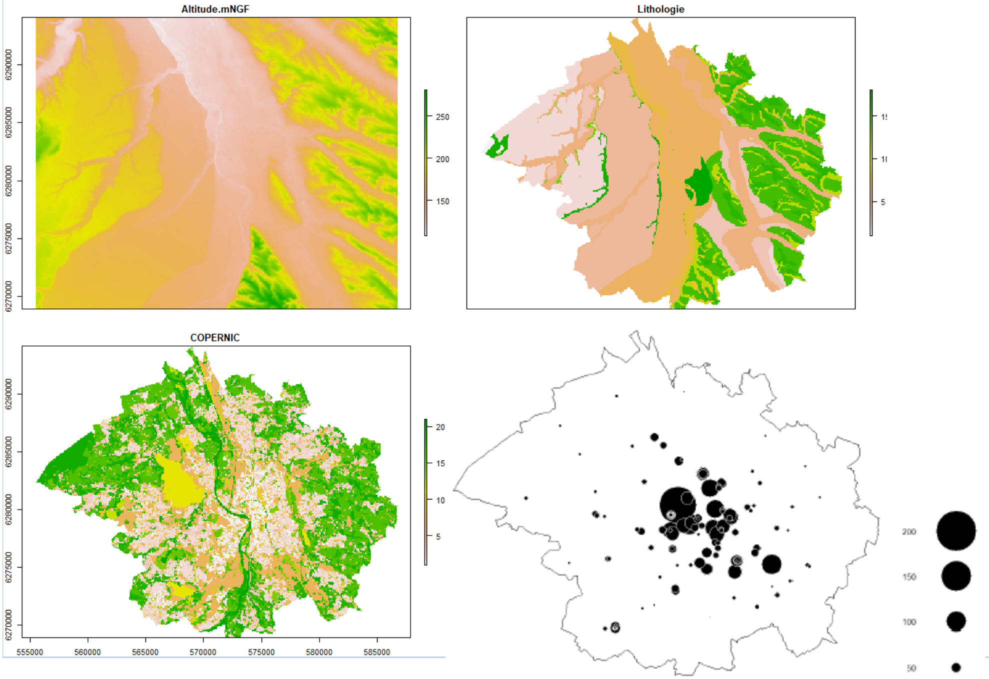
X

Punctual observations + spatial predictors + Expert knowledge



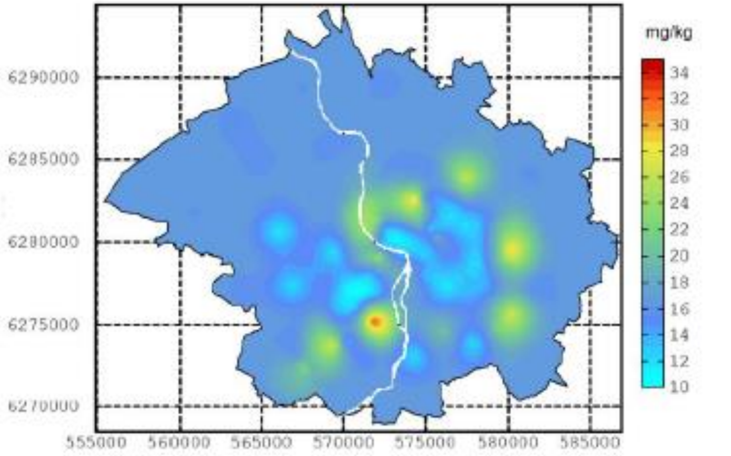
Y

Map of pollutant concentration at Toulouse [1]



g

= geostatisical model (with expert choices: top cut, censured data replacement, variogram choice)
Or
ML (with fewer expert choices)



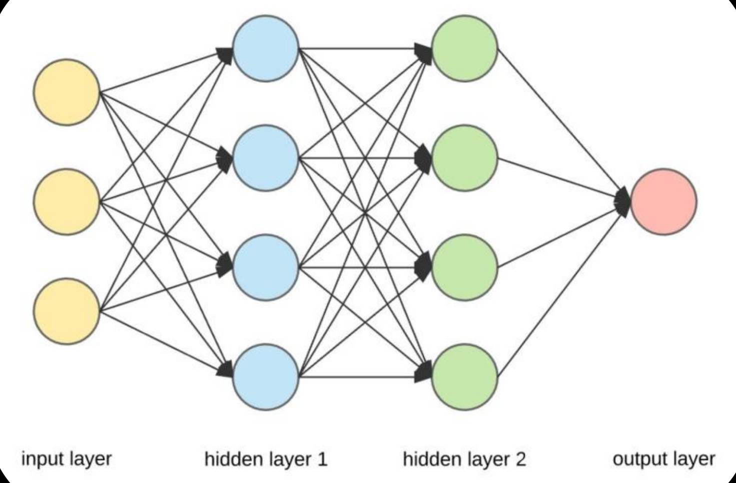
And many more….

[1] Belbeze et al. (2019)

Géosciences pour une Terre durable
brgm

Mineral prospectivity

X

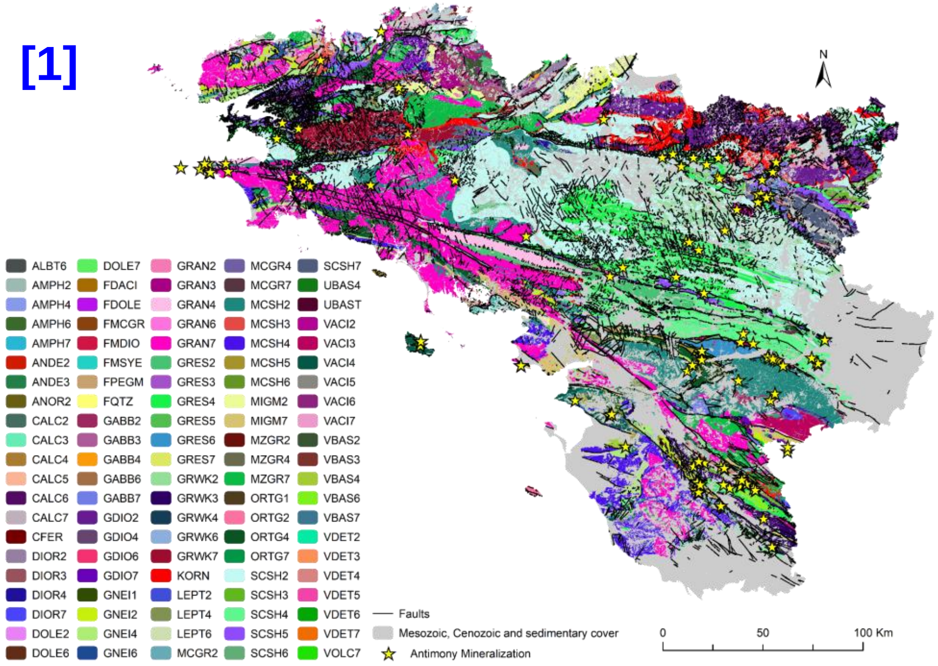


Y

Punctual observations (mineralization) + spatial predictors (geological map, geophysical measurements, etc.)

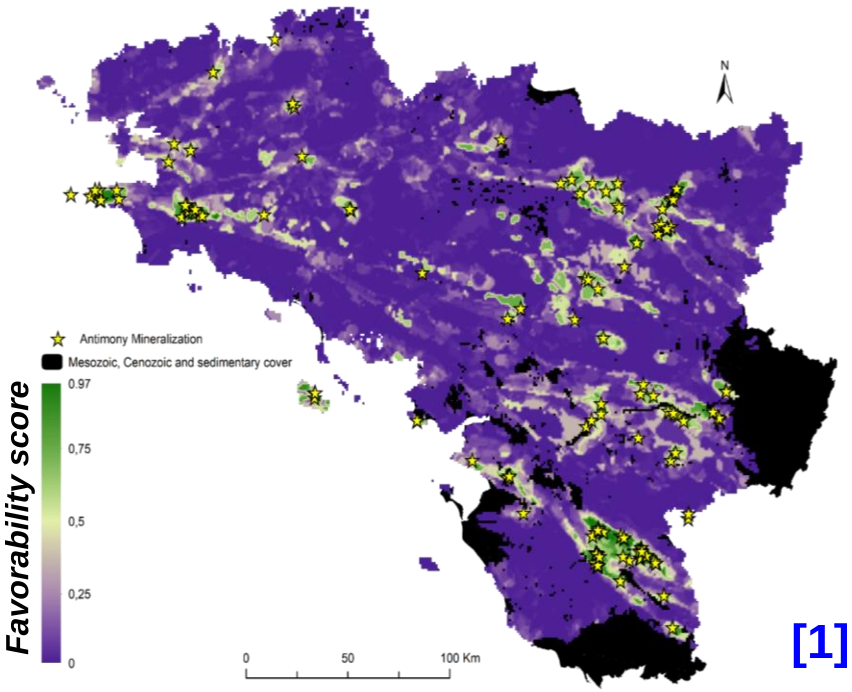
Favorability map (~ probability of mineralization)

[1]



g

Machine/deep learning method



[1]

[1] Vella and co-authors, 2022

Géosciences pour une Terre durable
brgm

Risk assessment

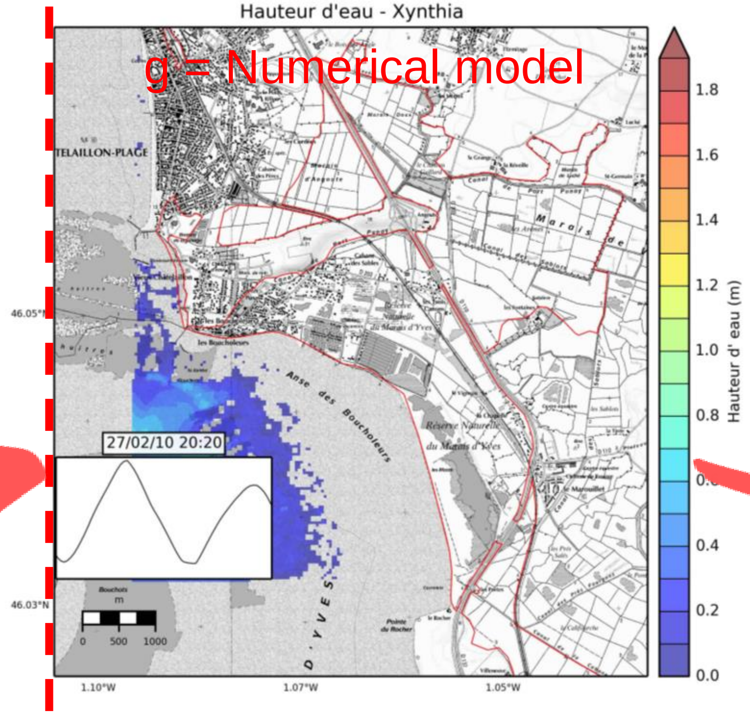
X

Multiple time series describing the offshore forcing conditions (wave, water levels, wind)

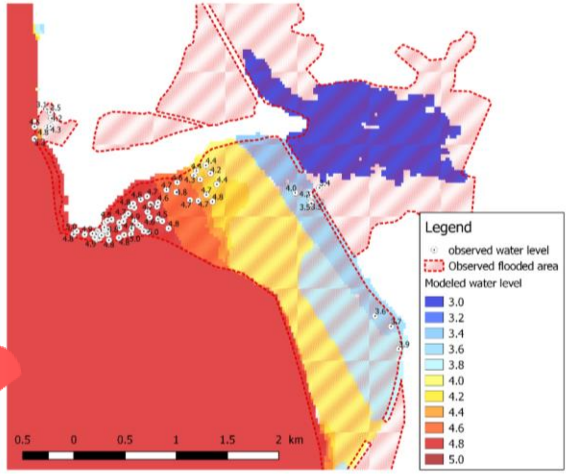
Y

Map of maximum water height induced by marine flooding [1]

+ spatial parameters (bathymetry, Manning coef., etc.)



Boundary conditions



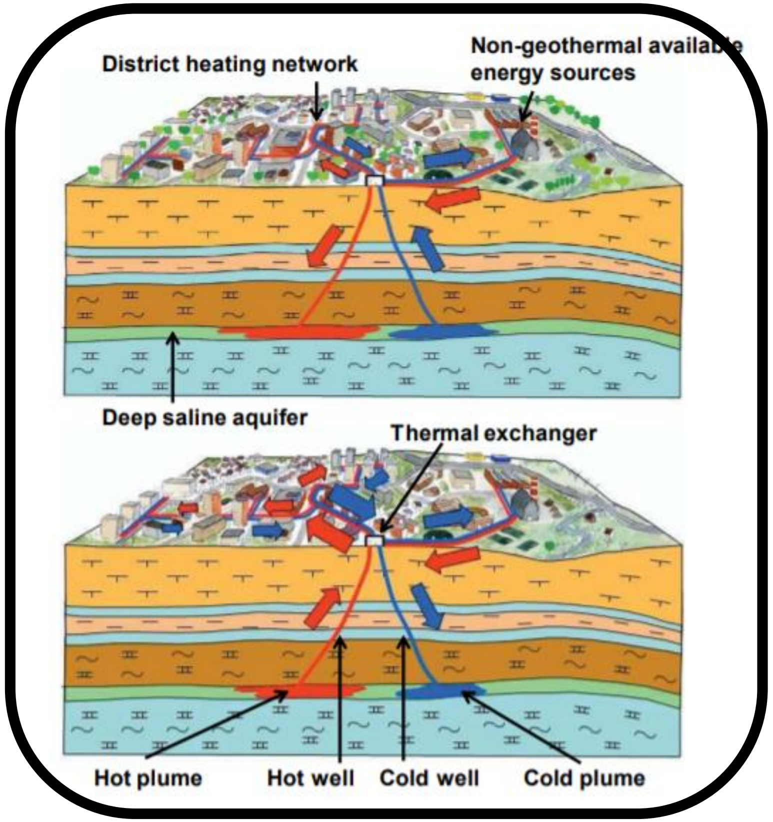
[1] Pedreros, Idier and co authors

Géosciences pour une Terre durable
brgm

Geothermal activities

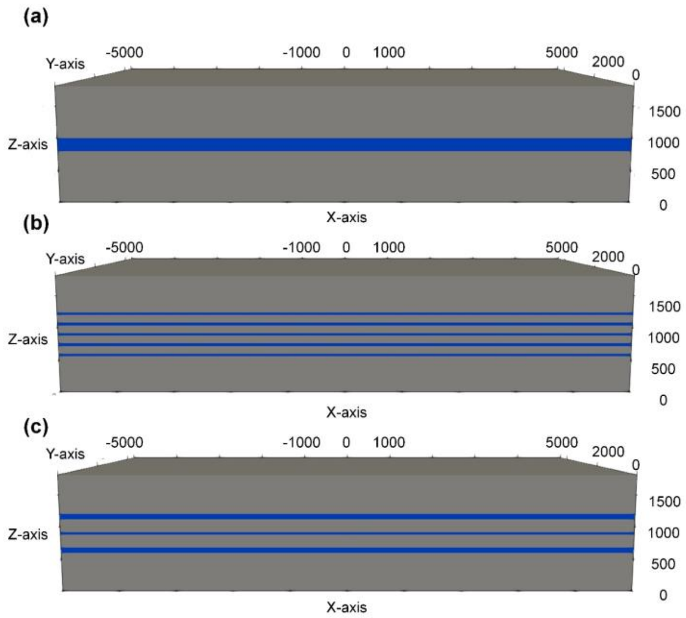
X

Characteristics of rock formations (permeability, porosity, etc.)
Geometry of the domain,
Reservoir architecture



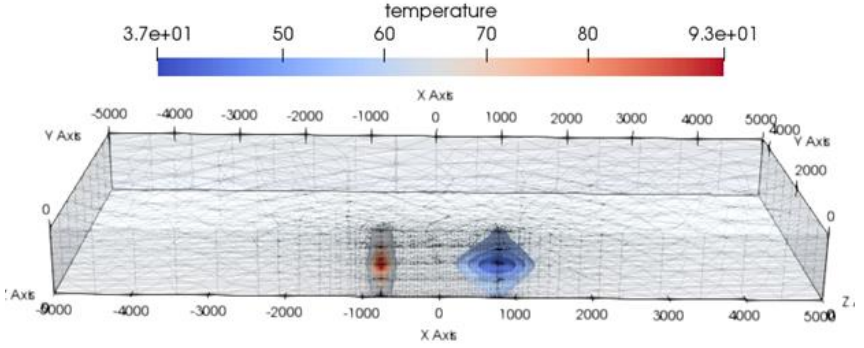
Y

Time and space evolution of temperature at depth
[1]



g

Numerical model



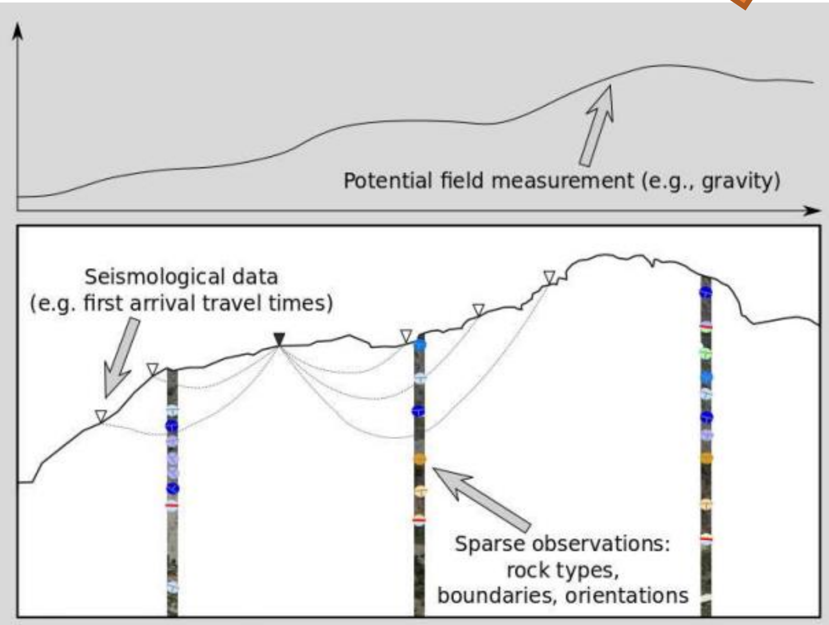
[1] Armandine les Landes, Maragna and co authors

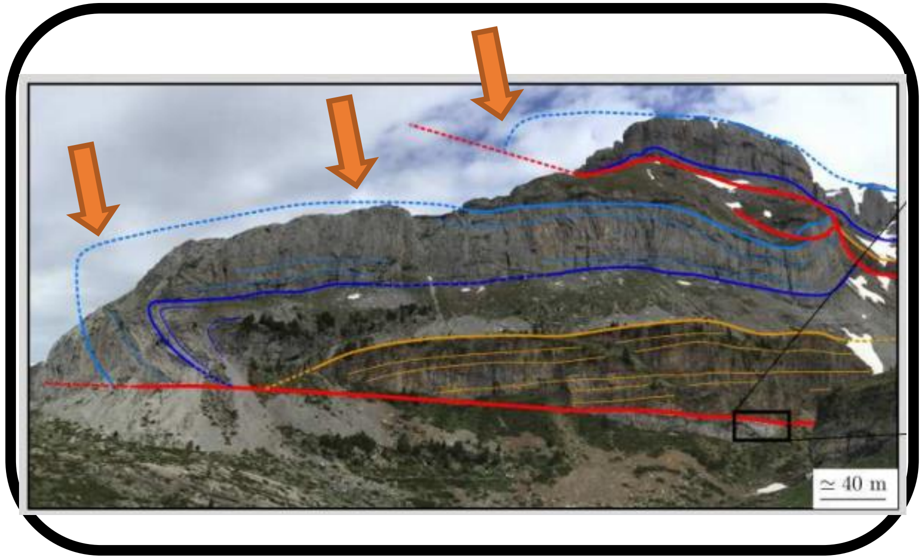
Géosciences pour une Terre durable
brgm

Geomodelling

X

Borehole (punctual) measurements
Geophysical imaging (spatial)
Field observations (interpretation)



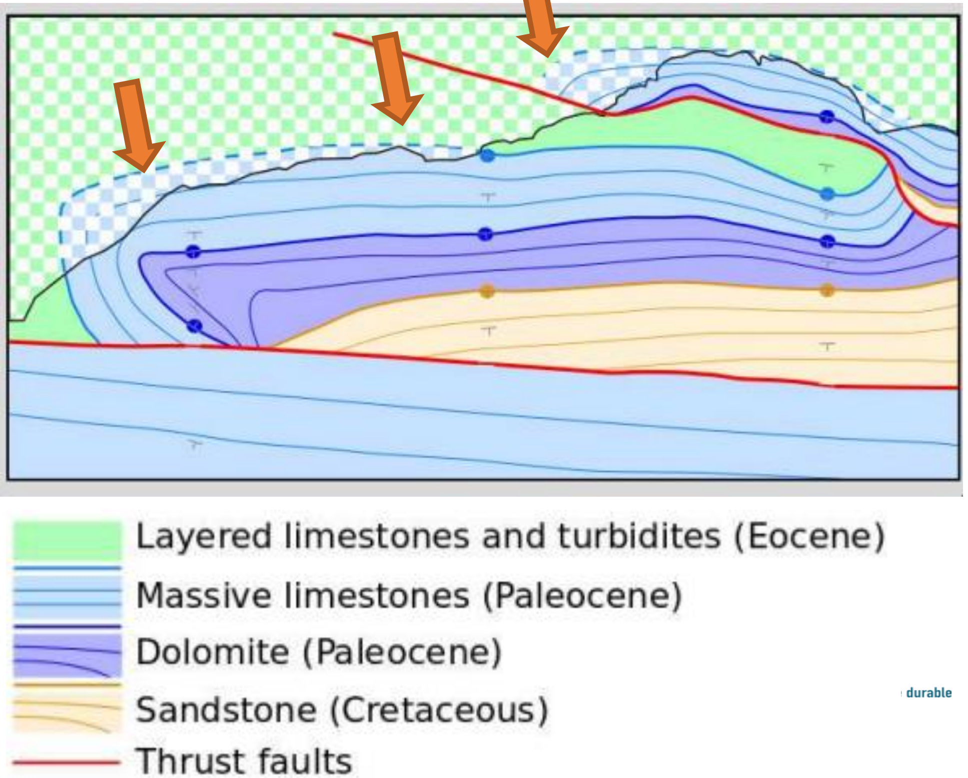


g

= Geometric Model (aka Geological Model)
+ experts' interpretation

Y

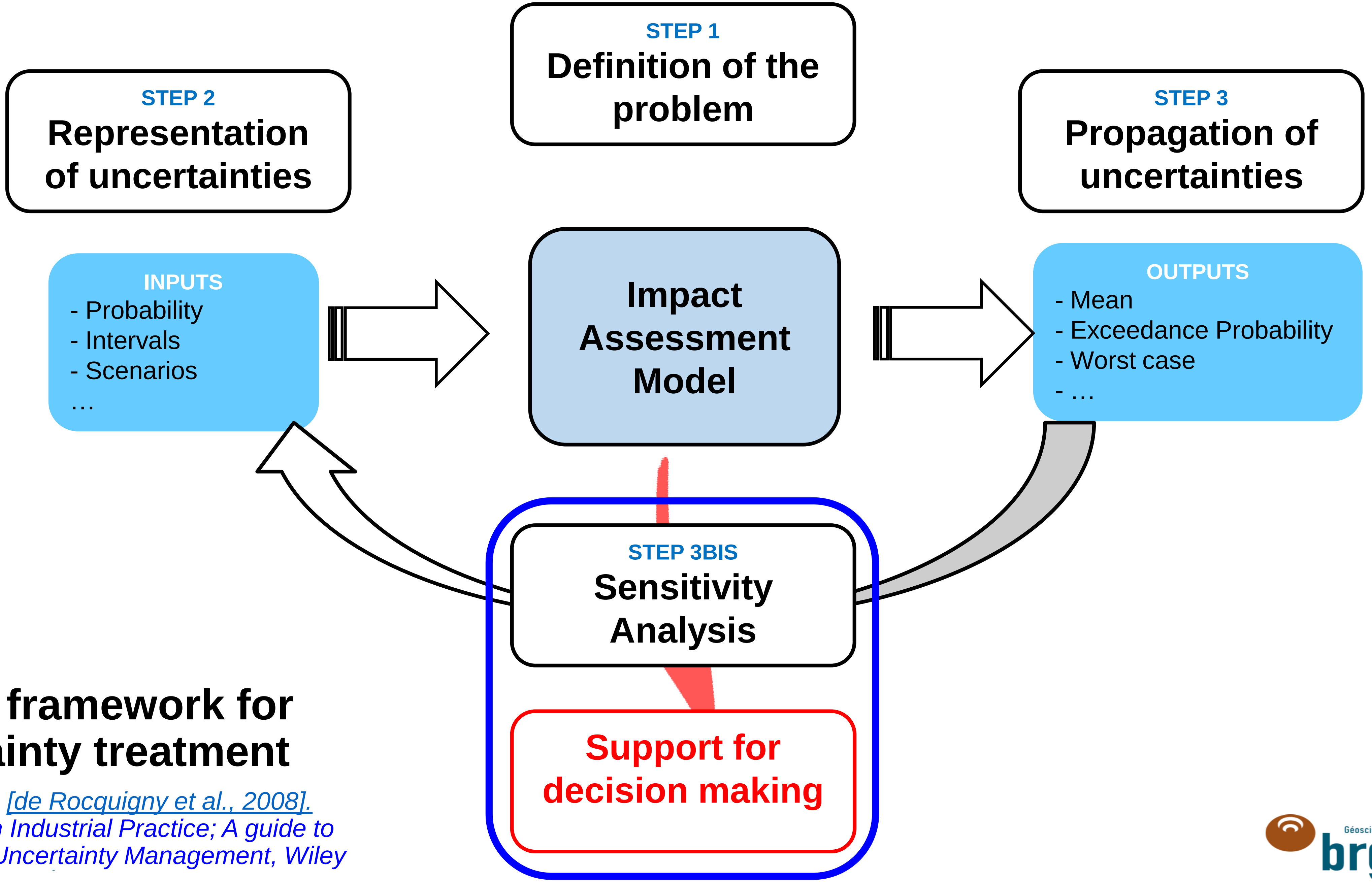
Map of the structures (fault) and rock formations [1]



[1] adapted from Wellmann & Caumon (2018)

Outline

- Context of ‘prediction’ at BRGM
- Current practices based on Uncertainty Quantification tools
- Towards explainable machine learning and open questions



Typical framework for uncertainty treatment

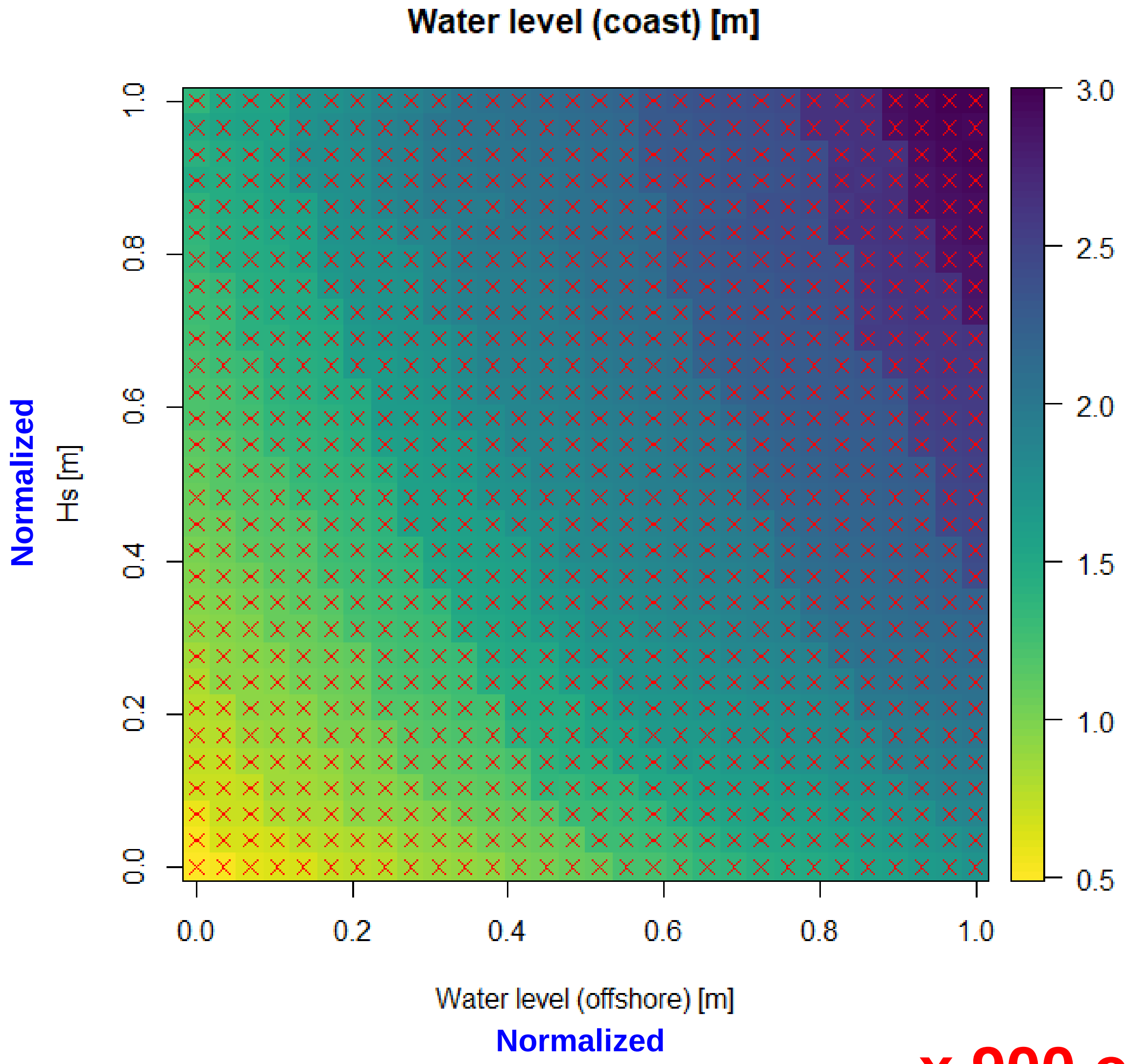
Adapted from [de Rocquigny et al., 2008]. Uncertainty in Industrial Practice; A guide to Quantitative Uncertainty Management, Wiley

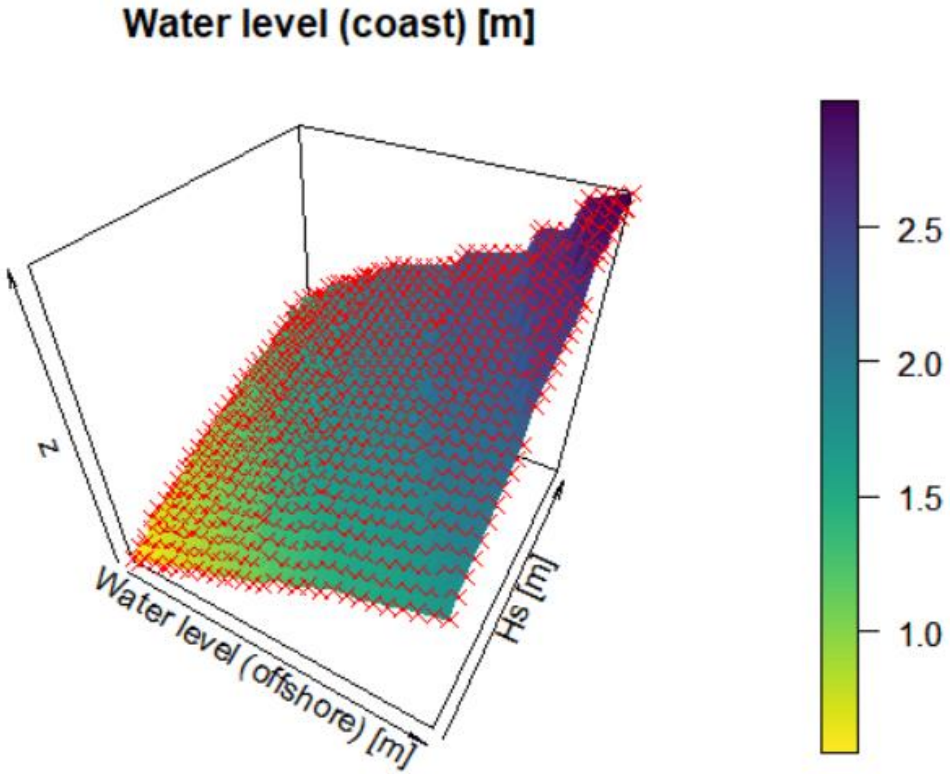
Géosciences pour une Terre durable
brgm

"For every dollar that is spent trying to quantify uncertainty, we should spend 10 dollars collecting and analyzing data that would reduce uncertainty".

Gail Atkinson (2004 World Conference on Earthquake Engineering)

Parametric analysis (‘One-at-a-Time’)



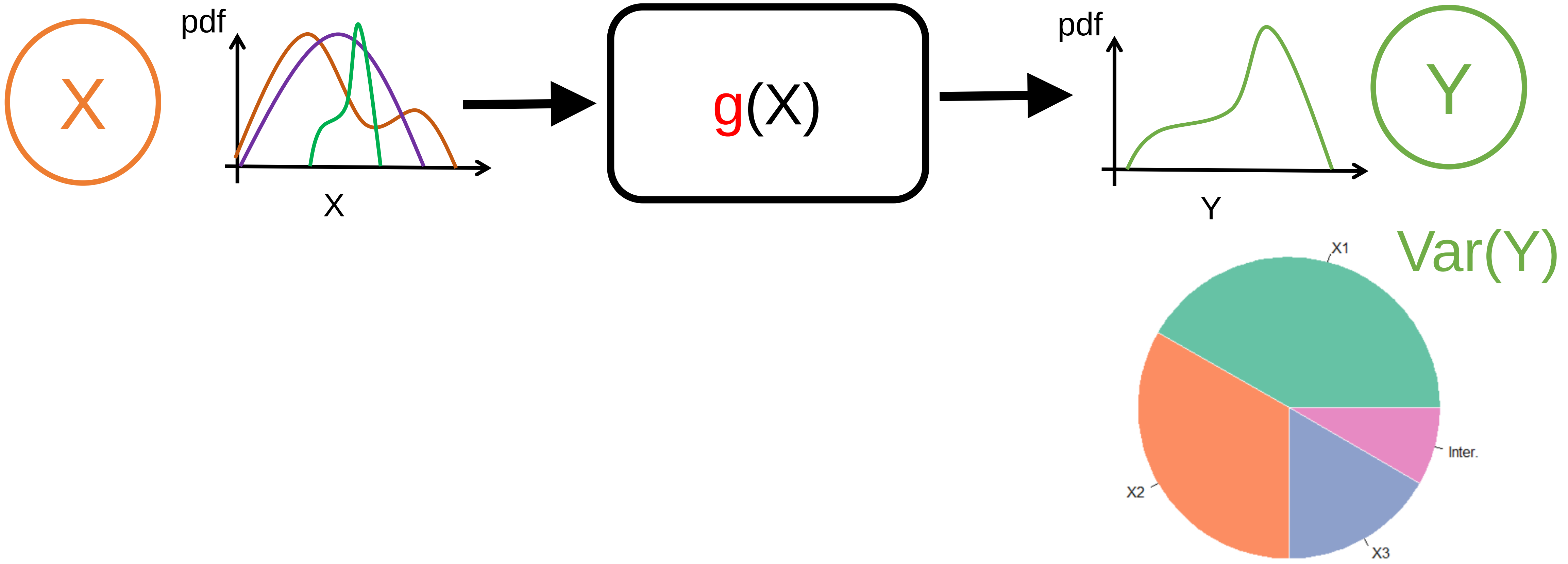


x 900 computer experiments

[1] Rohmer & Idier, NHESS (2012)

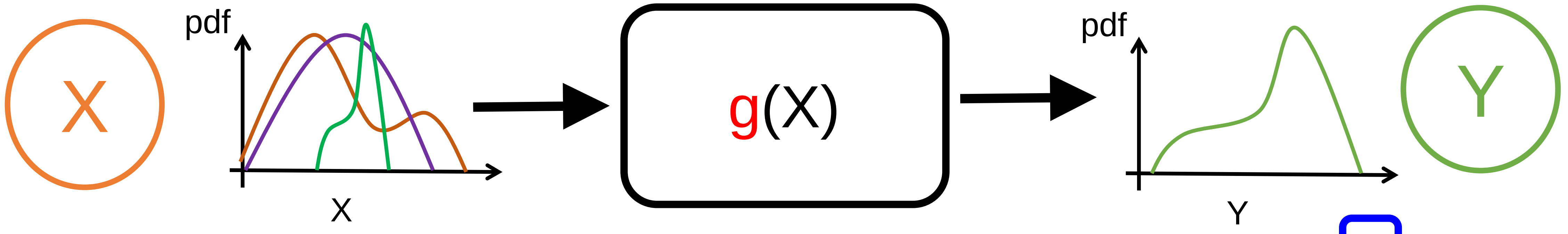
Géosciences pour une Terre durable
brgm

Variance-based global sensitivity analysis [1,2]



[1] Sobol' 1993; [2] Saltelli et al. (2008)

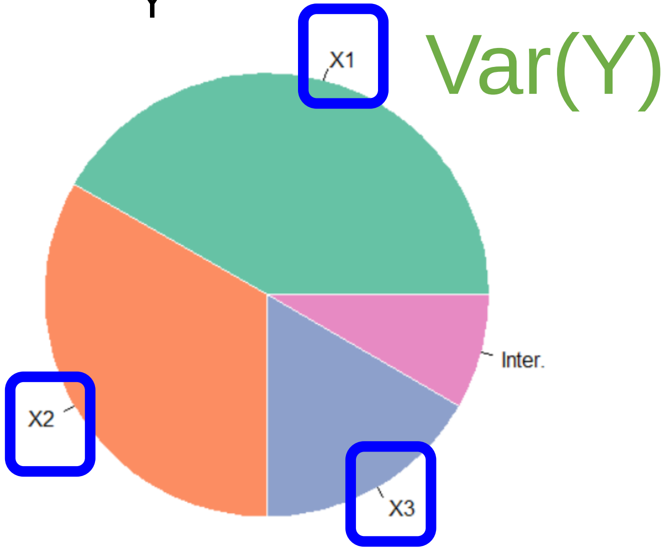
Variance-based global sensitivity analysis [1,2]



Sensitivity index of 1st order (main effect):

$$S_i = \frac{V(E(Y|X_i = x_i^*))}{V(Y)}$$

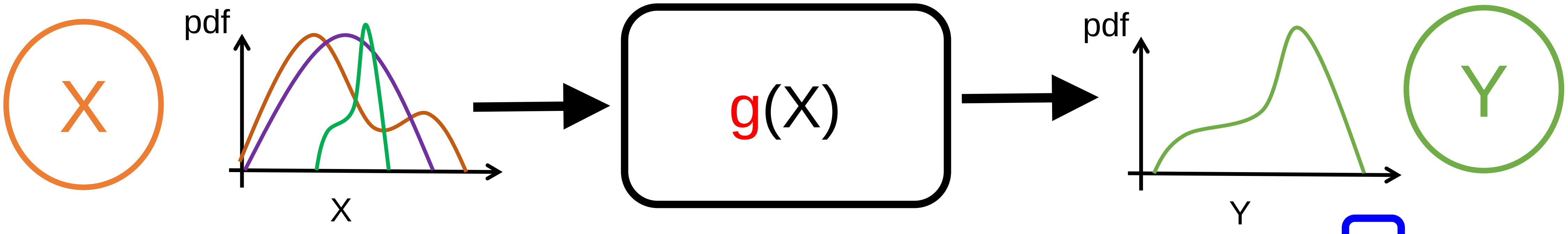
Importance ranking

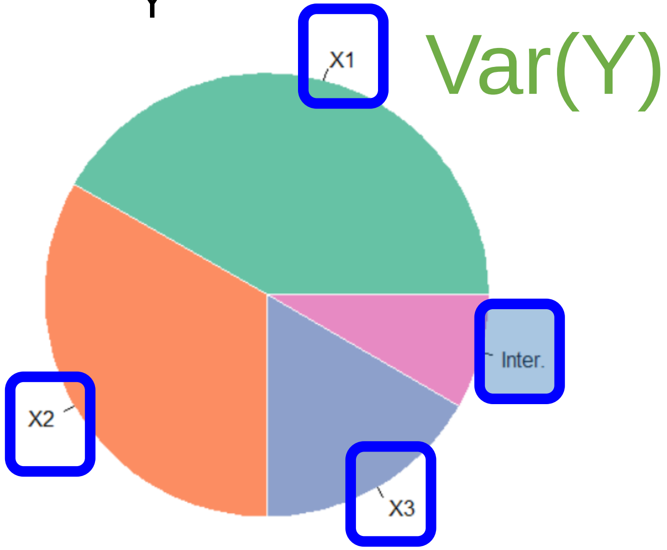


[1] Sobol' 1993; [2] Saltelli et al. (2008)

[1] Sobol' 1993; [2] Saltelli et al. (2008)

Variance-based global sensitivity analysis [1,2]





Sensitivity index of 1st order (main effect):

$$S_i = \frac{V(E(Y|X_i = x_i^*))}{V(Y)}$$

Importance ranking

Total sensitivity index

$$S_{Ti} = 1 - \frac{V(E(Y|X_{-i}))}{V(Y)}$$

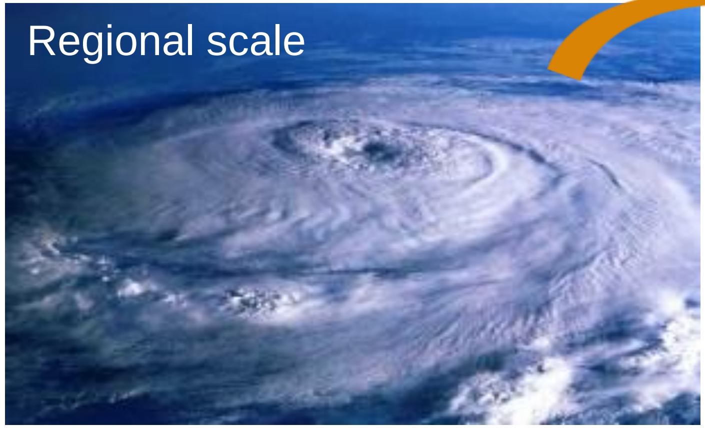
Main effects + interactions

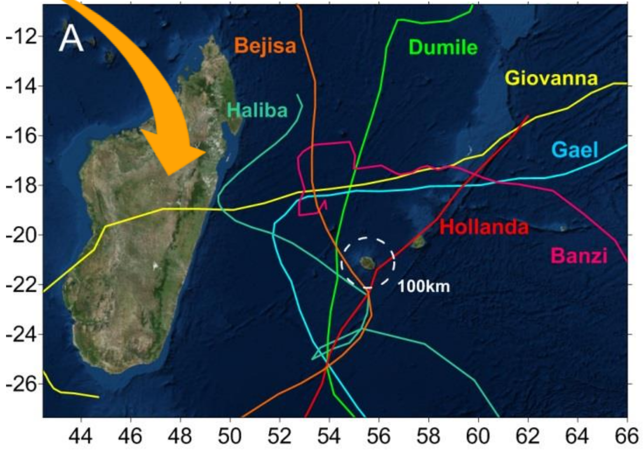
Factors' fixing

where $X_{-i}=(X_1,\ldots, X_{i-1}, X_{i+1},\ldots X_d)$

[1] Sobol' 1993; [2] Saltelli et al. (2008)

Case study in marine flooding [1]



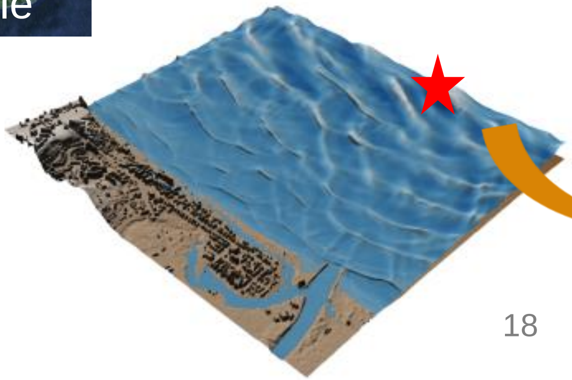


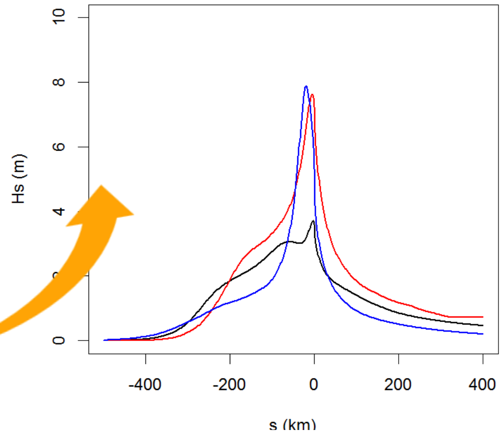
X: cyclone characteristics

Max. wind speed Vm;
Radius of max. wind $R_{m;}$
Shift around the central pressure δP;
Forward speed V_f
Track angle θ;
Landfall position x_o

g: numerical model approximated by a machine-learning model (Gaussian Process Regression)

Sainte-Suzanne city





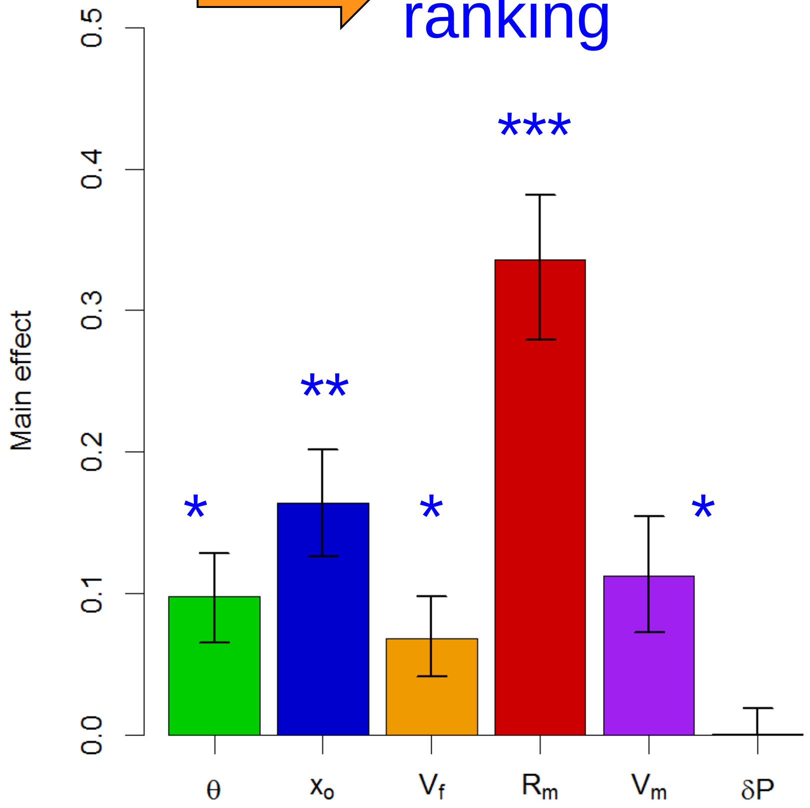
Y: wave significant height at the coast

[1] Rohmer et al. Nat. Haz. (2016)

X: cyclone characteristics

Max. wind speed Vm;
Radius of max. wind $R_{m;}$
Shift around the central pressure $\delta P;$
Forward speed V_f
Track angle $\theta;$
Landfall position x_o

Importance ranking



[1] Rohmer et al. Nat. Haz. (2016)

X: cyclone characteristics

Max. wind speed Vm;

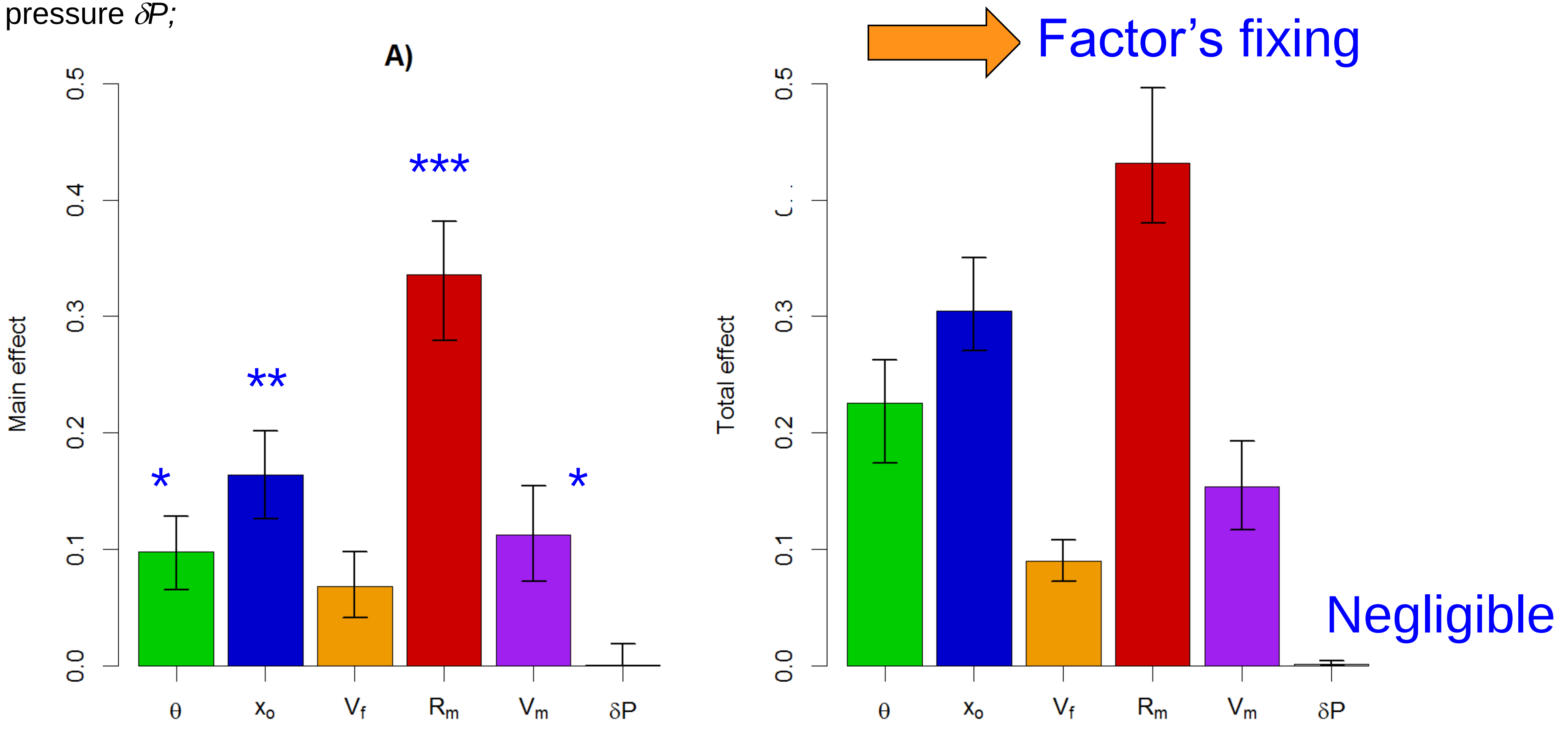
Radius of max. wind R_{m};

Shift around the central pressure δP;

Forward speed V_f

Track angle θ;

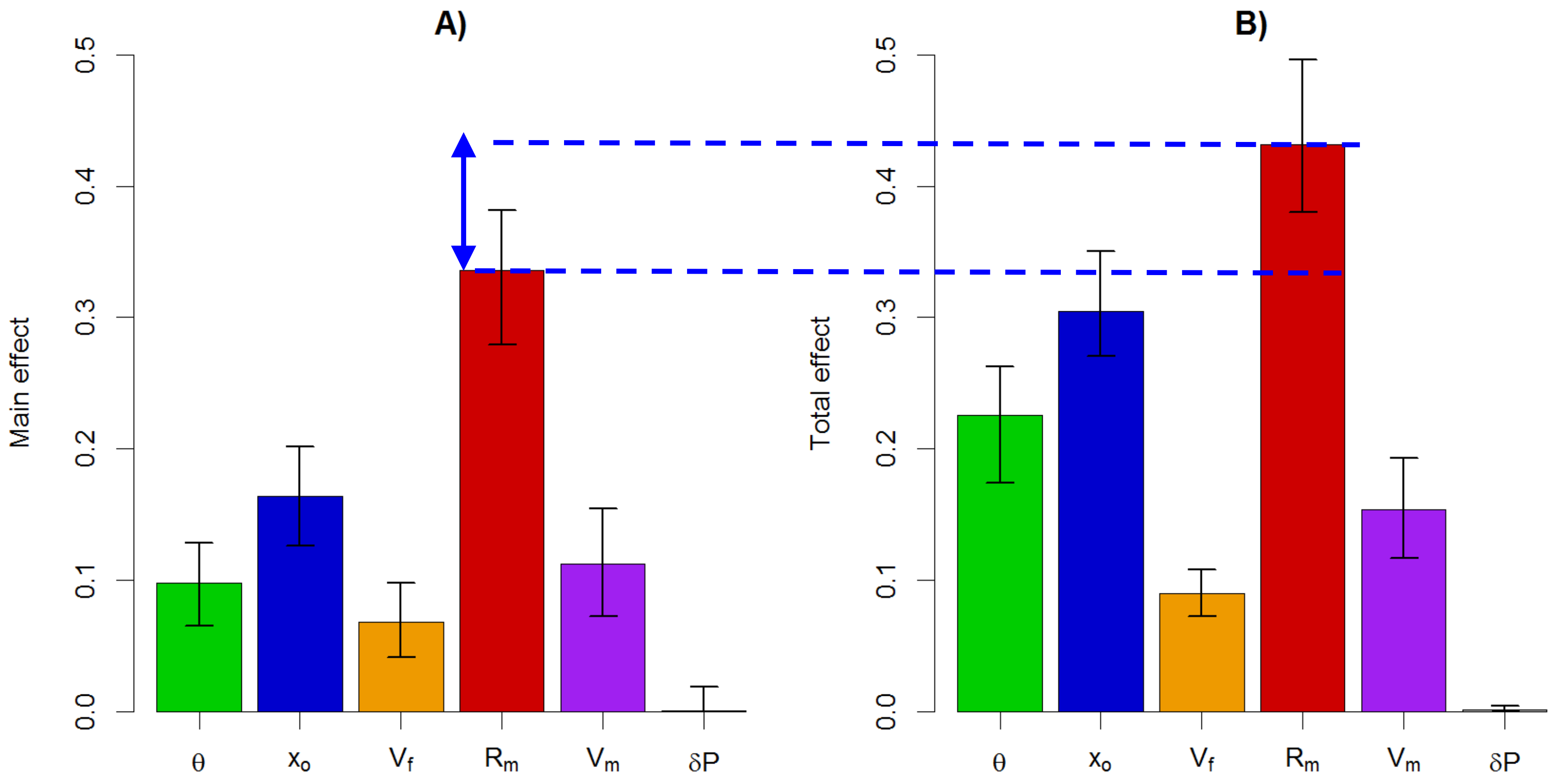
Landfall position x_o



[1] Rohmer et al. Nat. Haz. (2016)

X: cyclone characteristics

Max. wind speed Vm;
Radius of max. wind $R_{m;}$
Shift around the central pressure δP;
Forward speed V_f
Track angle θ;
Landfall position x_o



Understanding - Non-additive g function
Structure - Interaction effects

Some key challenges

- Computational burden → Use of ML-based surrogate models **[1]**
- Inputs' dependency → Shapley effects **[2]**
- Beyond variance → Moment-independent **[3]**
- Complex inputs/outputs → adapted algorithms **[4]**

Links to

Groupement de Recherche MASCOT-NUM
cnrs
Méthodes d'Analyse Stochastique pour les COdes et Traitements NUMériques

CIROQUO
Consortium Industrie Recherche pour l'Optimisation et la QUantification d'incertitude pour les données Onéreuses

[1] Rasmussen & Williamson (2006); **[2]** Iooss & Prieur (2019); **[3]** da Veiga (2015); **[4]** Gamboa et al. (2017)

Outline

- Context of prediction at BRGM
- Current practices based on Uncertainty Quantification tools
- Towards explainable machine learning and open questions

Motivation for ‘increased’ explainability of the geomodels

- **High stakes decisions**
  - **Early warning** systems and **Crisis** management
  - Planning for the future in the context of **climate change**
  - **Design and optimize** of subsurface systems (heat, CO2 storage, geothermal activities)
  - Identify **anomalies** (pollutant, reservoir fluid, etc.),
  - Etc.

Motivation for ‘increased’ explainability of the geomodels

- **High stakes decisions**
  - **Early warning** systems and **Crisis** management
  - Planning for the future in the context of **climate change**
  - **Design and optimize** of subsurface systems (heat, CO_2 storage, geothermal activities)
  - Identify **anomalies** (pollutant, reservoir fluid, etc.),
  - Etc.

- **Stress testing ‘scientific knowledge’**
  - **Understanding the ‘why’** of the predictions may force to think **‘out of the box’**
  - A path towards new **scientific discovery** (?)

Motivation for ‘increased’ explainability of the geomodels

- **High stakes decisions**
  - **Early warning** systems and **Crisis** management
  - Planning for the future in the context of **climate change**
  - **Design and optimize** of subsurface systems (heat, CO2 storage, geothermal activities)
  - Identify **anomalies** (pollutant, reservoir fluid, etc.),
  - Etc.

- **Stress testing ‘scientific knowledge’**
  - **Understanding the ‘why’** of the predictions may force to think ‘out of the box’
  - A path towards new **scientific discovery** (?)

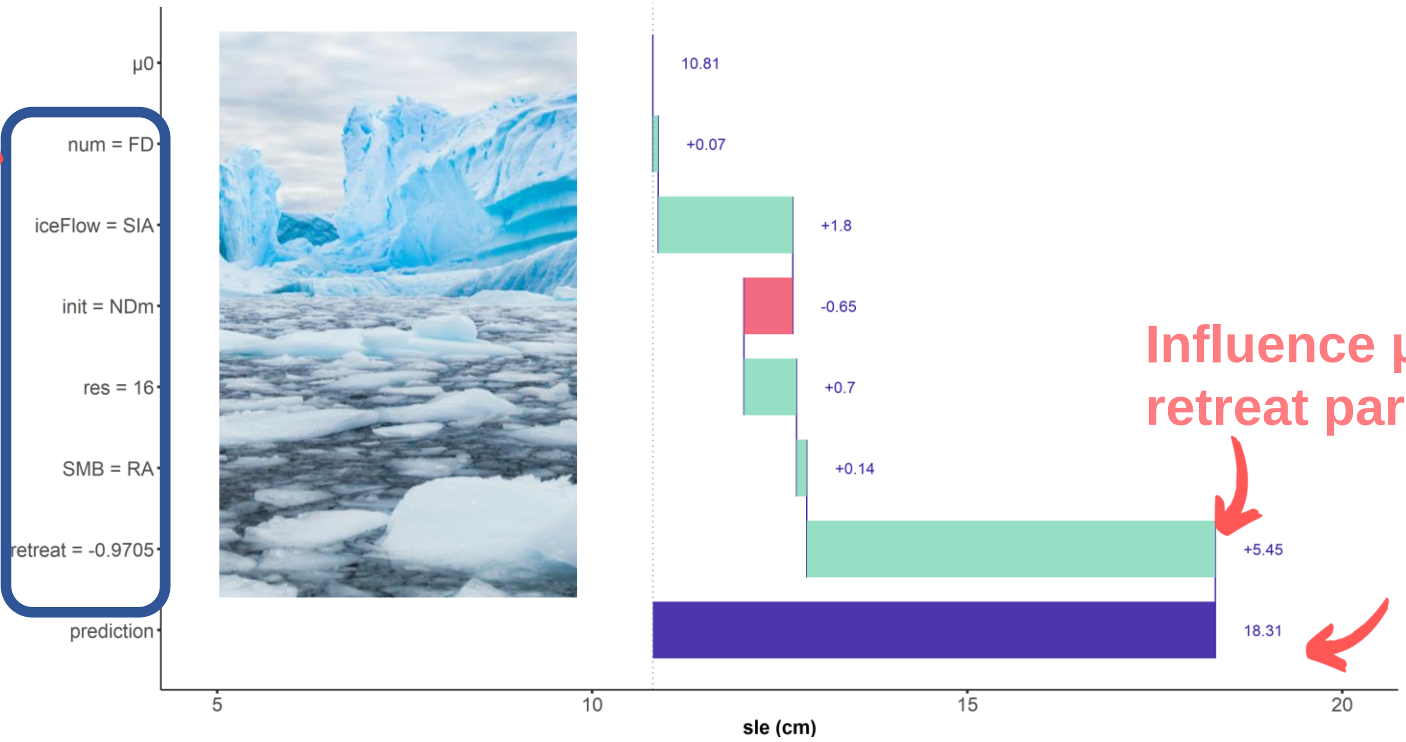
- **Convince modelers to improve widely-used practices**
  - **‘Keep control’:** a model is sometimes preferred if it can be more easily interpreted all along the different stages of the modelling/processing chain

Testing the benefits of SHAP [1]

Application to sea level change due to climate change [2]

$$\text{sea level}^{(m)} = \mu_0 + \mu_{\text{Retreat para}^{(m)}} + \mu_{\text{SMB}^{(m)}} + \mu_{\text{Numerics}^{(m)}} + \mu_{\text{Initialisation}^{(m)}} + \mu_{\text{iceflow}^{(m)}} + \mu_{\text{Resolution}^{(m)}}$$

Assumptions of the numerical model



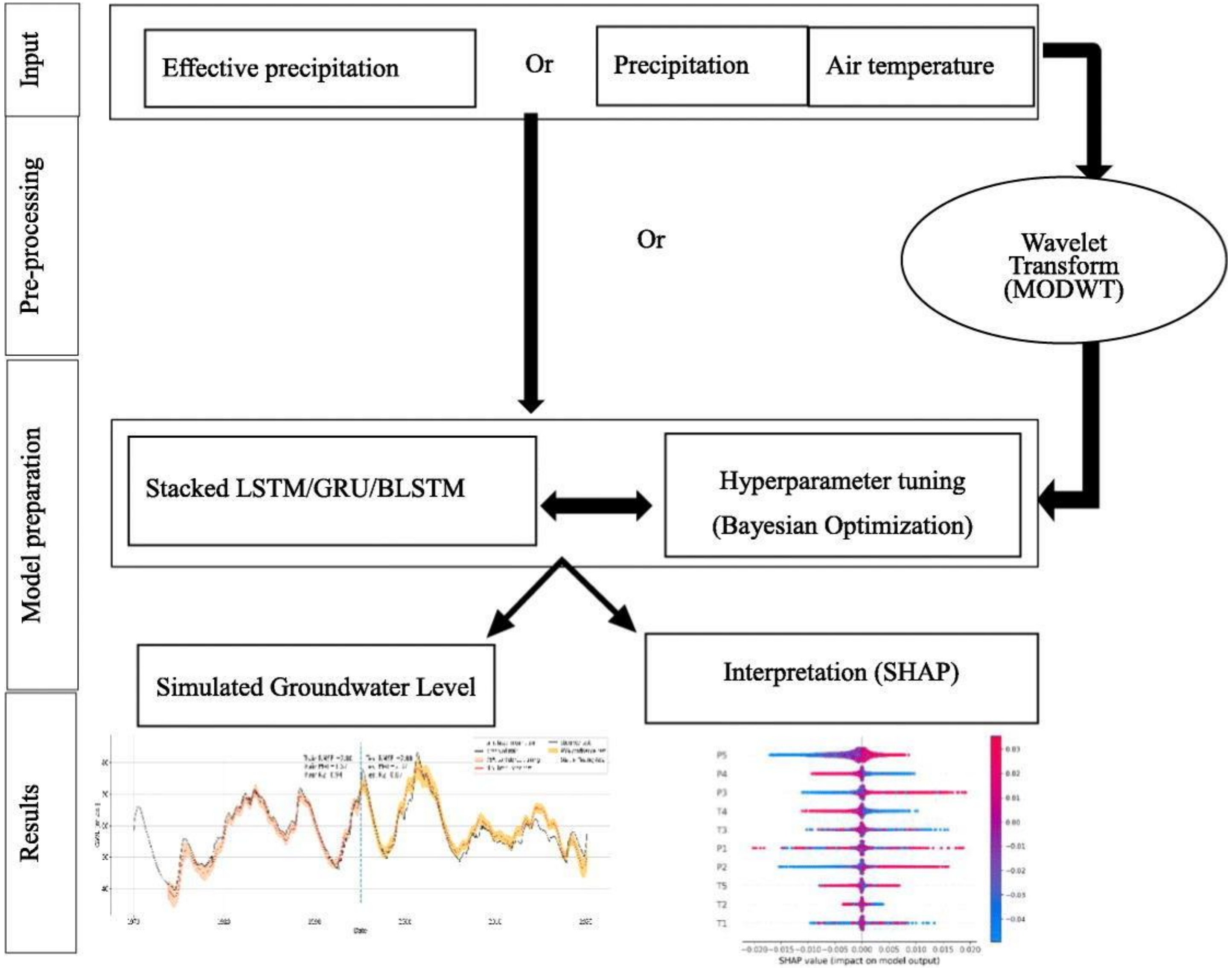
Sea level change (cm SLE)

Influence μ of retreat parameter

Sea level at 2100 for the given configuration

[1] Lundberg & Lee NeurIPS (2017) [2] Rohmer et al. Cryosphere (2022)

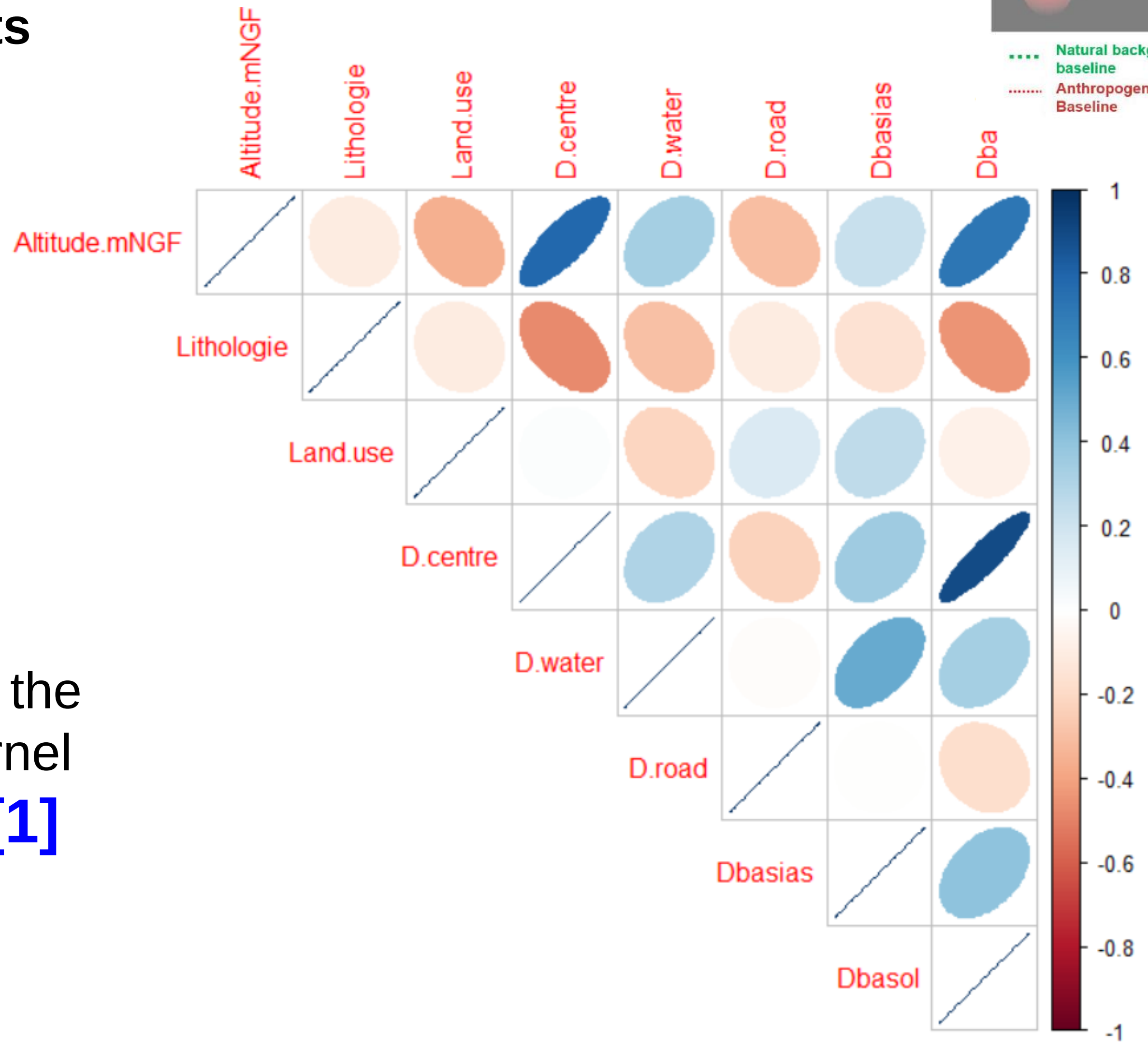
Other initiatives are emerging [1]

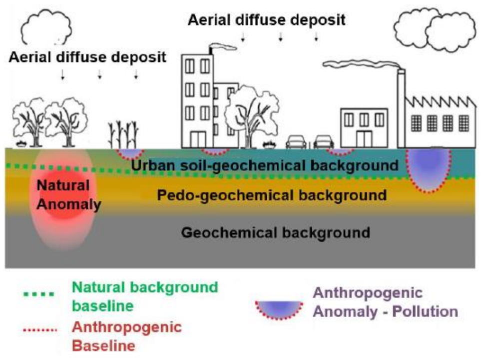


[1] Chidepudi et al. Sc. Tot. Env. (2023)

Open question 1: dependence

Matrix of linear (Pearson's) correlation coefficients

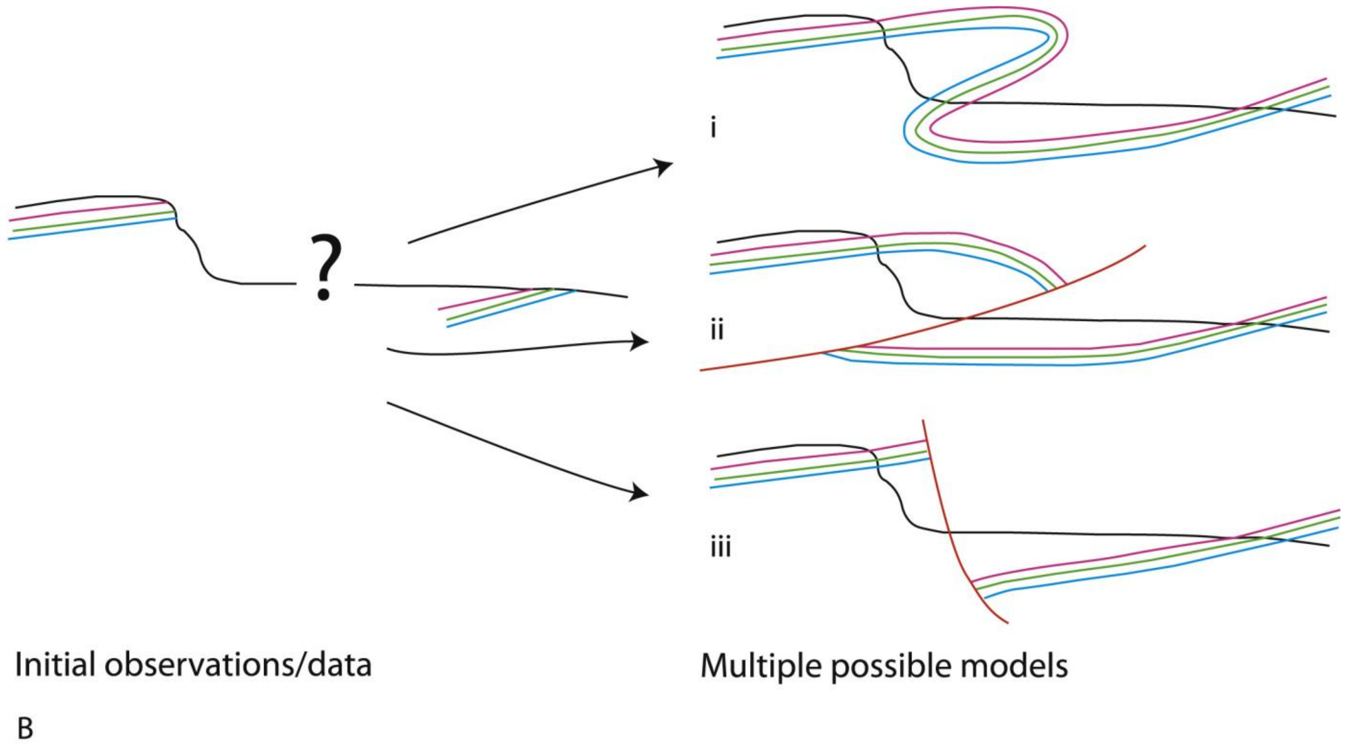




Need to correct the widely-used kernel SHAP method [1]

[1] Aas et al., AI (2021)

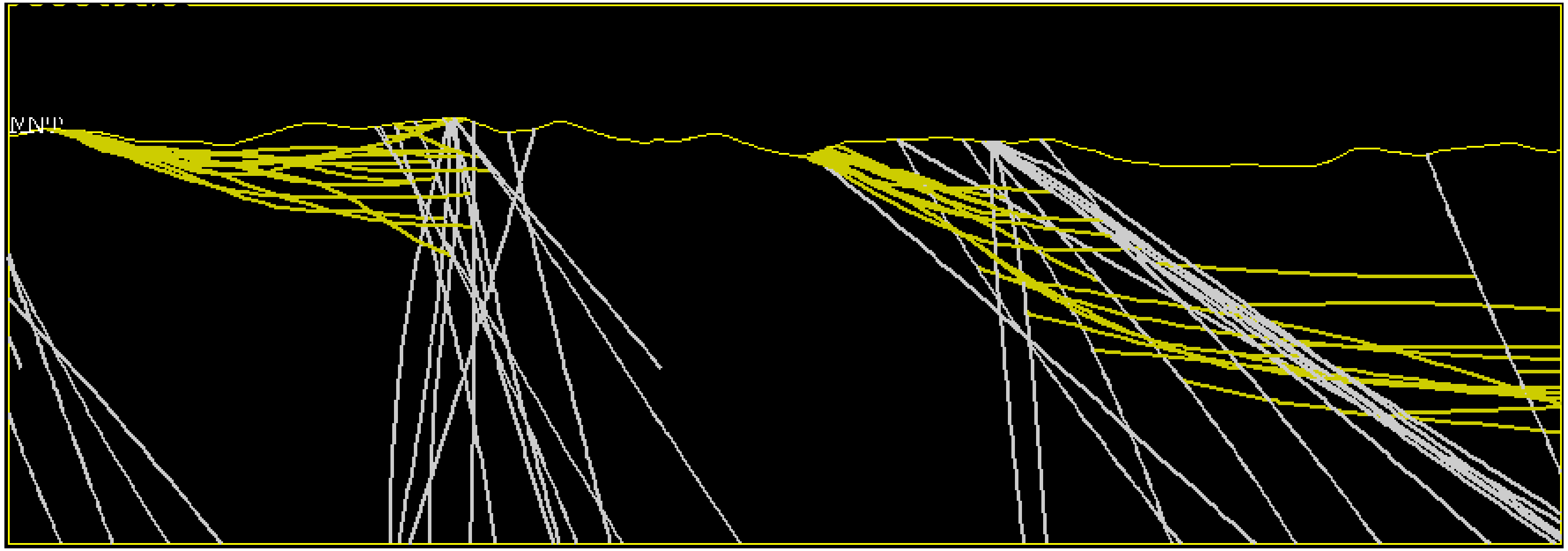
Open question 2: expert interpretation [1]



[1] Bond J. of Struct. Geol. (2015)

Open question 2: expert interpretation [1]

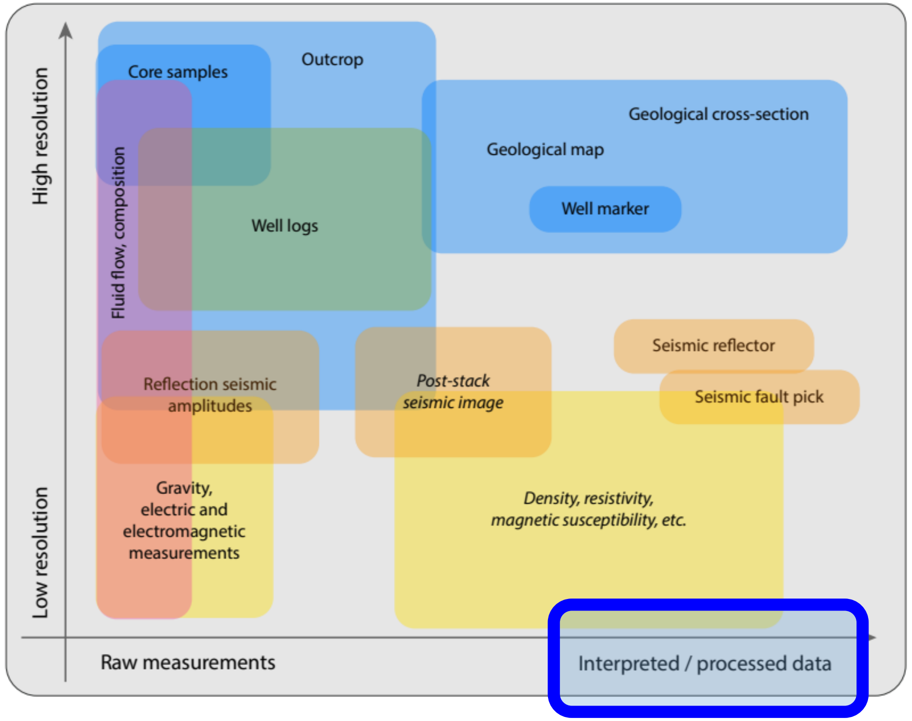
Geo-Models from different training sessions



Some Xs already hold a part of interpretation.
Depending on the expert, it can vary…

[1] Courrioux et al. (2015)

Open question 3: integrating multiple types of data

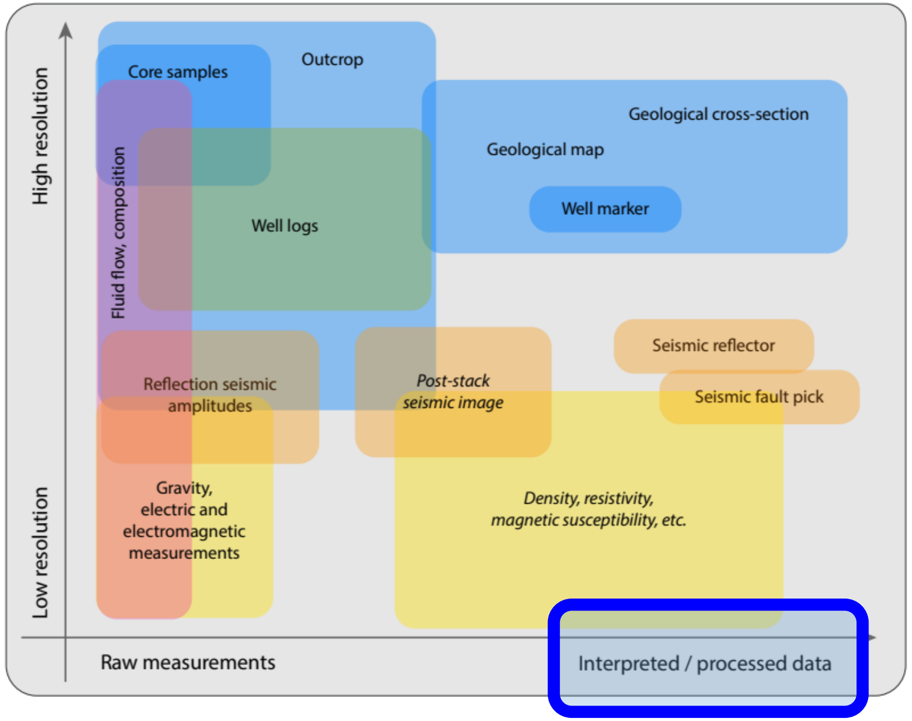


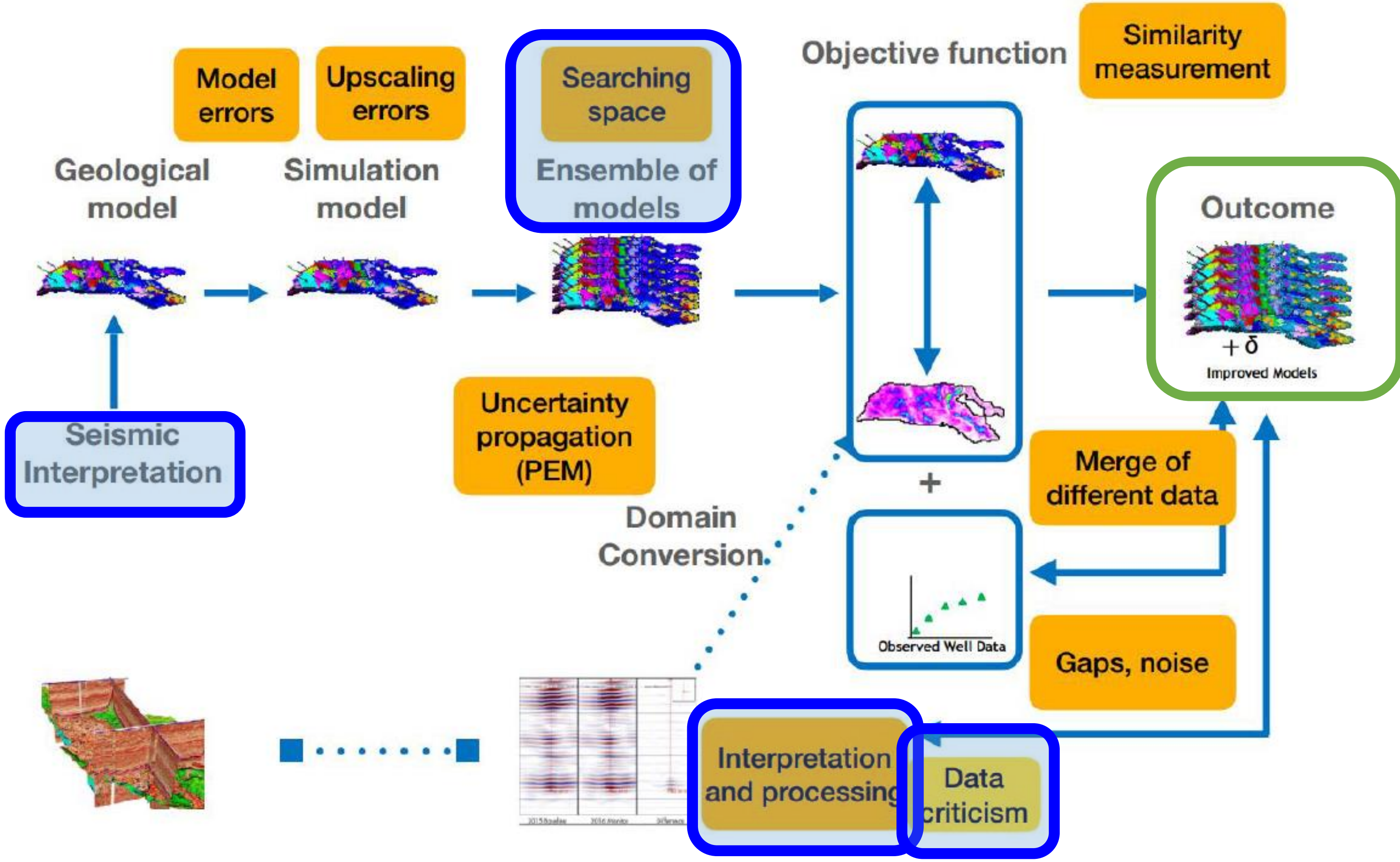
Typical Earth data used in geomodeling [1]

[1] Wellmann & Caumon, 2018

Géosciences pour une Terre durable
brgm

Open question 3: integrating multiple types of data





Typical Earth data used in geomodeling [1]

Typical workflow for data assimilation in exploitation phase [2]

[1] Wellmann & Caumon, 2018 [2] Chassagne (2023)

Summary

Diversity of 'prediction' contexts

- Data, prediction models, type of decision

UQ(SA) tools have provided some key insights,
BUT a deeper analysis is needed for:

- **High stake** decisions
- **Helping the modellers** in their current practices
- **Criticize existing frameworks** / settings / theories

Summary

Diversity of 'prediction' contexts

- Data, prediction models, type of decision

UQ(SA) tools have provided some key insights,
BUT a deeper analysis is needed for:

- High stake decisions
- Helping the modellers in their current practices
- Criticize existing frameworks / settings / theories

Key questions:

- **Complexity** of the predictor variables (in particular dependence, high dim.)
- Interplay with **expert interpretation**
    - Processing of predictor variables
    - Necessary for model construction in a context of data / information sparsity

Thank you for your attention!

We acknowledge financial funding by ANR-HOUSES
(grant number: **ANR-22-CE56-0006**)
https://anrhouses.github.io/