CLUSTERING METHODS FOR DECISION-MAKING

APPLICATION TO NATURAL & TECHNOLOGICAL (NATECH) RISKS

Y. RICHET

2025-01

Crédits F. Dubray / Ouest France 2021

Crédits IRSN

DECISIONS FACING NATECH RISKS

RISKS MANAGEMENT

NATURAL AND/OR TECHNOLOGICAL (NATECH)

- Crisis situations
  - High **uncertainties** are intrinsic to accidental situations
  - **Anticipation** : numerical simulations are a mandatory to predict potential consequences and protect population
  - Strong **time constraints**
- *Operational* decisions
  - Need for **clear and concise** information
  - Communication of **evaluation products** to **decision makers** *should* account for uncertainties

L'Aquila earthquake scientists win appeal

Six seismologists and official had been convicted over reassurance issued to residents before fatal quake in 2009

Orages en Corse : Météo-France invoque une situation « difficilement prévisible »

L'activation tardive de la vigilance orange, alors que les rafales ont atteint 200 km/h par endroits, illustre la difficulté à traduire les probabilités de phénomènes météo en un système d'alerte crédible.

Le Monde avec AFP

Publié le 18 août 2022 à 21h34, modifié le 24 juillet 2023 à 16h29 · Lecture 3 min.

Crisis : **quick** decisions under **uncertainties**

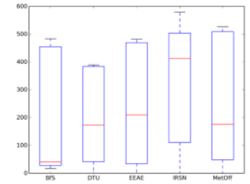
UNCERTAINTY ANALYSIS & INTERPRETABILITY

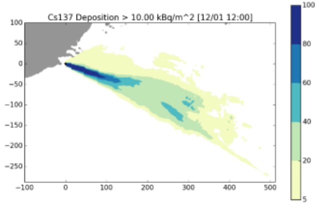
Uncertainty models for quantitative analysis or decision-making

- Quantification of input / model uncertainties
- Sensitivity analyses/indices (Sobol, Shapley, HSIC, ...)
- "Envelope" of trust, probability of threshold exceedance…
- Confidence level in evaluations
- **Identification of representatives/prototypes**

✓ Should be decision-oriented, incl. practical information (e.g. population, agriculture…)

✓ If possible, avoid interpretation bias

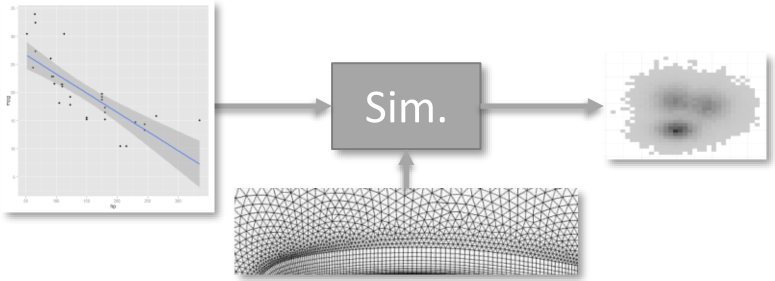




Simplicity / Interpretability

UNCERTAINTIES QUANTIZATION

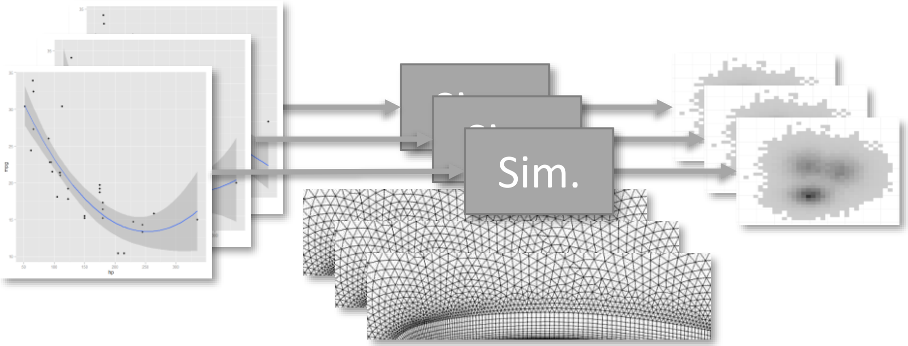
Simulations:



… a suitable support for propagation of uncertainties

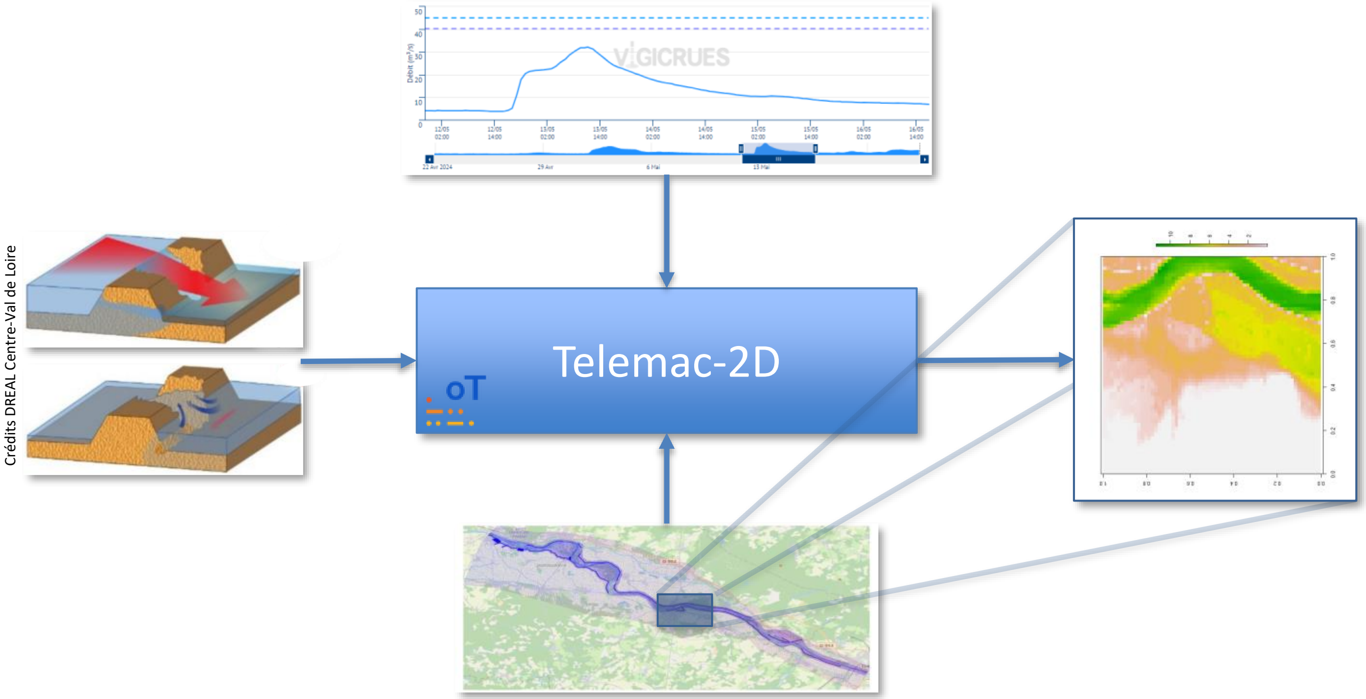
Ex. :

- Monte Carlo / random sampling
- Sets
- Quantiles / delta
- [Fuzzy logic]
- [Experimental Calibration]



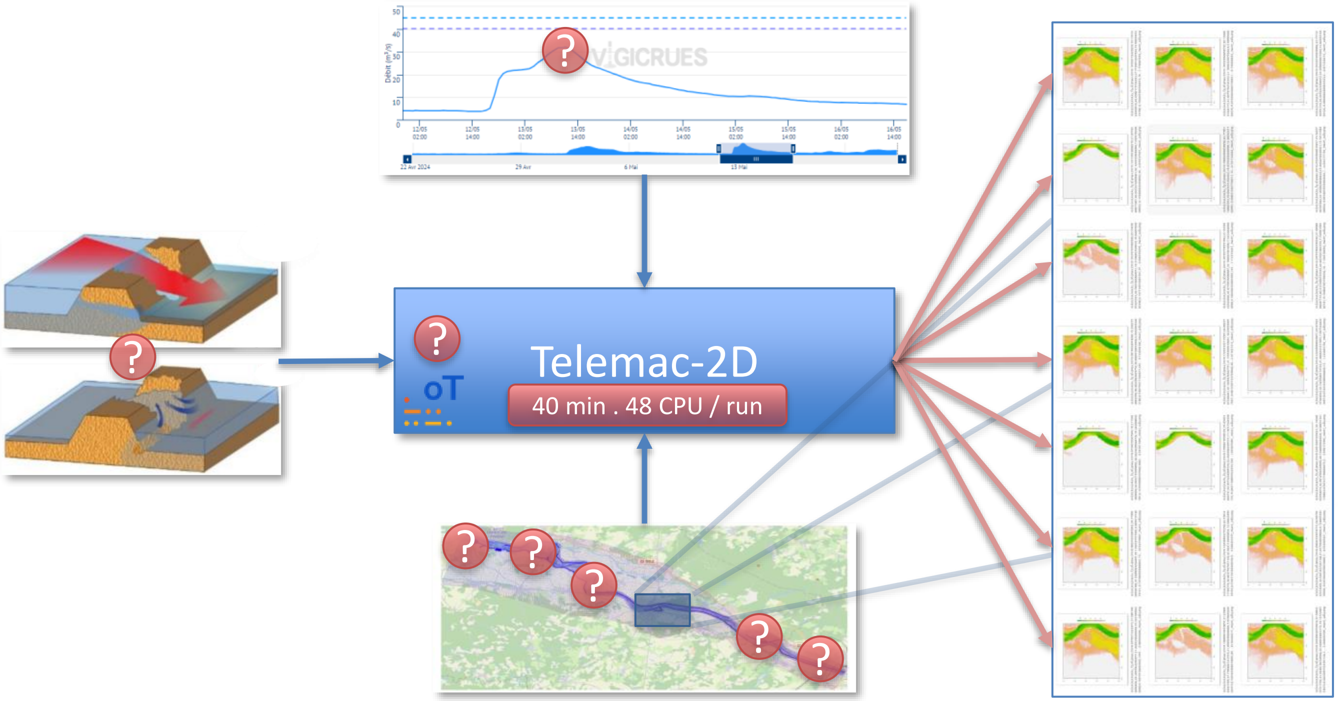
MODELLING HIGH-DIMENSIONAL DATA

[Ex.] Flooding

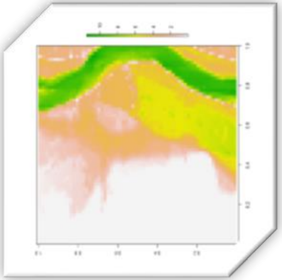


[Ex.] Flooding

Crédits DREAL Centre-Val de Loire



How to include uncertainties in decision making ?



Challenge 1: high-dimension input / output

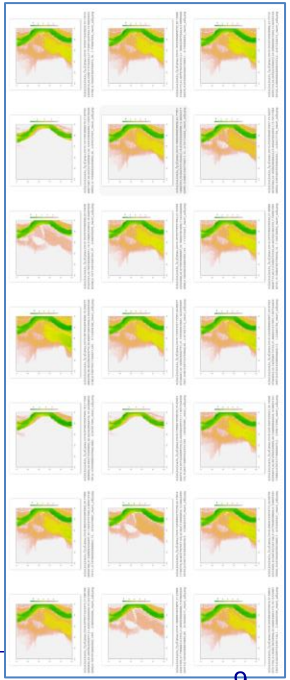
- Spatio-temporal physical fields
- Interactions / correlations between variables
- Use of appropriate dimension reduction methods

Challenge 2: Computational cost of physical models

- Use of metamodel as surrogate model

Challenge 3: interpretability of output

- Postage stamp ? Too many maps , Probability maps ? Complex interpretation
- Scenario-based approach: “best estimate” vs. “worst case
- Use of clustering methods



Challenge 4: representativity for extreme (& hopefully rare) events

- Use controlled/weighted sampling

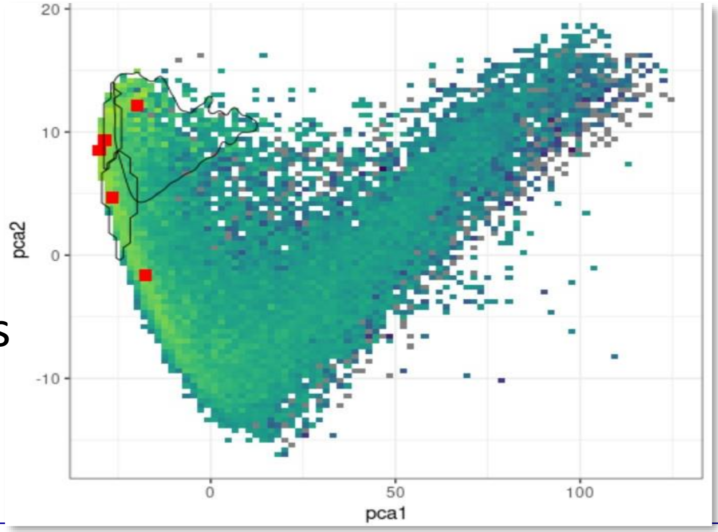
1. High-dim. Input/Output

Complex *(ie. non scalar)* numerical results:

- Physical quantities: spatial (lat,lon), temporal (t)
- Non-linear output / operational consequences

… supported by a **dimension reduction**:

- Unsupervised / ~weighted / supervised
- Non-significant "latent" space
- …but keep some desirable math. properties



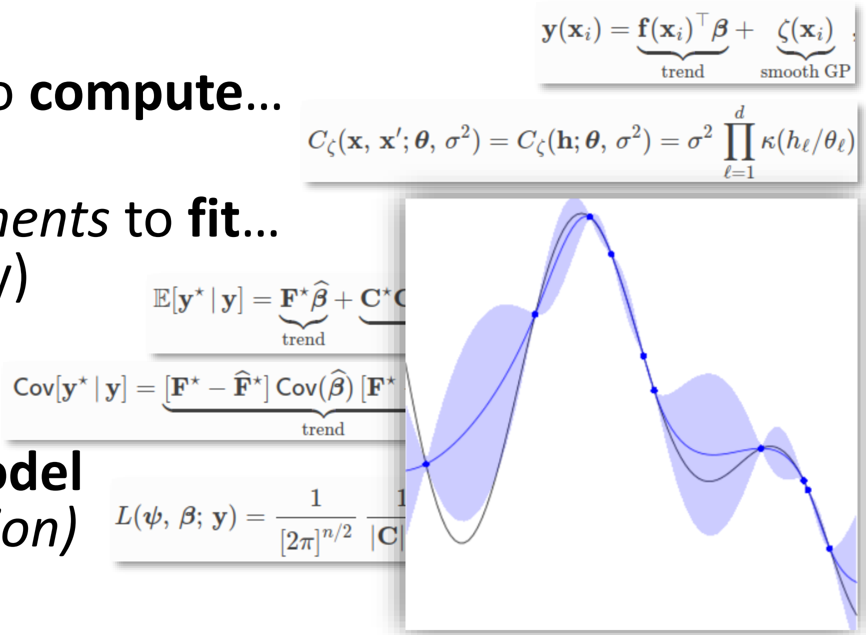
Credits: Charlie Sire PhD

2. Computational cost

Parametric modelling to assess physical behaviour:

- *High Performance Computing* to **compute**…
- …a numerical *Design of Experiments* to **fit**… (but still curse of dimensionality)
- …a *response surface*

= Training of a surrogate / **metamodel**
(*ex. Gaussian Process Regression*)



Credits: https://libkriging.readthedocs.io

2.' METAMODELLING

Parametric modelling to assess physical behaviour:
metamodel safety requirements

- **Compliance** with the state of the art in metamodel:
  learning set features, metamodel properties,
  targeted fitting error, training-testing-validation,
- **Robustness** of performance between the *learning* domain and the domain of *use* of the metamodel,
- **Explainability** that provide *understanding* of the phenomena and *completeness* of the information provided between input data and output results,
- **Transparency** to ensure that all information regarding the metamodel is accessible.

Ref: SYSPASSCOI 2024 working group

3. INTERPRETABILITY BY SPARSE SAMPLING

Objective: sparse & synthetic sampling

- *Some* **prototypes** / centroids (on projected output)
- Weighted / probabilistic classes (by input measure)
- "Real" <-> "Latent" space projection

Algorithm 2 Prototype Maps Algorithm

Require: $(y(x^i))_{i=1,\dots,n_{\text{train}}}$, f_X, g, minDistance, ℓ

Sample $(\tilde{x}^k)_{k\in\{1,\dots,n_{\text{maps}}\}}$ i.i.d. of density function g

Compute $(\hat{y}(\tilde{x}^k))_{1\le k\le n_{\text{maps}}}$ from $(y(x^i))_{i=1,\dots,n_{\text{train}}}$ (GP & FPCA)

Compute $\left(\frac{f_X(\tilde{x}^k)}{g(\tilde{x}_k)}\right)_{1\le k\le n_{\text{maps}}}$

Initialize $\Gamma_\ell^{[0]} \leftarrow \{\gamma_0^{[0]},\dots,\gamma_\ell^{[0]}\} \in \mathcal{Y}^\ell$

while $\|\Gamma_\ell^{[k+1]} - \Gamma_\ell^{[k]}\| >$ minDistance **do**

$\gamma_j^{[k+1]} \leftarrow \hat{E}_{n_{\text{maps}}}(\Gamma_\ell^{[k]}, j, \hat{y}), j = 1,\dots,\ell$

$k \leftarrow k+1$

end while

$\hat{\Gamma}_\ell^\star = \Gamma_\ell^{[k]}$

Compute $\hat{P}_{\tilde{n}}(\hat{\Gamma}_\ell^\star, j, \hat{y}), j = 1,\dots,\ell$

Output: $\hat{\Gamma}_\ell^\star$ and $\hat{P}_{\tilde{n}}(\hat{\Gamma}_\ell^\star, j, \hat{y})$

Algorithm 1 Lloyd's algorithm

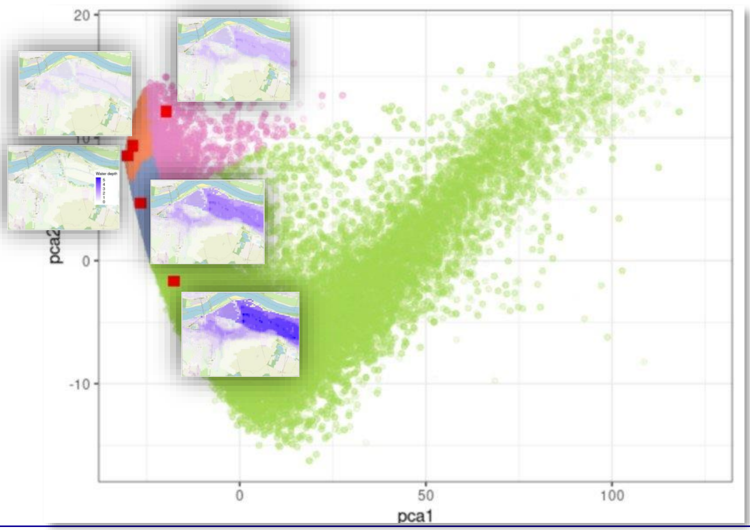
$\Gamma_\ell^{[0]} \leftarrow \{\gamma_1^{[0]},\dots,\gamma_\ell^{[0]}\}$, $k \leftarrow 0$

while stopping criterion not met **do**

$\gamma_j^{[k+1]} \leftarrow \mathbb{E}\left[Y(X) \mid Y(X) \in C_j^{\Gamma_\ell^{[k]}}\right], j \in \{1,\dots,\ell\}.$

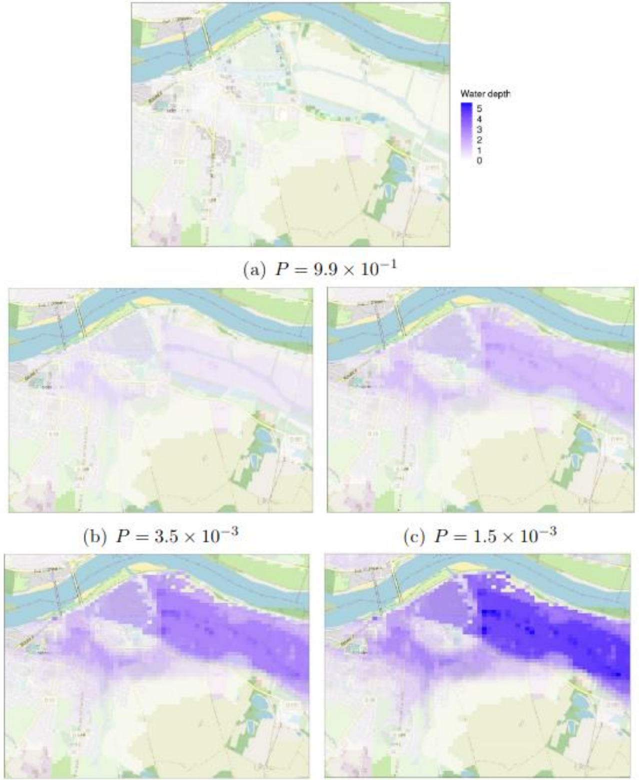
$k \leftarrow k+1$

end while



Credits: Charlie Sire PhD

3. Interpretability by sparse sampling



(a) $P = 9.9 \times 10^{-1}$

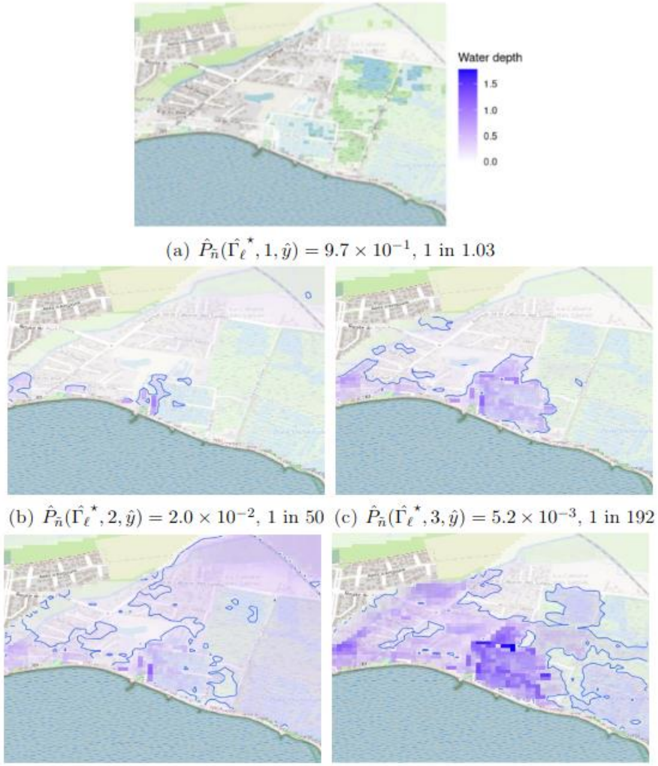
(b) $P = 3.5 \times 10^{-3}$

(c) $P = 1.5 \times 10^{-3}$

(d) $P = 8.9 \times 10^{-4}$

(e) $P = 5.8 \times 10^{-4}$

Credits: Charlie Sire PhD



(a) $\hat{P}_{\tilde{n}}(\hat{\Gamma}_\ell^\star, 1, \hat{y}) = 9.7 \times 10^{-1}$, 1 in 1.03

(b) $\hat{P}_{\tilde{n}}(\hat{\Gamma}_\ell^\star, 2, \hat{y}) = 2.0 \times 10^{-2}$, 1 in 50

(c) $\hat{P}_{\tilde{n}}(\hat{\Gamma}_\ell^\star, 3, \hat{y}) = 5.2 \times 10^{-3}$, 1 in 192

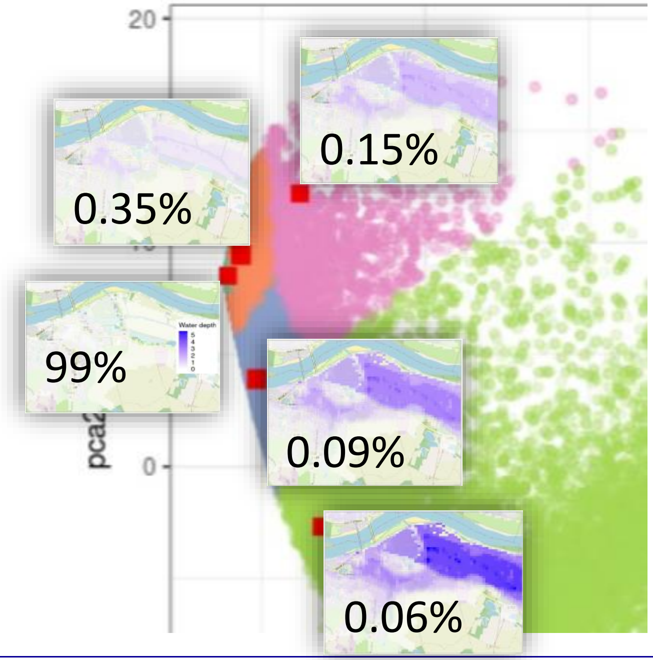
(d) $\hat{P}_{\tilde{n}}(\hat{\Gamma}_\ell^\star, 4, \hat{y}) = 6.2 \times 10^{-3}$, 1 in 161

(e) $\hat{P}_{\tilde{n}}(\hat{\Gamma}_\ell^\star, 5, \hat{y}) = 1.4 \times 10^{-3}$, 1 in 714

4. Rare/extreme events support

Objective: keep extreme events available to support decision

- Robust classes centroids
- Robust probability of classes
- Backward "Latent" -> "Real" space projection on input

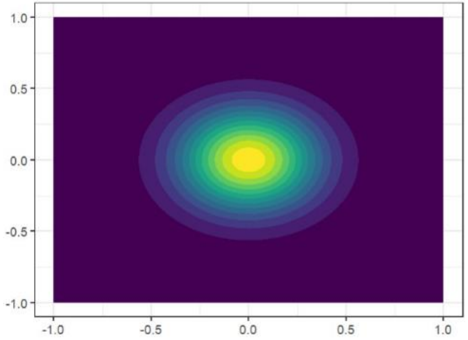


Credits: Charlie Sire PhD

4. Rare/extreme events support

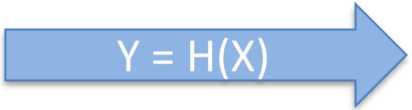
Objective : robust classes (centroid & probability/weight)

Example: analytical 2-dim. model



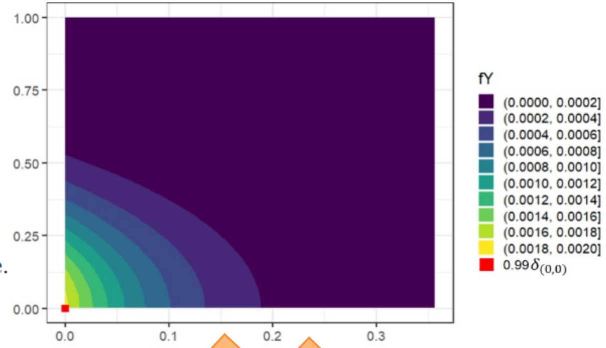
$$\begin{cases} X_i \sim \mathcal{N}_t(0, 0.25^2, -1, 1), i = 1, 2 \\ X_1 \text{ and } X_2 \text{ independent} \end{cases}$$

Credits: Charlie Sire PhD

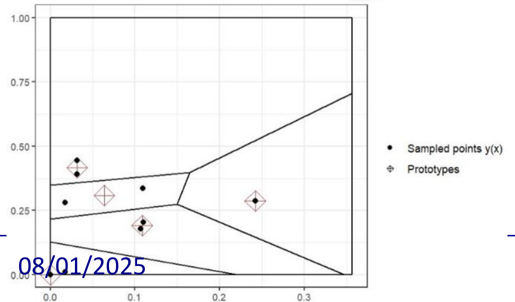


$$H(x) = \begin{cases} (0,0) \text{ if } |x_1| \leq \alpha \\ (|x_1| - \alpha, |x_2|) \text{ otherwise.} \end{cases}$$

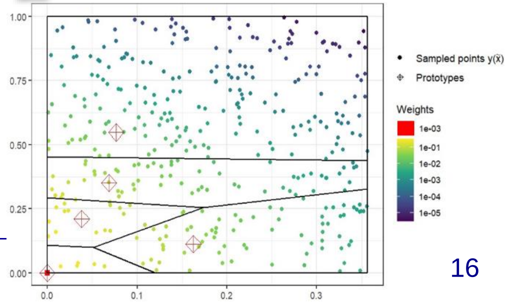
$$P(H(X) = (0,0)) = 0.99.$$



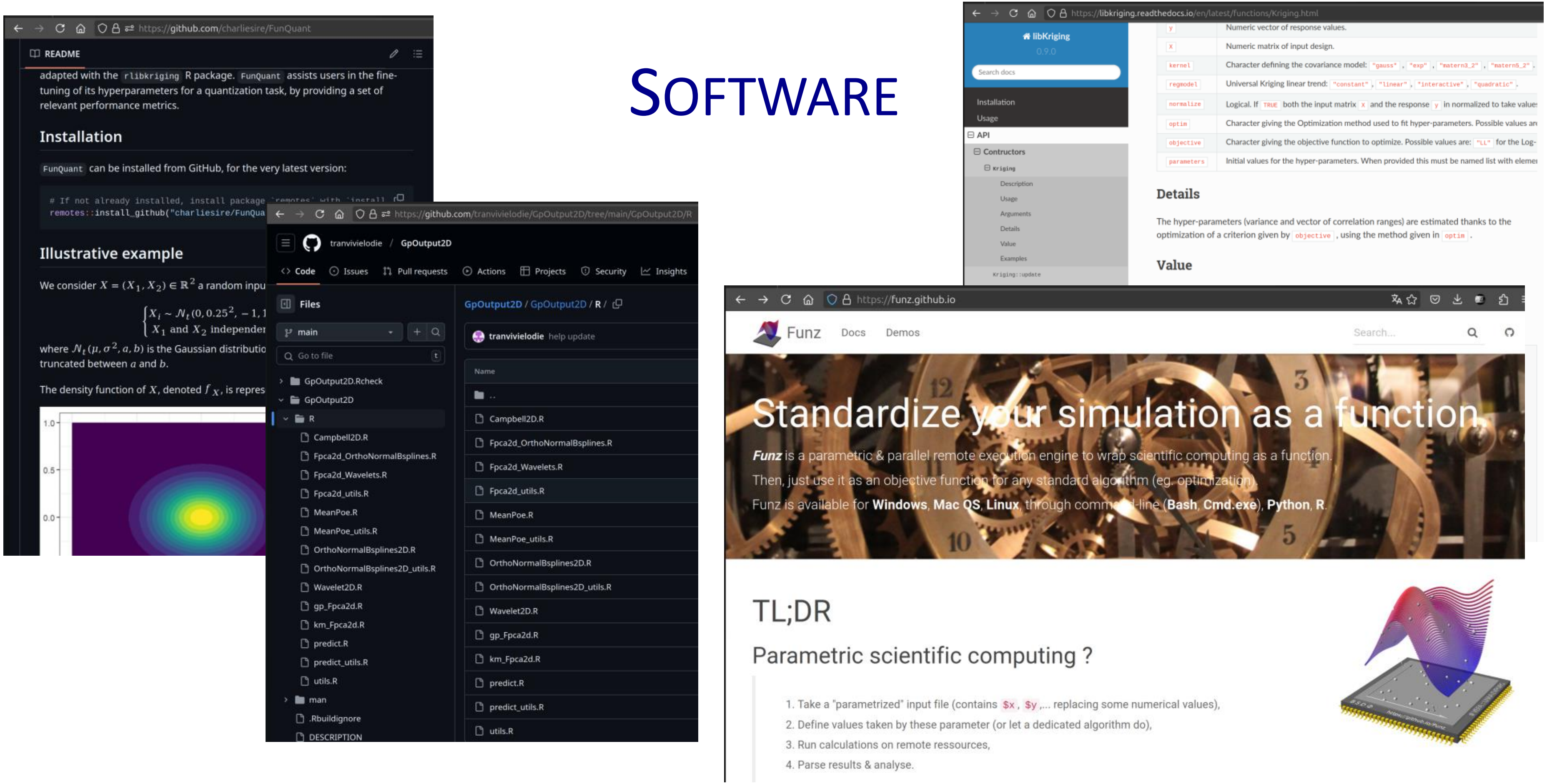
without importance sampling:



with importance sampling:



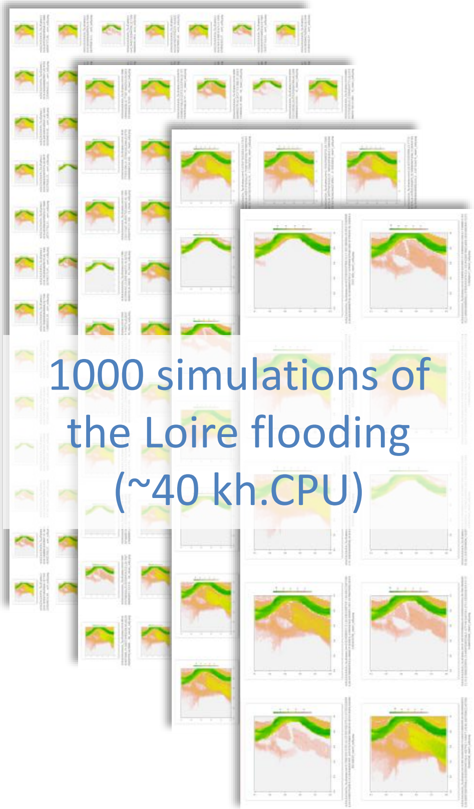
SOFTWARE


https://github.com/charliesire/FunQuant
README
adapted with the rlibkriging R package. FunQuant assists users in the fine-tuning of its hyperparameters for a quantization task, by providing a set of relevant performance metrics.
Installation
FunQuant can be installed from GitHub, for the very latest version:
Illustrative example
https://github.com/tranvivielodie/GpOutput2D/tree/main/GpOutput2D/R
tranvivielodie / GpOutput2D
Code Issues Pull requests Actions Projects Security Insights
Files
main
Go to file
GpOutput2D.Rcheck
GpOutput2D
R
Campbell2D.R
Fpca2d_OrthoNormalBsplines.R
Fpca2d_Wavelets.R
Fpca2d_utils.R
MeanPoe.R
MeanPoe_utils.R
OrthoNormalBsplines2D.R
OrthoNormalBsplines2D_utils.R
Wavelet2D.R
gp_Fpca2d.R
km_Fpca2d.R
predict.R
predict_utils.R
utils.R
man
.Rbuildignore
DESCRIPTION
GpOutput2D / GpOutput2D / R /
tranvivielodie help update
Name
https://libkriging.readthedocs.io/en/latest/functions/Kriging.html
libKriging
Installation
Usage
API
Contructors
Kriging
Details
The hyper-parameters (variance and vector of correlation ranges) are estimated thanks to the optimization of a criterion given by objective, using the method given in optim.
Value
https://funz.github.io
Funz Docs Demos
Standardize your simulation as a function
Funz is a parametric & parallel remote execution engine to wrap scientific computing as a function.
Then, just use it as an objective function for any standard algorithm (eg. optimization).
Funz is available for Windows, Mac OS, Linux, through command-line (Bash, Cmd.exe), Python, R.
TL;DR
Parametric scientific computing ?
1. Take a "parametrized" input file (contains $x, $y,... replacing some numerical values),
2. Define values taken by these parameter (or let a dedicated algorithm do),
3. Run calculations on remote ressources,
4. Parse results & analyse.

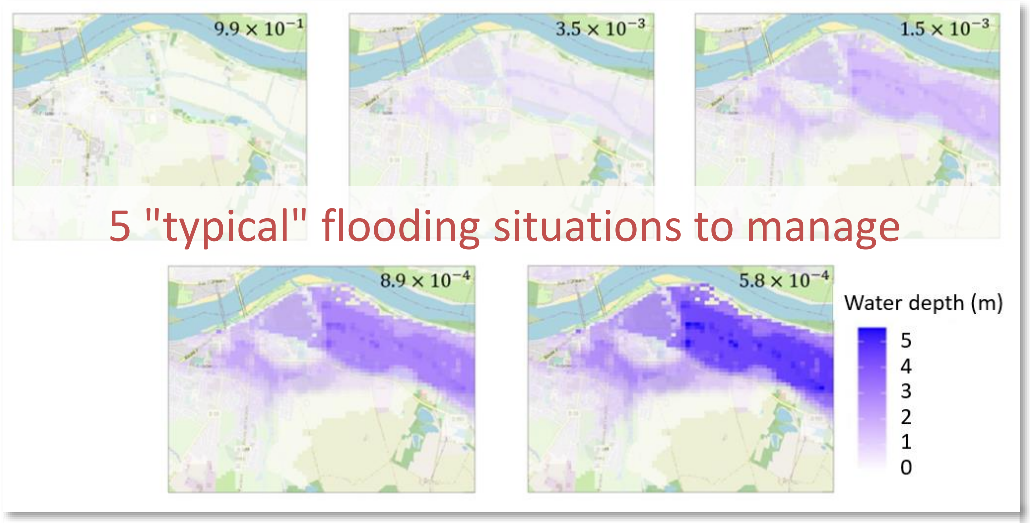
… INTERPRETING FOR DECISION-MAKING

[Ex.] Flooding risk

Open Access: https://github.com/IRSN/data_LoireSully-TelemacV8



1000 simulations of the Loire flooding (~40 kh.CPU)



5 "typical" flooding situations to manage

Soutenance de thèse
Doctorat de l'université de Toulouse

Vers l'utilisation d'ensembles météorologiques pour la dispersion à courte distance de radionucléides en cas de rejets accidentels dans l'atmosphère

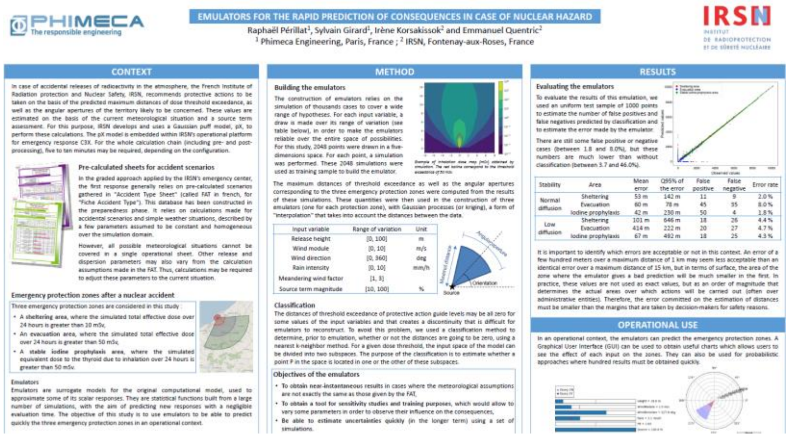
Youness El-Ouartassy

Membres du jury

V. Thouret	Présidente
E. Blayo et Y. Roustan	Rapporteurs
L. Soulhac, M. Rochoux et Y. Richet	Examinateurs
M. Plu et I. Korsakissok	Directeurs de thèse
L. Descamps	Co-encadrant

08 décembre 2023

Centre National de Recherches Météorologiques, Université de Toulouse, Météo-France, CNRS, Toulouse, France
Institut de Radioprotection et de Sûreté Nucléaire, PSE-SANTE/SESUC/BMCA, Fontenay-aux-Roses, France



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

Quantization methods for the visualization of the flooding risk

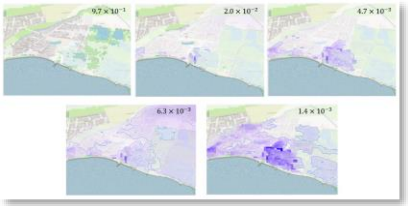
Charlie SIRE[1,2,3]

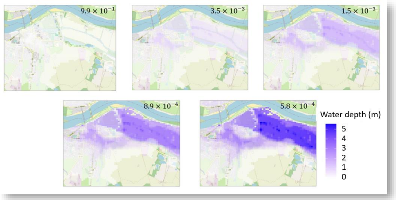
Supervisors: R. LE RICHE[3], D. RULLIERE[3], J. ROHMER[2], L. PHEULPIN[1], Y.RICHET[1]

[1]IRSN

[2]BRGM

[3]Mines Saint-Etienne and CNRS,LIMOS





WHAT'S NEXT ?

Major challenges remain in

- application field:
  - Communicating **uncertainties to decision** makers
  - Integrating **operational constraints** early in expertise
- ... and (still) in math. modeling tools:
  - Draw a path toward **standards in high-dim projections** (ex. similar to metamodel requirements)
  - Integrate **intrinsic physical properties** in metamodels (ex. "Physics Informed **", so mitigate **interpret-/explain-ability** tenet)